

Miscellaneous Tips for Writing and Formatting

© Jeffrey Scott Vitter

College of Science
Purdue University
West Lafayette, IN 47907–2067

October 12, 2005

Abstract

This report is an ongoing compendium of my writing and formatting rules that I got tired of repeating over and over again (often to the same students!). Now that it's written in plain English, there's no longer any excuse for not following these rules *to the letter*.

In the following sections I list some useful tips and observations (read “rules” if you're one of my students) for technical writing with $\text{T}_{\text{E}}\text{X}$ and $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$. My plan is to update this report on a regular basis. I've included just a few rules so far; no doubt more will be coming.

An up-to-date version of this report is always available on my World Wide Web page <http://www.science.purdue.edu/jsv/>. Feel free to distribute this report to anyone, but only for noncommercial use. I welcome any suggestions, corrections, or comments that you might have; you can send me email at jsv@purdue.edu. Every suggestion of yours that I put into effect earns you 2.56 ITL (with payments rounded off to the nearest dollar).

1. General Writing and Organization

- ▶ Always write in a top-down manner. I highly recommend *version-by-version refinement*, which is a simple mechanism to force yourself into proper organization. Start with Version 1 as a rough outline of the main points and topics in the order of treatment. On the side make a list of all the little detailed ideas and points that have come to you in the process, but you don't yet know where they should go. (Don't put them in the outline yet; you're not ready for that stage.)

Version 2 should be an elaboration of Version 1, in which you identify the section titles and specify the main ideas of each section. Insert the ideas and points you want to make into the relevant sections, so that you know where they belong. Version 3 should be a further elaboration of Version 2 with specific subideas and a clear story line. Version 4 should be close to the final version, if not the final version itself.

- ▶ Don't start writing a paper any other way. *DON'T START WRITING A PAPER ANY OTHER WAY!!!* Even though it seems like you're doing extra or redundant work to go through the version-by-version refinement process, in reality it will save you incredible amounts of time, and the resulting paper will be organized and coherent. (Theoretically, the extra percentage of work is bounded by a geometrically decreasing series, so there's only a constant factor of extra overhead. From a practical point of view, failure to use version-by-version refinement will cost

you much more time because of rewrites, and the result will likely be a mess anyway, not to mention that I won't be happy.)

- ▶ Go for clarity and simplicity. Review your writing to streamline it. Eliminate unneeded words, phrases, and sentences.
- ▶ State the main contributions clearly in the abstract. The abstract and the introduction of the paper are very important in conference submissions. (Some referees seem to form their opinions based only on the first two pages.) Explain the problem and background, state the contributions, and explain their importance.
- ▶ Use active voice in your writing as much as possible. Passive voice is to be avoided. (Or rather, try not to use passive voice excessively.) It's good style to write in first person, using the group "we," which helps promote use of active voice, but avoid use of "I." (Ignore the fact that I'm using "I" all the time in this report. Do as I say, not as I do)
- ▶ Place the verb close to the subject. Otherwise, in long sentences, confusion will reign. Consider this example: "Any implication that the cause of the confusion that swept the city was rooted in a lack of planning and not in poor coordination is not tenable." Here's a better rewrite with the verb close to the subject: "It is untenable to assert that the confusion that swept the city resulted from a lack of planning and not from poor coordination." Even better: "The confusion that swept the city did not result solely from lack of planning; a more critical component was the breakdown in coordination."
- ▶ Here's an important pointer: The two stress points of a sentence are at its beginning and its end. The end generally gets the most emphasis, so try to reserve the end of the sentence for its most important element. Structure each sentence to move from the known concept to the unknown, from the given idea to the new, from the background to the result. Put transitional and background words or phrases (such as "Therefore," "Thus," "However," "In the last section," etc.) at the beginning of the sentence, saving the important content for the end.
- ▶ You'll probably store what you write in files on your computer. If so, name your files in a logically meaningful way. Move old files to a special directory and give them a new name that reflects the last date of alteration. Do not keep garbage files in the main directory; get rid of them or put them in a special directory (e.g., called `MiscJunk`).

