DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.
Abstract

Recently, research in unsupervised learning has gravitated towards exploring statistical-computational gaps induced by sparsity. A line of work initiated by Berthet and Rigollet has aimed to explain these gaps through reductions to conjecturally hard problems from complexity theory. However, the delicate nature of average-case reductions has limited the development of techniques and often led to weaker hardness results that only apply to algorithms that are robust to different noise distributions or that do not need to know the parameters of the problem. We introduce several new techniques to give a web of average-case reductions showing strong computational lower bounds based on the planted clique conjecture for planted independent set, planted dense subgraph, biclustering, sparse rank-1 submatrix, sparse PCA and the subgraph stochastic block model. Our results demonstrate that, despite the delicate nature of average-case reductions, using natural problems as intermediates can often be beneficial, as is the case in worst-case complexity. Our main technical contribution is to introduce a set of techniques for average-case reductions that: (1) maintain the level of signal in an instance of a problem; (2) alter its planted structure; and (3) map two initial high-dimensional distributions simultaneously to two target distributions approximately under total variation. We also give algorithms matching our lower bounds and identify the information-theoretic limits of the models we consider.

Thesis Supervisor: Guy Bresler
Title: Assistant Professor of Electrical Engineering & Computer Science
Acknowledgements

First and foremost I would like to thank my advisor, Guy Bresler, for his guidance and faith in our research projects. Guy has been an incredible advisor and introduced me to a broad range of areas of research. His optimism, encouragement and perpetual willingness to discuss research has been a crucial part of my experience in graduate school and this thesis. I would also like to thank my co-advisees, Dheeraj Nagaraj and Mina Karzand, for always being happy to discuss research and graduate school in general. In addition, I would like to thank Philippe Rigollet, Ankur Moitra, Elchanan Mossel, Piotr Indyk, Yury Polyanskiy, Yash Deshpande, Julien Clancy, Jerry Li, Frederic Koehler, Vishesh Jain, Jon Weed, Andrej Risteski and Jason Altschuler for helpful comments and guidance with this thesis and in general over the past two years. Thank you David Arthur and James Rickards, who are always happy to talk about math no matter the occasion. Thanks to Irene, Amanda and Nikhil for being a great trivia group and always willing to see a movie or have a meal. Thank you Dan, Revanth, Jeff, Hugo, Sebastian and Neil for your support and all of the good times. Thanks Manon, Marie, Louis, Louis and Qi for making IDSS such a fun place to be. I’d also like to thank the MIT pick-up basketball community for being so welcoming and providing a great diversion from research. I would like to give a huge thanks to my parents, who are endlessly encouraging and have always supported me in whatever I do.
Contents

1 Introduction .............................................. 11
   1.1 Overview .......................................... 13
   1.2 Hardness Results from an Algorithmic Perspective .... 16
   1.3 Prior Work ......................................... 19
   1.4 Notation ........................................... 22

2 Summary of Results ....................................... 25
   2.1 Detection and Recovery Problems ....................... 25
   2.2 Problem Formulations ................................ 27
   2.3 Our Results ......................................... 31
   2.4 Our Techniques ...................................... 33

3 Average-Case Reductions under Total Variation ......... 39

4 Densifying Planted Clique and Planted Independent Set .. 43
   4.1 Detecting Planted Generalized Diagonals ............... 43
   4.2 Planted Clique Lifting ................................ 44

5 Rejection Kernels and Distributional Lifting .......... 51
   5.1 Rejection Kernels .................................... 51
   5.2 Distributional Lifting ................................ 60

6 Planted Dense Subgraph and Biclustering ............... 67
   6.1 Poisson Lifting and Lower Bounds for Low-Density PDS .. 67
   6.2 Gaussian Lifting and Lower Bounds for High-Density PDS and BC .. 70
   6.3 Lower Bounds for General PDS ......................... 80
7 Reflection Cloning and Subgraph SBM
   7.1 Reflecting Cloning and Rank-1 Submatrix .................................. 85
   7.2 Sparse Spiked Wigner Matrix ......................................................... 92
   7.3 Subgraph Stochastic Block Model .................................................. 97

8 Random Rotations and Sparse PCA .................................................... 103

9 Algorithms and Information-Theoretic Thresholds ............................ 111
   9.1 Biclustering, Planted Dense Subgraph and Independent Set ............... 111
   9.2 Rank-1 Submatrix, Sparse Spiked Wigner and Subgraph SBM ............. 119
   9.3 Sparse PCA and Biased Sparse PCA ............................................... 131

