

# Mathematical Writing

Xie Zhilin

November 20, 2025

## 1 Introduction: Why We Bother with Mathematical Writing?

The word *bother* in the title is deliberate. Mathematical writing is not just a skill. It is an art that requires dedication and practice. Good writing can clarify complex ideas, making them accessible to a broader audience. Conversely, poor writing can obscure even the most brilliant concepts.

Paul Halmos once noted that mathematicians who merely *think* great theorems have not completed their job. They must, like a painter revealing a canvas, effectively *communicate* those ideas to others.

This lecture is about that communication. We will focus on practical, actionable principles to make our technical writing clearer, more elegant, and more effective.

Our work will be guided by two fundamental goals, also from Halmos

**Do organize.**

**Do not distract.**

## 2 Core Rules for Technical Writing

This section will be divided into two parts. The first part summarizes general principles for writing mathematics clearly and correctly. The second part provides case studies to illustrate how to apply these principles in practice.

### 2.1 General Principles

The following rules are adapted from Donald Knuth's *Mathematical Writing* and tailored for our context.

- (a) **Separate symbols with words.** Symbols in different formulas should be separated by words.

**Bad:** Let  $f(x) = x^\lambda$ ,  $\lambda > 0$ .

**Good:** Let  $f(x) = x^\lambda$ , where  $\lambda > 0$ .

- (b) **Start sentences with words.** Avoid starting sentences with symbols.

**Bad:**  $f(x)$  is continuous on  $[a, b]$ .

**Good:** The function  $f(x)$  is continuous on the interval  $[a, b]$ .

- (c) **Use words to connect symbols.** Use words like “*is*”, “*equals*”, “*implies*” instead of symbols like  $\Rightarrow$  and  $\iff$ .

**Bad:**  $f$  is continuous  $\Rightarrow f$  is integrable.

**Good:** If  $f$  is continuous, then  $f$  is integrable.

- (d) **Complete statements preceding theorem.** Statements before theorems (algorithms, propositions, etc.) should be complete sentences or ended with a colon.

**Bad:** We have the following theorem

**Theorem.**  $DH(x) = \delta(x)$ .

**Good:** We have the following theorem.

**Theorem.** The Heaviside function satisfies  $DH(x) = \delta(x)$ .

- (e) **Self-contain theorems.** Theorems should be stated in a way that they can be understood without referring to previous text.

**Bad:** **Theorem.**  $DH(x) = \delta(x)$ .

**Good:** **Theorem.** Let  $H(x)$  be the Heaviside function defined by  $H(x) = 0$  for  $x < 0$  and  $H(x) = 1$  for  $x \geq 0$ . Then its distributional derivative satisfies  $DH(x) = \delta(x)$ , where  $\delta(x)$  is the Dirac delta function.

- (f) **“We” is better.** Use “we” instead of “I” or passive voice.
- (g) **Don’t omit “that”.** Always include the word “that” in sentences when it helps clarify the meaning.

**Bad:** Assume  $f \in L^1(\mathbb{R}^n)$ .

**Good:** Assume that  $f$  is in  $L^1(\mathbb{R}^n)$ .

Moreover, don’t say “which” when “that” is better. We only use “which” after a comma or a preposition.

- (h) **Use different words and structures.** Avoid repeating the same words or sentence structures in close proximity. Vary your language to maintain reader interest.

**Bad:** We first prove the lemma. Then we prove the theorem. Finally, we prove the corollary.

**Good:** We begin by proving the lemma. Next, we establish the theorem. Lastly, we derive the corollary.

But use parallelism when parallel concepts are being discussed. Here is an example from Jost:

*The Ricci curvature is the average of sectional curvatures, and the scalar curvature is the average of Ricci curvatures.*

- (i) **Connect formulas with words.** Don’t use the homework style to write such as listing formulas one after another without any explanation. Always connect formulas with words.
- (j) **Define new things twice.** When introducing a new concept, definition, or notation, try to state them in complementary ways twice to reinforce understanding. And all of variables used should be clearly defined when they are first introduced.
- (k) **Capitalize correctly.** Capitalize the first word of a sentence, proper nouns, and important terms in titles and headings. Avoid unnecessary capitalization of common nouns. But capitalize names of theorems, lemmas, and definitions.

