J. Fixed Point Theory Appl. 11 (2012) 43–63
DOI 10.1007/s11784-012-0071-6
Published online March 6, 2012

Journal of Fixed Point Theory and Applications

How to write a 21st century proof

Leslie Lamport

To D. Palais

Abstract. A method of writing proofs is described that makes it harder to prove things that are not true. The method, based on hierarchical structuring, is simple and practical. The author's twenty years of experience writing such proofs is discussed.

Mathematics Subject Classification (2010). 03B35, 03F07.

Keywords. Structured proofs, teaching proofs.

> In addition to developing the students' intuition about the beautiful concepts of analysis, it is surely equally important to persuade them that precision and rigor are neither deterrents to intuition, nor ends in themselves, but the natural medium in which to formulate and think about mathematical questions.
>
> Michael Spivak, *Calculus* [7]

1. Introduction

Some 20 years ago, I published an article titled *How to Write a Proof* in a festschrift in honor of the 60th birthday of Richard Palais [5]. In celebration of his 80th birthday, I am describing here what I have learned since then about writing proofs and explaining how to write them.

As I observed in the earlier article, mathematical notation has changed considerably in the last few centuries. Mathematicians no longer write formulas as prose, but they use symbolic notation such as $e^{i\pi}+1=0$. On the other hand, proofs are still written in prose pretty much the way they were in the 17th century. The proofs in Newton's *Principia* seem quite modern. This has two consequences: proofs are unnecessarily hard to understand, and they encourage sloppiness that leads to errors.

Making proofs easier to understand is easy. It requires only the simple application of two principles: structure and naming. When one reads a sentence in a prose proof, it is often unclear whether the sentence is asserting a new fact or justifying a previous assertion; and if it asserts a new fact,

one has to read further to see if that fact is supposed to be obviously true or is about to be proved. Adding structure makes it clear what the function of each assertion is. If an assertion follows from previously stated facts, in a prose proof it is often unclear what those facts are. Naming all facts makes it easy to tell the reader exactly which ones are being used.

Introducing structure has another important benefit. When writing a proof, we are continually deciding how detailed an explanation to provide the reader. Additional explanation helps the reader understand what is being shown at that point in the proof. However, in a prose proof, the additional length makes it harder to follow the logic of the complete proof. Proper structuring allows us to add as much detailed explanation as we like without obscuring the larger picture.

Eliminating errors from proofs is not as easy. It takes precision and rigor, which require work. Structuring proofs makes it possible to avoid errors; hard work is needed to make it probable.

One mistake I made in the earlier article was to advocate making proofs both easier to read and more rigorous. Learning both a new way to write proofs and how to be more precise and rigorous was too high a barrier for most mathematicians. I try here to separate the two goals. I hope that structuring proofs to be easier to read will make mathematicians more aware of how sloppy their proofs are and encourage greater precision and rigor. But the important goal is to stop writing 17^{th} century prose proofs in the 21^{st} century.

Another mistake I made was giving the impression that I know the best way of writing proofs. I claim to have a *better* way to write proofs than mathematicians now use. In the 21^{st} century, it is not hard to improve on 17^{th} century proofs. What I describe here is what I have learned in over 20 years of writing structured proofs. I am sure my way of writing proofs can be improved, and I encourage mathematicians to improve it. They will not do it by remaining stuck in the 17^{th} century.

The way I now write proofs has profited by my recent experience designing and using a formal language for machine-checked proofs. Mathematicians will not write completely formal proofs in the next 20 years. However, learning how to write a formal proof can teach how to write ordinary, informal ones. Formal proofs are therefore briefly discussed in Section 4, and the appendix contains a formal, machine-checked proof of the example introduced in Section 2.

I am writing this at the end of an era. For millenia, mathematics has been recorded on the processed remains of dead plants and animals. By the end of the 21^{st} century, mathematical books and articles will be read almost exclusively on electronic devices—or perhaps using even more advanced technology. This will enable the use of hypertext, which is the natural medium for structured proofs. However, in the next decade or two, mathematics will still be printed on paper or disseminated as static electronic images of printed pages. I therefore concentrate on how to write proofs as conventional printed text, with only brief mention of hypertext.

Corollary If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval.

Proof Let a and b be two points in the interval with $a < b$. Then there is some x in (a, b) with

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

But $f'(x) > 0$ for all x in (a, b), so

$$\frac{f(b) - f(a)}{b - a} > 0.$$

Since $b - a > 0$ it follows that $f(b) > f(a)$. ∎

FIGURE 1. Spivak's corollary and proof.

2. An example

To illustrate how to structure a proof, let us convert a simple prose proof into a structured one that is easier to read. The example I have chosen is adapted from the proof of a corollary to the Mean Value Theorem from Spivak's *Calculus* [7, page 170]; it is shown in Figure 1. (The original corollary covered both the case of increasing and decreasing functions, but Spivak proved only the increasing case; that proof is copied verbatim.) For compactness, I refer to the corollary and proof as Spivak's.

I chose this proof because it is short and simple. I would not normally structure such a simple four-sentence proof because the mathematically sophisticated readers of my papers would have no trouble understanding it. However, Spivak was writing for first-year calculus students. Moreover, seeing that even this simple proof is not so easy to understand should make it clear that structuring is needed for more complicated proofs.

We structure this proof as a sequence of named statements, each with a proof, and we adopt the conventional approach of using sequential numbers as statement names. The first numbered statement is the first sentence of the prose proof:

1. Let a and b be two points in the interval with $a < b$.

When we choose something in a proof, we have to prove that it exists. The proof of this statement should therefore prove the existence of points a and b in the interval with $a < b$. However, this can't be proved. The corollary assumes an interval, but it doesn't assume that the interval contains two different points. The book's definition of an interval allows a single point and the empty set as intervals. We can't prove that the interval contains two distinct points because it needn't.

A mathematician, examining the entire proof, will realize that we don't really have to prove that the interval contains two distinct points. However,

Corollary If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval.

