Homework Guide
CMS/ACM 104, Fall 2014

A manifesto: The research is only half the battle. Your job as a scientist or engineer is not complete until you have presented your results and discussed them in a clear and compelling manner, addressing both the advantages and disadvantages of your approach. Some researchers—to their own discredit—regard this step as irrelevant or burdensome. Therefore, one goal of this class is to give you practice in communicating. Once you have developed good habits, you will find that you can present your work elegantly and effectively without extra effort. Not only will your results make the greatest possible impact, but you will quietly convince your colleagues of your thoughtfulness and professionalism. To that end, I request that you consider the guidelines below as you prepare your assignments. These rules are not intended to create unnecessary work for you; they are central to the practice of good science.

Collaboration and citations: First, try to solve all the problems yourself. At the same time, you are *encouraged* to discuss assignments with other students. Moreover, the use of additional resources from the library or the internet is *recommended*. With this freedom comes responsibility. You are *required* to do the following.

- You must prepare the actual assignment that you submit by yourself.
- You must list on your assignment the names of other students you have collaborated with to solve the problems.
- You must provide complete citations to any external resources that you have used to complete an assignment.
- You must write all computer code for your assignment by yourself and include a copy in your homework solution.

Failure to abide by these rules is considered *academic dishonesty*, which will be referred to the Board of Control (BoC) or the Graduate Honor Council (GHC).

Basic guidelines: Those who read your work (probably) cannot read your mind. Attention to the following issues facilitates the transfer of information.

- Please write in complete, grammatical English sentences.
- You should answer each question completely but concisely. You should not be so terse that your work is cryptic, but the length of your assignment should not be comparable with that of a 19th century Russian novel. Use common sense.
- Equations intimidate everybody, including the graders, who are instructed to ignore unclear work. An unannotated calculation does not constitute an appropriate response to a question. At the minimum, you need a sentence at the beginning to say what you are doing and a sentence at the end to summarize your conclusion.
- Computing one example does not constitute a mathematical proof.
- Retain an appropriate number of significant digits, and provide units for all quantities.
- Prepare your assignments neatly. Pages must be in order. Leave ample margins. Mass erasures and crossouts are inappropriate. Secure pages with one staple in the upper corner.
- Do not waste paper. When appropriate, put multiple plots on one page. Print in duplex.

Preparation of graphs: Graphs should be numbered, and you should refer to the graph in the text of your answer. How else does the reader know that it exists and that she should look for it as support for your argument? All graphs must be labeled completely and correctly. See Figure 1 for an example.

- Provide a title that states what the figure represents.
- Each axis must be labeled with the name of the variable, its symbol, its units (unless unitless), and (if necessary) its scale.
- Axes should have a reasonable number of tick marks, labeled with appropriate numerical values. Choose the limits of the axes so that the plot highlights the region of interest.
- Add a legend that notes every data series and its corresponding marker. The legend may also include functions (such as regression lines) that are plotted along with their line styles. The legend should not eclipse any of the data.
- Provide a short caption that summarizes what information the figure conveys. It should include equations for regression lines if these do not appear in the legend. It would not be amiss to state the values of other statistics, such as correlation coefficients, that aid interpretability.
- Ideally, graphs should be comprehensible even when reproduced in black and white. Use (simple) markers to distinguish data series. Use line styles or weights to distinguish functions. When color must be deployed, be aware that a significant proportion of the population cannot distinguish red and green.
- For quantized data, you may want to jitter the data values slightly so that data points are not plotted on top of each other. This method has the advantage that local data density can be surmised from a glance.
- Tufte (qtd. in Wainer 1984) proposes a valuable principle for presenting data: Maximize the *data density index*, "the number of numbers plotted per square inch." A related concept is the *data-ink ratio*, which measures the amount of information presented relative to the total ink spilled. In other words, do not put anything in a graph or table that neither transmits information nor assists comprehension. Some ideas for maximizing these quantities:
  - Data typically consist of point measurements. When you "connect the dots," you may be creating visual clutter.
  - Regression lines carry very little information. They should usually be plotted *underneath* the data points, unless the data would efface the line completely. When regression lines are plotted on top of data, one should take care not to obscure the data.
  - Placing a grid in the area of the plot is not recommended. If you deem the grid necessary, you should plot it using hairlines *underneath* the data.
  - You may wish to consider whether another display method is superior to a graph.
- Computer-generated labels are preferable, but legible hand-written labels are acceptable.