2. Grammar and Style

- ▶ If ever in doubt, refer to the beginning section of the nice book by Knuth et al. on mathematical writing [1]. It gives a short and useful list of dos and don'ts worthy of memorization. And of course, be sure to read Strunk and White's classic text on the elements of style [3].
- ▶ Avoid use of "filler" words or phrases that carry no real meaning, such as
 - "actually,"
 - "basically,"
 - "essentially,"
 - "it is important to note that" (a phrase only for exceptional situations),
 - "note that,"
 - "now,"
 - "therefore,"
 - "thus."

Filler words can serve a useful purpose; an example is the use of “thus” or “therefore” to signal a conclusion. But they’re easy to overuse, so minimize their use.

- ▶ Certain words are awkward in formal papers, and there are better substitutes:
 - Replace “a lot of” with “many” or “much.”
 - Spell out contractions. (OK, I’m breaking that rule in this report too.)
 - Symbols and abbreviations should be reserved for mathematical formulas. Avoid their use within text. Instead, spell out the corresponding words. For example, replace “ \exists ” by “there exists,” “ \forall ” by “for all,” “ \Rightarrow ” by “implies,” and “ \ni ” and “s.t.” by “such that.”
- ▶ Use a colon to introduce a list of items or a mathematical formula only if the preceding sentence is complete. A colon is inappropriate when the list or formula is a required part of the natural flow of a sentence. For example, in the preceding two rules, look at how the two lists of bulleted items are introduced. In particular, in the former rule, there is no colon after

“Avoid use of ‘filler’ words or phrases that carry no real meaning, such as”

but it would be OK to append the phrase (with colon) “such as the following:” or better yet to use the alternate sentence

“Avoid the following ‘filler’ words or phrases that carry no real meaning:”

before the bulleted items.

- ▶ In a list of three or more items, phrases, or clauses (like this one), put a comma between each pair. Modern journalistic conventions have deteriorated and often leave out the last comma, but avoid any such temptation to do likewise.
- ▶ Commas and periods are generally placed inside the closing quotation mark, even when the logical grouping would suggest otherwise, as in the list of bulleted items at the beginning of this section. (Computer scientists like myself often have trouble with this rule!)
- ▶ Unless the clauses are very short, put a comma immediately before a coordinating conjunction (such as “and”) that connects two independent clauses, and if the second clause starts with an introductory phrase or clause like this one does, don’t put a comma immediately after the conjunction.
- ▶ Follow this example of not putting a comma before “and” when the sentence (or clause) has one subject and uses two different verbs.
- ▶ An exception to the previous rule occurs when the sentence is long and a comma would improve readability, but in those cases you’d be well advised to break the sentence into two separate sentences.
- ▶ Another logical inconsistency: it’s always fine to put a comma before “but” even when there are compound verbs. For example, writing is most effective when original and innovative, but still requires attention to basic mechanics.
- ▶ Use “that” (with no comma before it) to introduce a dependent clause that identifies what you’re talking about. If instead the purpose of the clause is only to give complementary information, which is not needed for identification, then use “which” and separate the clause by commas from the rest of the sentence. Reread this rule for examples.
- ▶ Never use “this” or “that” as nouns. Use them instead as adjectives by specifying this *what* or that *what*.

- ▶ Use a hyphen between words that combine to form a single adjective. Example: The $O(n)$ -I/O algorithms execute $O(n)$ I/Os. (If we forget the hyphen, then we’re talking about a set of I/O algorithms that are $O(n)$ in number.) Another example: I/O-efficient algorithms are the main goal of the study of I/O efficiency. Some adjective phrases that have (happily) become heavily used no longer need hyphens: A few years ago, I used to work on “external-memory algorithms,” but now I simply call them “external memory algorithms.”

Hyphens are also used in some compound words, like “mother-in-law,” but avoid them whenever possible (the hyphens, that is), unless the compound word is used as an adjective.

