Elements of Mathematical Style

A work in progress [1]

Bruce Francis
Department of Electrical and Computer Engineering
University of Toronto

February 10, 2015

[1]I thank Murray Wonham and Audrey Kertesz for reading this work and offering helpful comments.

Contents

Preface

This is a very brief handbook on writing documents with mathematical content: reports, theses, journal papers, and conference papers. The handbook is written primarily for graduate students in control and similar subjects.

To write your theorems correctly, you need to know elementary logic, namely, propositional calculus. The main point is that variables (ε, δ, etc.) must be quantified ("there exists," "for all") and the quantifiers must be in the right order.

I decided to write this handbook after years—in fact decades—of reading and correcting theses. Now when students ask me to be on their thesis examination committees, I tell them to go and read this handbook before giving me their theses. It reduces the amount of red ink.

Chapter 1

Preliminaries

1. Set. Examples of sets: the set $\mathbb{R}$ of real numbers; the set $\mathbb{R}^{m\times n}$ of $m \times n$ real matrices; the set $\mathcal{F}(U,V)$ of functions from the set U to the set V. A set is frequently defined using notation like this:

   $$S = \{x : x \in \mathbb{R}, x > 0\}.$$

   In words this says, S is the set of all x such that x belongs to the real numbers and x is positive. Thus a colon means "such that," a comma means "and," and $\in$ means "belongs to" or "is a member of." Notice that the symbol x is a dummy variable and that we have exactly the same set if we write

   $$S = \{\rho : \rho \in \mathbb{R}, \rho > 0\}.$$

   So, preceding the colon is a symbol for a variable, and following the colon are descriptors about that variable. A variation is

   $$S = \{x \in \mathbb{R} : x > 0\},$$

   where one of the descriptors of x precedes the colon. Another common form of writing the set is to use a vertical line instead of a colon:

   $$S = \{x | x \in \mathbb{R}, x > 0\}.$$

   Suppose you want the set of real numbers greater than a number a. You could write

   > Let $a \in \mathbb{R}$ and define
   >
   > $$S = \{x : x \in \mathbb{R}, x > a\}.$$

   Or you could write

   > Define
   >
   > $$S = \{x : x \in \mathbb{R}, x > a\}, \quad a \in \mathbb{R}.$$

In words: S is the set of all x such that x belongs to the real numbers and x is greater than a, where a belongs to the set of real numbers. But this is wrong:

$$S = \{x : x \in \mathbb{R}, x > a, a \in \mathbb{R}\}.$$

In fact, this latter set is all of $\mathbb{R}$, for it says S is the set of all x such that x belongs to the real numbers, x is greater than a, and a belongs to the set of real numbers too, i.e., a is not fixed.

2. If U and V are two sets, a **function** from U to V is a rule that assigns to every element of U an unambiguous element of V—there cannot be two possible different values $f(u)$ for the same u in U. The terms function, mapping, and transformation are synonymous. The notation

$$f : U \longrightarrow V$$

means that f is a function from U to V. The set U is called the **domain** of f and V is called the **codomain**. It is sometimes useful to avoid giving a name to the function, and then we use the "mapsto" symbol $\mapsto$. For example, the function $x \mapsto x^2$ from $\mathbb{R}$ to itself takes a real number x and squares it. This function could alternatively be written

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = x^2.$$

Notice that the symbol for the function is f whereas, strictly speaking, $f(x)$ denotes the value of f at the point x. That being understood, it is common to say "the function $f(x)$."

3. The meaning of "if." The word "if" is very common in mathematical writing. One says, for example, "If f is a continuous function on a closed, bounded interval $[a, b]$, then f has a maximum value on that interval." We say that continuity of the function and the interval being closed and bounded are *sufficient* conditions for the existence of a maximum.

4. The phrase "only if" is a little trickier. Consider this: "A sequence of real numbers is convergent only if it is bounded." We say that boundedness is a *necessary* condition for convergence. This is the same as saying, a sequence cannot converge unless it is bounded, or an unbounded sequence cannot converge. The statement does not say that if a sequence is bounded, then it converges; that would be saying that boundedness is sufficient for convergence.

5. It will be convenient to use the abbreviation iff for "if and only if." As an example use, let A be a square real matrix and consider the homogeneous equation $Ax = 0$, where x and 0 are vectors. The vector $x = 0$ is always a solution of the equation, but there may be other solutions. Here's a result about uniqueness:

> The equation $Ax = 0$ has the unique solution $x = 0$ iff the determinant of A is nonzero.

