Literate programming and reproducible research

Professor Eric S Fraga, UCL

6 June 2018
EGL2018, Essex

v1.33 4th June 2018

UCL

Outline

- Introduction
  - The aim & challenge
  - Literate programming
  - Emacs
- org mode
  - Programming
  - Writing
- Conclusions
  - Summary

Reproducible research

To ensure that our research is reproducible both by ourselves and by others.

Coding, processing, writing

In doing research,

1. we all write programs . . .
2. which generate results . . .
3. which need to be processed . . .
4. and which should be disseminated.

Currently, we use different tools for each step.

Tools

Workflow:

- coding IDE, MATLAB editor, `vi`, `notepad`, ...
- results `.txt`, `.xls`, `.dat`
- processing spreadsheet, `R`
- dissemination `Word`, `PowerPoint`, LaTeX, `beamer`
- project management ?

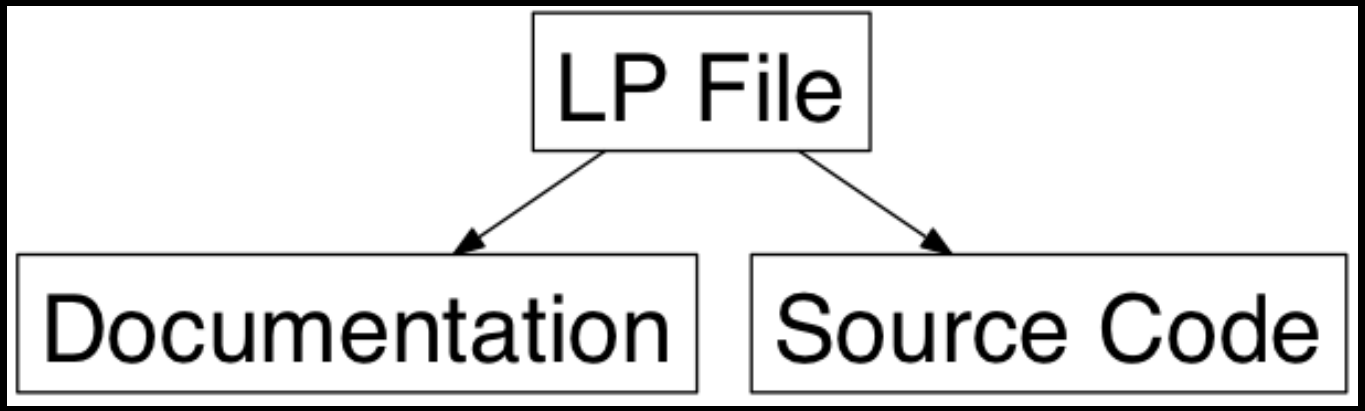
leading to constant transfer of data from one place to another.

Definition

> *Literate programming is a programming paradigm introduced by Donald Knuth in which a program is given as an explanation of the program logic in a natural language, such as English, interspersed with snippets of macros and traditional source code, from which a compilable source code can be generated.*

`https://en.wikipedia.org/wiki/Literate_programming`

Code and documentation



Source

Example: the Strawberry algorithm

```
* Strawberry -- main entry point...
  * overview...
  * function definition...
  * initialisation...
  * constrained problem?...
  * initial population...
  * iterate...
    * prune similar solutions to encouraged diversity...
    * rank solutions...
    * periodic output...
    * select and propagate...
    * COMMENT apply local search...
    * create new population...
  * end loop and finish function...
```

https://www.ucl.ac.uk/~ucecesf/strawberry.html

Editor

- originally written in 1976
- content aware editing
- fully extensible in `Emacs LISP`
- self-documenting with comprehensive help system
- large eco-system of packages