1. Let a and b be two points in the interval with $a < b$.
   PROOF: ???
2. There is some x in $(a,\ b)$ with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.
   PROOF: By 1, the corollary's hypothesis, and the Mean Value Theorem.
3. $f'(x) > 0$ for all x in $(a,\ b)$.
   PROOF: By the hypothesis of the corollary and 1.
4. $\dfrac{f(b) - f(a)}{b - a} > 0$
   PROOF: By 2 and 3.
5. $f(b) > f(a)$
   PROOF: By 1, which implies $b - a > 0$, and 4.

FIGURE 2. Adding structure to Spivak's proof.

this is not obvious to a beginning calculus student. A proof is not easy to understand if we must read the entire proof to understand why its first sentence is justified. Let us just note this problem for now; we will fix it later.

The second statement of the structured proof comes from the second sentence of the prose proof:

2. There is some x in $(a,\ b)$ with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.

Spivak gives no justification for this assertion; the "Then" that begins the sentence alerts readers that it follows from facts that preceded it. The reader must deduce that those facts are: a and b are in the interval, f is differentiable on the interval, and the Mean Value Theorem. Instead of forcing the reader to deduce them, we make the proof easier to understand by stating these facts in the following proof of the statement:

> PROOF: By 1, the corollary's hypothesis, and the Mean Value Theorem.

The prose proof's third sentence, "But $f'(x) > 0 \ldots$" asserts two facts, so let's turn them into steps 3 and 4 of the structured proof. The final sentence could also be viewed as two statements, but I find it more natural to turn it into the assertion $f(b) > f(a)$ and its proof. Adding these statements and their proofs, we obtain the simple structured proof of Figure 2.

Examine the structured version. Except for the $b - a > 0$ in the proof of step 5, all the explanation in the steps' proofs is missing from the original. Adding that explanation makes the proof easier to understand. It might

seem excessive to a mathematician, who can easily fill in those missing justifications. However, a web search reveals that Spivak's text, while highly regarded for its rigor, is considered very difficult for most students. Those students would appreciate the additional explanation.

Adding the explanations missing from Spivak's proof does make the structured proof somewhat longer, occupying about 40% more vertical space in the typeset version I am now viewing. However, the extra length does not obscure the proof's structure. It is easy to ignore the subproofs and read just the five steps. I find the structure more apparent in the rewritten version than in the original, and I think others will too—especially once they get used to reading structured proofs.

The advantage of the structured version becomes much more obvious with hypertext. When books are read as hypertext, readers will first see the corollary with no proof. They will click or tap on the corollary or some adjacent icon to view the proof. With the structured version, they will then see just the five steps—essentially the same proof as Spivak's, except with the omission of "Since $b - a > 0$ it follows that". They can then view and hide the proofs of individual steps as they choose. (They can also choose to view the complete proof.)

We are not through with the proof. We cannot prove step 1. This problem is a symptom of a glaring omission in the proof. Has the reader observed that neither Spivak's proof nor its structured version provides the slightest hint of why the proof actually proves the corollary? Readers must figure that out by themselves. Leaving such an important step to the reader does not make the proof easy to understand.

This gaffe is impossible in the proof style that I propose. The style requires that the last step of every proof be a statement of what it is that the proof is trying to prove—in this case, the statement of the corollary. Instead of repeating what we're trying to prove, which has already been stated, we write Q.E.D. (that which was to be shown) instead. A structured proof of the corollary cannot end as in Figure 2, with an assertion that is not what we are trying to prove.

The reader may have surmised that what is missing from Spivak's proof is an explanation of why we should choose a and b. His first sentence, and our first step, should have been:

> If suffices to assume that a and b are two points in the interval with $a < b$ and prove $f(b) > f(a)$.

The justification of that step is that it follows from the definition of an increasing function. Step 5 can then be replaced by:

5. Q.E.D.

   PROOF: Step 1 implies $b - a > 0$, so 4 implies $f(b) - f(a) > 0$, which implies $f(b) > f(a)$. By 1, this proves the corollary.

We can further improve the proof by making more use of naming. Let's give the nameless interval in the statement of the corollary the name I. Also,

Corollary If $f'(x) > 0$ for all x in an interval I, then f is increasing on I.

1. It suffices to assume
   1. a and b are points in I
   2. $a < b$

   and prove $f(b) > f(a)$.

   PROOF: By definition of an increasing function.

2. There is some x in $(a,\ b)$ with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.

   PROOF: By assumptions 1.1 and 1.2, the hypothesis that f is differentiable on I, and the Mean Value Theorem.

3. $f'(x) > 0$ for all x in $(a,\ b)$.

   PROOF: By the hypothesis of the corollary and assumption 1.1.

4. $\dfrac{f(b) - f(a)}{b - a} > 0$.

   PROOF: By 2 and 3.

5. Q.E.D.

   PROOF: Assumption 1.2 implies $b - a > 0$, so 4 implies $f(b) - f(a) > 0$, which implies $f(b) > f(a)$. By 1, this proves the corollary.

FIGURE 3. Fixing the proof.

step 1 asserts two assumptions: that a and b are in I and that $a < b$. Let's name those assumptions so we can refer in the proof to the exact one that is being used. The result is the proof in Figure 3.

Transforming Spivak's proof to the structured proof in Figure 3 was quite simple. We could have done it differently—for example, by making $b - a > 0$ a separate step, or by removing a bit of the explanation and using this proof of step 5:

> PROOF: Assumption 1.2 implies $b - a > 0$, so 4 implies $f(b) > f(a)$. By 1, this proves the corollary.

With any sensible choices, the resulting structured proof would be easier to understand than Spivak's unstructured proof.

3. Hierarchical structure

Writing the structured proof of Figure 3 forced us to write a justification for each step. Having to do this helps catch errors. However, it is not enough to ensure error-free proofs. In fact, our structured proof contains an important omission.

