Technical Preparation of OR Manuscripts: Dos and Don'ts

Mirko Janc, Ph.D.

Publishing Technologist, INFORMS
mirko.janc@informs.org

LaTeX and Word

- Most files generated by LaTeX and MS Word. Very few other file formats
- Disks used whenever possible
- Proper housekeeping of files
  - minimize number of files
  - s.t. others can figure out what is what
- Table files may be separate
- Art files *should* be separate
- Main text should, if possible, be in ONE file. No sections or even subsections scattered over separate files, please. Avoid using a LaTeX master file that calls in a large number of smaller files.

A notation list may be required to facilitate the editing and production workflow. Identifying proper symbols and indicating whether they are roman or italic is crucial for clarity.

Every submission *must* be accompanied by a table of contents file indicating what the other files in the package are and possibly giving additional hints.

Accompanying art files should be those that were really used to produce the article. Thus, original Excel files, PowerPoint files, etc. should also be supplied.

Don't forget the PDF!

- PDF format—de facto standard for transfer of formatted content
- PDF file necessary to view and print the submitted article. All fonts used must be embedded.
- Q: Why the Word file is not enough?

  A: You may not realize that certain features (especially symbols) appear properly on your screen because of your particular configuration and the fonts available on your system. What you see as

$$\sum_{k=1}^{n} a_k T_k(x) \quad \text{or} \quad \Phi \succ \Theta,$$

  on the reviewer's or publisher's screen may easily appear as

$$\square_{k=1}^{n} a_k T_k(x) \quad \text{or} \quad \Phi\ \phi\ \Theta,$$

  if the summation sign and the succession symbol are taken from a "weird" font on your system.
- With a properly formed PDF, we see exactly what you intended us to see (and hopefully saw yourself). Matching source files are still necessary for production and sometimes to decipher your real intent.

What can go wrong with fonts?

- To be on the safe side, it is best to use the standard TimesNewRoman in Word; Type1 (not bitmap!) Computer Modern fonts in LaTeX.
- With the usual free TeX distributions (MiKTeX, TeXLive, ...) the bitmap fonts are what you get: High-quality Type1 Computer Modern fonts are also freely available (on CTAN). You may need some help in installing them, depending on your platform and your understanding of where the fonts reside and how they make it into your PDF file.
- Q: How do I know whether my fonts in LaTeX are bitmap or Type1? A: That's easy! A PDF based on bitmap fonts looks fuzzy and symbols appear broken in Acrobat Reader (prior to version 6)—you just cannot miss it.
- If your source is a Word file, use Distiller, not PDFWriter. Additional products are nowadays available for direct writing of PDFs, but graphics may not look as intended.
  - When installing the Adobe/Microsoft PostScript printer driver, you also get Acrobat Distiller settings as if it were a real printer.
  - Just "print to Distiller." Set joboptions = PressOptimized; the system will rather quit than let you proceed without proper fonts embedded. Be aware that, in Adobe's verbiage, PrintOptimized = good to print on a laser printer, whereas PressOptimized = good to print on a hi-res device with NO fonts mismatch tolerated.
- For Word users that have lots of heavy math, the only right way to produce it is to use *MathType*. *MathType* installs its own fonts. For users of *Mathematica*: It supplies its own collection of math fonts, so be aware (= beware) of such situations.

Why do we need all these files?

- PDF is the most robust file type to view the formatted content
- We print the PDF file to get the hard copy to be copyedited
- The accompanying LaTeX or Word files should be the exact files that produced the submitted PDF. They are used in composition and should match the version approved and edited for language and journal style.
  - Pssst! If your PDF carries figures (line-art) in proper vector format, cutting them out of that file is by far the easiest way to get the "raw figures" that will be adjusted for final use both in the printed and online article.
- Version control is a big issue! Additional revised files should never be provided tacitly without proper documentation explaining why and what changes were made (*"Yeah, our proof of Theorem 3 in the version of March 2001 is incorrect, but we'd like to keep that 'received date' for priority issues!"*)
- Art updates, even if performed only at the production editor's request as adjustments to conform with the style, should be clearly identified (e.g., "variables italicized").
- Added, updated, and removed references (compared to the initial submission) should be also documented.

