Reproducible research with Emacs org-mode

Ivan Markovsky

Plan

Problems we want to address

Solution via Emacs org-mode

Trying it out

Reporting computational results

1. setup and run calculations
   - 1.1 prepare simulation files and data
   - 1.2 execute the simulation files

2. present the results in pdf document or on webpage
   - 2.1 explain what / how is done ($\leftrightarrow$ 1.1)
   - 2.2 copy-paste results from Matlab in a document

3. make the results reproducible (readme file)
   - 3.1 make the code and the data available
   - 3.2 explain how to run the code on the data (make file)

4. redo steps 1.2, 2, 3 every time 1.1 is modified

Why making the results reproducible?

1. comparing your method with another method
   - was the code available?
   - asking the authors, have you got it?
   - if you got it, was it clear how to use it?

2. applying your method on examples from a paper
   - was the data available?
   - were all details about the simulation setup given?

3. sharing your code with someone else
   - was it easy to find it?
   - did you remember how to use it and how it works?

Writing better code documentation

- link algorithms to software implementation
  - writing formulas in comments is tedious
  - pasting code and results into text is tedious
  - tedious means time consuming and error-prone
- we need papers with more implementation details as well as code with better explanations
- can be done at the same time (literate programming)

Tools available

- matlab publish
  - m-file with text in comments
  - export to pdf (via LaTeX) and html
  - results are inserted in the document
  - no literate programming support
  - issues with functions calling

- noweb / nuweb
  - mix LaTeX and any sort of code
  - external program extracts code and tex files
  - results are not inserted in the document

- emacs + org-mode
  - combines literate programming and reproducibility
  - does not need additional software
  - works with any language

Plan

Problems we want to address

Solution via Emacs org-mode

Trying it out

Outline

- code, text/formulas, and results are in one source file
- markup language separates code, data, and text
- code can be run and output is automatically inserted
  - the editor interacts with the OS
  - calling the compiler and capturing the output
- source is readable and can be exported (pdf, html)

Emacs in a nutshell

- written in 1976 at MIT AI Lab by Stallman and Steele
- extensible in Emacs-lisp
  - every aspect of the editor is customizable
  - any functionality can be added
- operates in modes that provide features
  - Matlab, C/C++, LaTeX, ... modes
  - provides editing functions, syntax highlighting, ...
- has integration with the OS
  - you can do everything you need from within emacs

Emacs Org-mode

- written by Carsten Dominik in 2003
- document organization via fold-able outline structure
- task management, links, tables, . . .
- markup language: code, LaTeX, . . . can be embedded
- code execution and results capture
- export engine

Code evaluation

- example:

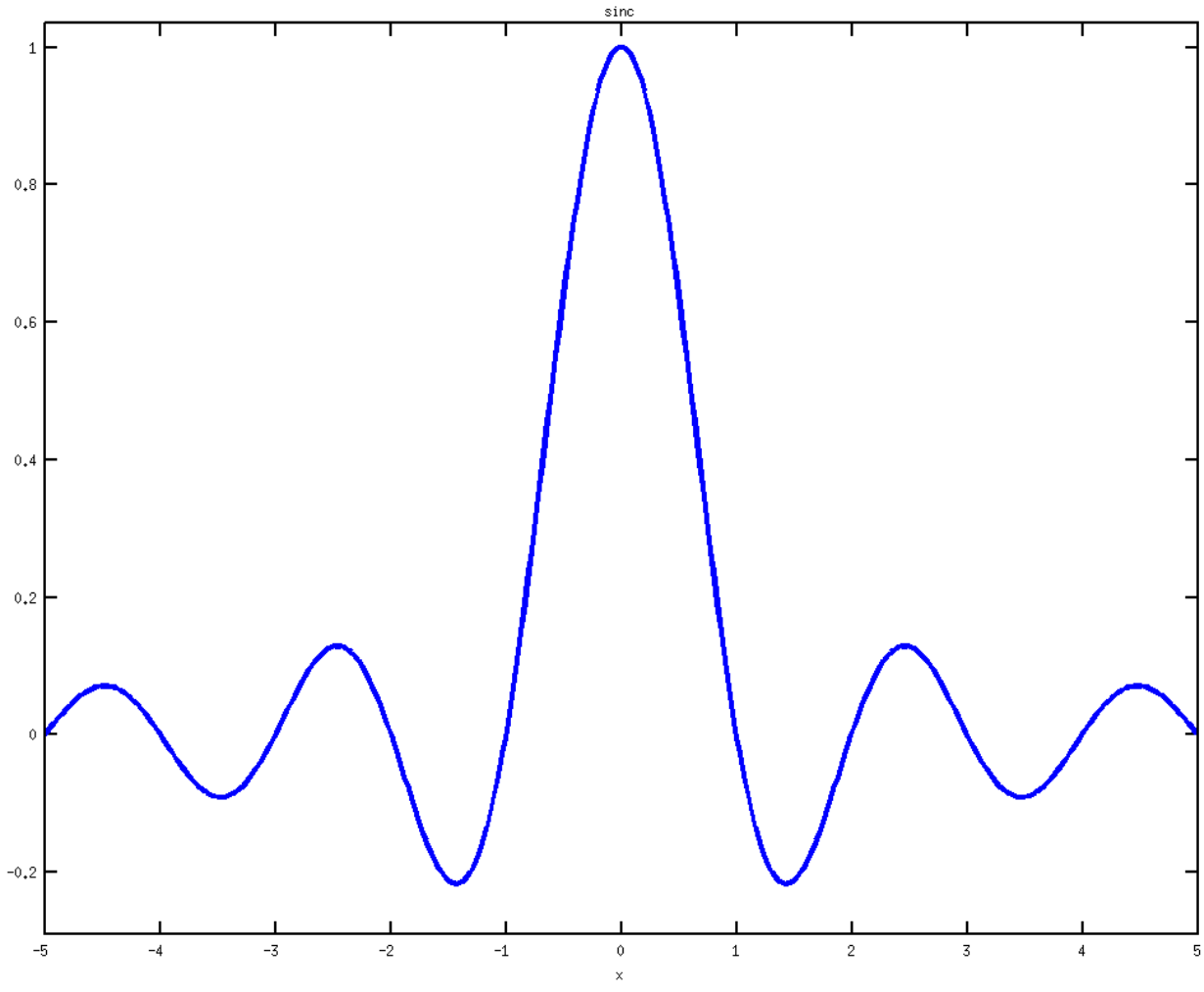
```
[pi exp(1) sqrt(2)]
```

3.1416	2.7183	1.4142

- code is tagged
  - see source file
- can be evaluated
  - place the cursor within the block
  - and type `Ctrl+c, Ctrl+c`
- options control the export
  - see the manual
- the value of `ans` after evaluation is included

Figure output

```
ezplot('sinc', [-5, 5])
print -dpng f1.png, ans = 'f1.png'
```



Literate programming

- uses the noweb syntax
  - see, user manual
  - options control the export, see noweb options

- `data` block

```
a = 1; b = 2;
```

- `computation` block

```
a + b
```

- putting them together

```
<<data>>
<<computation>>
```

```
3
```

Plan

Problems we want to address

Solution via Emacs org-mode

Trying it out

Trying it out

- publish is build in Matlab
- noweb / nuweb require installation
- emacs with org-mode
  - export to pdf and html works "out of the box"
  - in Windows, matlab integration needs tweaking
  - Egon and I are still trying to set it up . . .
- references
  - Donoho's original paper on reproducible research
  - journal of statistical software's paper
  - I used ideas from this presentation