

# MATH 166: READING ASSIGNMENT OVERVIEW

ETHAN LEVIEN

## 1. INSTRUCTIONS FOR PAPER ASSIGNMENT

**1.1. Selecting a paper.** Selecting which papers to read can be a difficult task when you are new to research, especially in interdisciplinary fields where the relevant papers are distributed across many different sub-communities. How you select a paper will depend heavily on where you are in your career. If have been doing research for a few years and want to present on a paper closely related to your current research, that is okay. In that case, you probably have an idea of which papers are relevant to your ongoing projects. Perhaps there are some papers that recently came out following up on your work.

If you are a new grad student, or are looking to dip into a new topic, then I recommend you find a paper which is well established in the literature. By this I mean that it is highly cited for how long it has been published. Citation counts are not everything and many great papers are poorly cited, but they are a reflection of the influence a paper had on subsequent scientific thinking. Thus, these papers tend to be nice entry points. One strategy for finding such papers is to search for a topic you are interested in on google scholar and read the introductions of some recent papers. If they are good papers, they should summarize the literature and point you to some more foundational papers in the field.

I will provide additional guidance on how to select papers in class, but to help you get started, see the list in Section 2. You might also ask your advisor for suggestions. If nothing on that list speaks to you I am happy to help find something else, provided it is not too far from my area of expertise.

**1.2. Tips for reading.** Here is a rough guide for reading scientific papers. Full disclosure: I've never consciously followed this myself, but it does seem to follow how I approach reading papers.

- **Initial reading** In an initial reading of a scientific text it can be easy to get hung up on things we don't understand. I encourage you to do an initial reading through the entire paper without getting hung up on technical details. If something doesn't make sense after a couple minutes of thought and some googling, simply mark it as such and move in. After an initial reading, you'll have a better sense for the bigger picture and what points earlier in the paper may be worth diving into. We often find that papers have many technical points which are irrelevant to our intentions or the results. You should also get a sense for how the paper is situated within the literature, i.e. what key papers come before. After the first reading, try to summarize the results on a conceptual level in your own words.
- **Deep dive into technical points** Now return to the paper for a second reading in which you determine what technical points you need to understand to arrive at the results. Pick at least one of these and really dive into them. You likely won't be able to dive deeply into EVERY technical detail, but that's okay. This may involve one or more of the following:
  - Writing code to run a simulation described in the paper. In some cases it may be necessary to start with a simplified version of the simulation, for example starting with the 1D case when the simulations in the paper are performed in 3D. That's okay. The point is to at least make some progress towards reproducing the results in the paper.
  - Learning the mathematical background needed to understand a calculation. For example, if Hamilton-Jacobi equations are essential in order to understand the results in your paper, you could read a chapter of a textbook or review article on Hamilton-Jacobi equations and connect it to the paper. It's okay to incorporate other readings into your presentation!
- **Connect to big picture** Reflect on the connection between the technical details and the big picture of the paper. For example, are there hidden assumptions that emerge only after you tried

to reproduce a simulation or a calculation? What are the implications for their key results? After carrying out a calculation, do you better understand what is challenging about the problem and why a new approach was needed? I've found that this process of connecting technical aspects of a paper to the conceptual points is where new research directions often emerge.

- **Future work** What questions do you have that are not contained in the paper? What steps might you take to answer those questions? Did you get what you expected from this paper? It's okay to be critical. Next take a look at papers which have cited this paper and read a few of the abstracts – have they answered the questions you have posed?

**1.3. Report.** Your report will be a 3-5 page write up of your experience reading the paper. The report should NOT be a summary of the results in the paper (although it should include an overview), but rather a documentation of your understanding. In it, you should answer the questions:

- Why did you select this paper?
- What was the authors motivation to write this paper and how is it related to the existing literature?
- What things did you initially not understand?
- What have you done to develop the background to read the paper?
- What progress have you made in understanding this paper
- What are some interesting questions that are not answered within the paper?

## 2. PAPERS

This is list of papers that might be good entry points into the literature and are consistent with the theme of the course (really they are just topics I happen to be interested in and feel I can provide guidance on). You might try reading the intro of a few in your area of interest and checking out some of the references. **These are NOT meant to be the most important or influential papers, rather they are just a random selection that I happen to be curious about.** I am happy to provide more personalized guidance.

