

# How to deal with underfull and overfull boxes

Gady Kozma

July 17th, 2004

## 1 What are we talking about?

Like all wordprocessing software,  $\LaTeX$ , and therefore  $\text{LyX}$ , will encounter from time to time a situation where it is unable to divide text into lines or divide the paper into pages in an aesthetic manner. Unlike some other wordprocessors,  $\LaTeX$  will complain about it, by printing messages like `overfull hbox` and `underfull vbox`. More notably<sup>1</sup>, its reaction to the first problem is not to create a line with enormous white gaps but to create a line that sticks out of the margin.

Unfortunately,  $\text{LyX}$  doesn't have a very convenient mechanism to find the places in the document where  $\LaTeX$  thought there were problems of this kind. Thus you must use either of the following methods: choose `View > Latex Log-file` and search for the error messages, or, usually easier, choose `File > Export > Latex` and use a terminal to run  $\LaTeX$  directly, i.e. type the command `latex myfile.tex`.  $\LaTeX$  will spew out mysterious error messages, among them the infamous `overfull hbox` and `underfull vbox` messages.

Always remember that these problems are *very* sensitive to font, font size, and page size. Therefore if this is not the final version of the document, if at all possible, just ignore them. The time to handle them is when your work has taken on its final form. And don't worry about this too much. It shouldn't take you more than a few minutes work per 10 pages to fix even a complicated document.

Before using all the sophisticated techniques discussed here always remember it might be easier to just change the formulation of the text a little. Hey, don't give me that look! *Così Fanni Tutti*.

## 2 Overfull hboxes

An "hbox" is an object that  $\LaTeX$  tries to fill horizontally, typically a line. It's "overfull" when  $\LaTeX$  cannot push all the things it wants to into the line. The

---

<sup>1</sup>A third difference between  $\LaTeX$  and some other word processors which is perhaps less notable, but important nonetheless, is the fact that  $\LaTeX$  breaks lines in a more-than-average intelligent manner. Specifically, if  $\LaTeX$  encounters a problem in a certain line, it can rearrange all lines leading up to it. As a result, it is less likely that  $\LaTeX$  will have problems breaking any line after the second line of a given paragraph.

reason is typically a long word or a formula at the end which, if pushed to the next line would make the line too spacey. So  $\LaTeX$  calls for help. Here is where you step in.

Before you start, check if the complaint is for real.  $\LaTeX$  will sometimes complain about hardly noticeable problems. If you cannot spot the problem immediately, one solution is as follows: in the **Extra options** box of the **Document** tab of **Layout > Document** add the word `draft`.<sup>2</sup> In most document classes (in particular the standard and AMS classes) this will cause a small black box to be printed to the right of the problematic line. If you cannot see the problem — good for you. Ignore it and move on to the next error. Otherwise proceed reading.

The second easiest solution (after ignoring the problem) is being sloppy about it.<sup>3</sup> This is done by the  $\LaTeX$  commands `\begin{sloppypar}` and `\end{sloppypar}`. To enter them, change to  $\TeX$  mode (a.k.a Evil Red Text) — use **Insert >  $\TeX$**  — and write `\begin{sloppypar}` in the beginning of the paragraph and `\end{sloppypar}` at the end. Your lines will contain too much space, but at least they won't stick out of the margins. If you don't want to do that on a paragraph by paragraph basis, just put `\sloppy` in your preamble (use **Layout > Preamble**, or, in LyX Qt, use **Layout > Document** and move to the Preamble section) or where you want the sloppy mode to start (use  $\TeX$  mode). The opposite of `\sloppy` is `\fussy`, the default behavior. Both commands take effect from the beginning of the paragraph (or part thereof, if you have a display formula) where they are found.

If the problem is a long word which  $\LaTeX$  cannot, or would not, hyphenate, the simplest solution is to add a hyphenation mark where you think the word should be broken. Use **Insert > Special Character > Hyphenation Point**. If you want the word to be broken but do not want a hyphen to appear (for example, you are breaking a URL<sup>4</sup>) you need to add the  $\TeX$  command `\linebreak`. Notice that this is **not** the same as **Insert > Special Character > Linebreak**! This will force  $\LaTeX$  to break in that point. Usually, however, you would want to use `\linebreak[n]`, which only recommends to  $\LaTeX$  to break the line there rather than forces it to do so. The number  $n$  is the level of recommendation, from 0 to 4, with `\linebreak[0]` being the weakest (and usually what you want) and `\linebreak[4]` the strongest, and equivalent to just `\linebreak`. Using this method instead of simply `\linebreak` will mean that you'll have less work the *next* time you change font.

