How to Read a Paper

Sam Hopkins, MIT

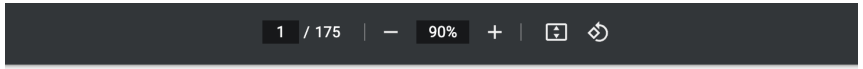
With thanks to: Tim Roughgarden, Srinivasan Keshav, Let-All Organizers

Resources:

https://cs.stanford.edu/~rishig/courses/ref/paper-reading-overview.pdf

https://cs.stanford.edu/~rishig/courses/ref/paper-reading-technical.pdf

http://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/07/paper-reading.pdf



Reducibility and Statistical-Computational Gaps from Secret Leakage

Matthew Brennan * Guy Bresler †

June 30, 2020

arXiv:2005.08099v2 [cs.CC] 28 Jun 2020

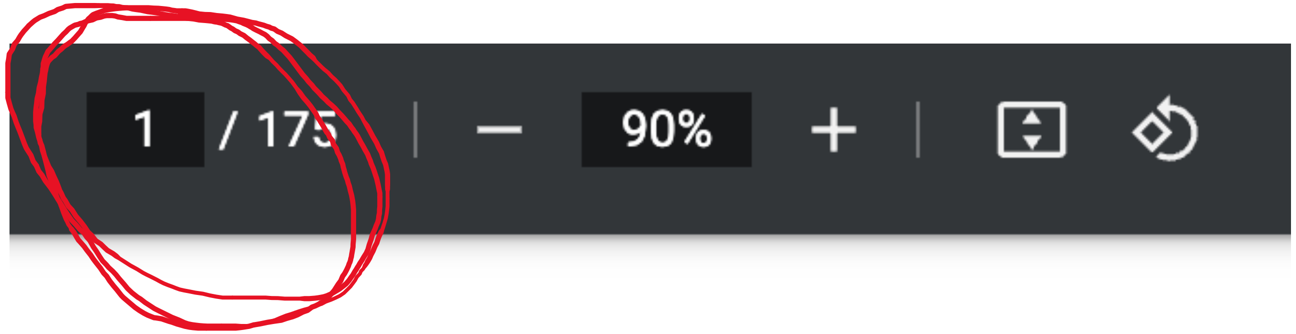
Abstract

Inference problems with conjectured statistical-computational gaps are ubiquitous throughout modern statistics, computer science, statistical physics and discrete probability. While there has been success evidencing these gaps from the failure of restricted classes of algorithms, progress towards a more traditional reduction-based approach to computational complexity in statistical inference has been limited. These average-case problems are each tied to a different natural distribution, high-dimensional structure and conjecturally hard parameter regime, leaving reductions among them technically challenging. Despite a flurry of recent success in developing such techniques, existing reductions have largely been limited to inference problems with similar structure – primarily mapping among problems representable as a sparse submatrix signal plus a noise matrix, which is similar to the common starting hardness assumption of planted clique (PC).

The insight in this work is that a slight generalization of the planted clique conjecture – secret leakage planted clique (PC_ρ), wherein a small amount of information about the hidden clique is revealed – gives rise to a variety of new average-case reduction techniques, yielding a web of reductions relating statistical problems with very different structure. Based on generalizations of the planted clique conjecture to specific forms of PC_ρ, we deduce tight statistical-computational tradeoffs for a diverse range of problems including robust sparse mean estimation, mixtures of sparse linear regressions, robust sparse linear regression, tensor PCA, variants of dense k-block stochastic block models, negatively correlated sparse PCA, semirandom planted dense subgraph, detection in hidden partition models and a universality principle for learning sparse mixtures. This gives the first reduction-based evidence supporting a number of statistical-computational gaps observed in the literature [Li17, BDLS17, DKS17, CX16, HWX15, BBH18, FLWY18, LSLC18, RM14, HSS15, WEAM19, ASW13, VAC17].