- ▶ En-dashes “–” (obtained in \TeX by a double hyphen --) are used to indicate a range (as in pages 1–4) and in various contexts where a little more emphasis than a hyphen is needed. Examples: Call me at (765) 494–1730 if you disagree or write me at zip code 47907–2067. This work was not supported by my NSF grant IIS–0415097.
- ▶ Em-dashes “—” (obtained in \TeX by a triple hyphen ---) are used for emphasis—usually to separate one part of a sentence from another. Don’t put any blank spaces before or after an em-dash.
- ▶ Spell out all positive numbers less than 10 when used as an adjective. Use the numeral when used as a noun. Example: Rule 1 for writing a paper is to always do at least three or four versions. You should never need 10 versions.
- ▶ Use a spell checker periodically to get rid of obvious errors. Be sure you spell-check before distributing a version of the paper.

3. \TeX and \LaTeX Basics

- ▶ A good template to use for \LaTeX is in the file `template.tex` in my `latex` directory. I have some useful macros (definitions) in the style file `jsv1.sty`.
- ▶ Be sure to use `\label` and `\ref` for referring to equations and sections, since their numbering may change when new material is added. It’s convenient to have label names that indicate the type and location of the label; for example, just after the `\section` command that started this section, I included the \LaTeX code `\label{sec:TeXbasics}`, so that I could produce the reference “Section 3” elsewhere in this report by writing `Section~\ref{sec:TeXbasics}`.
- ▶ Use `\cite` for references. \BibTeX is highly recommended. Several large \BibTeX bibliographies are available on the Web, such as at <http://theory.lcs.mit.edu/~dmjones/hbp/> and <http://liinwww.ira.uka.de/bibliography/>.
- ▶ The \LaTeX macro `\emph{ ... }` is used to emphasize important terms when they are first introduced, as well as words in the text that deserve *special* emphasis. For example, the last sentence ends with “... deserve `\emph{special}` emphasis.” (An older way of getting italics is `{\em special\}`, but the simpler `\emph` syntax is preferred. The italic correction `\/` is called for with `\em` in this particular case but is never needed with `\emph`.)
 Nested uses of `\emph` undo one another. So if you use `\emph` inside a `theorem` environment (in which the text is normally italicized), the emphasized text will be upright, not italicized.
- ▶ \LaTeX also defines macros `\textrm`, `\textit`, `\textsl`, `\textbf`, `\textsf`, `\textsc`, and `\texttt`, which format text enclosed in braces into roman, italics, slant, boldface, sans serif, small caps, and typewriter font.

- ▶ The `\verb` macro formats its argument verbatim into typewriter font so that even the special characters like `\` and `&` are treated like ordinary characters. The argument to `\verb` must be enclosed in a pair of matching characters (other than a space, letter, or asterisk) that do not appear elsewhere in the argument; for example, the previous sentence starts with “The `\verb#\verb# macro ...`” The `verbatim` environment can be used to simulate multiline typewriter text including the carriage returns.
- ▶ Names of identifiers (or variables or program modules) having more than one letter should be formatted so that they appear in italics with the proper spacing. Feel free to use my macro `\id` for that purpose. Single-letter identifiers are formatted correctly in math mode without need for any macros. My style file `jsv1.sty` has macros called `\idrm`, `\idit`, `\idbf`, `\idsf`, `\idtt`, and `\idcal` for formatting math text into roman, italics, boldface, sans serif, typewriter, and calligraphic font. The macro `\id` is shorthand for `\idit`. The argument to `\idcal` must contain only uppercase letters and spaces. A space must be preceded by `\` to be recognized, and underscores are denoted by `_`.

The macros work correctly in both text and math modes, and they automatically use the right size in subscripts and superscripts. Example: The \LaTeX code

```
We can insert key~\idcal{K} into the hash table \id{table\_ptr} by means
of the function call $\id{hash\_insert}(\idcal{K}, \id{table\_ptr}, M)$,
where $M \leq M\_idrm{max}$ is the size of the table.
```

produces the following:

“We can insert key \mathcal{K} into the hash table `table_ptr` by means of the function call `hash_insert(\mathcal{K} , table_ptr, M)`, where $M \leq M_{\max}$ is the size of the table.”