10 Conclusions and Future Directions ............................................... 143
List of Figures

1.1 Graph of average-case reductions for detection problems showing tight statistical-computational gaps given the planted clique conjecture. ................................................. 14

1.2 Algorithms for sparse PCA with \( d = \Theta(n) \), \( k = \Theta(n^\beta) \) and \( \theta = \Theta(n^{-\alpha}) \). Lines represent the strongest guarantees of each algorithm. The line marked as previous reductions shows the strongest previously known planted clique lower bounds for sparse PCA when \( k \lesssim \sqrt{n} \). No planted clique lower bounds were known for \( k \gg \sqrt{n} \). 17

2.1 Parameter regimes by problem plotted as signal vs. sparsity. Sparsity is \( k = \Theta(n^\beta) \). First labels characterize detection and second labels characterize exact recovery. Recovery is not considered for SSBM and weak recovery is considered for SPCA and BSPCA. In Easy (E) regimes, there is a polynomial-time algorithm. In Hard (H) regimes, the PC or PDS conjecture implies there is no polynomial-time algorithm. In Impossible (I) regimes, the task is information-theoretically impossible. Hardness in black regions is open. ................................................................. 35

2.2 Types of hardness regimes by SNR in Theorem 2.3.1. TYPE I and TYPE III each have two variants A and B depending on the Easy regime when \( k \ll n^{1/2} \). ............................. 36

2.3 Classification of regimes for each problem as in Theorem 2.3.1. For each problem, \( k \) is in the regime \( k = \Theta(n^\beta) \) where \( \beta \in (0, 1) \) is a constant. ................................. 37

4.1 Planted clique lifting procedure in Lemma 4.2.1. .................................................. 46

5.1 Rejection kernel in Lemma 5.1.1 ............................................................... 52

5.2 Distributional lifting procedure in Theorem 5.2.1. ............................................. 62

6.1 Poisson lifting procedure in Lemma 6.1.1. ....................................................... 68

6.2 Gaussian lifting procedure in Lemma 6.2.1. ..................................................... 71

6.3 Reduction to biclustering in Lemma 6.2.2. ..................................................... 74
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>Reduction to biclustering recovery in Lemma 6.2.3.</td>
<td>77</td>
</tr>
<tr>
<td>6.5</td>
<td>Reduction to the general regime of planted dense subgraph in Lemma 6.3.1.</td>
<td>81</td>
</tr>
<tr>
<td>7.1</td>
<td>Reflection cloning procedure in Lemma 7.1.</td>
<td>86</td>
</tr>
<tr>
<td>7.2</td>
<td>Reduction to rank-1 submatrix in Lemma 7.2.</td>
<td>90</td>
</tr>
<tr>
<td>7.3</td>
<td>Reduction to sparse spiked Wigner matrix in Lemma 7.2.</td>
<td>94</td>
</tr>
<tr>
<td>7.4</td>
<td>Reduction to the subgraph stochastic block model in Lemma 7.3.</td>
<td>97</td>
</tr>
<tr>
<td>8.1</td>
<td>Random rotation procedure in Lemma 8.</td>
<td>104</td>
</tr>
<tr>
<td>8.2</td>
<td>Reductions to SPCA_D when ( k \gtrsim \sqrt{n} ) and ( k \lesssim \sqrt{n} ) in Lemmas 8.0.3 and 8.0.4 and reduction to SPCA_R in Theorem 8.0.5.</td>
<td>106</td>
</tr>
<tr>
<td>9.1</td>
<td>Exhaustive search algorithm for sparse rank-1 submatrix recovery in Theorem 9.2.2.</td>
<td>122</td>
</tr>
<tr>
<td>9.2</td>
<td>Algorithm for sparse rank-1 submatrix recovery from [CLR15] and in Theorem 9.2.5.</td>
<td>129</td>
</tr>
<tr>
<td>9.3</td>
<td>SDP algorithm for weak ( \ell_2 ) estimation in sparse PCA from [WBS16] and in Theorem 9.3.9.</td>
<td>139</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The field of statistics is undergoing a dramatic conceptual shift, with computation moving from the periphery to center stage. Prompted by the demands of modern data analysis, researchers realized two decades ago that a new approach to estimation was needed for high-dimensional problems in which the dimensionality of the data is at least as large as the sample size. High-dimensional problems are inherently underdetermined, often precluding nontrivial rates of estimation. However, this issue typically disappears if the underlying signal is known to have an appropriate structure, such as low rank or sparsity. As a result, high-dimensional structured estimation problems have received significant attention in both the statistics and computer science communities. Prominent examples include estimating a sparse vector from linear observations, sparse phase retrieval, low-rank matrix estimation, community detection, subgraph and matrix recovery problems, random constraint satisfiability, sparse principal component analysis and covariance matrix estimation. Although structural assumptions can yield nontrivial estimation rates, the statistically optimal estimators for these problems typically entail an exhaustive search over the set of possible structures and are thus not efficiently computable. Conversely, all known efficient algorithms for these problems are statistically suboptimal, requiring more data than strictly necessary. This phenomenon has led to a number of conjectured statistical-computational gaps for high-dimensional problems with structure. This raises an intriguing question: how are these gaps related to one another and are they emerging for a common reason?