We introduce a number of new average-case reduction techniques that also reveal novel connections to combinatorial designs based on the incidence geometry of $\mathbb{F}_r^t$ and to random matrix theory. In particular, we show a convergence result between Wishart and inverse Wishart matrices that may be of independent interest. The specific hardness conjectures for PC_ρ implying our statistical-computational gaps all are in correspondence with natural graph problems such as k-partite, bipartite and hypergraph variants of PC. Hardness in a k-partite hypergraph variant of PC is the strongest of these conjectures and sufficient to establish all of our computational lower bounds. We also give evidence for our PC_ρ hardness conjectures from the failure of low-degree polynomials and statistical query algorithms. Our work raises a number of open problems and suggests that previous technical obstacles to average-case reductions may have arisen because planted clique is not the right starting point. An expanded set of hardness assumptions, such as PC_ρ, may be a key first step towards a more complete theory of reductions among statistical problems.

*Massachusetts Institute of Technology. Department of EECS. Email: brennanm@mit.edu.
†Massachusetts Institute of Technology. Department of EECS. Email: guy@mit.edu.



Goal: Provide you with tools to read papers **productively and efficiently**

Assumption: theory papers

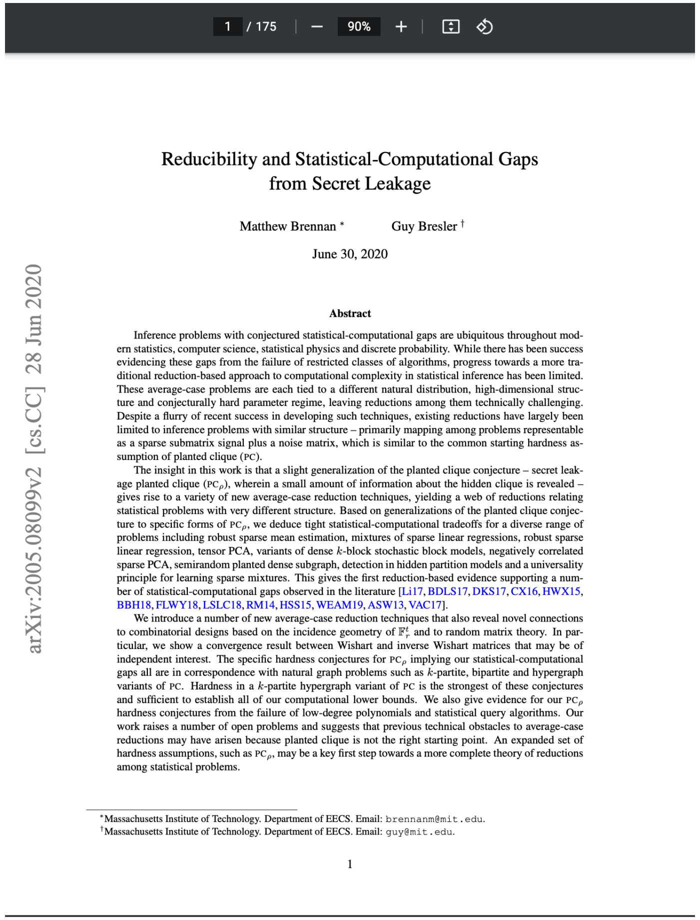
Part 1: General Principles

Part 2: Iterative Refinement & The Big Questions

Part 3: Reading Proofs

Part 1: General Principles

1. Papers are not novels


1 / 175 – 90% +

arXiv:2005.08099v2 [cs.CC] 28 Jun 2020

Reducibility and Statistical-Computational Gaps
from Secret Leakage

Matthew Brennan [*] Guy Bresler [†]

June 30, 2020

Abstract

Inference problems with conjectured statistical-computational gaps are ubiquitous throughout modern statistics, computer science, statistical physics and discrete probability. While there has been success evidencing these gaps from the failure of restricted classes of algorithms, progress towards a more traditional reduction-based approach to computational complexity in statistical inference has been limited. These average-case problems are each tied to a different natural distribution, high-dimensional structure and conjecturally hard parameter regime, leaving reductions among them technically challenging. Despite a flurry of recent success in developing such techniques, existing reductions have largely been limited to inference problems with similar structure – primarily mapping among problems representable as a sparse submatrix signal plus a noise matrix, which is similar to the common starting hardness assumption of planted clique (PC).