The best way I know to eliminate errors is to imagine that there is a curious child sitting next to us. Every time we write an assertion, the child asks: *Why*? When we wrote the proof of step 2, the child would ask "Why does step 2 follow from the assumptions and the Mean Value Theorem?" To answer, we would point to Spivak's statement of the Mean Value Theorem.[1]

Theorem 4 (The Mean Value Theorem). *If f is continuous on $[a, b]$ and differentiable on (a, b), then there is a number x in (a, b) such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

Step 2 is identical to the conclusion of the theorem, but the child would ask why the hypotheses hold. In particular, why is f continuous on $[a, b]$? We would answer, "Because f is differentiable on I." "But why does that imply f is continuous?" We would turn back 36 pages in the book and point to Spivak's Theorem 1, which asserts that differentiability implies continuity. "You didn't tell me you needed Theorem 1." The child is right. If that result is stated as a numbered theorem, surely its use should be mentioned. The proof of step 2 should be:

> Proof: By the Mean Value Theorem and 1.2, since 1.1 and the hypothesis of the corollary imply that f is differentiable on $[a, b]$, so Theorem 1 implies it is continuous on $[a, b]$.

How much detail is necessary? For example, why do 1.1 and the hypothesis of the corollary, which asserts that f is differentiable on I, imply that f is differentiable on $[a, b]$? The proof is assuming the fact that a and b in the interval I implies that $[a, b]$ is a subset of I. Should this also be mentioned?

If you are writing the proof to show someone else that the theorem is correct, then the answer depends on the sophistication of the reader. A beginning student needs more help understanding a proof than does a mathematician.

If you are writing the proof for yourself to make sure that the theorem is correct, then the answer is simple: if the truth of a statement is not completely obvious, or if you suspect that there may be just the slightest possibility that it is not correct, then more detail is needed. When you write a proof, you believe the theorem to be true. The only way to avoid errors is to be ruthlessly suspicious of everything you believe. Otherwise, your natural desire to confirm what you already believe to be true will cause you to miss gaps in the proof; and every gap could hide an error that makes the entire result wrong.

Our proof of step 2 is a prose paragraph. As with any prose proof, every detail we add to it makes it harder to follow. In ordinary mathematical writing, the only solution to this problem would be to state and prove the step as a separate lemma. However, making each such subproof a lemma

[1]The astute reader will notice that this theorem assumes the unstated hypothesis $a < b$. When introducing the notation (a, b), the book states that $a < b$ "is almost always assumed (explicitly if one has been careful, implicitly otherwise)."

2. There is some x in $(a,\ b)$ with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.

   2.1. f is differentiable on $[a,\ b]$.
   PROOF: By 1.1, since f is differentiable on I by hypothesis.

   2.2. f is continuous on $[a,\ b]$.
   PROOF: By 2.1 and Theorem 1.

   2.3. Q.E.D.
   PROOF: By 2.1, 2.2, and the Mean Value Theorem.

FIGURE 4. An expanded proof of step 2.

would submerge the interesting results in a sea of lemmas. With structured proofs there is a simple solution: replace the paragraph with a structured proof. Figure 4 shows the result of doing this for the proof of step 2 above.

With hypertext, there is no problem adding extra detail like this. We could add enough levels to reduce the reasoning to applications of elementary axioms. The reader can stop opening lower levels of the proof when satisfied that she understands why the statement is true. With proofs on paper, extra details kill trees. But while not as convenient as hypertext, the use of indentation makes it easy for the reader to skip over details that do not interest her.

4. A Language for structured proofs

TLA$^+$ is a formal language designed for specifying and reasoning about algorithms and computer systems [3]. It includes a standard formalization of ordinary mathematics based on first-order logic and Zermelo–Fraenkel set theory. TLA$^+$ contains constructs for writing proofs that formalize the style of structured proofs that I advocate. The TLA$^+$ Toolbox is a program with a graphical interface for writing and checking TLA$^+$ specifications and proofs. It provides the type of hypertext viewer of structured proofs that I expect eventually to be commonplace.

This section describes the TLA$^+$ proof constructs that are relevant for ordinary mathematics. With one possible exception, these constructs can be written informally so that their meanings are obvious to a reader who has never before seen structured proofs. However, it is as silly to express logical proof structures in words as it is to express equations in words. When mathematicians leave the 17$^{\text{th}}$ century and begin writing structured proofs, I trust that they will adopt compact notation to replace phrases like "We now consider the case in which."

4.1. The proof steps of TLA$^+$

Simple assertions. The most common type of proof step is a simple assertion. Steps 2–5 of the proof in Figure 3 are such assertions. A simple assertion is

a mathematical formula. For example, the statement of the corollary can be written in TLA$^+$ as

$$\begin{aligned} &\forall f : \forall I \in \mathit{SetOfIntervals} : \\ &\qquad (\forall x \in I : d(f)[x] > 0) \;\Rightarrow\; \mathit{IsIncreasingOn}(f,\, I) \end{aligned} \tag{1}$$

where *SetOfIntervals* is defined to be the set of all intervals of real numbers, $d(f)$ is the derivative[2] of f, square brackets are used for function application, $\Rightarrow$ is logical implication, and *IsIncreasingOn* is defined by

$$\mathit{IsIncreasingOn}(f,\, I) \;\triangleq\; \forall\, x,\, y \in I : (x < y) \Rightarrow (f[x] < f[y])$$

If asked to formalize the statement of the corollary, most mathematicians would probably write something like (1), except with unimportant notational differences.

A Q.E.D. step is a simple assertion of the formula that is the proof's current goal.

ASSUME/PROVE. The statement of the corollary is actually not a simple assertion. We can replace f and I by other variables without changing the meaning of the formula (1). However, if we changed f and I just in the corollary's statement, then Spivak's proof would make no sense because the variables f and I that appear in it would be meaningless. The statement of the corollary can be expressed in TLA$^+$ as this ASSUME/PROVE statement:

ASSUME NEW f, NEW $I \in \mathit{SetOfIntervals}$, $\forall x \in I : d(f)[x] > 0$
PROVE $\mathit{IsIncreasingOn}(f,\, I)$

This ASSUME/PROVE asserts the truth of formula (1). It also declares the goal of the corollary's proof to be the PROVE formula and allows the assumptions in the ASSUME clause to be assumed in the proof. The assumption NEW x introduces a new variable x, implicitly asserting that the conclusion is true for all values of x. The assumption NEW $x \in S$ is equivalent to the two assumptions NEW x, $x \in S$.

An ASSUME/PROVE can appear as a proof step as well as the statement of the theorem. The PROVE formula is the goal of the step's proof, and the ASSUME clause's assumptions can be used only in that proof. We can consider an ordinary assertion to be an ASSUME/PROVE with an empty ASSUME clause.