Art files

- Traditionally, publishers call it "art." Outside publishing professions, probably no one would ever think that "art" is a proper word to describe a pie chart, or a piece of the first quadrant of the *x*-*y* plane filled with mathematically computed exact positions of 100,000 little crosses, out of which even the most careful visual inspection cannot find more than 763.
  - Did you ever notice that sometimes an apparently simple chart/graph is 1.8 Mb in size and takes 40 seconds to render even on a 2.8 GHz Pentium 4?
  - Yes? OK. That is a result of mathematical overkill in some graphic outputs of, say, simulation results. Bear in mind that the online reader will face the same frustrating problem!

Back to the dry matter!

- Most graphs are produced in Excel, PowerPoint, Word (by using its simple drawing commands), fig2dev, LaTeX picture environment, and statistical packages like SAS.
- Nowadays, scans are submitted only if their sources are available only as printed in a publication from the prehistoric, pre-electronic era (say, before 1995).
  - Note this before it is too late: Do not save scans of line-art as jpeg! Use only high-res (600dpi and up) tiff, if scanning is an issue at all.

Art files (2)

- INFORMS journals are printed in two columns (20 picas wide = $3\frac{1}{3}$ in) (with exception of MOR).
- If at all possible, avoid stretching the graphs too wide, unless the wealth of information shown really requires a two-column spread.
  - When processing art, we do our best to fit them in one column
  - Horizontal compressing the type (down to about 80%) is among the tools
- Make sure that the graph labels are consistent in size and overall appearance, including uppercasing and lowercasing. Variables should be italic, but the digits and parentheses should be upright:

BAD: $(a_2 + b_2)$ GOOD: $(a_2 + b_2)$

- Sizes of all labels, including numbers on the axes, should be appropriate. Ideal final value is 8pt for the prevalent lettering in a graph. If the axis numbers are too small, the graph has to be too big to keep the labels legible.
- Do not letterspace the labels, and make sure that when inserting the figure into your document you do not squeeze it or stretch it. Generally, just scaling is OK.
- Having labels in a sans serif font, if math is present, is not a good idea:

l_I-l_{II}+$l_{|I|}$ is actually the same formula as $l_I - l_{II} + l_{|I|}$.

Art files (3)—Color issues

- INFORMS journals do not accept color, except when authorized by the editor-in-chief. As drawing programs have a variety of options, it may be tempting to overuse color—to the extent that details on its grayscale version become indistinguishable.
- Hatch patterns should be avoided at all cost. If coloring (including gray) is necessary, only solid colors should be used. Colors that look obviously different on screen may turn out as very similar in grayscale! Your mother's old red sweater and the green leaves behind her on that old b/w photograph invariably look dark, almost black.
  - Grayscale solid "color" is normally measured in percent from 0, being white, to 100, being black. To distinguish them, at least 15 percentage points of difference is recommended.
  - If lines on a graph are colored, they should be shown quite thick (1pt min, 2pt is even better), so that the level gray could be seen as a feature, not as scarcity of black paint.
  - Standard way to avoid the color problem is to use:
    - Dash-dot-dot type of lines, and
    - Little marks as squares, circles, and triangles along the lines.
- Original art files (say, in Excel, PowerPoint, etc.) should be included. Cutting them and pasting into Word often turns the original WMF format into a low-res bitmap that looks bad and, generally, can not be improved. Thus, the originals are necessary for further processing. (If the pasting went OK, we would be very glad to cut the line-art from your PDF in no time!)

Math

Let us now move to the question of representing math formulas. This is a point of division showing why math typesetting is so much more complicated and demanding than typesetting of plain text.

For clarity, we will keep using Helvetica as the font for the metalanguage comments, whereas the samples will be set in Times.