This statement contains two logical implications, the "if" part and the "only if" part. The "if" part is this: The equation has a unique solution if the determinant of A is nonzero. That is, $\det(A) \neq 0$ is a sufficient condition for there to be a unique solution. The "only if" part is this: The equation has a unique solution only if the determinant of A is nonzero. That is, if $\det(A) = 0$, then there is not a unique solution.

6. A quantifier is a mathematical symbol indicating the amount or quantity of the variable or expression that follows. There are two quantifiers:

   $\exists$ denotes the existential quantifier meaning "there exists" (or "for some").

   $\forall$ denotes the universal quantifier meaning "for all" or "for every."

7. The symbol for logical negation is $\neg$.

8. Logical conjunction (**and**) is denoted by $\wedge$ or by a comma, and logical disjunction (**or**) is denoted by $\vee$. The negation of $P \wedge Q$ is $\neg P \vee \neg Q$. The negation of $P \vee Q$ is $\neg P \wedge \neg Q$.

9. The final logic operation is denoted by the symbol $\implies$, which means "implies" or "if ... then." The statement $P \implies Q$ is logically equivalent to the statement $\neg Q \implies \neg P$. The truth table for the logical implication operator is

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

   Writing out the truth table for $P \wedge \neg Q$ will show you that it is (logically equivalent to) the negation of $P \implies Q$.

Chapter 2

Definition Statements

Logic notation is a great aid in precision, and therefore in a clear understanding of mathematical concepts. This chapter provides a brief introduction to mathematical statements using logic notation.

Example

Let $x = \{x_n\}_{n\geq 1}$ be a sequence of real numbers. We want to define what it means for x to be bounded. In words, x is bounded if the absolute values of all its components are less than some bound. The key phrases are "all its components" and "some bound." The phrase "all its components" suggests x_n for every n. Clearly, there must be one bound for all n.

Here's the definition in mathematical form that the sequence x is bounded:

$$(\exists B \geq 0)(\forall n \geq 1)\ |x_n| \leq B. \tag{2.1}$$

This statement is parsed from left to right. In words, (2.1) says this: There exists a nonnegative number B such that, for every positive integer n, the absolute value of x_n is bounded by B. Notice in (2.1) that the two quantifier phrases, $\exists B \geq 0$ and $\forall n \geq 1$, are placed in brackets and precede the term $|x_n| \leq B$. This latter term has n and B as variables in it that need to be quantified. We cannot say merely that $\{x_n\}_{n\geq 1}$ is bounded if $|x_n| \leq B$ (unless it is known or understood what the quantifiers on n and B are). In general, all variables in a statement need to be quantified.

As an example, the sequence $x_n = (-1)^n$ of alternating $+1$ and -1 is obviously bounded. Here are the steps in formally proving (2.1) for this sequence:

Take $B = 1$.

Let $n \geq 1$ be arbitrary.

Then $|x_n| = |(-1)^n| = 1$. Thus $|x_n| = B$.

The order of quantifiers is crucial. Observe that (2.1) is very different from saying

$$(\forall n \geq 1)(\exists B \geq 0)\ |x_n| \leq B, \tag{2.2}$$

which is true of **every** sequence. Let's prove, for example, that the unbounded sequence $x_n = 2^n$ satisfies (2.2):

Let $n \geq 1$ be arbitrary.

Take $B = 2^n$.

Since $|x_n| = 2^n$, so $|x_n| = B$.

The sequence $\{x_n\}_{n\geq 1}$ is not bounded if, from (2.1),

$$\neg(\exists B \geq 0)(\forall n \geq 1)\ |x_n| \leq B.$$

This is logically equivalent to

$$(\forall B \geq 0)(\exists n \geq 1)\ |x_n| > B.$$

Study how this statement is obtained term-by-term from the previous one: $\exists B \geq 0$ changes to $\forall B \geq 0$; $\forall n \geq 1$ changes to $\exists n \geq 1$; and $|x_n| \leq B$ is negated to $|x_n| > B$. The order of the variables being quantified (B then n) does not change.

Example

This example discusses the definition of bounded function. The first attempt is incorrect because it leaves out the quantifiers:

> A real-valued function $x(t)$ of the real variable t is **bounded** if $|x(t)| \leq B$.