The insight in this work is that a slight generalization of the planted clique conjecture – secret leakage planted clique (PC_ρ), wherein a small amount of information about the hidden clique is revealed – gives rise to a variety of new average-case reduction techniques, yielding a web of reductions relating statistical problems with very different structure. Based on generalizations of the planted clique conjecture to specific forms of PC_ρ, we deduce tight statistical-computational tradeoffs for a diverse range of problems including robust sparse mean estimation, mixtures of sparse linear regressions, robust sparse linear regression, tensor PCA, variants of dense k-block stochastic block models, negatively correlated sparse PCA, semirandom planted dense subgraph, detection in hidden partition models and a universality principle for learning sparse mixtures. This gives the first reduction-based evidence supporting a number of statistical-computational gaps observed in the literature [Li17, BDLS17, DKS17, CX16, HWX15, BBH18, FLWY18, LSLC18, RM14, HSS15, WEAM19, ASW13, VAC17].

We introduce a number of new average-case reduction techniques that also reveal novel connections to combinatorial designs based on the incidence geometry of $\mathbb{F}_r^t$ and to random matrix theory. In particular, we show a convergence result between Wishart and inverse Wishart matrices that may be of independent interest. The specific hardness conjectures for PC_ρ implying our statistical-computational gaps all are in correspondence with natural graph problems such as k-partite, bipartite and hypergraph variants of PC. Hardness in a k-partite hypergraph variant of PC is the strongest of these conjectures and sufficient to establish all of our computational lower bounds. We also give evidence for our PC_ρ hardness conjectures from the failure of low-degree polynomials and statistical query algorithms. Our work raises a number of open problems and suggests that previous technical obstacles to average-case reductions may have arisen because planted clique is not the right starting point. An expanded set of hardness assumptions, such as PC_ρ, may be a key first step towards a more complete theory of reductions among statistical problems.

[*] Massachusetts Institute of Technology. Department of EECS. Email: brennanm@mit.edu.
[†] Massachusetts Institute of Technology. Department of EECS. Email: guy@mit.edu.

1

≠

1. Papers are not novels

- Do not read papers **linearly**!
- Often can **ignore or skip large sections**
- OTOH, often need to read **key sections/paragraphs/sentences many times**

2. Goals and Context

What do you want to **get out of the paper**?

Understand the main problem?

Understand the main result?

Potentially adopt a proof technique?

How much **background** do you have in the area?

None? May need to read a textbook or consult related work first

Lots? May be able to skip introduction entirely

(And may possibilities in-between)

3. Don’t read papers (😉)

- Course notes
- Blog posts
- Twitter threads
- Recorded lectures from courses, conferences, seminars
- Slides
- PhD theses
- Survey articles

Are often more approachable resources
(recent result? Need details? You may be out of luck…)


Berkeley
UNIVERSITY OF CALIFORNIA
Powered by Zoom
37:11 / 1:06:50

Where to look for resources

- Authors’ websites
- Simons TV (https://simons.berkeley.edu/videos)
- Other regular recorded seminars/workshop venues (TCS+, MIFODS, BIRS,…)
- Recorded conference talks (STOC, FOCS, COLT, NeurIPS, ICML, COLT, ALT, …)
- **Google is your friend**

Part 2: Iterative Refinement & The Big Questions

Basic workflow

$$\begin{tikzcd} \text{Read} \arrow[rr, bend left] & & \text{Think} \arrow[ll, bend left] \end{tikzcd}$$

The Big Questions (Roughgarden)

1. **What problem is the paper trying to solve?**
2. **Why is the problem interesting?**
3. **What is the primary contribution?**
4. **How did they do it?**
5. **What are the key take-aways?**

What Problem is the Paper Trying to Solve?