- Generally, variables should be italic. Superscripts and subscripts, if they are actually acronyms, should be roman, unless they have no more than two letters, in which case italic is OK. Expected value η_{ev} and calculated value η_{cv} should satisfy the inequality $|\eta_{ev} - \eta_{cv}| < \varepsilon$. However, $K_{\max} = \max\{|L_{\text{upper}}|, |L_{\text{lower}}|\}$. Digits should never be italic: $a_{11} + b_{22}$ is really bad; only $a_{11} + b_{22}$ is OK. In LaTeX most of this italic vs. roman automatically happens. This is not case in Word, unless you have a proper setup and strictly adhere to MathType rules—whenever you key math—not just when the math is too complicated to be keyed as a part of text.
- In OR areas, often long "wordy" identifiers are used as variables. You should make a careful choice. If in some case you want your variables to be $PotentialCustomer$ and $ActualCustomer$, and also want to put a sum $\sum_{i=1}^{K} |PotentialCustomer_i - ActualCustomer_i|^2$ as an inline formula, it may totally destroy your paragraph visually. It is a better idea to present it as $\sum_{i=1}^{K} |PCust_i - ACust_i|^2$. If you decide you cannot live without such variables, use `\mathit{Pcust}`, not just `PCust` in math mode. Using $offset$ is definitely a better choice than using $offset$. Spacing in math is fine tuned for strings of variables, not for typing words; although $stop = post$ is certainly true in a commutative group.

Math (2)

- When chosing notation, never use the exactly same letter (including its being italic or roman) for a concept and its numerical value in any sense. Let us consider customers A and B. If we denote their willingness to pay by A and B, we have $|A-B|^{\alpha}=O(t^2)$. A couple of lines further, there may be a theorem. Because the style is theorem-italic, we may have: *If customers A and B act independently, we have* $|A-B|^{\alpha+1}=O(t^2)$. Do you see my point? Variables and customer identifiers look exactly the same.
- Another big math issue are math operators. Those predefined are well known and they behave as expected: the text is roman, and it is offset by a thin space from the left and right neighbor, unless the right neighbor is an opening delimiter. For example, $2\sin x\cos y=\sin(x+y)+\sin(x-y)$. The left side is just `2\sin x\cos y`. But what about those not predefined? Let's take the notorious "arg max."
  - If you put `\mbox{arg\,max}`, limits will never go under it (as expected), even in display. If you use `\mbox{arg}\max`, the limits will center under "max." The proper way is to define `\def\argmax{\mathop{\rm arg\,max}}`. Here is the difference:

$$\arg\max_{x\in A}\{f(x,\beta)\} \quad \text{vs.} \quad \arg\max_{x\in A}\{f(x,\beta)\} \quad \text{vs.} \quad \arg\max_{x\in A}\{f(x,\beta)\}$$

  - If you want an operator that should not have limits, say the European cotangent ctg, you define `\def\ctg{\mathop{\rm ctg}\nolimits}`

Math (3)

- For maximum, the usual math operator is max not Max, unless you have special reasons and can make a case for the second case. Similarly, the natural logarithm is ln, not Ln, or LN (OR people probably have the multiform complex Ln rarely in mind).
- There is a big number of usual operators in related fields that are not predefined. Among them are cl (closure), ri (relative interior), conv (convex hull), supp (support), Hom (set of homomorphisms), id (identical mapping), E (mathematical expectation, also $\mathbf{E}$, $\mathbf{E}$, or $\mathbb{E}$). (Use definitions as explained before to have your own commands `\cl`, `\ri`, `\conv`, etc.)
- Although a number of books about TeX suggest the use of "script ell" ℓ, it makes sense only if you absolutely need two distinctive varieties of "ell." In a normal serif italic font the difference between the "letter ell" and the "digit one" should be sufficient $l \neq 1$. A side condition $k, \ell, m = 1, 2, \ldots$ should rather be $k, l, m = 1, 2, \ldots$.
- Unfortunately, the lowercase letter "vee" in most cases looks almost the same as the Greek letter "nu." (v, ν). To emphasize the distinction, for some fonts we have a special rounded "vee" (and "w") $v \neq \nu$.
- If you tend to use many math accents, bear in mind that the available choice for tildes, hats, and bars jumps from the too small to the too big. For example $\overline{V}, \bar{V}, \overline{a}, \bar{a}$. We have our custom versions to interpolate between—let us illustrate it on just one case $\overline{V}, \overline{V}, \overline{V}, \bar{V}$, $\widetilde{V}, \widetilde{V}, \widetilde{V}, \tilde{V}$, or $\widehat{V}, \widehat{V}, \widehat{V}, \hat{V}$.
- For long expressions that should be "hatted" it is much better to use $(abc)\hat{}$, than $\widehat{abc}$.