In the last few years, several lines of work have emerged to make rigorous the notion of what is and what is not achievable statistically by efficient algorithms. In the seminal work of [BR13a], a conjectured computational-statistical gap for sparse principal component analysis (PCA) was shown to follow from the planted clique conjecture. This marked the first result basing the hardness of a natural statistics problem on an average-case hardness assumption and produced a framework for
showing statistical-computational gaps by approximately mapping in total variation. This subsequently led to several more reductions from the planted clique conjecture to show computational-statistical gaps for problems including submatrix detection/biclustering [MW15b], submatrix localization [CLR15], planted dense subgraph [HWX15], RIP certification [WBP16], sparse PCA and sparse canonical correlation analysis [WBS16, GMZ17]. We draw heavily from the framework for average-case reductions laid out in these papers. More recently, focus has shifted to showing unconditional hardness results for restricted models of computation and classes of algorithms. An exciting line of work has emerged surrounding applications of the Sum of Squares (SOS) semidefinite programming hierarchy to problems with statistical computational gaps. SOS Lower bounds have been shown for planted clique [BHK+16] and for sparse PCA [KNV15, MW15a, HKP+17]. Tight computational lower bounds have also been shown in the statistical query model for planted clique and planted random $k$-SAT [FGR+12, FPV15].

One reason behind this focus on showing hardness in restricted models of computation is that average-case reductions are inherently delicate, creating obstacles to obtaining satisfying hardness results. As described in [Bar17], these technical obstacles have left us with an unsatisfying theory of average-case hardness. Reductions in worst-case complexity typically take a general instance of a problem $A$ to a structured instance of a problem $B$. For example, a classic reduction from 3SAT to INDEPENDENT-SET produces a very specific type of graph with a cluster of seven vertices per clause corresponding to each satisfying assignment such that two vertices are connected if together they yield an inconsistent assignment. If such a reduction were applied to a random 3SAT instance, the resulting graph instance would be far from any natural graph distribution. Unlike reductions in worst-case complexity, average-case reductions between natural decision problems need to precisely map the distributions on instances to one another without destroying the underlying signal in polynomial-time. The delicate nature of this task has severely limited the development of techniques and left open reductions between decision problems that seem to be obviously equivalent from the standpoint of algorithm design. For example, it remains unknown whether refuting random constraint satisfaction problems with $10m$ clauses is equivalent to refuting those with $11m$ clauses or whether the planted clique conjecture at edge density 1/2 implies the conjecture at edge density 0.49. There are also a variety of negative results further demonstrating new obstacles associated with average-case complexity, including that it likely cannot be based on worst-case complexity [BT06a]. For more on average-case complexity, we refer to the survey of [BT+06b].

In order to overcome these average-case difficulties, prior reductions have often made assumptions on the robustness of the underlying algorithm such as that it succeeds for any noise distributions from a fixed class as in [BR13a, WBS16, CLR15]. This corresponds to composite vs.
composite hypothesis testing formulations of detection problems, where the composite null hypothesis $H_0$ consists of the class of noise distributions. Other reductions have shown hardness for precise noise distributions but for algorithms that do not need to exactly know the parameters of the given instance [MW15b, GMZ17]. This typically corresponds to simple vs. composite hypothesis testing where the composite alternative $H_1$ consists of models defined by varying parameters such as the sparsity $k$ or signal strength. The strongest prior reduction from planted clique is that to the sparsest regime of planted dense subgraph in [HWX15]. A lower bound is shown for a simple vs. simple hypothesis testing variant of the problem, with each consisting of a single distribution. However, the community in their formulation of planted dense subgraph was binomially distributed and therefore still assumed to be unknown exactly to the algorithm. Prior reductions have also shown hardness at particular points in the parameter space, deducing that an algorithm cannot always perform better than a conjectured computational barrier rather than showing that no algorithm can ever perform better. For example, prior reductions for sparse PCA have only shown tight hardness around the single parameter point where the signal is $\theta = \tilde{O}(1)$ and the sparsity is $k = \tilde{\Theta}(\sqrt{n})$. Simplifying parameters in their reductions, both [BR13a] and [GMZ17] approximately map a planted clique instance on $n$ vertices with clique size $k$ to an instance of sparse PCA with $\theta \approx \tilde{O}(k^2/n)$ which is only tight to the conjectured barrier of $\theta^* = \Theta(\sqrt{k^2/n})$ when $k = \tilde{\Theta}(\sqrt{n})$.