Tables: Tables should be numbered and referenced in the text. Similar with the case for graphs, tables should also be labeled completely. See Table 1 for an example.

- Tables require titles and captions.
- All appropriate rows and columns must be labeled with the name of the variable and its symbol, with units and scale if necessary.
- Use scientific notation to improve numerical displays. Decimal points should always be aligned in a column.
- Never display insignificant digits. In fact, tables are typically clearer when you round to as few digits as possible.
- You may also wish to consider whether another display method is superior to a table.

Resources: The following references discuss some of the issues involved in presenting information.

- E. R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed., Graphics Press, 2001.
- H. Wainer, "How to present data badly," *Amer. Statistician*, Vol. 38, No. 2, May 1984.
- D. Solow, *How to Read and Do Proofs*, 6th ed., Wiley, 2013.

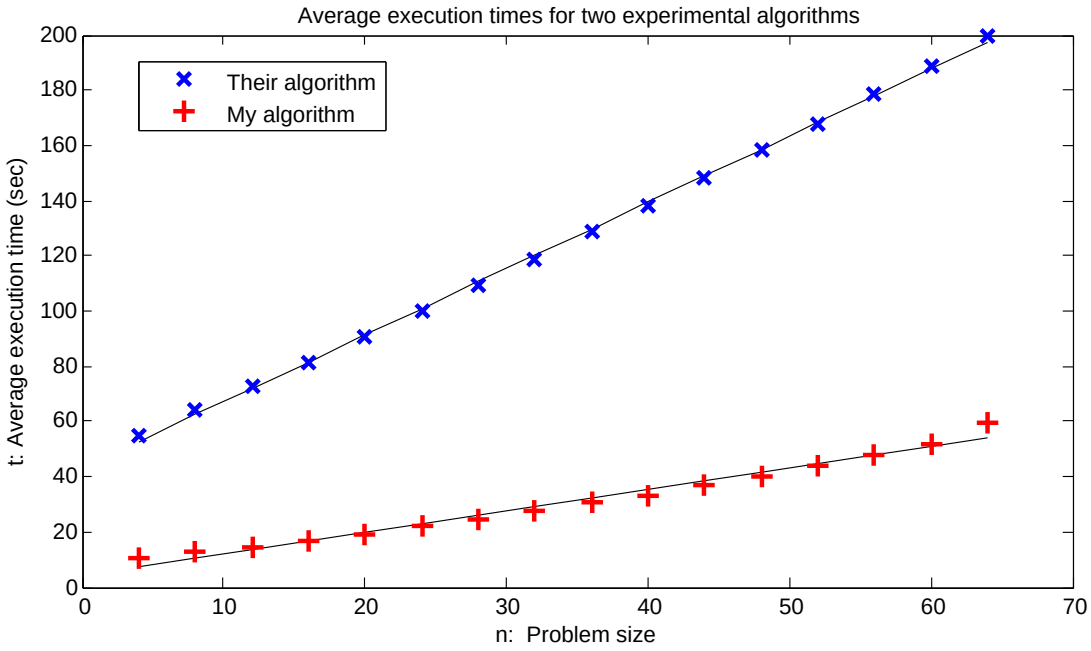

FIGURE 1. EXECUTION TIMES FOR TWO ALGORITHMS. Each data point represents the average of 1000 trials. The equations of the regression lines for their algorithm and my algorithm are respectively $t = 2.4n + 43$ and $t = 0.77n + 4.5$.

TABLE 1. PACKING POINTS ON THE SPHERE: The packing diameter ρ of a set of N points on the sphere is the least angle subtended by a pair of distinct points. This table compares the largest known packing diameter of N points (Sloane 2008) with the packing diameter of the best configuration produced by an experimental algorithm (Tropp et al. 2007).

	Packing diameter (ρ)		
Points (N)	Best known	Calculated	Difference
4	109.471°	109.471°	0.001°
5	90.000	90.000	0.000
6	90.000	90.000	0.000
7	77.870	77.869	0.001
8	74.858	74.858	0.001
9	70.529	70.528	0.001
10	66.147	66.140	0.007
11	63.435	63.434	0.001
12	63.435	63.434	0.001
13	57.137	57.136	0.001
14	55.671	55.670	0.001
15	53.658	53.620	0.038
16	52.244	52.243	0.001