`https://www.gnu.org/software/emacs/`

org mode

Text (it's all text) based mode for

- writing & dissemination
- project management
- literate programming

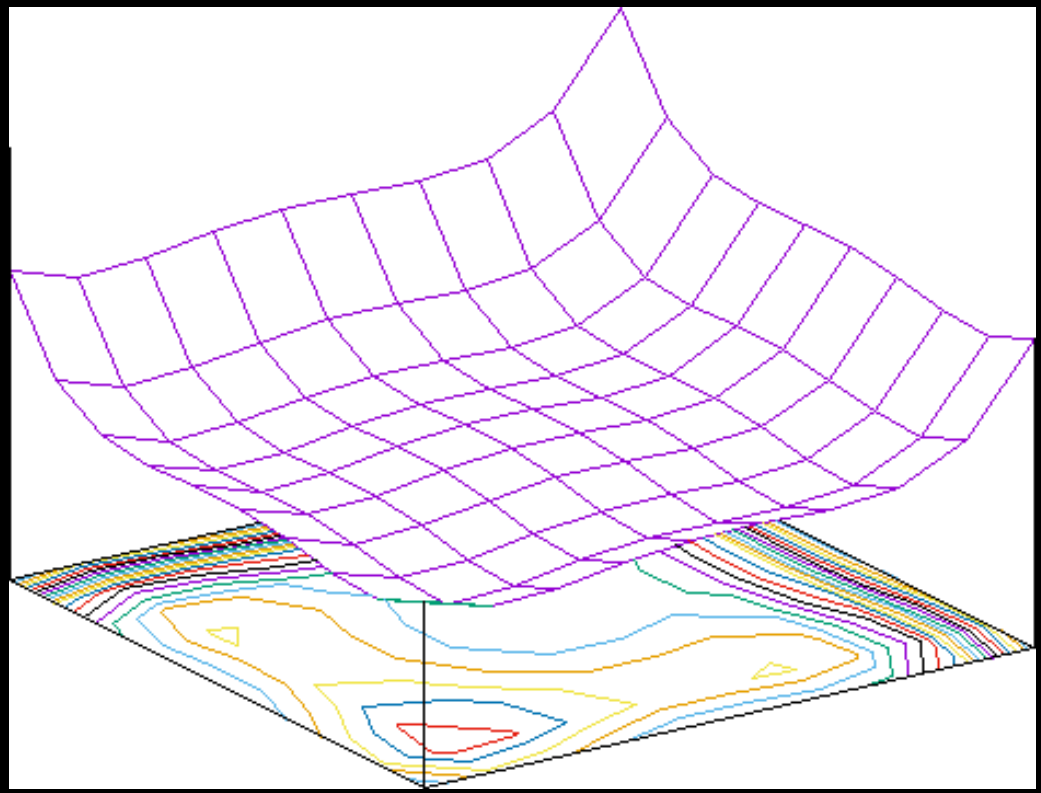
`https://orgmode.org/`

Example objective function

$$\min_x z = \sum_{i=1}^{n} x_i^4 + x_i^3 - 2x_i^2$$

$$x \in [a, b] \cap \mathbb{R}^n$$

Plot of objective function



Tangling

Create an octave file with the objective function using tangling:

```
function [z g] = f(x)
  z = sum(x.^4 + x.^3 - 2*x.^2);
  g = 0;  % unconstrained
endfunction
```

Code segments

Specify the parameters for the optimisation problem:

```
n = 2;
x0 = rand(n,1);
a = -2*ones(n,1);
b = 2*ones(n,1);
```