- (l) **Never abuse subscripts.** Don't get carried away with subscripts, especially when dealing with a set that doesn't need to be indexed. For example, the definition below is troublesome, since its subscripts give the false impression that  $Y$  depends on the choice of indices.

*Let  $X = \{x_1, \dots, x_n\}$  be a set. And  $Y = \{x_{i_1}, \dots, x_{i_m}\}$  is a subset of  $X$ .*

- (m) **Spell out small numbers.** Spell out numbers from zero to nine in words, but not when used as labels or in equations.

**Bad:** There are 3 cases to consider.

**Good:** There are three cases to consider.

- (n) **The first one places the first.** The opening paragraph and the first sentence of a section are crucial. They set the tone and context for what follows. Make them clear and engaging. Please emphasize the main objectives and results right at the beginning.

**Bad:** A core concept in harmonic analysis is the Fourier transform.

**Good:** Fourier transform plays a central role in harmonic analysis.

In Knuth's words,

*The worst way to start is with the sentence of the form “An  $x$  is  $y$ .”*

- (o) **Avoid long strings of symbols.** Long strings of symbols are hard to read and understand. Break them up with words or rephrase them.

**Bad:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a function in  $L^1(\mathbb{R}^n)$  with Fourier transform  $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by  $\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i x \cdot \xi} dx$  for all  $\xi \in \mathbb{R}^n$ .

**Good:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a function in  $L^1(\mathbb{R}^n)$ . Its Fourier transform  $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}$  is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i x \cdot \xi} dx$$

for all  $\xi \in \mathbb{R}^n$ .

## 2.2 Case Studies

Here is a bad example taken from my handwritten notes.

**(Bad) Theorem 1.**  $u^*$  is the minimal point of functional  $I$ , then

$$\int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds - L_{p^i}(t, u^*(t), \dot{u}^*(t)) = C$$

**Proof.**  $\varphi \in C_0^1(J, \mathbb{R}^n)$ ,

$$\begin{aligned} & \Rightarrow I(u^* + \epsilon\varphi) - I(u^*) \geq 0 \\ \lim_{\epsilon \rightarrow 0} \frac{I(u^* + \epsilon\varphi) - I(u^*)}{\epsilon} &= \int_{t_0}^{t_1} \sum_{i=1}^n (L_{u^i}(s, u^*(s), \dot{u}^*(s))\varphi^i(s) + L_{p^i}(s, u^*(s), \dot{u}^*(s))\dot{\varphi}^i(s)) ds \\ &= \int_{t_0}^{t_1} \sum_{i=1}^n \left( \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds - L_{p^i}(t, u^*(t), \dot{u}^*(t)) \right) \dot{\varphi}^i(t) dt \\ &\geq 0 \blacksquare \end{aligned}$$

replace  $\epsilon$  with  $-\epsilon$ , we have

$$\begin{aligned} & \int_{t_0}^{t_1} \sum_{i=1}^n \left( \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds - L_{p^i}(t, u^*(t), \dot{u}^*(t)) \right) \varphi^i(t) dt \leq 0 \\ & \Rightarrow \int_{t_0}^{t_1} \sum_{i=1}^n \left( \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds - L_{p^i}(t, u^*(t), \dot{u}^*(t)) \right) \dot{\varphi}^i(t) dt = 0 \end{aligned}$$

by du Bois-Reymond lemma, we have

$$\int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds - L_{p^i}(t, u^*(t), \dot{u}^*(t)) = C$$

□

Let's analyze the issues with this example:

- It starts with a symbol instead of a word.
- The theorem statement is incomplete and lacks context. It does not define the functional  $I$  or the space in which  $u^*$  resides.
- The proof starts abruptly without a clear introduction or explanation of the strategy.

