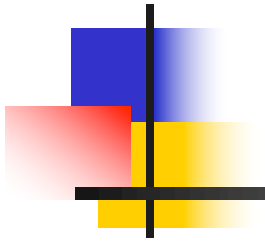
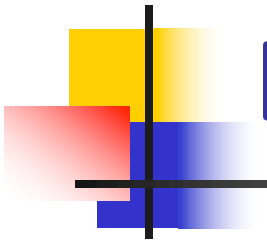
Simulation development: best practices



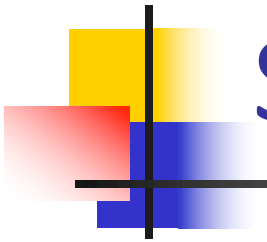


Designing the simulator

- before coding, clarify for the simulator
  - assumptions
  - inputs (e.g., parameters, files, etc)
  - outputs (e.g., performance metrics, files, etc)
  - data structures
  - algorithms
- concentrate on the question you want to address with your simulations

Simulation code

- comment your code as much as you can
  - helpful not only for the others but also for the programmer
- print the simulator input parameters
  - helpful to recover the settings adopted in the simulation when analyzing the data
- debug very carefully under very simple toy cases
- for long simulations, useful to plot some real-time information to understand if the simulator keeps running as expected



Simulation architecture

- distinguish between
  - code for the simulation
  - code for the output processing and graph plotting
- the simulation generates output files with the data to be processed
  - typically, run just once the simulator and then run multiple times the code for processing and plotting

Input parameters

- choose carefully the input parameters to get meaningful results, i.e., in a reasonable range
  - avoid extreme/degenerate cases
  - requires to get an a-priori intuition of how the original system works

Validation of the simulator

- the most difficult and most important step
- run and study the results under
  - toy cases or corner cases in which the behavior can be easily predicted
  - cases for which some analytical formula is known

Simulation duration

- how long to run each simulation? how many runs?
- the only correct answer: it depends
- only by checking the confidence intervals it is possible to know if the accuracy is enough and thus if the duration was enough
- much much better short simulations than long ones!!!
  - given that the desired accuracy is achieved

Simulation duration

- two phases
  - simulation development, debugging and validation
    - very short simulations (few seconds/minutes) to get immediate results
    - preliminary results are very useful for debugging and validation
  - “in production” simulation
    - enough long simulations (seconds/minutes/hours/day) to reach the desired accuracy
    - to get the results for the report, the thesis, the scientific paper, etc.

Common mistakes

- too long simulations
  - use confidence intervals to understand the accuracy
- too complex simulation model with respect to the required study
  - useless/counterproductive to keep any detail in the model that has no relevance in the final result
- too complex scripting pipes
  - if anything gets wrong, the output is unpredictable and it is hard to understand that a problem occurred

Graphs

- any graph must **always** report the label on the x-axis and y-axis, with the corresponding unit of measurement
- in the case of multiple curves, each of them must be labeled
- curve style
  - line-points more appropriate for simulation/experimental data
  - lines more appropriate for theoretical curves