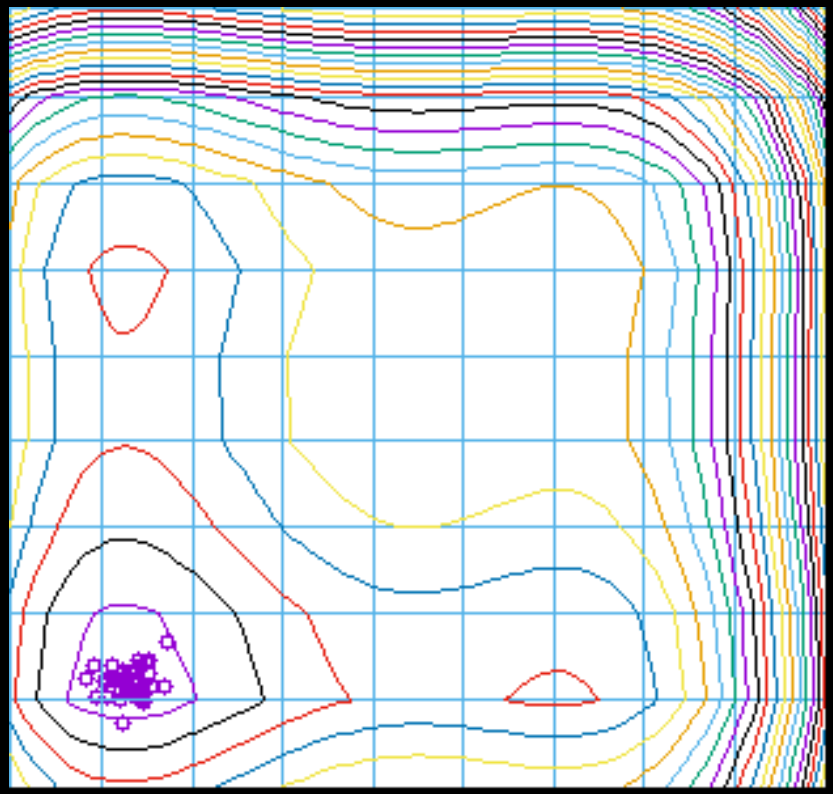
Bringing bits together

The code to include the problem setup directly and solve the problem:

```
clear
format short
n = 2;
x0 = rand(n,1);
a = -2*ones(n,1);
b = 2*ones(n,1);
[x y] = strawberry(x0, a, b, @f, 10, 10);
[x;y]'
```

Plotting results

Using data in table of results on previous slide:



Processing results

Statistical analysis of results obtained above:

Statistic	Value
Best	-5.666
Average	-5.582
Worst	-5.255
Standard deviation	0.094

Outlines

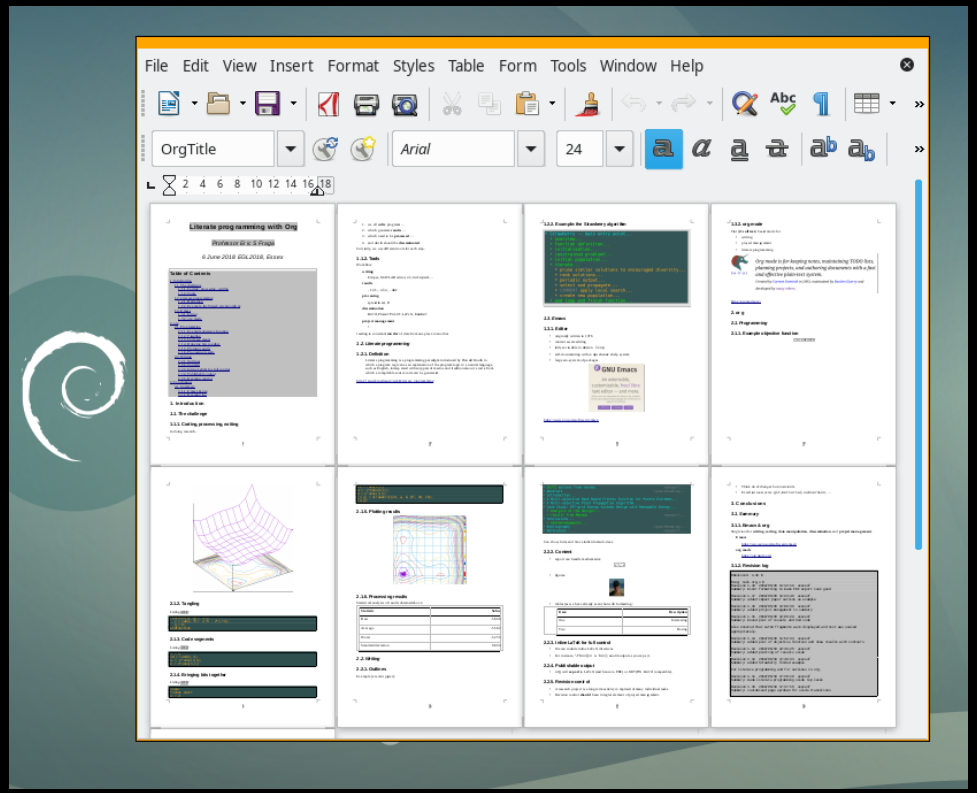
Example (a recent paper):

```
* [6/7] actions from review
* Abstract                                                          :ig
* Introduction...
* A Multi-objective Rank Based Fitness Function for Pareto E
* A Multi-objective Plant Propagation Algorithm...
* Case Study: Off-grid Energy Systems Design with Renewable
  * Analysis of the designs...
  * results from Mayowa
* Conclusions...
  * Acknowledgements...
* Bibliography                                                      :ig
* References
```

Can show, hide, and move individual sub-trees.