- ▶ \LaTeX 's list-making environments `itemize`, `enumerate`, etc. often leave excessive vertical space between items. A simple fix is to insert my macro `\cramped` before the first `\item`, as I did in the two examples near the beginning of Section 2.
- ▶ \LaTeX puts extra space before the start of an environment if there is a blank line in the input before the `\begin{ ... }` command that starts the environment. The extra space is appropriate for environments like `theorem` and `definition` that logically start a new paragraph, but not usually for list-making environments. Delete the blank lines in the latter case.
- ▶ Use `~` and `_` and `\@` appropriately to prevent bad line breaks and for proper spacing in \TeX . (The `_` character in the preceding sentence is meant to denote a space.) Example: The \LaTeX code

```
Prof.~Vitter wants all Comp.\ Sci.\ students to write English well,
not just those from the U.S\@. The \LaTeX\ code~\verb#U.S\@.#
instructs \LaTeX\ to treat the previous “U.S.” not only as initials
but also as the end of a sentence, so that more spacing is added.
```

produces the following:

“Prof. Vitter wants all Comp. Sci. students to write English well, not just those from the U.S. The \LaTeX code `U.S\@.` instructs \LaTeX to treat the previous “U.S.” not only as initials but also as the end of a sentence, so that more spacing is added.”

Another example:

Let I denote the subset of n points in the interior of circle C . (Refer to Figures 2, 3, and 4.)

produces the following:

“Let I denote the subset of n points in the interior of circle C . (Refer to Figures 2, 3, and 4.)”

The use of \sim in this example is to prevent an isolated symbol or number from appearing at the beginning of a line of text. There is no \sim before `n points`, because `n` is linked to `points` and thus it would not be isolated if it appeared at the start of a line.

4. Figures and Tables

- It’s easy to include pictures or displays in a \LaTeX document. If the picture is in a postscript file, include `\usepackage{epsfig}` in the preamble and use the command `psfig` to insert the picture. For example,

```
\begin{figure}[h]
\centerline{\psfig{figure=picture1.eps,width=2.9in}}
\caption{Here’s a sample figure of something near and dear to my heart.}
\label{fig:figexample}
\end{figure}
```

reads in the picture of a disk platter in the file `picture1.eps` and produces Figure 1. The `h` argument to `\begin{figure}` inserts the figure immediately after where it occurs in the \LaTeX file. A `b` argument would cause the figure to be “floated” to the bottom of the current page (or if it doesn’t fit on the current page, it goes to the next page on which it fits). The argument `t` would instead float the figure to the top of the page. The width argument in the `\psfig` command sets the actual width for the picture, and the picture in `picture1.eps` gets scaled appropriately. You can also scale to a certain height, or even to a certain height and width simultaneously.

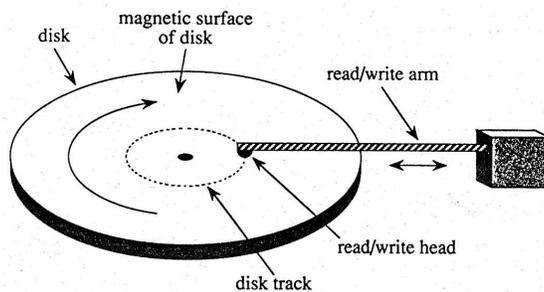


Figure 1: Here’s a sample figure of something near and dear to my heart.

- Use `\begin{table}` instead of `\begin{figure}` if you want a table instead of a figure. Tables can also be floated to the top or bottom of a page.