It is not obvious how to write an ASSUME/PROVE step in ordinary mathematical prose. When an assumption is introduced in a prose proof, it is assumed to hold until some unspecified later point in the proof. (One reason prose proofs are hard to understand is that it can be difficult to figure out the scope of an assumption.) As in the statement of the corollary, an ordinary prose assertion is interpreted as an ASSUME/PROVE when it is the statement of the theorem to be proved. If it appears as a proof step, we can try indicating the ASSUME/PROVE structure by writing "If we assume ..., then we can prove ...". That and the hierarchical structuring may be good enough to convey the intended meaning. However, it seems safer to introduce ASSUME

[2]For reasons irrelevant to ordinary mathematics, prime ($'$) has a special meaning in TLA$^+$.

and PROVE as keywords and explain their meaning to the reader. Fortunately, ASSUME/PROVE steps are not common in informal proofs.

SUFFICES. Step 1 in Figure 3 can be written in TLA$^+$ as this SUFFICES step:

SUFFICES ASSUME NEW $a \in I$, NEW $b \in I$, $a < b$
PROVE $f[b] > f[a]$

This step asserts the truth of the formula

$$(\forall\, a, b \in I : (a < b) \Rightarrow (f[b] > f[a])) \Rightarrow \mathit{IsIncreasingOn}(f, I) \qquad (2)$$

where the hypothesis of the implication (2) is the assertion of the ASSUME/PROVE, and the conclusion $\mathit{IsIncreasingOn}(f, I)$ of (2) is the proof's current goal. The step changes the current goal of the proof to be the PROVE formula $f[b] > f[a]$; and it allows the assumptions $a \in I$, $b \in I$, and $a < b$ of the ASSUME clause to be assumed in the rest of the proof. These assumptions do not apply to the proof of the step itself, and the NEW variables a and b are meaningless in that proof.

A proof by contradiction of a simple assertion F begins with the step

SUFFICES ASSUME $\neg F$
PROVE FALSE

A SUFFICES step can also have the form SUFFICES F for a formula F. Since a formula is an ASSUME/PROVE with no assumptions, this step asserts that F implies the proof's current goal, and the step's proof must prove this assertion. The step changes the current goal to F.

The SUFFICES construct is not necessary; any SUFFICES can be eliminated by restructuring the proof. For example, we could eliminate the "It suffices to" from the proof of Figure 3 by using the following top-level structure:

1. ASSUME 1. a and b are points in I.
   2. $a < b$
   PROVE $f[b] > f[a]$
2. Q.E.D.
   PROOF: By 1 and definition of an increasing function.

The proof of the original step 1 appears in the proof of the new Q.E.D step; steps 2–5 of the original proof become the proof of the new step 1.

As this example illustrates, removing a SUFFICES adds one level to the proof. Proofs are generally easiest to read if each level contains about 4 to 10 steps. The SUFFICES step eliminates the kind of two-step proof that would otherwise be needed to prove the corollary.

Step 1 of Figure 3 shows that a SUFFICES is easily expressed with informal prose, even if it is a SUFFICES ASSUME/PROVE.

PICK. If we were to expand the proof of step 4 of Figure 3, it would look like this:

4.1. Pick x in (a, b) with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.

PROOF: Such an x exists by step 2.

4.2. Q.E.D.

PROOF: By 4.1 and 3.

Step 4.1 is expressed formally by a step of the form

$$\text{PICK } x \in S : P(x)$$

This step introduces the new variable x and asserts that $P(x)$ is true. The step's proof must show that there exists an x satisfying $P(x)$.

There is no problem expressing PICK in prose.

CASE. If F is a formula, the step CASE F is an abbreviation for

ASSUME F
PROVE Q.E.D.

where Q.E.D. stands for the formula that is the current goal. An ordinary proof by cases ends with a sequence of CASE steps followed by the Q.E.D. step. Here is a typical example:

1. PICK $n \in S : \ldots$
2. CASE $n \geq 0$
3. CASE $n < 0$
4. Q.E.D.

   PROOF: By 1, 2, and 3.

In general, the Q.E.D. step's proof cites the CASE statements and shows that the cases are exhaustive—which in this example is presumed to be trivial because S is a set of numbers.

It is easy to express a CASE statement in prose—for example:

We now consider the case in which $n \geq 0$.

However, why not just write CASE $n \geq 0$? Readers will understand what it means.

Definitions. It is often convenient to give some expression a name in part of the proof. TLA$^+$ allows definitions as proof steps. A single step can contain multiple definitions. The definition is in effect for the rest of the proof's current level.

4.2. Hierarchical numbering

Figure 4 introduced a naming scheme in which, for example, 3.3.4 is the 4th step in the proof of step 3.3. The most obvious problem with this scheme is that step names get longer as the proof gets deeper. For steps beyond level 4, the names are too hard to read and take up too much space.

A less obvious problem is that any step can be mentioned anywhere in the proof. One can refer to step 2.4.1 from inside the proof of step 3.3.4. Such a reference should not be allowed because it violates the hierarchical structuring of the proof. It is illegal to use step 2.4.1 in the proof of step 3.3.4 if step 2 or step 2.4 is an ASSUME/PROVE, because step 2.4.1 has then been proved under assumptions that may not hold for the proof of 3.3.4. Even

if there is no ASSUME/PROVE making the reference illegal, such a reference makes the proof harder to read. The proof of step 3.3.4 should refer only to the following steps: 1, 2, 3.1, 3.2, 3.3.1, 3.3.2, and 3.3.3.

In TLA$^+$, the 4$^{\text{th}}$ step of a level 3 proof is named $\langle 3\rangle 4$. A proof can have many different steps named $\langle 3\rangle 4$. However, at any point in the proof, only one of them can be referred to without violating the hierarchical structure. (Proving this is a nice little exercise.) Any valid reference to step $\langle 3\rangle 4$ refers to the most recent preceding step with that name. This numbering scheme is used in the TLA$^+$ proof of Spivak's corollary in the appendix.