More about this can be found in <http://dont.really.go.there/this/is/just/to/demonstrate/how/this/works>.

---

<sup>2</sup>Note that this will have other effects. For example, all pictures will be replaced by empty rectangles in the output.

<sup>3</sup>The paragraph with the footnote mark, for example, if set like that. Notice the extra wide space in this line! This is due to sloppily pushing the text `\end{sloppypar}` to the next line.

<sup>4</sup>If you enter a URL, the easiest method is **Insert > URL** which in most  $\LaTeX$  installations also takes care of breaking lines for you. Note though that the  $\LaTeX$  system in [arxiv.org](http://arxiv.org) is an exception and won't break URLs across lines, so if you plan to submit your paper there, plan ahead.

## 2.1 Inline formulas

Formulas inside you text pose more of a problem as  $\LaTeX$  is typically shy of breaking formulas, and many authors consider breaking formulas across lines an ugly habit<sup>5</sup>. The first option you should consider is probably to make the formula into a display formula (**Insert**  $\triangleright$  **Math**  $\triangleright$  **Display Formula**). If you want to keep the formula inline, you can try the `\linebreak` mechanism above. Notice that to create evil red text in math mode you don't need to select it from a menu, it is activated automatically by the `\` sign and terminated by the `[` — the `[0]` will not appear in red but don't worry about it. In many cases it works inside formulas. The main case where it doesn't is inside grouped parenthesis, by which I mean parenthesis entered using the parenthesis button in the math panel (**Insert**  $\triangleright$  **Math**  $\triangleright$  **Math Panel**).  $\LaTeX$  will simply ignore `\linebreak` commands inside a grouped parenthesis. Here is an example:

I love long formulas in the middle of my text, like  $\int 1/(t \log t \log \log t \log \log \log t) dt = \log \log \log \log x$ . But when I write  $\log(\sin x + \cos x + \tan x)$  the grouped parenthesis always makes my formulas stick to the right.

The solution is to replace them with regular parenthesis. You would probably want to enlarge them, though. To do that, just write `\big` before each parenthesis. If `\big` is not big enough, just try `\Big`, `\bigg` or `\Bigg`. Table 1 shows how they look.

regular	$\{\} \langle \rangle \lceil \rceil \uparrow \downarrow \lfloor \rfloor$
<code>\big</code>	$\{\} \langle \rangle \lceil \rceil \uparrow \downarrow \lfloor \rfloor$
<code>\Big</code>	$\{\} \langle \rangle \lceil \rceil \uparrow \downarrow \lfloor \rfloor$
<code>\bigg</code>	$\{\} \langle \rangle \lceil \rceil \uparrow \downarrow \lfloor \rfloor$
<code>\Bigg</code>	$\{\} \langle \rangle \lceil \rceil \uparrow \downarrow \lfloor \rfloor$

Table 1: parenthesis and friends in various sizes.

Apropos table 1, one should be careful to distinguish between the symbols  $\langle$  and  $\rangle$  use for scalar product and bra-ket notation and the simple bigger/smaller signs  $\langle \rangle$ . Not only are they different in appearance,  $\TeX$  will also space them differently. There is no GUI for entering the symbols  $\langle$ ,  $\rangle$ , nor for the symbols  $\lceil$ ,  $\rceil$ ,  $\lfloor$ ,  $\rfloor$  — from the math panel you can only insert the grouped version, not the “stand alone” version. I simply enter them using their  $\TeX$  shortcuts and  $\text{LyX}$  understands and transforms them to the appropriate characters automatically. The shortcuts are, respectively, `\langle`, `\rangle`, `\lfloor`, `\lceil`, `\backslash` or `\`, `\rceil`, `\rfloor`.

<sup>5</sup>Section 2.4 is for you guys. Naturally, using the techniques there will only aggravate overfull hbox problems.

If you use inverted parenthesis, as for example in the old-fashioned style  $]0, 1[$ , you need to tell  $\LaTeX$  which is left and which is right in order for it to get the spacing right. You do that by replacing the various `\big` constructs with `\bigl` and `\bigr`. For example

$I = ]0, 1[$  looks much better than  $I =]0, 1[$ . Notice how the `=` and the `]` are cramped together in the second example, which has the right and left `\big` commands reversed.

There is also a `\bigm` construct for notations such as  $\mathbb{P}(X | Y)$ .