The next attempt has quantifiers but in the wrong order and therefore it is incorrect:

> A real-valued function $x(t)$ of the real variable t is **bounded** if for all t there is a real B such that $|x(t)| \leq B$.

The next attempt has quantifiers but not in the best places:

> A real-valued function $x(t)$ of the real variable t is **bounded** if there is a real B such that $|x(t)| \leq B$ for all t.

The next one is best:

> A real-valued function $x(t)$ of the real variable t is **bounded** if there exists a real number B such that, for all t, $|x(t)| \leq B$.

Notice that the last one has, as it should, the same structure as the logic statement that $x(t)$ is bounded:

$$(\exists B)(\forall t)\ P(B, t).$$

Example

Here's the definition that the sequence $\{x_n\}_{n\geq 1}$ converges to 0:

$$(\forall \varepsilon > 0)(\exists N \geq 1)(\forall n \geq N)\ |x_n| < \varepsilon.$$

This says, for every positive ε there exists a positive N such that, for every n that is greater than or equal to N, $|x_n|$ is less than ε. In words: x_n is arbitrarily small if n is sufficiently large. The quantifier $(\forall n \geq N)$ has two variables in it for simplicity. The longer form is

$$(\forall \varepsilon > 0)(\exists N \geq 1)(\forall n \geq 1)\ n \geq N \implies |x_n| < \varepsilon. \tag{2.3}$$

The structure of (2.3) is

$$(\forall \varepsilon)(\exists N)(\forall n)\ P(N, n) \Longrightarrow Q(n, \varepsilon).$$

The negation is therefore

$$(\exists \varepsilon)(\forall N)(\exists n)\ P(N, n) \wedge \neg Q(n, \varepsilon),$$

that is,

$$(\exists \varepsilon > 0)(\forall N \geq 1)(\exists n \geq 1)\ n \geq N,\ |x_n| \geq \varepsilon.$$

Some people write (2.3) in this way,

$$(\forall \varepsilon > 0)(\exists N(\varepsilon) \geq 1)(\forall n \geq 1)\ n \geq N(\varepsilon) \implies |x_n| < \varepsilon,$$

to emphasize that N may depend on ε. This is entirely unnecessary, since the left-to-right order already allows N to depend on ε.

You might read (2.3) like this: "For all positive ε there exists a positive N ...." Strictly speaking this is correct, but it doesn't sound quite right since by saying "for all" you may seem to be emphasizing "all" to the extent that someone may think that N doesn't depend on ε. Better to be safe and say "for every."

Example

The function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous at the point $x = a$ iff

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x)\ |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon. \tag{2.4}$$

The negation is therefore

$$(\exists \varepsilon > 0)(\forall \delta > 0)(\exists x)\ |x - a| < \delta,\quad |f(x) - f(a)| \geq \varepsilon.$$

Example

This example concerns stability theory. Consider the differential equation $\dot{x} = f(x)$, where x is a vector-valued function of time t. Assume $x = 0$ is an equilibrium point, i.e., $f(0) = 0$. Here's the definition in logic format that the origin is a **stable** equilibrium point:

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x(0))\ \|x(0)\| < \delta \Longrightarrow (\forall t \geq 0)\ \|x(t)\| < \varepsilon.$$

Here's the same thing in words:

> For every $\varepsilon > 0$ there exists $\delta > 0$ such that if the state starts in the ball of radius δ, it will remain forever in the ball of radius ε.

Here's a simplified verbal version:

> The solution will remain arbitrarily close to the origin if it starts sufficiently close to the origin.

Example

Here's a definition of the concept of safety:

> Given a set of admissible states $\mathcal{X} \subset \mathbb{R}^n$, a set of initial states $\mathcal{X}_0 \subset \mathcal{X}$, a set of unsafe states $\mathcal{X}_u \subset \mathcal{X}$, and a set of admissible inputs $\mathcal{U} \subset \mathbb{R}^m$, the **safety property** for
>
> $$\dot{x} = f(x, u), \quad x \in \mathbb{R}^n, \quad u \in \mathcal{U}$$
>
> holds if there exists no $T \geq 0$ with input trajectory $u : [0, T] \longrightarrow \mathcal{U}$ such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$, $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