4.3. Equational proofs

A different kind of structuring is provided by equational reasoning. An equational proof consists of a sequence of relations and their proofs, from which one deduces a relation by transitivity—for example, we prove

$$a + 1 \;\leq\; b^2 - 3 \;=\; \sqrt{c - 42} \;\leq\; d \tag{3}$$

and deduce $a + 1 \leq d$. I find such proofs to be elegant, and I use them when I can. Figure 5 shows a more sophisticated type of equational reasoning using set inclusion; it is a slightly modified version of a proof from a paper of which I was an author [1].

Equational proofs have a serious drawback when printed on paper: there seems to be no good way to display hierarchical proofs of the individual relations. On paper, equational reasoning works well only as a lowest-level proof, where the proof of each relation is very short—as in Figure 5. TLA$^+$ provides the following way of writing the sequence of relations (3) in a proof:

$\langle 1\rangle 1.\;\; a + 1 \;\leq\; b^2 - 3$
$\langle 1\rangle 2.\;\; @ \;=\; \sqrt{c - 42}$
$\langle 1\rangle 3.\;\; @ \;\leq\; d$

where the @ symbol stands for the preceding expression. However, this lacks the visual simplicity of (3). There is no problem displaying hierarchically structured equational proofs with hypertext.

4.4. Comments

Like any computer-readable language, TLA$^+$ allows comments to aid human readers. With hypertext, comments can be attached to any part of a proof, to be popped up and hidden as the reader wishes. On paper, comments in arbitrary places can be distracting. However, there is one form of comment

PROOF: Π	$= \mathcal{C}(\Pi \wedge L_1)$	[Hypothesis 1]
	$\subseteq \mathcal{C}(\Pi \wedge L_2)$	[Hypothesis 2 and monotonicity of closure]
	$\subseteq \mathcal{C}(\Pi)$	[monotonicity of closure]
	$= \Pi$	[Hypothesis 1, which implies Π closed]
This proves that $\Pi = \mathcal{C}(\Pi \wedge L_2)$.		

FIGURE 5. An example of equational reasoning.

that works well on paper for both formal and informal proofs: a proof sketch that comes between a statement and its proof. The sketch can explain the intuition behind the proof and can point out the key steps. Proof sketches can be used at any level in a hierarchical proof, including before the highest-level proof. A reader not interested in the details of that part of the proof can read just the proof sketch and skip the steps and their proofs.

4.5. Completely formal proofs

In principle, the proof of a theorem should show that the theorem can be formally deduced from axioms by the application of proof rules. In practice, we never carry a proof down to that level of detail. However, a mathematician should always be able to keep answering the question *why?* about a proof, all the way down to the level of axioms. A completely formal proof is the Platonic ideal.

Most mathematicians have no idea how easy it is to formalize mathematics. Their image of formalism is the incomprehensible sequences of symbols found in *Principia Mathematica*. The appendix contains formal TLA$^+$ definitions of intervals, limits, continuity, and the derivative, assuming only the definitions of the real numbers and of ordinary arithmetic operations. I expect most readers will be surprised to learn that this takes only 19 lines. The appendix also contains a TLA$^+$ proof of Spivak's corollary that has been checked by the TLAPS proof system [6].

Formalizing mathematics is easy, but writing formal, machine-checkable proofs is not. It will be decades before mechanical proof checkers are good enough that writing a machine-checked proof is no harder than writing a careful informal proof. Until then, there is little reason for a mathematician to write formal mathematics.

However, there is good reason for teaching how to write a formal proof as part of a standard mathematics education. Mathematicians think that the logic of the proofs they write is completely obvious, but our examination of Spivak's proof shows that they are wrong. Students are expected to learn how to write logically correct proofs from examples that, when read literally, are illogical. (Recall the first sentence of Spivak's proof.) It is little wonder that so few of them succeed. Learning to write structured formal proofs that a computer can check will teach students what a proof is. Going from these formal proofs to structured informal proofs would then be a natural step.

Is it crazy to think that students who cannot learn to write proofs in prose can learn to write them in an unfamiliar formal language and get a computer to check them? Anyone who finds it crazy should consider how many students learn to write programs in unfamiliar formal languages and get a computer to execute them, and how few now learn to write proofs.

For reasons mentioned in the appendix, TLA$^+$ and its TLAPS prover are not ideal for teaching mathematics students. However, the structured proofs of TLA$^+$ make it the best currently available language for the task that I know of. It should be satisfactory for writing proofs in some particular domain such as elementary group theory.

5. Experience

I am a computer scientist who was educated as a mathematician. I discovered structured proofs through my work on concurrent (multiprocess) algorithms. These algorithms can be quite subtle and hard to get right; their correctness proofs require a degree of precision and rigor unknown to most mathematicians (and many computer scientists). A missing hypothesis, such as that a set must be nonempty, which is a trivial omission in a mathematical theorem, can mean a serious bug in an algorithm.

Proofs of algorithms are most often mathematically shallow but complicated, requiring many details to be checked. With traditional prose proofs, I found it impossible to make sure that I had not simply forgotten to check some detail. Computer science offers a standard way to handle complexity: hierarchical structure. Structured proofs were therefore an obvious solution. They worked so well for proofs of algorithms that I tried them on the more mathematical proofs that I write. I have used them for almost every proof of more than about ten lines that I have published since 1991. (The only exceptions I can find are in a paper in which the proofs served only to illustrate how certain formal proof rules are used.)

My earlier paper on structured proofs described how effective they are at catching errors. It recounted how only by writing such a proof was I able to rediscover an error in a proof of the Schroeder–Bernstein theorem in a well-known topology text [2, page 28]. I recently received email from a mathematician saying that he had tried unsuccessfully to find that error by writing a structured proof. I asked him to send me his proof, and he responded:

> I tried typing up the proof that I'd hand-written, and in the process, I think I've found the fundamental error ... I now really begin to understand what you mean about the power of this method, even if it did take me hours to get to this point!

It is instructive that, to find the error, he had to rewrite his proof to be read by someone else. Eliminating errors requires care. Structured proofs make it possible, not inevitable.