- ▶ Tables are somewhat cumbersome in L^AT_EX using the tabular environment, so use some good templates to design them. Several good examples appear in [2]. In the example below, note the use below of `\begin{center}` to center the table. The macro `\O` provides horizontal spacing equal to that of a digit, so that the numbers line up correctly.

```

\begin{center}
\def\O{\phantom{0}}
\begin{tabular}{|c|c|c|c|} \hline
\textit{Error Norm}& \textit{Exact} & \textit{Prob. Counting} &
\textit{Static} \\ \hline
 $\|e^{\mathrm{abs}}\|_1$  & 52.13 & 52.42 & 52.13 \\
 $\|e^{\mathrm{abs}}\|_2$  & 88.70 & 88.80 & 88.70 \\
 $\|e^{\mathrm{rel}}\|_1$  & \O0.17 & \O0.17 & \O0.17 \\
 $\|e^{\mathrm{comb}}\|_1$  & 16.71 & 16.72 & 16.71 \\ \hline
\end{tabular}
\end{center}

```

Here's what's produced:

<i>Error Norm</i>	<i>Exact</i>	<i>Prob. Counting</i>	<i>Static</i>
$\ e^{\mathrm{abs}}\ _1$	52.13	52.42	52.13
$\ e^{\mathrm{abs}}\ _2$	88.70	88.80	88.70
$\ e^{\mathrm{rel}}\ _1$	0.17	0.17	0.17
$\ e^{\mathrm{comb}}\ _1$	16.71	16.72	16.71

- ▶ It's sometimes useful in T_EX and L^AT_EX to insert a zero-width `\strut` into text in order to produce the proper vertical spacing when the text is enclosed in a box or abutted to other items in a vertical list. L^AT_EX often needs extra help. For example, I've found it useful to define macros `\tallstrut` and `\tallerstrut` in my style file `jsv1.sty` to add extra spacing above the first line of boxed tables made with L^AT_EX's tabular environment. Here's an example where I slightly modify the last example by adding `\tallstrut` immediately after the first two instances of `\hline`:

<i>Error Norm</i>	<i>Exact</i>	<i>Prob. Counting</i>	<i>Static</i>
$\ e^{\mathrm{abs}}\ _1$	52.13	52.42	52.13
$\ e^{\mathrm{abs}}\ _2$	88.70	88.80	88.70
$\ e^{\mathrm{rel}}\ _1$	0.17	0.17	0.17
$\ e^{\mathrm{comb}}\ _1$	16.71	16.72	16.71

5. Formatting Mathematics

- ▶ Avoid using `\frac` for fractions that appear in text style or when the resulting characters will be too small, such as in exponents, subscripts, and superscripts. For example, use $(N/B)\exp(z/t)$ or $(N/B)e^{z/t}$ in text, but not $\frac{N}{B}e^{\frac{z}{t}}$. It's marginally OK to use $\frac{N}{B}e^{z/t}$, but one of the previous two choices is preferred. This distinction would more obvious if the $e^{z/t}$ term were replaced by something like $\log N$. In that case, $(N/B)\log N$ is definitely better than $\frac{N}{B}\log N$; otherwise,

the term that is mathematically the most significant (namely, $\frac{N}{B}$) is dwarfed in type size by the less significant term ($\log N$). In display format, use

$$\frac{N}{B}e^{z/t} \quad \text{and} \quad \frac{N}{B}\log N. \quad (1)$$

- ▶ Be sure to include proper punctuation in sentences involving math formulas like $E = mc^2$. Periods always go outside of math mode when the formula is in text, like in the last sentence, but inside math mode when in math display format, as in (1).
- ▶ Long sets and sequences should be specified differently when they appear in text than when they are displayed, so as to allow a line break in the text if needed. Example: The L^AT_EX code

The set `\{ x_1$, x_2, \dots,~$x_n\}` is in text style.
 But in display style, it should be `$$\{ x_1, x_2, \ldots, x_n \}.$`

produces the following:

“The set $\{x_1, x_2, \dots, x_n\}$ is in text style. But in display style, it should be

$$\{x_1, x_2, \dots, x_n\}.”$$

Always use `\dots` for an ellipsis in text mode. In math mode, `\dots` and `\ldots` are interchangeable, and `\cdots` results in centered dots.