- Could be a well-defined mathematical/algorithmic problem:
  - *Show that planted clique reduces to robust sparse linear regression*

- Or, could be less well-defined:
  - *Find a sound mathematical model for observed empirical phenomenon X*

Why is the Problem Interesting?

- Progress can lead to new algorithms?
- Progress leads to improved understanding of observed phenomenon?
- Progress leads to improvements in practice?
- …

- All research can be criticized – ”glass half full” is important

What is the Primary Contribution?

- New algorithm?
- New lower bound?
- New model?
- …

- Often captured by a theorem or a definition or a combination of a small number thereof

Pass 1: First Contact

So you found a paper you might want to read:

- Popped up on twitter/arxiv/etc
- Showed up list of accepted papers
- Your friend sent it to you
- Etc.,

The 5-30 Minute Assessment

Goals:

1. **What problem is the paper trying to solve?**
2. **Why is the problem interesting?**
3. **What is the primary contribution?**
4. **How did they do it?**
5. **What are the key take-aways?**

The 5-30 Minute Assessment

1. Read the 1-2 paragraphs
   Skip the abstract! Typically written for experts

2. Read the main subject headers/table of contents, look for anything called **main question** or similar

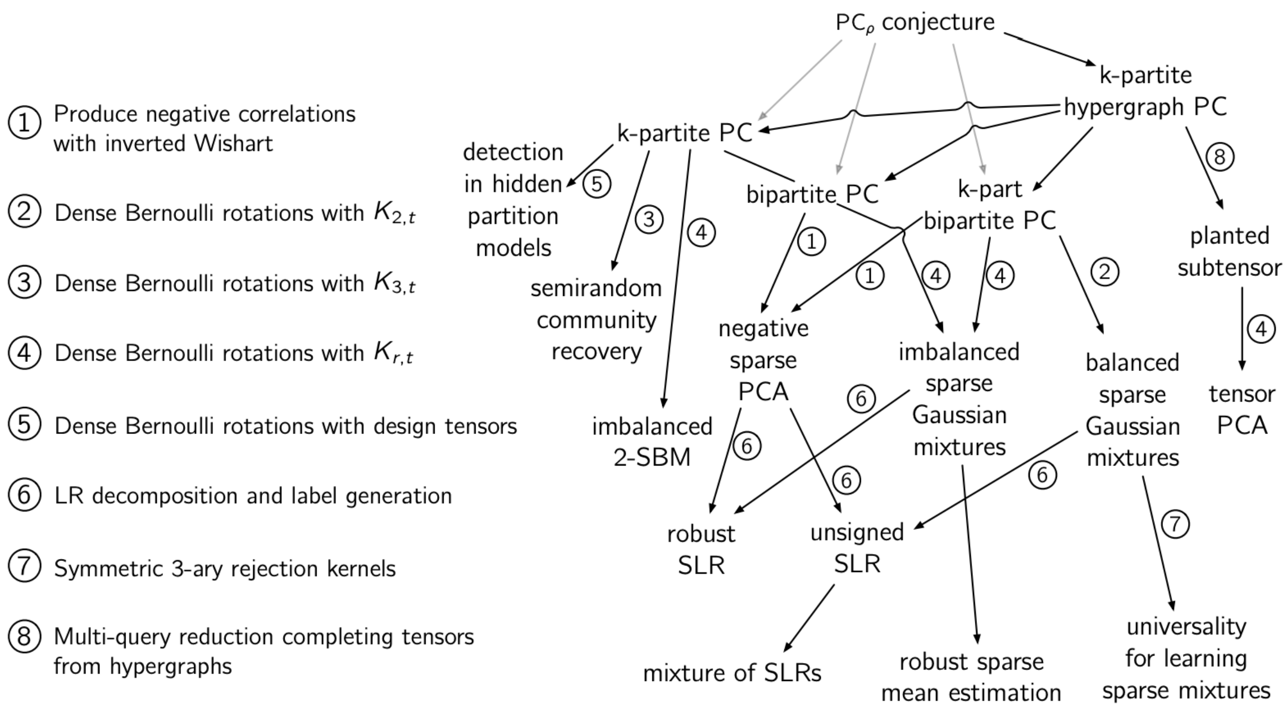
3. Read the conclusions (not always useful in theory papers)