The idea in this definition is that an unsafe situation has occurred if $x(t)$ has entered the unsafe set $\mathcal{X}_u$ at some time; hence, the system is unsafe if $x(t)$ can enter $\mathcal{X}_u$ for some $x(0)$ and some u. Of course, the case $x(0) \in \mathcal{X}_u$ must be ruled out, and that is why the set $\mathcal{X}_0$ is needed. Similarly, there may be control values we may wish to rule out, so bring in a set $\mathcal{U}$ of permitted control values. The reason for $\mathcal{X}$ is unclear, so let's leave that set out. This leads us to the tentative definition that the system is **unsafe** if

$$(\exists x(0))(\exists T > 0)(\exists u)\ (\forall t, 0 \leq t \leq T)\ u(t) \in \mathcal{U}, x(0) \in \mathcal{X}_0, x(T) \in \mathcal{X}_u.$$

Of course, it is understood that u must be as stated before: bounded and piecewise continuous. Are any other quantifiers required? The definition is about the system and the sets $\mathcal{X}_0, \mathcal{X}_u, \mathcal{U}$, so these should not be quantified. Also, $x(T)$ is uniquely determined by $x(0), T, u$, so it should not be quantified. With these thoughts, here's a proposed correction:

> Consider the system
>
> $$\dot{x} = f(x, u)$$

where f maps $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^n$. Assume the input u is bounded and piecewise continuous. Assume also that f is sufficiently regular that solutions $x(t)$ exist for all $x(0)$ and all such inputs. Let $\mathcal{X}_0 \subset \mathbb{R}^n$ be a set of allowed initial states, $\mathcal{X}_u \subset \mathbb{R}^n$ a set of states deemed to be unsafe, and $\mathcal{U} \subset \mathbb{R}^m$ a set of admissible input values. The system is **unsafe** if there exist $x(0)$, $T > 0$, and u, with $x(0)$ in $\mathcal{X}_0$ and $u(t)$ in $\mathcal{U}$ for all $0 \leq t \leq T$, such that $x(T)$ is unsafe, i.e., $x(T) \in \mathcal{X}_u$. The system is **safe** if it is not unsafe, i.e., such $x(0)$, T, u do not exist.

Chapter 3

Theorem Statements

Theorem statements are similar: They have to be grammatically correct according to the grammar of mathematics.

Example

The following statement is a theorem in *Complex Variables and Applications*, by Churchill and Brown:

> **Theorem** *If a function is analytic at a point, then its derivatives of all orders are also analytic at that point.*

Let us write the statement in logic form so that we're sure we understand it. The variables in the statement are the function, say f, the point, say a, and the order, say m. Then "f is analytic at a" defines a proposition $P(f,a)$, and "the m^{th}-derivative of f is analytic at a" defines a proposition $Q(f,m,a)$. Thus the theorem statement has the form

$$(\forall f)(\forall a)\ P(f,a) \Longrightarrow (\forall m)\ Q(f,m,a).$$

Example

Here's a theorem statement related to a result in control theory.

> **Theorem** *Consider system (1). Assume (A,B) is controllable and (C,A) is observable. Let $x(0)$ be arbitrary. There exists an LTI controller, with input $y_{ref} - y$ and output u, such that the feedback system is stable and the tracking error goes to zero as t goes to infinity.*

Let us analyze this statement by putting it in logic notation. We have to quantify all variables. What are the variables? Let's say that system (1) is a number of state-space equations, parametrized by matrices $A, B, \ldots$ and certain types of inputs. Is "system (1)" a variable? The sensible thing is to fix the structure of the equations but not the actual matrices. So our logic statement would start like this:

$$(\forall A, B, C, \dots)(A, B) \text{ controllable and } (C, A) \text{ observable } \implies .$$

After this there will be the term ($\exists$ LTI controller) and since this term comes after the quantification on $A, B, C, \dots$ the implication is that the controller depends on $A, B, C, \dots$, which is natural. What about the phrase "Let $x(0)$ be arbitrary." There's only one way to quantify that, $(\forall x(0))$. So it seems the theorem is of this form:

$$(\forall A, B, C, \dots)(A, B) \text{ controllable and } (C, A) \text{ observable } \implies (\forall x(0))(\exists \text{ LTI controller}) \dots .$$