Math (4)

- What to do with bold in math? There are two very distinctive cases: (1) particular bold symbols (vectors, matrices, or similar), and (2) all bold style in figure captions, abstracts, etc. In Case 2, INFORMS has always been "politically correct." Math should have same appearance in a bold environment, as in regular roman environment. Our special math font setup makes it automatically happen: **... if** $a = b \neq a' + b'$**, then ...**. Generally, authors should not worry about such matters.
- Particular bold symbols are usually set upright. Our house style for lowercase bold Greek does not follow that rule (it probably should!) because it helps with readability. $\mathbf{\Phi} = (\Phi_1, \ldots, \Phi_n)$ or $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ (better: $\mathbf{\mu} = (\mu_1, \ldots, \mu_n)$). When set in subscripts, at a smaller point size, the "uprightness" helps with readability.
  - A common error: $\mathbf{a_i} + \mathbf{b_i}$ (bold is not turned off for subscripts: `\bf a_i+b_i`). It should be $\mathbf{a}_i + \mathbf{b}_i$ (`{\bf a}_i+{\bf b}_i`).
  - Another challenge: In standard LaTeX setting, `\bf` does not affect Greek letters. To make them bold, the clumsy `\mbox{\boldmath$\lambda$}` is used, and even then the upright bold version of lowercase Greek letters is simply unavailable. Authors often do not recognize this misfeature, so only by analyzing their TeX file we see that they meant the symbol to be bold. `\bf A\bf\lambda\bf B` gives $\mathbf{A}\lambda\mathbf{B}$, instead of $\mathbf{A\lambda B}$. Nevertheless, the source code (and mathematical common sense) indicate that lambda was supposed to be bold.

Math (5)

There is a number of other details in math that we are going to address now.

- Exponential function e^t is also often written as $\exp t$. In cases when the exponent is a complex expression, the exp-notation is preferred: $e^{\int_0^t f(z,u)\,du}$ looks better if presented as $\exp\left(\int_0^t f(z,u)\,du\right)$.
- Summation sign $\sum$ is undoubtedly derived from the Greek letter sigma Σ, but they should not be used interchangeably. Thus, $\sum_i X_i$, not $\Sigma_i X_i$. A similar situation is with the product sign $\prod$ versus the uppercase Greek letter pi Π.
- Fractions are often a sticky point for people with no strong mathematical background. In a double-spaced manuscript, it is OK for the authors to leave a full-sized fraction amidst a paragraph: $\frac{\partial F(x,y)}{\partial x}$. Such a construct spreads the lines and should somehow be made flatter. For fractions like "one half" it is OK to use the case fraction $\frac{1}{2}$, but applying it to a more elaborate fraction just does not look right: $\frac{\partial F(x,y)}{\partial x}$.
  - Authors, of course, know that the above is not really a fraction; it is only customarily written as a fraction. Such "fractions" involving the "partial" symbol ∂ should be "slashed" or "shilled" (old printers' terms) as $\partial F(x,y)/\partial x$, or $(\partial/\partial x)F(x,y)$.
  - For real fractions, a common editorial problem is whether and where to add parens around the slashed parts. The rule of thumb is that the fraction bar in $\frac{a}{b}$ acts as very powerful glue. If slashed, the full "translation" is $((a)/(b))$. It is customary to delete unnecessary parentheses (parens, for short).

Which parens to delete is context-dependent. Here are some examples:

- $\frac{a}{b}$ becomes a/b; $\frac{a}{b}c$ becomes $(a/b)c$; $c\frac{a}{b}$ becomes ca/b or $c(a/b)$ (if extra emphasis is required).
- $\frac{a}{b+c}$ becomes $a/(b+c)$, but $\frac{a}{b+c}d$ becomes $(a/(b+c))d$.
- $\frac{a+m}{b+n}$ becomes $(a+m)/(b+n)$, but $\frac{a+m}{b+n}\log\left(1+\frac{m}{n}\right)$ becomes $((a+m)/(b+n))\log(1+m/n)$. Using brackets sometimes enhances clarity: $[(a+m)/(b+n)]\log(1+m/n)$.