### 2.1. Data science.

#### 2.1.1. *Machine learning.*

- [1] I believe this was the first paper to identify, in general, the so-called “double descent” phenomena. Double descent refers to the plot of generalization error vs. the ratio of data points and model complexity.
- [2] A more recent physics-y paper which explains WHY the double descent phenomena occurs in terms of geometry and random matrix theory and analogies to statistical mechanics. The transition between the bias-variance trade-off regime and the over parameterized machine learning regime can be understood as a phase transition.
- [3] This paper observes a different type (I think) of phase transition in many algorithms which happens to be related to projections of random polytopes. Very interesting math!

#### 2.1.2. *Bayesian statistics.*

- [4] A fundamental question in Bayesian statistics is how to select priors. It concerns how this can optimally be done when there is finite data.
- Try looking at Andrew Gelman's google scholar page.

#### 2.1.3. *Computation.*

- [5] An introduction to Hamiltonian Markov chain Monte Carlo, the most widely used Markov chain Monte Carlo method in bayesian statistics. See also a more technical introduction by the same author.

### 2.2. Biology.

2.2.1. *Ecology and evolution.*

- [6] A central question in biology is to what extent we can predict the fate of a mutation: will it go extinct, coexist with its host or takeover the population (fixate)? This paper concerns the question of how environmental fluctuations influence the chance for a mutation to fixate.
- [7] A very interesting paper, which was one of the first to incorporate biophysical constraints of gene regulation into evolution dynamics models. A major interest of mine!
- [8] This is a classic paper on the topic of phenotypic switching, the phenomena where genetically identical organisms switch between different phenotypes (e.g. growth rates) to persist in uncertain environments. Their results are closely related to the classical notion of bet-hedging [9].
- [10] Another paper on bet-hedging which concerns the relationship between so-called within and between generation hedging. The idea is that organisms can hedge their bets against variation in time or space.

2.2.2. *Single-cell physiology.*

- [11] Influential paper concerning the question of how cells maintain homeostasis of their sizes and understanding the relationship between variation in cell-cycle duration, growth rate and size.

2.3. **Probability theory and stochastic processes.**2.3.1. *Stochastic hybrid systems.*

- [12] This paper concerns what happens when you have a linear ODE, but the matrix on the right hand side is randomly switching.
- [13] And this is about switching PDEs

2.3.2. *Chemical reaction network theory.*

- [14] Seminal paper proving an important theorem which states that a wide class of chemical reaction networks have “simple behavior” in the long run. This is an extension of the deficiency zero theorem, which makes a similar statement for deterministic chemical reaction networks.

## REFERENCES

- [1] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, jul 2019.
- [2] Jason W Rocks and Pankaj Mehta. Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models. *Physical Review Research*, 4(1):013201, 2022.
- [3] David Donoho and Jared Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–4293, 2009.
- [4] Henry H Mattingly, Mark K Transtrum, Michael C Abbott, and Benjamin B Machta. Maximizing the information learned from finite data selects a simple model. *Proceedings of the National Academy of Sciences*, 115(8):1760–1765, 2018.
- [5] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. 01 2017.
- [6] Ivana Cvijović, Benjamin H. Good, Elizabeth R. Jerison, and Michael M. Desai. Fate of a mutation in a fluctuating environment. *Proceedings of the National Academy of Sciences*, 112(36):E5021–E5028, Aug 2015.
- [7] Michael Lässig. From biophysics to evolutionary genetics: statistical aspects of gene regulation. *BMC Bioinformatics*, 8(Suppl 6):S7, 2007.
- [8] Edo Kussell and Stanislas Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.
- [9] John. L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, Jul 1956.
- [10] Sebastian J Schreiber. Unifying within-and between-generation bet-hedging theories: an ode to jh gillespie. *The American Naturalist*, 186(6):792–796, 2015.
- [11] Ariel Amir. Cell size regulation in bacteria. *Physical Review Letters*, 112(20), 2014.
- [12] Sean D Lawley, Jonathan C Mattingly, and Michael C Reed. Sensitivity to switching rates in stochastically switched odes. *arXiv preprint arXiv:1310.2525*, 2013.
- [13] Sean D Lawley. Boundary value problems for statistics of diffusion in a randomly switching environment: Pde and sde perspectives. *SIAM Journal on Applied Dynamical Systems*, 15(3):1410–1433, 2016.
- [14] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. 03 2008.