Publishable output

org will export to LaTeX (and hence to PDF) or ODT (MS Word compatible).



UCL

Project management

Support for tasks, scheduling, appointments:

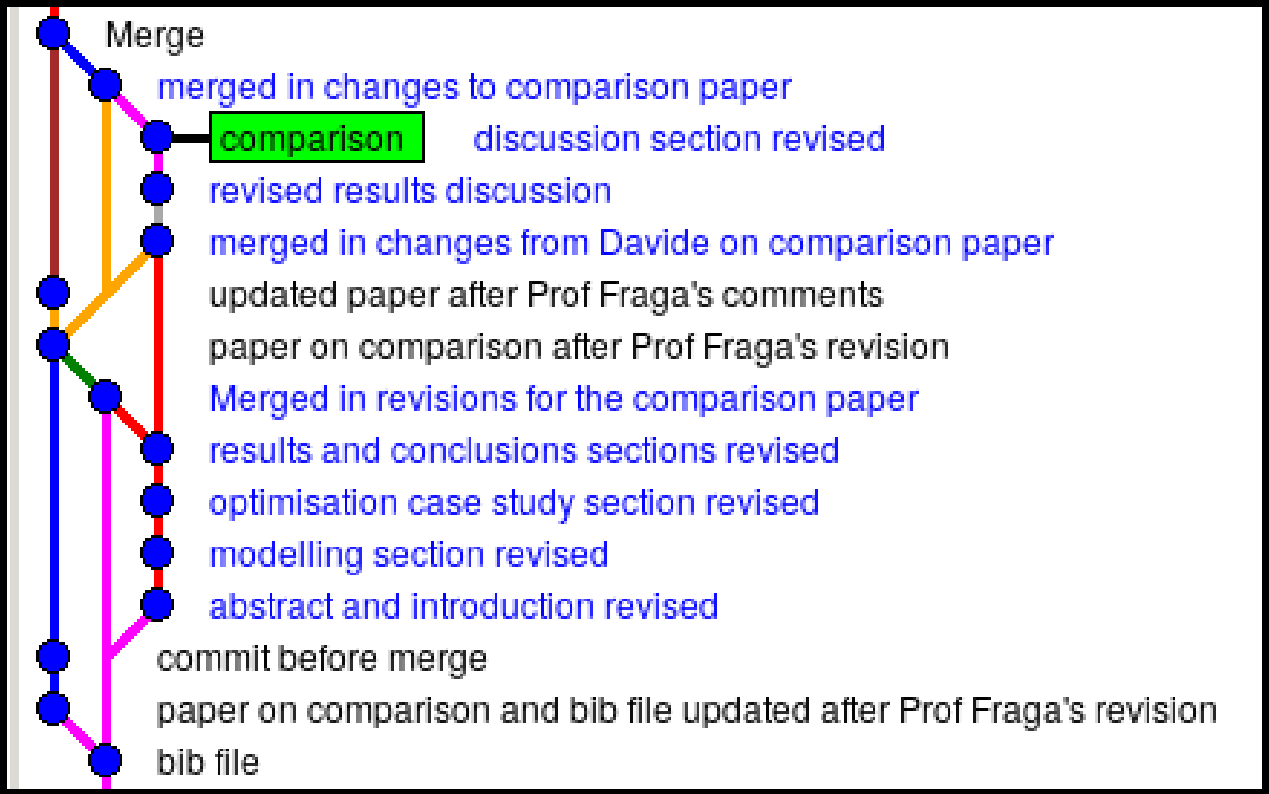
[3/4] prepare and give presentation on literate programming

- ☒ collect images
- ☒ write slides
- ☒ book hotel
- ☐ give presentation

Revision control

- A research project is a long term activity comprised of many individual tasks.
- Revision control should (must) be an integral element of project management.
- Think track changes but on steroids and which works for data as well.
- Excellent tools exist: `git`, `mercurial`, `subversion`, `...`

Example of revision control



Emacs & org

Single tool for writing, coding, data manipulation, data provenance, dissemination, and project management.

