Conducting Numerical Experiments

Tapabrata Ray
Multidisciplinary Design Optimization Group
University of New South Wales, Canberra

During your research, you will be developing algorithms and implementing them. You will be conducting a lot of numerical experiments and its extremely important you adopt a very structured and an efficient approach.

(a) You can opt to use MATLAB, Java, C++ or Python. It's your choice.
(b) If you need a lot of experiments to be done or computation that is likely to take long, think ahead how you would go about utilizing all cores of your machine. We do have 72, 40 and 32 core windows machines and ones with GPUs. Think about how you can use them to speed up things. If you have slow implementations, you will face serious problems in completing tasks on time.
(c) Most of the algorithms that you would be developing are likely to involve stochastic components or parameters. Benchmarking is an essential element of our research and you would need lot of runs with various parameter settings to establish and assess performance.
(d) To do above, it is advisable that you write wrappers that will fire algorithms for all problems for all different parameter settings. This process has to be done automatically as its impossible to fire jobs one at a time. Structuring result folders is of paramount importance. It is also of great importance that you store all information that might be required for future analysis. If you have not stored them, re-running would be terribly costly. Store all run information and data, so that you can replicate results if required.
(e) Post processing is the next logical step that follows from the above. Once again all tables, all figures should be created automatically without user intervention. Please take care in setting figure legends, axis labels etc etc along with figure file names. MATLAB may be useful to do post processing or for the wrapper that fires all jobs. Store .fig and .epsc files. You may use matrix2latex to convert matrices in MATLAB to direct Latex tables. Be mindful of the precision used to store results of runs. As for MATLAB, store in .mat files rather than fixed format .dat files.
(f) Having epsc files would be useful for paper drafts while fig files might be useful for PPTs or internal reports.
(g) Before running experiments, make sure you have a clear idea on what you want to compare and with what. Think through if the results would allow you to claim what you intend to. Think about what statistical analysis would be required to establish some of your claims.
(h) UNSW data retention policy requires all results, code algorithms to be stored and use the external hard drive provided to you for backups. You can also use MDO-NAS and your own backup system. Backup is your responsibility and make sure you have your own arrangements to deal with it.
(i) Contact me if you need scripts that can automatically fire jobs and post process results. You may need to customize some of it to suit your purpose.

Food for Thought: Design of Algorithms

Context: Are you developing an approach to solve a problem, which no one has attempted to solve? Are you developing an efficient algorithm? Are you developing an algorithm with greater flexibility?

- What is novel in your approach? Why is it likely to alleviate existing problems?
- How are you going to compare your results? With what? Are they state of the art methods?
- You may consider running your algorithm with parameter settings that can match reported results say for the same number of function evaluations or same convergence condition.
- Are there too many parameters in your proposed algorithm? If so, how do you justify the performance of the algorithm? The more parameters you have, the less attractive is your algorithm. Seriously think about reducing them to bare minimum.
- Can you logically justify the schemes within your algorithm? Why is certain things done that way and not in some other manner?
- Its not about presenting your results are better, it's about why is the performance better i.e. which component is influencing the performance and how do you show through analysis that it is indeed so.
-