But this says the controller depends on $x(0)$, that is, if the initial plant state changes, the controller has to be redesigned. This must be wrong. The correct form is

$$(\forall A, B, C, \dots)(A, B) \text{ controllable and } (C, A) \text{ observable } \implies (\exists \text{ LTI controller})(\forall x(0)) \dots$$

In this case, the theorem statement should have been

> **Theorem** *Consider system (1). Assume (A, B) is controllable and (C, A) is observable. There exists an LTI controller, with input $y_{ref} - y$ and output u, such that the feedback system is stable and, for every $x(0)$, the tracking error goes to zero as t goes to infinity.*

Example

This example concerns the Nyquist criterion:

> **Theorem** *A feedback system is stable if and only if the contour Γ_L in the $L(s)$-plane does not encircle the $(-1, 0)$ point when the number of poles of $L(s)$ in the right-hand s-plane is zero.*

The phrase "when the number of poles ..." seems to be a quantifier of the form

$$(\forall L) \text{ the number of poles of } L(s) \text{ in the right-hand } s\text{-plane is zero} \dots$$

But this cannot be. The statement concerns a **given** feedback system with open-loop transfer function $L(s)$. Thus, $L(s)$ is not a variable and therefore can't be quantified. Then what does the subordinate clause "when ... " mean? In fact the clause should be replaced by an assumption. The next mistake concerns the ambiguity of the verb "encircle." What if the contour passes through the point $(-1, 0)$? Does that count as not encircling it? The correct statement (though not the full Nyquist criterion) is this:

> **Theorem** *Assume the open-loop transfer function $L(s)$ is stable. Then the feedback system is stable if and only if the contour Γ_L does not pass through or encircle the point $(-1, 0)$.*

Example

The next example concerns the converse Lyapunov theorem.

> **Theorem** *Let $x = 0$ be an equilibrium point for the nonlinear system*
>
> $$\dot{x} = f(t, x)$$
>
> *where $f : [0, \infty) \times D \longrightarrow \mathbb{R}^n$ is continuously differentiable, $D = \{x \in \mathbb{R}^n : \|x\| < r\}$, and the Jacobian matrix $\partial f/\partial x$ is bounded on D, uniformly in t. Let k, γ and r_0 be positive constants with $r_0 < r/k$. Let $D_0 = \{x \in \mathbb{R}^n : \|x\| < r_0\}$. Assume that the trajectories of the system satisfy*
>
> $$\|x(t)\| \le k\|x(t_0)\|\mathrm{e}^{-\gamma(t-t_0)}, \quad \forall x(t_0) \in D_0, \quad \forall t \ge t_0 \ge 0$$
>
> *Then there is a function $V : [0, \infty) \times D \longrightarrow \mathbb{R}$ that satisfies the inequalities*
>
> $$c_1\|x\|^2 \le V(t, x) \le c_2\|x\|^2$$
>
> $$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \le -c_3\|x\|^2$$
>
> $$\left\|\frac{\partial V}{\partial x}\right\| \le c_4\|x\|$$
>
> *for some positive constants c_1, c_2, c_3, and c_4.*

This theorem statement is self-contained, and therefore includes the setup, assumptions, notation—everything. There are a few things missing, such as assumptions about the differentiability of V and a quantifier on t in the latter three inequalities. The quantifiers that *are* given are not uniformly placed: "there is a function V" precedes the inequalities while "for some positive constants" follows them. Is t_0 quantified or not? It should be. Finally, by saying the system is nonlinear, the theorem may seem to be ruling out linear systems. But of course the theorem applies to the case where $f(t, x)$ is linear in x.

Many people like theorem statements to be self-contained, with all the definitions, notation, etc. Then they don't have to hunt for these things. But the danger is that you may not see what the main point of the theorem is. The above theorem's main point is that, if the origin is locally exponentially stable, then there exists a Lyapunov function. Here's an attempt to bring that out by introducing definitions before the theorem statement:

> The theorem to follow concerns the system
>
> $$\dot{x} = f(t, x), \tag{3.1}$$
>
> where $x(t)$ is a vector in $\mathbb{R}^n$. Three **preliminary assumptions**: The equation has a unique solution for every initial time $t_0 \ge 0$ and every initial state $x(t_0)$, and the solution exists for all $t > t_0$; the origin is an equilibrium, that is, for all $t \ge 0$, $f(t, 0) = 0$; the Jacobian matrix $\partial f/\partial x$ is bounded for sufficiently small $\|x\|$, uniformly in t. Next, some notation: For a positive number r, let $B(r)$ denote the open ball in $\mathbb{R}^n$ of radius r.
>
> The theorem concerns two key concepts: local exponential stability and local Lyapunov function. Let us define those concepts. The origin is **locally exponentially stable**

if, for every t_0, there exist positive k, γ, r such that, for every $x(t_0) \in D(r)$ and every $t \geq t_0$,

$$\|x(t)\| \leq k\|x(t_0)\|\mathrm{e}^{-\gamma(t-t_0)}.$$

A **Lyapunov function** for system (3.1) is a continuously differentiable function $V : [0, \infty) \times \mathbb{R}^n \longrightarrow \mathbb{R}$ having the property that there exist positive c_1, c_2, c_3 such that for all $t \geq 0$ and $x \in \mathbb{R}^n$

$$c_1\|x\|^2 \leq V(t, x) \leq c_2\|x\|^2$$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \leq -c_3\|x\|^2$$

$$\left\|\frac{\partial V}{\partial x}\right\| \leq c_4\|x\|.$$

A **local Lyapunov function** is defined similarly except that $V(t, x)$ is defined only for $\|x\|$ sufficiently small, and then the three inequalities are required to hold only for such x.

Theorem *Consider system (3.1) along with the three preliminary assumptions. If the origin is locally exponentially stable, then there exists a local Lyapunov function.*

Chapter 4

The Elements of Style

Here in the form of a list we offer some general comments about mathematical writing.

1. Write from the specific to the general.

   How does someone learn to be a control theorist? The typical route is via an undergraduate degree in electrical engineering. One begins with circuit theory in second year, then signals and systems, preferably with complex variables, followed by classical control and then state space theory. The route thus begins with a specific system—the electric circuit—and then generalizes. The curriculum would be disastrous in the reverse order: General mathematical systems followed by, say, linear time-invariant systems, and finally electric circuits as an example. For the same reason one should write from the specific to the general. Give the examples before the theorem statement and proof. Indeed, a well-designed example is probably more revealing than the full theorem statement.

2. Choose notation carefully and be consistent.

   In general, your notation should be simple and intuitive. Matrices are usually denoted by capital Roman letters; vectors by lower case Roman. There's no need to write a vector as $\vec{x}$ unless it's the usual convention, as in physics. Don't change alphabets for no reason: $x(t)$ and $\omega(t)$ shouldn't be the same sort of object. Otherwise, the reader has a harder time because the meanings of symbols are not transparent; the reader loses confidence in the writer.

   Two examples illustrate entrenched but potentially confusing notation. The first is **convolution**. In continuous time, this is usually written $y(t) = g(t) * u(t)$, which stands for

   $$y(t) = \int_{-\infty}^{\infty} g(t-\tau)u(\tau)d\tau.$$

   The right-hand side shows an integral operator, say C for c̲onvolution, acting on the function u and producing the function y: $y = Cu$. If we want to show the independent variable t, we should write $y(t) = (Cu)(t)$, which says that y at a specific time t depends on u at all times, not just t. But the notation $y(t) = g(t) * u(t)$ suggests that $y(t) = C(u(t))$, which is not the case. So the correct symbolic form of the convolution equation is $y(t) = (g * u)(t)$. Nevertheless, the notation $y(t) = g(t) * u(t)$ is entrenched, with the consequence that many students don't really get convolution.

Another example along the same lines is that of the Lie derivative. Consider a differential equation

$$\dot{x} = f(x),$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, and let h be a function $\mathbb{R}^n \longrightarrow \mathbb{R}^m$. Consider the derivative

$$\frac{d}{dt}h(x(t)),$$

where $x(t)$ satisfies the differential equation. The chain rule and standard notation yield

$$\frac{d}{dt}h(x(t)) = \frac{\partial h}{\partial x}(x(t))\frac{d}{dt}x(t) = \frac{\partial h}{\partial x}(x(t))f(x(t)).$$

The **Lie derivative** of h with respect to f is the rightmost expression,

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad g(x) = \frac{\partial h}{\partial x}(x)f(x).$$