- There are several instances where symbols are used without accompanying words, making it hard to follow the logic. And symbols such as  $\Rightarrow$  are used instead of words.
- The notation is inconsistent and sometimes unclear.
- The overall structure of the proof is disorganized, making it difficult to discern the main steps and conclusions.

We can improve this example by applying the rules discussed earlier. The following is a revised version.

**Revised Version.** Fix an interval  $J = [t_0, t_1]$ , an open region  $\Omega \subset \mathbb{R}^N$ . Fixed a continuously differentiable function  $L : J \times \Omega \times \mathbb{R}^N \rightarrow \mathbb{R}$  where  $L = L(t, u, p)$  for  $t \in J$ ,  $u \in \Omega$ , and  $p \in \mathbb{R}^N$ . We denote by  $C^1(J, \mathbb{R}^N)$  the set of continuously differentiable functions from  $J$  to  $\mathbb{R}^N$ . Let  $a, b$  in  $\mathbb{R}^N$  be fixed vectors and  $M$  be the set defined by

$$M = \{u \in C^1(J, \mathbb{R}^N) \mid u(t_0) = a, u(t_1) = b\}.$$

We consider the functional  $I : M \rightarrow \mathbb{R}$  defined by

$$I(u) = \int_{t_0}^{t_1} L(t, u(t), \dot{u}(t)) dt,$$

where  $\dot{u}(t)$  denotes the gradient of  $u$  at  $t$ , that is

$$\dot{u}(t) = \left( \frac{du^1}{dt}(t), \dots, \frac{du^N}{dt}(t) \right).$$

A *minimizer* of  $I$  is an element  $u^*$  in  $M$  such that there exists a neighborhood  $U$  of  $u^*$  under the  $C^1(J, \mathbb{R}^N)$  norm that satisfies

$$I(u^*) \leq I(u), \quad \forall u \in U \cap M.$$

We want to find a necessary condition for  $u^*$  to be a minimizer of  $I$ . Like optimization problems in finite-dimensional spaces, we consider variations of  $u^*$  in the direction of a test function  $\varphi = (\varphi^1, \dots, \varphi^i, \dots, \varphi^N)$  in  $C_0^1(J, \mathbb{R}^N)$ , the set of continuously differentiable functions from  $J$  to  $\mathbb{R}^N$  that vanish at the endpoints  $t_0$  and  $t_1$ . Assume

that when it satisfies  $|\epsilon| < \epsilon_0$  then  $u^* + \epsilon\varphi$  belongs to the neighborhood  $U$ . Since  $u^*$  is a minimizer of  $I$ , we have

$$I(u^* + \epsilon\varphi) - I(u^*) \geq 0.$$

We transform this problem into the minimization of the following function:

$$\min_{|\epsilon|<\epsilon_0} g_\varphi(\epsilon), \quad \text{where } g(\epsilon) = I(u^* + \epsilon\varphi)$$

Since  $g_\varphi(\epsilon)$  attains its minimum at  $\epsilon = 0$ , we have

$$g'_\varphi(0) = \lim_{\epsilon \rightarrow 0} \frac{I(u^* + \epsilon\varphi) - I(u^*)}{\epsilon} = 0.$$

Let us compute  $g'_\varphi(0)$ , we put the limit inside the integral since  $L$  is continuously differentiable

$$\begin{aligned} g'_\varphi(0) &= \lim_{\epsilon \rightarrow 0} \frac{I(u^* + \epsilon\varphi) - I(u^*)}{\epsilon} \\ &= \int_{t_0}^{t_1} \sum_{i=1}^N (L_{u^i}(t, u^*(t), \dot{u}^*(t))\varphi^i(t) + L_{p^i}(t, u^*(t), \dot{u}^*(t))\dot{\varphi}^i(t)) dt \end{aligned}$$