Testimonial I

> *By the age of 35 you should have realized that Emacs is the One True Editor and should have embraced it. If that's not the case - your life so far has been completely wasted.*

@bbatsov, 04:02 pm May 20, 2018

Testimonial II

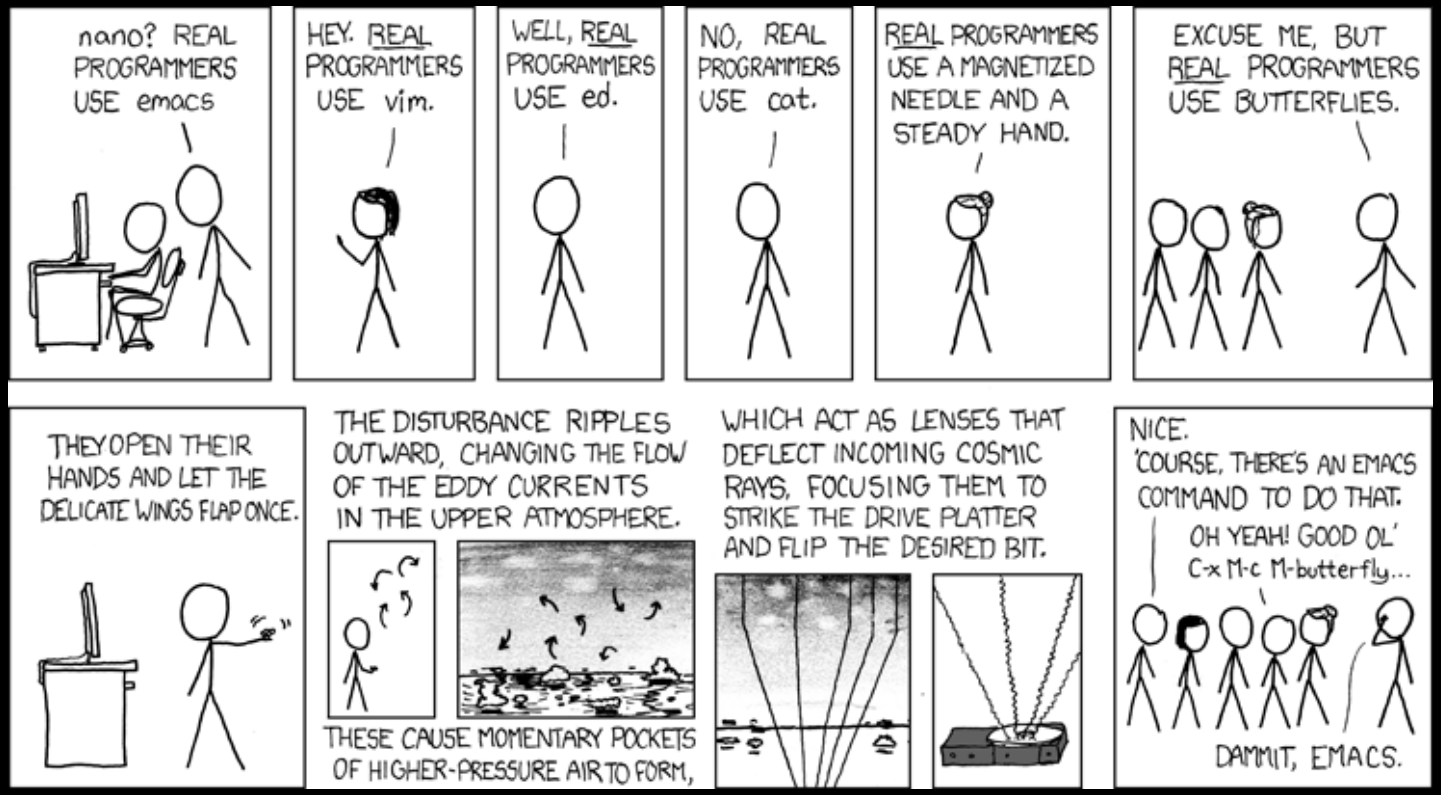
> *The advantages of plain text are hard to overstate, as is the advantage of having everything from plot notes to research material in a single (large) file under version control. And building up a novel from an outline is a natural process with org-mode.*

Bob Newell, `emacs.help` newsgroup, 2018-05-30.

Links

Emacs	https://www.gnu.org/software/emacs/
Complete computing environment	http://doc.rix.si/cce/cce.html
org mode	https://orgmode.org/
reproducible research	https://reproducibleresearch.net/links/
blog	reproducible research for management

And finally



https://www.xkcd.com/378/