The notation is $g = L_f h$, or if we want to display the independent variable, $g(x) = (L_f h)(x)$ This is commonly written $g(x) = L_f h(x)$, which may suggest that L_f acts on the vector $h(x)$, whereas it acts on the function h. Thus L_f is a linear operator that maps the function h to the function g. We observe that $(L_f h)(x)$ equals the directional derivative of h at the point x in the direction of $f(x)$:

$$\frac{\partial h}{\partial x}(x)f(x) = \left.\frac{d}{d\varepsilon}h(x + \varepsilon f(x))\right|_{\varepsilon=0}.$$

3. Write your definitions in causal format.

   Here's a noncausal definition:

   > Let $\{L_n\}$ be a sequence, where L_n belongs to $\mathrm{Hom}(\mathcal{X}, \mathcal{Y})$, where Hom stands for homomorphism, and where $\mathcal{X}$ and $\mathcal{Y}$ are vector spaces.

   This is annoying because the reader doesn't fully comprehend until the end of the sentence. Here's the causal form, in which the reader understands from the beginning:

   > Let $\mathcal{X}$ and $\mathcal{Y}$ be vector spaces and let $\mathrm{Hom}(\mathcal{X}, \mathcal{Y})$ denote the space of homomorphisms from $\mathcal{X}$ to $\mathcal{Y}$. For each integer n, let L_n belong to $\mathrm{Hom}(\mathcal{X}, \mathcal{Y})$.

4. Integrate mathematical phrases into the text so that it flows well.

   It is possible to write so that every mathematical phrase grammatically is a noun. Example:

   > In general $Ax = b$ is solvable iff b belongs to the column span of A.

   In this sentence, the phrase "$Ax = b$" is a noun, the subject of the verb "is." Likewise, "b" is the subject of the verb "belongs" and "A" is the object of the preposition "of." The equal sign is not treated as a verb. Alternatively, we could write the sentence in this form:

   > In general there exists an x such that $Ax = b$ iff b belongs to the column span of A.

In this case "Ax" is a noun, "$=$" a verb, and "b" another noun. Or we could even use fewer words but more symbols:

> In general there exists an x such that $Ax = b$ iff $b \in \text{Colspan } A$.

Clearly, "$\in$" is a verb. It is common and quite appropriate to use mathematical symbols in this way, especially to use the equal sign as a verb.

5. Punctuate displayed statements consistently.

   Here's some text with a displayed mathematical phrase:

   > The main result on the solvability of linear equations is
   > $$(\exists x)\ Ax = b \quad \Longleftrightarrow \quad \text{rank} \begin{bmatrix} A & b \end{bmatrix} = \text{rank } A.$$

   The displayed statement ends the sentence, hence the period. Two consecutive equations require only one period. Example:

   > We consider the usual state model:
   > $$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du. \end{aligned}$$

   A comma at the end of the first sentence is unnecessary.

6. Avoid dangling participles.

   It is so convenient in mathematical writing to use this construction: Substituting (3) into (1), the state equation becomes .... But it is grammatically wrong. Who is doing the substituting? Presumably the writer or the reader. But as the sentence is written, the only noun that could be doing the substituting is "state equation," and that's not what the writer wants to say. There are two possible fixes: Make "substituting" a noun: Substituting (3) into (1) results in the state equation ...; or put in the proper noun doing the substituting: Substituting (3) into (1), we arrive at the state equation ....

7. Don't put definitions or new notation in theorem statements.

   A theorem statement should be as succinct as possible. Put the preamble before the theorem statement so that the result itself stands out clearly.

8. Don't start a sentence with a mathematical symbol.

   It may conflict with the convention that sentences begin with capital letters.

9. Don't use a colon where it doesn't belong.

   You wouldn't write "My favourite movie is: The Bourne Identity." So don't write

   > The main result on the solvability of linear equations is:
   > $$(\exists x)\ Ax = b \quad \Longleftrightarrow \quad \text{rank} \begin{bmatrix} A & b \end{bmatrix} = \text{rank } A.$$

   But this is correct:

The main result on the solvability of linear equations is the following:

$$(\exists x)\ Ax = b \iff \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} = \operatorname{rank} A.$$

10. Use "that" and no comma before a restrictive clause, and "which" and a comma before a nonrestrictive clause.

    Examples:

    > The condition $K \neq 0$ is ruled out by our initial assumption, which we made only because we were going to divide by K.
    >
    > If a step, which is bounded, is applied to a stable system, the output is bounded too.
    >
    > If an input that is bounded is applied to a stable system, the output is bounded too.