We observe that the first term can be integrated by parts and its boundary terms vanish since  $\varphi$  is in  $C_0^1(J, \mathbb{R}^N)$ , that is

$$\int_{t_0}^{t_1} L_{u^i}(t, u^*(t), \dot{u}^*(t))\varphi^i(t) dt = - \int_{t_0}^{t_1} \left( \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds \right) \dot{\varphi}^i(t) dt$$

for each  $i = 1, \dots, N$ . Thus we have

$$g'_\varphi(0) = \int_{t_0}^{t_1} \sum_{i=1}^N \left( L_{p^i}(t, u^*(t), \dot{u}^*(t)) - \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds \right) \dot{\varphi}^i(t) dt = 0.$$

Since  $\varphi$  is an arbitrary function in  $C_0^1(J, \mathbb{R}^N)$ , we assume  $\varphi^j = 0$  for all  $j \neq i$ . Then for each  $i = 1, \dots, N$ , the equation above reduces to

$$\int_{t_0}^{t_1} \left( L_{p^i}(t, u^*(t), \dot{u}^*(t)) - \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s)) ds \right) \dot{\varphi}^i(t) dt = 0. \quad (2.1) \boxed{\text{eq:1}}$$

This equation is concerned with function  $\varphi^i$ . Thus we need a lemma to remove it.

**Lemma 2.1** (du Bois-Reymond). *Let  $f \in C([a, b], \mathbb{R})$ . If*

$$\int_a^b f(t)\dot{\varphi}(t)dt = 0, \quad \forall \varphi \in C_0^1([a, b], \mathbb{R}),$$

*then  $f(t)$  equals  $c$  for some constant  $c \in \mathbb{R}$  and for all  $t \in [a, b]$ .*

**Proof.** Define  $c$  and  $\varphi$  as follows

$$c = \frac{1}{b-a} \int_a^b f(t)dt, \quad \varphi(t) = \int_a^t (f(s) - c)ds.$$

Then  $\varphi$  belongs to  $C_0^1([a, b], \mathbb{R})$  and

$$\int_a^b f(t)\dot{\varphi}(t)dt = \int_a^b f(t)(f(t) - c)dt = \int_a^b (f(t) - c)^2 dt = 0.$$

Thus  $f(t)$  is  $c$  almost everywhere. Since  $f$  is continuous, we have  $f(t) = c$  for all  $t \in [a, b]$ .  $\square$

By applying du Bois-Reymond Lemma to equation (2.1), we obtain the following necessary condition for  $u^*$  to be a minimizer of  $I$ .

**Theorem 2.1** (Euler-Lagrange Equation). *If  $u^*$  is a minimizer of the functional  $I$ , then for each  $i = 1, \dots, N$ , it satisfies*

$$L_{p^i}(t, u^*(t), \dot{u}^*(t)) - \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s))ds = C,$$

where  $L_{u^i}$  and  $L_{p^i}$  denote the partial derivatives of  $L$  with respect to its second and third arguments, and  $C$  is a constant.

**Proof.** Let  $f$  defined by

$$f(t) = L_{p^i}(t, u^*(t), \dot{u}^*(t)) - \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s))ds.$$

By equation (2.1) and du Bois-Reymond Lemma, we have

$$f(t) = L_{p^i}(t, u^*(t), \dot{u}^*(t)) - \int_{t_0}^t L_{u^i}(s, u^*(s), \dot{u}^*(s))ds = C$$

for some constant  $C \in \mathbb{R}$  and all  $t \in [t_0, t_1]$ . This completes the proof.  $\square$

As we can observe, the revised version starts with a clear context and definitions, making it easier for readers to understand the theorem and its proof. The proof is structured logically, with each step explained in detail and connected with words. Symbols are used appropriately, and the overall presentation is more organized and reader-friendly.