There are corresponding constructs specific for `|` signs — the commands `\lvert` and `\rvert`. I cannot in honesty vouch for or against them. Check the documentation for the `amsmath` package which should come with your  $\TeX$  distribution — try e.g. `/usr/share/texmf/doc/latex/amsmath/amslldoc.dvi` — where these things are thoroughly explained.

## 2.2 Display formulas

Display formulas can also be too long, quite easily so in fact, and will create an `overfull hbox` warning. There isn't much to do about it except transform them into multi-line equations. Multi-line equations are covered quite fully in the *User's Guide* (as of writing this, in section 5.4).

As in the case of inline equations, you might need to break parenthesis across lines and run into various problems. The simplest solution is to use the “no parenthesis” square at the bottom right corner of the parenthesis panel (remember to uncheck the `\keep matched` box if you are using  $\text{LyX Qt}$ ), for example

$$f(n) = \pi \left( \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \right. \\ \left. \dots + \frac{n-2}{n-1} + \frac{n-1}{n} \right)$$

But what if there were no fractions on the bottom line? The parenthesis would come out too small. Obviously, one could solve this using the `\big`, `\Big`, `\bigg` etc. mechanism. But there is another solution, using the `\vphantom`  $\LaTeX$  command. Let us demonstrate with an example

$$\left( \sum_{i=1}^{\infty} \text{The parenthesis on this line are large due to the } \sum \text{ sign,} \right. \\ \left. \text{so lets add here a vphantom command with the same construct} \right).$$

For those reading the printed version, the lower line has a `\vphantom{\sum_{i=1}^{\infty}}` which doesn't appear at all, but makes the parenthesis large. You can put anything

you want inside the `{}` signs — whatever you put does not show but its height is considered when the line height is calculated.

There is nothing stopping you from using the `\vphantom` trick on inline formulas, but it really more suitable for using in large, complicated display formulas.

## 2.3 More on hyphenation

There are a number of issues you might want to know about if you have hyphenation problems. First of all, hyphenation is language sensitive. Really! You can change the default language of the document via the **Language** tab of the **Layout**  $\triangleright$  **Document** popup, or you can change a particular piece of text via the **Layout**  $\triangleright$  **Character** popup. A typical  $\LaTeX$  installation will support hyphenation in English, French, German, Spanish, Russian and a few others.

The second thing you have to know is that there is a package to help with hyphenation problems: `hyphenat`. The most popular uses of `hyphenat` are to turn hyphenation off completely, or to allow hyphenation of text in the `typewriter` font,<sup>6</sup> and it has some advanced tools in addition. More likely than not, it is a part of your  $\LaTeX$  distribution. Read the documentation, it could be in, for example, `/usr/share/texmf/doc/latex/styles/hyphenat.dvi`.

Finally, there are special cases where you want special effects when the line is broken — not just any old `hyphen`. For these unusual cases, you need the `\discretionary` command. Since this is a bit off topic, I've put the full description in an appendix.

## 2.4 More on breaking inline formulas

$\LaTeX$  (and  $\TeX$ ) will break inline formulas around equality-like signs, e.g. `=`, `\approx`, `\leq` etc., and less willingly around binary operators like `+`, `\cdot`, `\wedge`. To make  $\LaTeX$  less or more likely to break formulas in that way, do the following: in the latex preamble add the following commands:

```
\relpenalty=n
\binoppenalty=m
```

where *n* and *m* are two numbers that tell  $\LaTeX$  how bad is it to break formulas around equality-like signs and binary operators respectively. The default values (for both  $\TeX$  and  $\LaTeX$ ) are 500 and 700. The value 10000 is used as  $\infty$  so if you set either to 10000  $\TeX$  will never break your formulas in that kind of position.

## 2.5 Underfull hboxes

These are quite rare, but can happen, e.g. when you use forced linebreaks. Here is an example:

---

<sup>6</sup>By default, any text in typewriter font is never hyphenated. URLs are only a specific case.

Wow, why is this line like that?  
Oh, because I asked so explicitly. Silly me.

Similarly, the use of `sloppypar` can cause `underfull hbox` errors.

### 3 Underfull vboxes

A vbox is something that  $\LaTeX$  tries to fill vertically, typically a page. For some unknown reason, when  $\LaTeX$  cannot fill the page because the last object is too large, it will not do as it does for lines and go over the margins, but it will prefer to put the offending object on the next page. Hence the error will usually be `underfull vbox` and the visible symptom will be pages which are too short, have too much space between the text and the footnotes or that have too much space between paragraphs, formulas etc.