Another example: The L^AT_EX code

The sequence `x_1, x_2,~\dots\` is in text style;
 but in display style, it should be `$$x_1, x_2, \ldots. $$`

produces the following:

“The sequence x_1, x_2, \dots is in text style; but in display style, it should be

$$x_1, x_2, \dots.”$$

- ▶ Use `\langle` and `\rangle`, not `<` and `>`, to get the math angle brackets “ \langle ” and “ \rangle .”

6. Becoming a T_EXpert

T_EX is a finely tuned machine for producing beautifully formatted documents. We’ve already seen some basic guidelines for using T_EX in Sections 3 and 5. Below are some fine points about how get the most out of T_EX and L^AT_EX.

- ▶ Use `\bigl` and `\bigr` in math expressions that appears in text, but `\left` and `\right` in math displays. (There’s also `\bigm` for middle delimiters as well as plain ol’ `\big`.) In some L^AT_EX environments, `\bigl`, `\bigr`, etc. don’t have any effect in text style, but you can fix that problem by including `\usepackage{amsmath}` in the preamble. Example: The L^AT_EX code

Write `$0\bigl((N^2 (\log N) + \sqrt{m} \,) / 2 \bigr)$` in text style,
 not `$0((N^2 (\log N) + \sqrt{m} \,) / 2)$`. In displays, write
`$$0\left(\frac{1}{2} \left(N^2 (\log N) + \sqrt{m} \right)\right)`
`\quad\mbox{or}\quad 0\left(\frac{N^2 (\log N) + \sqrt{m}}{2}\right).$$`

produces the following:

“Write $O((N^2(\log N) + \sqrt{m})/2)$ in text style, not $O((N^2(\log N) + \sqrt{m})/2)$. In displays, write

$$O\left(\frac{1}{2}(N^2(\log N) + \sqrt{m})\right) \quad \text{or} \quad O\left(\frac{N^2(\log N) + \sqrt{m}}{2}\right).”$$

Another example: The L^AT_EX code

```
Write  $\bigl\{ \bigl( x(t), y(t) \bigr) \bigm| 0 \leq t \leq n \bigr\}$ 
instead of  $\{ (x(t), y(t)) \mid 0 \leq t \leq n \}$  since  $x(t)$ 
already has a parenthesis and the larger delimiters make it easier to read.
```

produces the following:

“Write $\{(x(t), y(t)) \mid 0 \leq t \leq n\}$ instead of $\{(x(t), y(t)) \mid 0 \leq t \leq n\}$ since $x(t)$ already has a parenthesis and the larger delimiters make it easier to read.”

It would have been even worse in the latter case if `|` were used instead of `\mid`, since

```
 $\{ (x(t), y(t)) \mid 0 \leq t \leq n \}$ 
```

gives the wrong spacing around the `|` relation:

```
“ $\{(x(t), y(t)) \mid 0 \leq t \leq n\}$ ”
```

- In text style, enclose a tall or deep math expression or subexpression within `\smash{ ... }` to make T_EX ignore its height. Otherwise, T_EX may add extra space between adjacent lines in the text, which can look awkward, for example, when there is a really deep expression like $x_{h(y_{f_1})}$ on one line and a really tall expression like $2^{2^{2^n}}$ on the next. Often the lines won’t interfere with one other if the extra space is eliminated by use of `\smash{ ... }`, and the resulting look (e.g., $2^{2^{2^n}}$) will be much improved. Very tall or deep expressions are best put in displays.

References

- [1] Donald E. Knuth, Tracy L. Larrabee, and Paul M. Roberts. *Mathematical Writing*. Mathematical Association of America, 1989.
- [2] Leslie Lamport. *A Document Preparation System L^AT_EX*. Addison-Wesley, 1994.
- [3] William Strunk, Jr. and E. B. White. *The Elements of Style*. Macmillan Publishing Company, New York, 3rd edition, 1979.