These assumptions leave a subtle disparity between the existing average-case lower bounds for many problems and algorithmic upper bounds. Many algorithmic results assume a canonical generative model or implicitly assume knowledge of parameters. For example, even in the recent literature on robust algorithms for problems with sparsity in [BDLS17, Li17], the setup is in the context of specific canonical generating models, such as the spiked covariance model for sparse PCA. Even when corrupted by adversarial noise, the spiked covariance model is far in distribution from many sub-gaussian formulations of sparse PCA. Despite existing average-case lower bounds, hardness for the canonical generative models for many problems has remained open. This includes biclustering with a flat planted $k \times k$ submatrix selected uniformly at random in gaussian noise, sparse PCA with a $k$-sparse principal component chosen uniformly at random to have entries equal to $\pm 1/\sqrt{k}$ and planted dense subgraph with deterministic community size.

1.1 Overview

The aim of this paper is threefold: (1) to demonstrate that a web of average-case reductions among problems with statistical-computational gaps is feasible even for showing strong computational lower bounds; (2) to introduce a number of new techniques for average-case reductions between
problems; and (3) to fully characterize the computationally hard regime of several models. The graph of our reductions is shown in Figure 1.1. Our new lower bounds are as follows.

- **Planted Independent Set**: We show tight lower bounds for detecting a planted independent set of size $k$ in a sparse Erdős-Rényi graph of size $n$ with edge density $\Theta(n^{-\alpha})$.

- **Planted Dense Subgraph**: If $p > q$ are the edge densities inside and outside of the community, we show the first lower bounds for the general regime $q = \Theta(n^{-\alpha})$ and $p - q = \Theta(n^{-\gamma})$ where $\gamma \geq \alpha$, matching the lower bounds predicted in [CX16]. Our lower bounds apply to a deterministic community size $k$, resolving a question raised in [HWX15].

- **Biclustering**: We show lower bounds for Gaussian biclustering as a simple hypothesis testing problem to detect a uniformly at random planted flat $k \times k$ submatrix. Our alternative reduction matches the barriers in [MW15b], where a computational lower bound was shown for a composite hypothesis testing variant of biclustering. We show hardness for the natural simple hypothesis testing problem where the $k \times k$ submatrix is chosen uniformly at random and has equal entries.

- **Sparse Rank-1 Submatrix**: We show that detection in the sparse spiked Wigner model has a different computational threshold from biclustering when $k \gg \sqrt{n}$. Surprisingly, we are able to obtain tight lower bounds matching these different detection thresholds with different reductions from planted clique.

- **Sparse PCA**: We give a reduction between rank-1 submatrix and sparse PCA to obtain tight lower bounds in the less sparse regime $k \gg \sqrt{n}$, when the spectral algorithm is optimal over the SDP. This yields the first tight characterization of a computational barrier for sparse
1.1. **OVERVIEW**

PCA over an entire parameter regime. We also give an alternate reduction recovering the lower bounds of [BR13a] and [GMZ17] in the canonical simple hypothesis testing variant of sparse PCA.

- **Biased Sparse PCA:** We show that any assumption on the sparse principal component having a constant fraction more or fewer positive entries than negative entries yields a detection-recovery gap that is not present in sparse PCA.

- **Subgraph Stochastic Block Model:** We introduce a model where two small communities are planted in an Erdős-Rényi graph of the same average edge density. Parallel to the difference between biclustering and sparse rank-1 submatrix when \( k \gg \sqrt{n} \), we show that detection in this model is much harder than in planted dense subgraph when \( k \gg \sqrt{n} \).

Our lower bounds for planted independent set, the general regime of planted dense subgraph, rank-1 submatrix, sparse PCA when \( k \gg \sqrt{n} \), biased sparse PCA and the subgraph stochastic block model are novel. As previously mentioned, lower bounds for sparse PCA when \( k \ll \sqrt{n} \), for biclustering and for planted dense subgraph in the sparsest regime were previously known. In each of these cases, we strengthen the existing lower bounds to the apply to the canonical generative model. We show computational lower bounds for simple vs. simple hypothesis testing in all cases other than for sparse PCA, rank-1 submatrix and the subgraph stochastic block model all in the regime \( k \gg \sqrt{n} \). This is a consequence of our underlying reduction technique, reflection cloning, and appears unavoidable given our methods. However, we do show that the distribution we reduce to is in some sense close to the canonical generative model.