Over the years, I have published quite a few papers with structured proofs. For proofs with the level of detail typical of mathematics papers, I cannot remember any reader commenting on the proof style unless explicitly asked to. Structured proofs are easy enough to read that one forgets about the form and concentrates on the content. However, I have also published some papers with long, excruciatingly detailed proofs. I am reluctant to publish a paper with a short proof if I would not have been able to find the correct result without writing a longer one. Here are a referee's comments on one such proof.

> The proofs ... are lengthy, and are presented in a style which I find very tedious. I think the readers ... are going to be more interested in understanding the techniques and how they can apply them, than they will be in reading the formal proofs. A problem with the proofs

> is that they do not clearly distinguish the trivial manipulations from the nontrivial or surprising parts. ... My feeling is that informal proof sketches ... to explain the crucial ideas in each result would be more appropriate.

I think the referee would have found the proofs much more tedious if written in a conventional prose style, but my coauthor and I could have made the proof easier to read had we used the proof sketches described in Section 4.4 above. I do agree that the proofs were too long and detailed for the journal's readers. Today, the obvious approach would be to put the long proof on the Web and publish a proof sketch with a link to the real proof. The Web was in its infancy when the paper was published, but the editor agreed to publish the proofs as a separate appendix available only online. None of the first reviewers had read the proofs, so the editor found another referee to do that. When asked how he or she found the proof style, the referee responded:

> I have found the hierarchical structuring of proofs to be very helpful, if read top-down according to the suggestions of the authors. In fact, it might well be the only way to present long proofs ... in a way that is both detailed (to ensure correctness) and readable. For long proofs, I think that describing the idea of the proof in a few words at the beginning (if appropriate) would help make them more understandable. ... But in general, I found the structured approach very effective.

6. Objections to structured proof

When lecturing about structured proofs, I have heard many objections to them. I cannot recall any objection that I found to be based on a rational argument; they have all been essentially emotional. Here are three common ones.

They are too complicated.

In lectures, I usually flash on the screen one of my multipage structured proofs. People have reacted by saying that the structuring makes the proof too complicated, as if replacing the numbering and indentation by prose would magically simplify the proof. One mathematician described how she had explained a beautiful little proof by Hardy to an audience of nonmathematicians, and that the audience could not possibly understand my proofs. She apparently believed that structuring Hardy's tiny proof would turn it into multiple pages full of obscure symbols.

It is an unfortunate fact that being rigorous requires filling in missing details, which makes a proof longer. As we saw with Spivak's corollary, a structured proof makes it easier to see what is missing; this would lead mathematicians to correct the omissions, resulting in longer proofs. Fortunately, structuring allows us to add those details without making the proof any harder to read as hypertext, and only a little harder to read on paper.

They don't explain why the proof works.

Mathematicians seem to think that their proofs explain themselves. I cannot remember reading a mathematician's proof that was both a proof and an explanation of why the proof works. It is hard enough to make the structure of a prose proof clear; doing it while also providing an intuitive explanation is a formidable task.

I suspect that this objection is based on confusing a proof sketch with a proof. Proof sketches are fine, but they are not proofs. Mathematicians sometimes precede a proof with a proof sketch, but there is no easy way to relate the steps of the proof with the proof sketch. The ability to add proof sketches at any level of a structured proof makes it possible to provide a much clearer explanation of why a proof works.

A proof should be great literature.

This is nonsense. A proof should not be great literature; it should be beautiful mathematics. Its beauty lies in its logical structure, not in its prose. Proofs are more like architecture than like literature, and architects do not use prose to design buildings. Prose cannot add to the beauty of $e^{i\pi} + 1 = 0$, and it is a poor medium for expressing the beauty of a proof.

7. Beginning

When I started writing structured proofs, I quickly found them to be completely natural. Writing nontrivial prose proofs now seems as archaic to me as writing

> The number e raised to the power of i times π, when added to 1, equals 0.

Imagine how many errors we would make performing algebraic calculations with equations written in prose. Writing proofs in prose is equally error prone. As I reported in my earlier paper, anecdotal evidence indicates that a significant fraction of published mathematics contains serious errors. This will not change until mathematicians understand that precision and rigor, not prose, are the natural medium of mathematics, and they stop writing 17^{th} century proofs.

Fortunately, it is not hard to write 21^{st} century proofs. There is no need to wait until other mathematicians are doing it. You can begin by just adding structure to an existing proof, as we did with Spivak's proof. Start by rewriting a simple proof and then try longer ones. You should soon find this a much more logical way to write your proofs, and readers will have no trouble understanding the proof style.

Writing structured proofs is liberating. It allows you to concentrate on logical structure instead of sentence structure. You will no longer waste your time searching for different ways to say *therefore*. To help you typeset your structured proofs, a LaTeX package and an associated computer program are available on the Web [4].

Appendix: A formal proof

This appendix contains a TLA$^+$ formalization of Spivak's proof, preceded by formal definitions of the necessary concepts of differential calculus and statements of the two theorems used in that proof. This is all done in a TLA$^+$ module named *Calculus*.

Mathematicians will be struck by the absence of some common mathematical notation. For example, the open interval (a, b) is written $\mathit{OpenInterval}(a, b)$. No single syntax can capture the wide variety of notation used in mathematics, where (a, b) may be an interval or an ordered pair, depending on the context. A good formal language for math should permit such context-dependent notation. TLA$^+$ was not designed for use by mathematicians, so it provides conventional notation mainly for basic operators of logic and set theory.

TLAPS uses sophisticated algorithms for performing simple reasoning about integers. However, because it is still under development and real numbers are seldom used in algorithms, TLAPS does not yet provide any special support for reasoning about reals. The proof of Spivak's corollary therefore assumes without proof five very simple facts about real numbers.

The appendix shows the typeset version of the *Calculus* module, not its actual ASCII text. For example, the definition of *OpenInterval* that is typeset as

$$\mathit{OpenInterval}(a, b) \triangleq \{r \in \mathit{Real} : (a < r) \land (r < b)\}$$

appears in the module as

```
OpenInterval(a, b) == {r \in Real : (a < r) /\ (r < b)}
```

The LaTeX source for the typeset version was generated from the ASCII by a program, but it is possible that editing of the document introduced errors.