### 3 Writing Style and Artistry

Conventions introduced in Section 2 help us write a *correct* mathematical text. However, correctness alone does not guarantee clarity or elegance and what we need to convey our ideas effectively is a suitable *writing style*.

The reason why I use the word *suitable* is that writing style is often subjective and can vary based on the audience, context, and personal preferences. For instance, Professor Herbert Wilf himself edits two very different journals: *The Journal of Algorithms* and *The American Mathematical Monthly*. The former is a research journal focusing on algorithms, and results presented there are new, important and significant to this field. As a result, the writing style is formal, concise, and technical. There is little leeway for personal expression, and the emphasis is on clarity and precision. On the other hand, *The American Mathematical Monthly* is a journal that aims to present mathematical ideas in an accessible and engaging manner to a broader audience. The writing style here is more relaxed, conversational, and often includes anecdotes or historical context to make the content more relatable. (What is interesting is that this style perhaps attracts others to send "proof" of Fermat's Last Theorem to this journal!)

Though different styles play different roles, those good writing styles often share some common characteristics. Here are some general tips to develop a good mathematical writing style.

- (a) **Get the attention of readers immediately.** Snappy titles, engaging introductions, and clear statements of purpose can help capture the reader's interest from the outset.

For example, Hugh Thurston (not the famous William Thurston who can observe a 7-dimensional sphere in 8-dimensional space) began his paper with the sentence “Can a graph be continuous and discontinuous?”. This immediately piques the reader’s curiosity and sets the stage for the discussion on fractal geometry.

- (b) **Get everything up front.** Tell your readers in plain English what you are going to write about before you start writing about it. This helps readers understand the context and purpose of your work.
- (c) **Remember: Readers scan papers when they are reading.** Potential readers will skim through your paper to decide whether to read it in detail. Use bold face to emphasize and summarize key points, and use section headings to guide readers through the structure of your paper.

- (d) **Motivation is necessary, but too much is harmful.** Presenting examples that do not yield desired results can help motivate the need for new definitions or theorems. However, excessive motivation can distract readers from the main content. Strike a balance by providing just enough motivation to clarify the purpose of your work without overwhelming the reader.

As examples, we list several superb textbooks of mathematics.

*Functional Analysis* by Yosida. This classic textbook is known for its rigorous approach to functional analysis. Yosida's writing style is concise and precise, making complex concepts accessible to advanced undergraduate and graduate students. Yosida's clear explanations and structured presentation make it a valuable resource for students and researchers in the field.

*Algebra: Chapter 0* by Paolo Aluffi. This textbook takes a unique approach to teaching algebra by starting from category theory. Aluffi's writing style is engaging and intuitive, and he uses numerous examples to illustrate abstract concepts.

*Introduction to Smooth Manifolds* by John M. Lee. Lee's writing style is clear and thorough, providing detailed explanations and examples that help readers grasp the intricacies of differential geometry and smooth manifolds.

The chosen books are from three different major subfields of mathematics: analysis, algebra, and geometry. Each book is well-regarded for its clarity, depth, and pedagogical approach, making them excellent examples of good mathematical writing style. An interesting observation is that books in analysis seem to prioritize rigor and precision, while books in algebra and geometry often emphasize intuition and visualization. This reflects the different nature of these subfields and the types of thinking they require.

Besides writing style, another noticeable word is *art*. By Nils Nilson, writing can be both art and communication. A key word of art is *COMPOSITION*, we summarize it as a formula

$$\text{composition} = \text{organization} + \text{simplification}$$

But if writing is to be art, we must first master the craft of writing. We give an overview of some important aspects.

- (a) **Write, rewrite, and rewrite again.** Writing is a process that involves multiple drafts and revisions. The first draft is rarely perfect, and it is essential to revisit and refine your work to improve clarity, coherence, and style.