4. Look at figures (if any)

1 Introduction

Computational complexity has become a central consideration in statistical inference as focus has shifted to high-dimensional structured problems. A primary aim of the field of mathematical statistics is to determine how much data is needed for various estimation tasks, and to analyze the performance of practical algorithms. For a century, the focus has been on *information-theoretic* limits. However, the study of high-dimensional structured estimation problems over the last two decades has revealed that the much more relevant quantity – the amount of data needed by *computationally efficient* algorithms – may be significantly higher than what is achievable without computational constraints. These *statistical-computational gaps* were first observed to exist more than two decades ago [Val84, Ser99, DGR00] but only recently have emerged as a trend ubiquitous in problems throughout modern statistics, computer science, statistical physics and discrete probability [BB08, CJ13, JM15]. Prominent examples arise in estimating sparse vectors from linear observations, estimating low-rank tensors, community detection, subgraph and matrix recovery problems, random constraint satisfiability, sparse principal component analysis and robust estimation.

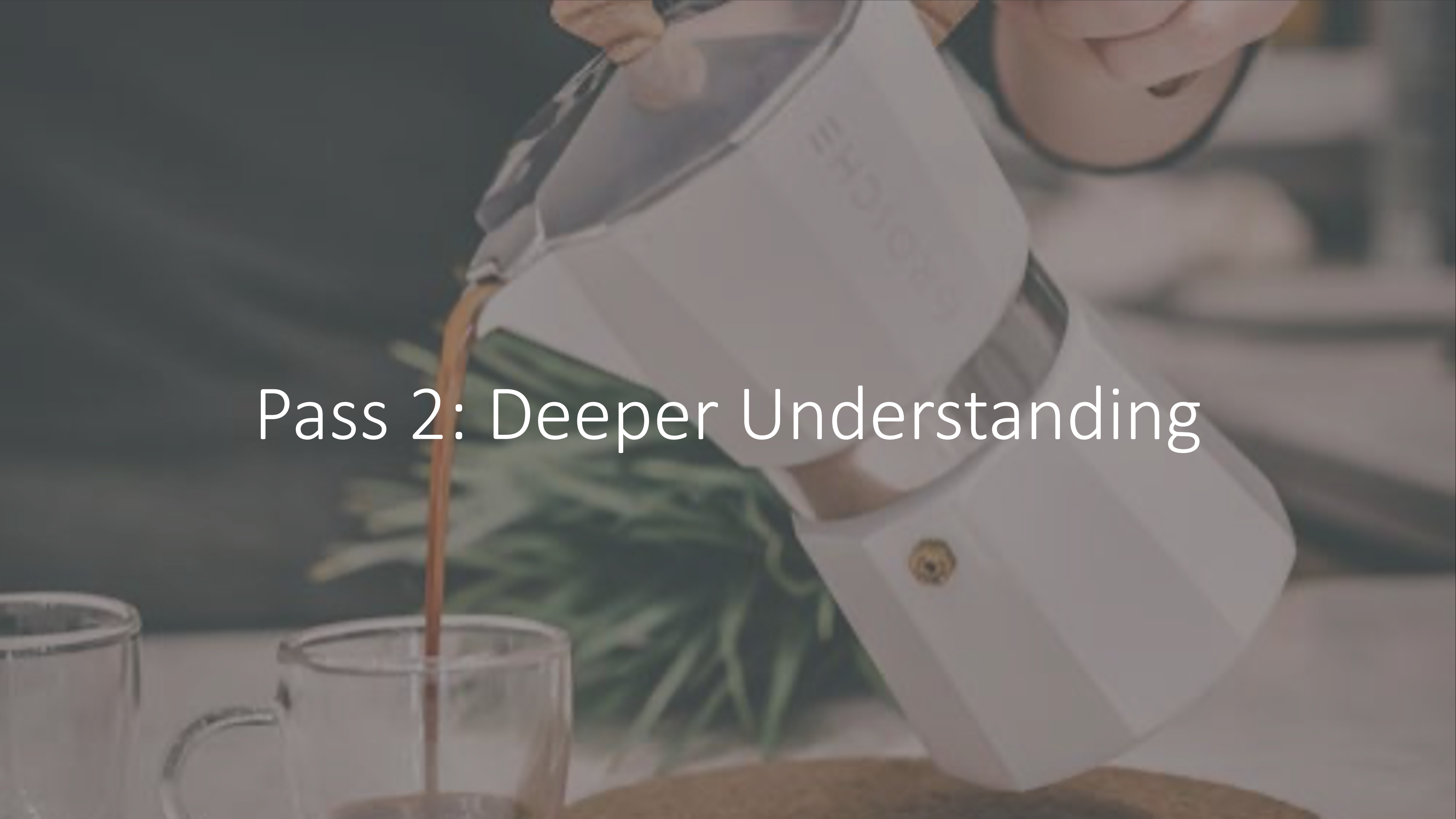
Question 1.1. *Can statistical-computational gaps in problems with different high-dimensional structures be related to one another through average-case reductions?*