The rest of the appendix consists of the *Calculus* module together with interspersed comments explaining some of the TLA$^+$ notation. These explanations and a thorough knowledge of calculus should allow you to figure out what's going on. However, you probably won't find it easy reading. This is hardly surprising, since the definition of the derivative, which is line 20 of the module, appears on page 127 of Spivak's book. With a few pages of explanation, I expect a mathematician would find the TLA$^+$ formulas as easy to read as Spivak's text.

MODULE *Calculus*

The following EXTENDS statement imports the standard *Reals* module that defines the set *Real* of real numbers and the usual operators on real numbers such as / (division) and $\leq$.

EXTENDS *Reals*

$\mathit{OpenInterval}(a, b) \triangleq \{r \in \mathit{Real} : (a < r) \land (r < b)\}$

$\mathit{ClosedInterval}(a, b) \triangleq \{r \in \mathit{Real} : (a \leq r) \land (r \leq b)\}$

In TLA$^+$, SUBSET S is the power set (the set of all subsets) of S.

$$SetOfIntervals \triangleq \{S \in \text{SUBSET } Real : \forall\, x,\, y \in S : OpenInterval(x,\, y) \subseteq S\}$$

UNION S is the union of all sets that are elements of the set S, and $[S \to T]$ is the set of functions with domain S and range a subset of T. Hence, the following defines *RealFunction* to be the set of all real-valued functions whose domain is a set of real numbers.

$$RealFunction \triangleq \text{UNION } \{[S \to Real] : S \in \text{SUBSET } Real\}$$
$$AbsoluteValue(a) \triangleq \text{IF } a > 0 \text{ THEN } a \text{ ELSE } -a$$
$$OpenBall(a,\, e) \triangleq \{x \in Real : e > AbsoluteValue(x - a)\}$$
$$PositiveReal \triangleq \{r \in Real : r > 0\}$$

TLA$^+$ uses square brackets for function application, as in $f[x]$. Parentheses are reserved for arguments of defined operators like *OpenInterval*.

$$\begin{aligned} IsLimitAt(f,\, a,\, b) \triangleq\ & (b \in Real) \land \\ & \forall\, e \in PositiveReal : \exists\, d \in PositiveReal : \\ & \quad \forall\, x \in OpenBall(a,\, d) \setminus \{a\} : f[x] \in OpenBall(b,\, e) \end{aligned}$$
$$IsContinuousAt(f,\, a) \triangleq IsLimitAt(f,\, a,\, f[a])$$
$$IsContinuousOn(f,\, S) \triangleq \forall\, x \in S : IsContinuousAt(f,\, x)$$

Mathematics provides no formal notation for writing a function. In TLA$^+$, $[x \in S \mapsto e(x)]$ is the function with domain S that maps x to $e(x)$ for every x in S. If f is a function, then DOMAIN f is its domain.

$$\begin{aligned} IsDerivativeAt(f,\, a,\, b) \triangleq\ & \\ & \exists\, e \in PositiveReal : \\ & \quad (OpenBall(a,\, e) \subseteq \text{DOMAIN } f) \land \\ & \qquad IsLimitAt([x \in OpenBall(a,\, e) \setminus \{a\} \mapsto (f[x] - f[a])/(x - a)],\, a,\, b) \end{aligned}$$
$$IsDifferentiableAt(f,\, a) \triangleq \exists\, b \in Real : IsDerivativeAt(f,\, a,\, b)$$
$$IsDifferentiableOn(f,\, S) \triangleq \forall\, x \in S : IsDifferentiableAt(f,\, x)$$

CHOOSE $x : P(x)$ is an arbitrary value x satisfying $P(x)$, if such a value exists; otherwise its value is unspecified. The CHOOSE operator is known to logicians as Hilbert's ε.

$$d(f) \triangleq [x \in \text{DOMAIN } f \mapsto \text{CHOOSE } y : IsDerivativeAt(f,\, x,\, y)]$$

The following THEOREM asserts the truth of the formula $\forall f : \ldots$ and names that formula *Theorem*1.

$$\begin{aligned} \text{THEOREM } Theorem1 \triangleq\ & \forall\, f \in RealFunction : \forall\, a \in Real : \\ & \quad IsDifferentiableAt(f,\, a) \Rightarrow IsContinuousAt(f,\, a) \end{aligned}$$

$$IsIncreasingOn(f,\, I) \triangleq \forall\, x,\, y \in I : (x < y) \Rightarrow (f[x] < f[y])$$

THEOREM $MeanValueTheorem \triangleq$
$\forall f \in RealFunction : \forall a, b \in Real :$
$(\quad (a < b)$
$\wedge IsContinuousOn(f, ClosedInterval(a, b))$
$\wedge IsDifferentiableOn(f, OpenInterval(a, b)))$
$\Rightarrow (\exists x \in OpenInterval(a, b) :$
$d(f)[x] = (f[b] - f[a])/(b - a))$

We assume without proof the following five trivial facts about real numbers. TLAPS easily proves the first four for integers.

PROPOSITION $Fact1 \triangleq \forall x \in Real : x - x = 0$
PROPOSITION $Fact2 \triangleq \forall x, y \in Real : (x \leq y) \equiv (x < y) \vee (x = y)$
PROPOSITION $Fact3 \triangleq \forall x, y \in Real : x - y \in Real$
PROPOSITION $Fact4 \triangleq \forall x, y \in Real : (x < y) \equiv (y - x > 0)$
PROPOSITION $Fact5 \triangleq \forall x, y \in Real : (y > 0) \wedge (x/y > 0) \Rightarrow (x > 0)$

Below is the corollary and its proof. Each step of the high-level proof formalizes the correspondingly-numbered step of the proof in Figure 3, where step $\langle 1 \rangle 1a$ has been added in the formal proof. The proof of step $\langle 1 \rangle 2$ formalizes the proof of Figure 4.

The lowest-level paragraphs of an informal proof are replaced with BY statements that say what facts and definitions (DEF) are used. (Assumptions in NEW declarations and in the statement of the corollary are automatically used by TLAPS.) The proof has been decomposed for the benefit of TLAPS, not for a human reader.