- (b) **Read as much as you can.** Reading widely in mathematics and other fields can expose you to different writing styles, techniques, and ideas. This can help you develop your own voice and improve your writing skills.
- (c) **Master the medium.** Writing needs a good vocabulary. There are issues other than pure language: indexes, tables, figures, and how to use them to best advantage. Learning to use typesetting systems like L<sup>A</sup>T<sub>E</sub>X can help you produce professional-quality documents.
- (d) **Simplify.** Strive for simplicity in your writing. Avoid unnecessary complexity, jargon, and convoluted sentences. Aim to convey your ideas in the clearest and most straightforward manner possible.
- (e) **Avoid recycling.** Reusing phrases, sentences, or paragraphs from previous works can lead to redundancy and a lack of originality. Each piece of writing should be fresh and tailored to its specific purpose and audience.

## 4 Writing in the Real World: Refereeing

Refereeing is an essential part of the academic publishing process. A paper submitted to an academic journal is usually passed to one or more referees by the editor. Each referee is expected to read the paper carefully, evaluate its quality, and provide constructive feedback to the editor and the authors.

Professor Donald Knuth once had an interesting experience of his now-famous research on “The Toilet Paper Problem”. It was first published in *The American Mathematical Monthly*. Donald joked to the editor (Paul Halmos) that many readers probably keep their copies in the bathroom. In response, Halmos commented “jokes are dangerous in our journal”, especially for these scatological ones. Finally, Donald agreed to change the section names, which originally were “An absorbing barrier”, “A process of elimination”, and “Residues” as innocuous bathroom humor, but he insisted on keeping the original title.

Refereeing is not like working as a real referee in a stadium, who judges the performance of athletes. Instead, refereeing is more like being a reader and author, the purpose of which is to improve the quality of a paper and encourage high standards of writing in the mathematical community. Here are some tips for effective refereeing:

- (a) **Be a teacher.** The author you criticize today will write another new version of the paper tomorrow. Your comments should help the author improve their work and learn from their mistakes. Therefore over-critical comments should be avoided.
- (b) **Giving references to related works.** If you know of relevant literature that the author has not cited, it is helpful to provide references. This can help the author situate their work within the broader context of the field and acknowledge prior contributions.
- (c) **Spend less time on grammar and typos.** While it is important to point out significant grammatical errors or typos that affect the clarity of the paper, spending too much time on minor issues can detract from the overall evaluation of the work. Focus on the content and structure of the paper first, and only address grammar and typos if they significantly impede understanding.

Here is another experience from Professor Knuth. In 1960, Knuth had just began to edit for *Communication of the ACM* and *Journal of the ACM*. He had no way to know which of his referees were good or bad. So he sent each referee a copy of his own paper on an interesting algorithm but written in a poor style. The result was that some referees

simply went through the paper line by line and listed their complaints point by point. Another made much more general comments about the overall structure and clarity of the paper. A third said that the paper contained little that was new and supplied a substantial bibliography of related work to refer to. From this experiment, we can see that good referees are not necessarily those who find the most mistakes, but those who provide constructive feedback that helps improve the quality of the paper.

## 5 Useful References

From esoteric to practical, here are some useful references for further reading on mathematical writing.

- (a) *The Oxford English Dictionary*. Famous for its historical approach to the English language, providing definitions and etymologies of words over time.
- (b) *The American Heritage Dictionary*. This dictionary also takes a historical perspective but includes more contemporary usage and examples.
- (c) *The Longman Dictionary of Contemporary English*. Instead of the historical words found in the mentioned two dictionaries, this dictionary focuses on current usage and contemporary meanings in very simple English.
- (d) *Webster's New Word Speller Divider*. People who don't spell well can use this book to check spelling.
- (e) *Roget's Thesaurus*. This book is a synonym dictionary that helps writers find alternative words and phrases to enhance their writing.
- (f) *Webster's Dictionary of Synonyms*. A resource filled with choice examples of synonyms and antonyms to enrich vocabulary.