The range of solutions to underfull vboxes is more restricted. Most notably it is to tell  $\LaTeX$  where to break pages yourself. This is covered in the *User's Guide* under “forcing page breaks” (as of the time of writing, this is subsection 6.4.4). Unfortunately, in many cases the offending object is a multi-line formula, and you cannot enter a pagebreak inside a formula using the `Layout`  $\triangleright$  `Paragraph` popup. A simple solution might be to break the formula into two formulas, possibly forcing a pagebreak between them, but actually there's a better way so read on.

It is possible to allow page breaks inside formulas, globally or on a formula-by-formula basis. To understand how, we first need to differentiate between  $\LaTeX$  formula environment and AMS formula environments. The (regular)  $\LaTeX$  formula environment are what you get when you hit `C-Return` inside a formula, or by choosing `Insert`  $\triangleright$  `Math`  $\triangleright$  `Eqnarray` environment. The AMS formula environment are supported in  $\LaTeX$  using `Insert`  $\triangleright$  `Math`  $\triangleright$  `AMS *`. Generally speaking the AMS environments are more flexible, provide nicer spacing and more options (see the AMS documentation already mentioned). The only real issue with using them are potential incompatibilities between the AMS package and the document class you might be using, which might be more common than you think, especially if you got the documentclass files from some European journal. All the documentclasses supported by  $\LaTeX$ , however, are fully compatible with the AMS package, so you should worry about that only if you plan to export your document as  $\LaTeX$  and do all the configuration work manually on that file.

For the  $\LaTeX$  formula environments you have only one option, to allow page breaks globally and completely. This is done by putting the command `\allowdisplaybreaks` in your  $\LaTeX$  preamble. For the AMS math environment you have the following options:

- You can put a `\displaybreak` or a `\displaybreak[n]` in a specific formula in a specific position where you allow the page to break.  $n$  is as for the `\linebreak` command.

- You can allow page breaks in all AMS equation environments globally. Put the command `\allowdisplaybreaks[n]` in your preamble where the parameter of permissiveness  $n$  is between 1 (not 0!) and 4. Note that `\allowdisplaybreaks[4]` is identical to `\allowdisplaybreaks` and will affect the  $\LaTeX$  math environments too.

See some examples in appendix B.

## 4 About this document

If you have any comments about this document I would love to hear them. I am especially interested in new  $\LaTeX$  tips and tricks! There always seem to be more. My e-mail is `gadykozma@hotmail.com`.

This document is in the public domain. Please do with it as you please.

## A The discretionary command

The discretionary command takes three arguments, i.e. its general format is `\discretionary{a}{b}{c}`. The third argument is the text to put in the usual case, namely the line is not broken there. The first and second are for breaking lines, and appear in the end of the first line and the beginning of the second. Thus the usual hyphenation point is equivalent to `\discretionary{-}{-}{-}` but the German `\ck` command is equivalent to `\discretionary{k-}{k}{ck}`. That's right, in German, sometimes when you break a `ck` across lines you want the broken text to be `k-k`.<sup>7</sup>

Here is a particularly freaky example. In the preamble, define the following:

```
\newcommand{\brmul}
{\discretionary{\mbox{\$, \cdot}}{ }{}}
```

Next, add a formula like that  $f(x) = x \log x \log \log x$  (if you are reading the printed version, there is a `\brmul` between the  $\log x$  and the  $\log \log x$ ) and play around with the surrounding text to force it to break between the  $\log x$  and the  $\log \log x$ . For example:

When using, inside the line, long formulas like  $f(x) = x \log x \cdot \log \log x$ , strange things could happen.

Voilà! A multiplication sign (`\cdot` is the code for “.”<sup>8</sup>) appears in the middle of the formula.

---

<sup>7</sup>I don't know the first thing about German, so all this should be taken with a mountain of salt.

<sup>8</sup>`\,` gives the standard space before the multiplication sign. What about the `\mbox` and the `\$` signs? Weird  $\LaTeX$  working mode issues. Don't ask me, I only fiddle until it works.

## B Page break examples

Gauss number =  $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$

$+ 11 + 12 + 13 + 14 + 15 + 16 + 17 + 18 + 19 + 20$

$+ 21 + 22 + 23 + 24 + 25 + 26 + 27 + 28 + 29 + 30$

this line contains a `displaybreak[3]` command

as you can see, in such a bad position it is not enough to force a break.

this line contains a `displaybreak` command



as you can see, this command is absolute.

$$\begin{aligned} &+ 31 + 32 + 33 + 34 + 35 + 36 + 37 + 38 + 39 + 40 \\ &+ \cdots + 100 = 5050 \end{aligned}$$