After the first assessment:

- Identify/revisit your goals
- Look for additional resources

And, if you decide to read the paper “for real”, make a cup of coffee…



Pass 2: Deeper Understanding

The 2^{nd} pass(es) – several hours

- Read the introduction in detail
- Read "technical overview"/ "proof sketch" or similar
- Read related work section

(More on this later)

- Pick apart key theorem statements: what do they say?
  - Identify interesting **examples**
  - Improvement over "obvious"/"baseline" approaches
  - Simplest nontrivial **special cases**
  - Probably still ignore proofs (but maybe read proof sketches)

- Identify concepts that can be "black-boxed" so you don't have to worry about them right now

After the 2^{nd} Pass(es)

1. **What problem is the paper trying to solve?**
2. **Why is the problem interesting?**
3. **What is the primary contribution?**
4. **How did they do it?**
5. **What are the key take-aways?**

Part 3: How did they do it?

a.k.a.: reading challenging technical proofs

Principles

- Proofs are trees
- Most proofs have **1-2 key ideas**: find them!
- Concepts >> details
  - But, **concepts can be quantitative**
- Prove it for yourself – treat the paper as a series of hints
- Chekov’s gun
  - Assumptions must be used! What happens when assumptions break down?
- Examples & special cases
  - **Simple simple simple**

Expanded Workflow (in no particular order)

- Read a line of math
- Write proof on paper/whiteboard
- Discuss with/explain to friend/colleague
- Work through small example
- Consult textbooks, referenced papers, course notes,…

Words of Warning & Encouragement

- Research papers, esp. conference versions, have a lot of **mistakes**
  - Most are typos/easily fixed! Try to fix for yourself

- Don’t get discouraged! Understanding technical details can take days, weeks, months…**even for senior researchers/area experts**!

- For most technical papers, only a few people **in the world** understand them at a deep level. Understanding technical paper = almost unique superpower!

Wrapping up

- What & how you read depends on your **goals**
- Always know what question(s) you're answering!
- Refine your understanding iteratively
- Reading proofs is mostly not about reading

- **Good luck!!**