- Large fractions are another source of trouble that, luckily, can be treated successfully. Authors should bear in mind that a fraction that is 5" long can not fit into a 3.33" line. Simple rewriting (by the author, if possible) is the best way to avoid typesetting and editorial nightmares.

$$\frac{1-a+a^2-a^3+a^4-a^5+a^6-a^7+a^8-a^9+a^{10}-a^{11}+a^{12}}{1+a}$$

should be rewritten by using one of the equivalents $\frac{a}{b}=a/b=ab^{-1}=b^{-1}a$. Thus, $(1-a+a^2-a^3+a^4-a^5+a^6-a^7+a^8-a^9+a^{10}-a^{11}+a^{12})(1+a)^{-1}$ is a “slim” equivalent of the above and offers a number of options to break it across lines.

- In displays, we generally use full-size fractions. In mixed cases, only experience and common sense can help. Rules are not strictly set but things can get ugly easily.

Math (6)

- Math operators like $\cup, \cap, \vee, \wedge$ have also their large counterparts that are used when applied to a family of objects. The small ones, being binary, stand between two objects. Thus, $A \cup B$, but $\bigcup_{\lambda\in\Lambda} A_\lambda$. In display it gets even bigger: $\displaystyle\bigcup_{\lambda\in\Lambda} A_\lambda$.
- Summations with huge stacked limits should be used with care and probably some shorthand notation introduced:

$$A_{jk} = \sum_{\substack{i=1\\ i\neq j\\ i\neq k}} R_{ijk} - \sum_{i \text{ odd}} Q_i \quad \text{should be} \quad A_{jk} = \sum_{i=1,\, i\neq j,\, i\neq k} R_{ijk} - \sum_{i \text{ odd}} Q_i$$

  The trick with "odd" is that `\mbox{ odd}` does not scale into the subscript size, whereas the command `\text{ odd}`, loaded with `\usepackage{amsmath}` does.
- Large chunks of unbreakable math should be avoided. What is OK for a poster is not OK for journal publication

$$f(x) = 2x + \left(\begin{array}{c} \text{terms} \\ \text{of higher} \\ \text{order} \end{array} \right) \quad \text{should be} \quad f(x) = 2x + \text{terms of higher order}$$

- For "cases," $f(x) = 0, \text{ if } x \geqslant 0; \ = 1, \text{ if } x < 0$, is a good inline equivalent of

$$f(x) = \begin{cases} 0, & \text{if } x \geqslant 0, \\ 1, & \text{if } x < 0. \end{cases}$$

Header information

A complete set of data must be submitted with an article. There may be some small journal-specific variations.

- Title of the article (including the short title for running heads if the full title is too long)
- Authors and their affiliations (temporary affiliations, if applicable), street addresses, e-mail addresses
- Corresponding author, Key words, Subject classification (if applicable)
- Abstract
  - Preferably one paragraph
  - Math should be avoided if at all possible (because of poor rendering in countless repackaged online versions where it will appear in addition to the print issue).
  - No references. If a reference is absolutely necessary, it should be self-contained, not pointing to the references of the article (the online reader who is not subscribed will not have the references available).
- Acknowledgments (not acknowledgements) as well as grants and permissions notes
- Dedication (if applicable)
- We supply the DOI (Digital Object Identifier) string. It will be included in all issues starting in 2004. It will look like 10.1287/mans.2004.0157 and serve for online linking. Slowly but steadily, DOI will appear in references too. It is a part of the ongoing effort of publishers around the globe to provide a standardized way of linking of millions of articles on Web.

Text and sectioning

- Throughout the body of the article, it must be clear to what level of sectioning a title belongs. Some our INFORMS journals discourage lower-level sectioning at all.
- In LaTeX terms, a scheme `\section`, `\subsection`, `\subsubsection`, `\paragraph`, and `\subparagraph` should be followed (aka SSSPS). The starred versions `\section*`, `\subsection*`, `\subsubsection*`, `\paragraph*`, and `\subparagraph*` denote the corresponding unnumbered versions of the above (mostly applicable from the third level down, if necessary).
- Loose titling styles that are OK in presentations (like this one), on the chalkboard, on posters, and within large verbose schemes, are unacceptable for the article body text. Article elements must be complete and their substructures properly nested.
  - Large displayed formulas with titles should have a lead-in text and their titles should conform to the above scheme—it should fit into the SSSPS scheme. For example (in LaTeX terms)
    ```
    \subsubsection*{Contraints of Type 1}
    <long displayed formula 1>
    \subsubsection*{Contraints of Type 2}
    <long displayed formula 2>
    ```
    etc.
  - If a long formal proof is interrupted by a subsection introducing a few lemmas necessary to complete the proof, it should be rephrased.