COROLLARY $Spivak \triangleq$ ASSUME NEW $f \in RealFunction$,
NEW $I \in SetOfIntervals$,
$IsDifferentiableOn(f, I)$,
$\forall x \in I : d(f)[x] > 0$
PROVE $IsIncreasingOn(f, I)$

$\langle 1 \rangle 1$. SUFFICES ASSUME NEW $a \in I$, NEW $b \in I$, $a < b$
PROVE $f[a] < f[b]$
BY DEF $IsIncreasingOn$

$\langle 1 \rangle 1$a. $(OpenInterval(a, b) \subseteq I)$
$\wedge (ClosedInterval(a, b) \subseteq I)$
$\wedge (OpenInterval(a, b) \subseteq ClosedInterval(a, b))$
$\langle 2 \rangle 1$. $OpenInterval(a, b) \subseteq I$
BY DEF $SetOfIntervals$
$\langle 2 \rangle 2$. $ClosedInterval(a, b) = OpenInterval(a, b) \cup \{a, b\}$
BY $\langle 1 \rangle 1$, $Fact2$ DEF $ClosedInterval$, $OpenInterval$, $SetOfIntervals$
$\langle 2 \rangle 3$. QED
BY $\langle 2 \rangle 1$, $\langle 2 \rangle 2$

$\langle 1 \rangle 2$. $\exists x \in OpenInterval(a, b) : d(f)[x] = (f[b] - f[a])/(b - a)$
$\langle 2 \rangle 1$. $IsDifferentiableOn(f, ClosedInterval(a, b))$

BY ⟨1⟩1a, ⟨1⟩1 DEF *IsDifferentiableOn*
⟨2⟩2. *IsContinuousOn*(*f*, *ClosedInterval*(*a*, *b*))
⟨3⟩1. SUFFICES ASSUME NEW $x \in \mathit{ClosedInterval}(a, b)$
PROVE $\mathit{IsContinuousAt}(f, x)$
BY DEF *IsContinuousOn*
⟨3⟩2. *IsDifferentiableAt*(*f*, *x*)
BY ⟨2⟩1 DEF *IsDifferentiableOn*
⟨3⟩3. QED
BY ⟨3⟩2, *Theorem*1 DEF *ClosedInterval*
⟨2⟩3. QED
BY ⟨1⟩1, ⟨1⟩1a, ⟨2⟩1, ⟨2⟩2, *MeanValueTheorem*
DEF *SetOfIntervals*, *IsDifferentiableOn*

⟨1⟩3. $\forall\, x \in \mathit{OpenInterval}(a, b) : d(f)[x] > 0$
BY ⟨1⟩1a

⟨1⟩4. $(f[b] - f[a])/(b - a) > 0$
⟨2⟩1. PICK $x \in \mathit{OpenInterval}(a, b) :$
$d(f)[x] = (f[b] - f[a])/(b - a)$
BY ⟨1⟩2
⟨2⟩2. $d(f)[x] > 0$
BY ⟨2⟩1 DEF *SetOfIntervals*
⟨2⟩3. QED
BY ⟨2⟩1, ⟨2⟩2

⟨1⟩5. QED
⟨2⟩1. $(a \in \mathit{Real}) \land (b \in \mathit{Real})$
BY DEF *SetOfIntervals*
⟨2⟩2. $(f[a] \in \mathit{Real}) \land (f[b] \in \mathit{Real})$
⟨3⟩1. SUFFICES ASSUME NEW $x \in \mathit{Real}$,
$\mathit{IsDifferentiableAt}(f, x)$
PROVE $f[x] \in \mathit{Real}$
BY ⟨2⟩1 DEF *IsDifferentiableOn*
⟨3⟩2. $\forall\, e \in \mathit{PositiveReal} : x \in \mathit{OpenBall}(x, e)$
BY *Fact*1 DEF *OpenBall*, *PositiveReal*, *AbsoluteValue*
⟨3⟩3. $x \in$ DOMAIN f
BY ⟨3⟩1, ⟨3⟩2 DEF *IsDifferentiableAt*, *IsDerivativeAt*
⟨3⟩4. QED
BY ⟨3⟩3 DEF *IsDifferentiableAt*, *IsDerivativeAt*, *RealFunction*
⟨2⟩3. $(f[b] - f[a] \in \mathit{Real}) \land (b - a \in \mathit{Real})$
BY ⟨2⟩1, ⟨2⟩2, *Fact*3
⟨2⟩4. $f[b] - f[a] > 0$
BY ⟨1⟩1, ⟨1⟩4, ⟨2⟩1, ⟨2⟩3, *Fact*4, *Fact*5
⟨2⟩5. QED
BY ⟨2⟩2, ⟨2⟩4, *Fact*4

Acknowledgement

In his courses on analysis and algebraic topology, Richard Palais taught me how mathematics could be made precise and rigorous, and thereby more beautiful.

References

[1] M. Abadi and L. Lamport, *An old-fashioned recipe for real time.* ACM Transactions on Programming Languages and Systems **16** (1994), no. 5, 1543–1571.

[2] J. L. Kelley, *General Topology.* The Univesity Series in Higher Mathematics, D. Van Nostrand Company, Princeton, NJ, 1995.

[3] L. Lamport, TLA—temporal logic of actions. A web page, a link to which can be found at URL http://lamport.org. The page can also be found by searching the Web for the 21-letter string formed by concatenating `uid` and `lamporttlahomepage`.

[4] L. Lamport, Useful LaTeX packages. http://research.microsoft.com/en-us/um/people/lamport/latex/latex.html. The page can also be found by searching the Web for the 23-letter string formed by concatenating `uid` and `lamportlatexpackages`.

[5] L. Lamport, *How to write a proof.* In: Global Analysis in Modern Mathematics, pp. 311–321. Publish or Perish, Houston, Texas, 1993. A symposium in honor of Richard Palais' sixtieth birthday. Also published in Amer. Math. Monthly **102** (1995), no. 7, 600–608.

[6] Microsoft Research-INRIA Joint Centre. Tools and methodologies for formal specifications and for proofs. http://www.msr-inria.inria.fr/Projects/tools-for-formal-specs.

[7] M. Spivak, *Calculus.* W. A. Benjamin, Inc., New York, 1967.

Leslie Lamport
Microsoft
1065 La Avenida
Mountain View, CA 94043
USA