11. Abbreviations of Latin phrases must be punctuated correctly.

    The Latin equivalent of "that is" is "id est." Its abbreviation therefore has a period after each letter. It is always followed by a comma. Example:

    > Assume $x = 0$ is an equilibrium point of the equation $\dot{x} = f(x)$, i.e., $f(0) = 0$.

    The Latin equivalent of "for example" is "exempli gratia." Its abbreviation therefore has a period after each letter. It is always followed by a comma. Example:

    > A convergent sequence is bounded, but the converse is not necessarily true, e.g., $x_k = (-1)^k$.

    The Latin phrase "et alii" means "and others." The phrase "et al." is the abbreviation. Notice that there is only one period.

12. The word "only" is a delimiter and should be placed immediately before that which it is delimiting.

    Examples:

    > She only walked to the store to buy milk. (Instead of also skipping or running, she only walked.)
    >
    > She walked only to the store to buy milk. (Instead of going also to the post office, she went only to the store.)
    >
    > She walked to the store only to buy milk. (Instead of also wanting to talk to the owner or meet a friend, she went only to buy milk.)
    >
    > She walked to the store to buy only milk. (Instead of also wanting to buy other things.)
    >
    > Wrong: A local Lyapunov function $V(x)$ is only defined for $\|x\|$ sufficiently small.
    >
    > Right: A local Lyapunov function $V(x)$ is defined only for $\|x\|$ sufficiently small.

13. Avoid the comma splice, where two independent clauses are separated merely by a comma.

14. Avoid using the word "any" in place of "every."

    Suppose someone asked you to prove that any polynomial function is continuous. You might reply, $f(x) = 1$ is a polynomial and it is continuous. But you couldn't be so flippant if you were asked to prove that **every** polynomial function is continuous.

15. Don't use "if and only if" in a definition.

    Example:

    > Write: A function f from a set X to a set Y is *injective* if the equation $f(x_1) = f(x_2)$ implies $x_1 = x_2$.
    >
    > Don't write: A function f from a set X to a set Y is *injective* if and only if the equation $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

    The phrase "if and only if" is reserved for logical equivalence and should not be used for equality by definition.

Chapter 5

Giving a Talk Using Computer Slides

The best talks I've ever seen were Steve Job's talks at the WWDC—he was a master.

1. The obvious point, which almost everyone forgets.

   When you give a talk using computer slides, you are talking while showing slides. You are speaking words while showing words, mathematics, images, movies, etc. If you are speaking words while showing words, what's going to happen? If you're speaking the same words as you're showing, there's no interference and the audience can follow you. But if the words you're speaking are not the same as the words you're showing, there is interference and the audience has to eliminate it either by not listening to you or by not reading the displayed words. This is such an obvious point, and yet hardly anyone makes slides accordingly. One solution would be to have no text on your slides: You would speak your presentation and illustrate it with the slides. That is, on the slides would go diagrams, pictures, movies, and mathematics—no text except for titles. Then there would be no interference. But people seem to want to write text on their slides. In that case, to avoid interference, **you must read the text to the audience**. At first you may feel awkward doing this, but, believe me, your audience will appreciate it. If there is text on your slides that you are not going to read, get rid of it.

2. Coordinate the slides and what you're saying.

   You can't prepare beautiful slides and then afterwards think about what you're going to say. Your slides should complement your speech.

3. Remove from your slides everything that is unnecessary.

   When you have finished preparing your talk, go over your slides and remove everything that the audience does not need to see. For example, your university logo and banners. And while you're at it, do you really need background colour on the pages and animated transitions? Aren't these distractions from your message?

4. Say as early as possible in your talk what problem you're going to present.

   Get the audience interested in your problem as soon as possible. Don't begin with boring generalities, literature review, etc. Say what your problem is so that the audience can fix on something specific. Do the generalities later—they won't be so boring.

5. If your talk has theorems, do the examples first.

6. Use a pointer.

   The audience must know at all times where you are on a slide. If your slide has several things on it, you must point to what part you're talking about. You can use the mouse/pointer, a laser-pointer, or a real pointer. The latter two require you to turn away from the audience, which is somewhat rude.

7. If your talk is at a conference, the purpose of your talk is to get people to read your paper in the proceedings.

   Thus, you don't need to give all your results.