Theorems

- Every research article with some level of formality in exposition makes use of the so-called theorem-like environments. They generally come in two flavors: those that are set in italic—emphasis commonly used for enunciations like theorems, lemmas, hypotheses, etc.; and those set in roman—definitions, remarks, examples, etc.
  - In enunciations everything tends to be set in italic. In Word it literally becomes so; in standard LaTeX an unwanted mixture of upright (some elements in math) and italic may occur. To illustrate this, let us revert to a serif font:
    roman: … Let f be a mapping [nonsingular outside $(x, y) = (0, 0)$]…
    all italic (Word): … *Let f be a mapping [nonsingular outside $(x, y) = (0, 0)$]…*
    standard LaTeX: … *Let f be a mapping [nonsingular outside* $(x, y) = (0, 0)$*]…*
    our theorem-italic: … *Let f be a mapping* [*nonsingular outside* $(x, y) = (0, 0)$]…
    Only the last style is actually correct. All delimiters, big punctuation (colon, semicolon, question mark, interjection mark) and digits(!) should stay upright.
  - Without our special fonts that do it automatically, it is a lot of handwork to fix; the best you can do is to leave the theorems roman—variables that should anyway be italic will nicely stand out—the copyeditors will greatly appreciate it.
- Ubiquitous additional comments after a theorem must be clearly separated from the text of the theorem. A sentence like “The reverse is also true,” may be, but also may be not a part of the theorem.
- If a theorem is expressed by using a number of terms that were not yet introduced, it is much better style to have a “where-list” before, and then state the theorem. The "where-list" should by no means be a part of such theorem.

Lists

- A generally accepted classification speaks about numbered, bulleted, and unnumbered lists. In LaTeX terms, the first two are `enumerate` and `itemize`.
- In addition to it there are (1) run-in lists and (2) where-lists, actually unnumbered lists of symbols to

  introduce math notation, and
  explain meaning of certain formulas,

  which is OK if their structure is logically consistent.
- All lists with longer elements, per our house style, finish up in paragraph style with the number, bullet, or other identifier as a lead-in into text.
- Lists of short items may be set to hang (at discretion of editorial staff, depending on circumstances).
- In cases of complex notation introduced in an article, nomenclature lists may be required—at least for reference—for the editorial staff.

Figures and Tables

- Figure and table captions should be kept short.
- Excessive text goes into figure notes and table notes.
- INFORMS house style generally allows only for three horizontal rules in tables:
  (1) above the table,
  (2) under the column heads to separate them from the table body, and
  (3) after the table body.
  Straddle rules in the column-head area are OK.
- Shading in tables is strongly discouraged! If it is used, shaded entries are customarily turned into bold.
- With additional complex two-dimensional elements (variable shading, zig-zag lines to separate areas of a table, etc), we use the table as art (cut out of the author-supplied PDF, for example) and accordingly adjust to fit.

References

- INFORMS house style for in-text references is a version of "author-year" with a number of specific "subrules," including et al. for more than two authors.
- Word users just key such references out
- In LaTeX, for users of `\cite`, it becomes a bit tricky. It is not as simple as having a proper style as `harvard.sty` or `chicago.sty`. It requires references to be in the shape to supply the style with necessary data: at least: (a) version for in-text citation, and (b) version for parenthetical citation.
  - Our mid-term project—to create such a style to reflect the in-house citation style.
  - Our long-term project—to create and regularly update a BiBTeX database of all articles published in INFORMS journals in recent "electronic" years for authors to use instead of typing the information. As a plus, we plan to form a database of often-cited general works—books, articles, proceedings, etc.—that our authors frequently cite, and create another BiBTeX database as a service to our authors.
  - Another mid-term project is to create a BiBTeX style to draw from a BiBTeX database and form a `.bbl` file that conforms to the requirements of the author-year style.
- Because of detailed referencing requirements in online versions of research articles, we will expect the authors to submit additional information about items that are not clearly a journal article or a book. Proper tagging of such items, even it is not visible in print, is very important for online HTML and XML archiving purposes.