Our results demonstrate that, despite the delicate nature of average-case reductions, using natural problems as intermediates can often be beneficial as in reductions in worst-case complexity. Our main technical contribution is to introduce several techniques for mapping problems approximately in total variation without degrading the underlying planted sparse structure. These techniques are:

- **Distributional Lifting:** A variant of graph lifts that iteratively maps to intermediate matrices with entries from chosen distributions. Varying the underlying distribution produces different relationships between the resulting edge density and size of a planted subgraph. This is introduced in Chapters 4, 5 and 6.

- **Rejection Kernels:** A general framework for a change in measure of the underlying noise distribution while preserving the planted sparse structure. This unifies ideas introduced in [HWX15], [MW15b] and [GMZ17]. This is introduced in Chapter 5.
• **Reflection Cloning:** A method of increasing the sparsity of a planted rank-1 structure in noise while preserving the level of signal in the planted structure. This is introduced in Chapter 7.

• **Random Rotations for Sparse PCA:** An average-case connection between the sparse spiked Wigner model and the spiked covariance model of sparse PCA. This is introduced in Chapter 8.

We consider two main variants of distributional lifting using matrices with Poisson and Gaussian entries as intermediates. Poisson and Gaussian lifting lead to two very different parameter scalings, which when combined fully characterize general planted dense subgraph. Reflection cloning can be viewed as a more randomness-efficient variant of Gaussian lifting that we use to show sharper lower bounds for the sparse spiked Wigner model and sparse PCA. We also give algorithms matching our lower bounds and identify the information-theoretic limits of the problems that we consider in Chapter 9.

Although the problems we consider and reduce from planted clique are similar in the sense that they can all be viewed as matrix problems with planted sparse structure, providing tight reductions in total variation that exactly preserve the conjectured computational boundaries proves to be a difficult task and is the main novelty of the present work. A simple example demonstrating this difficulty is our reduction from planted clique to planted independent set. Consider the following naive candidate reduction. Add edges independently to planted clique and then take the complement graph. This maps planted clique to a valid instance of planted independent set with the correct sparse edge density. However, the size of the planted independent set in this proposed reduction turns out to be far smaller than the best known polynomial time algorithms can detect, leaving a gap between the demonstrated and conjectured computational lower bounds. Another example of this lies in the difference between our reductions to biclustering and sparse rank-1 subamtrix. Biclustering is a specific instance of sparse rank-1 submatrix. The reduction we give for biclustering detection is based on a Gaussian variant of distributional lifting and shows tight lower bounds for biclustering. However, this biclustering lower bound turns out to be far below the conjectured lower bound for sparse rank-1 submatrix. We show that a different technique we introduce – reflection cloning – that does not apply for biclustering exactly reduces to this stronger lower bound.

### 1.2 Hardness Results from an Algorithmic Perspective

In this section, we motivate our computational lower bounds and techniques using algorithms for sparse PCA as an example. Consider the detection problem for sparse PCA where either
1.2. HARDNESS RESULTS FROM AN ALGORITHMIC PERSPECTIVE

**Figure 1.2:** Algorithms for sparse PCA with $d = \Theta(n)$, $k = \tilde{\Theta}(n^{\beta})$ and $\theta = \tilde{\Theta}(n^{-\alpha})$. Lines represent the strongest guarantees of each algorithm. The line marked as previous reductions shows the strongest previously known planted clique lower bounds for sparse PCA when $k \leq \sqrt{n}$. No planted clique lower bounds were known for $k \gg \sqrt{n}$.

$X_1, X_2, \ldots, X_n$ are sampled i.i.d. from $N(0, I_d)$ or are sampled i.i.d. from $N(0, I_d + \theta vv^T)$ for some latent $k$-sparse unit vector $v$ with nonzero entries equal to $\pm 1/\sqrt{k}$. The task is to detect which of the two distributions the samples originated from. For now assume that $d = \Theta(n)$. Now consider the following four algorithms:

1. **Semidefinite Programming:** Form the empirical covariance matrix $\Sigma = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T$ and solve the convex program

\[
\max_Z \quad \text{Tr}\left( \Sigma Z \right)
\]

\[
\text{s.t.} \quad \text{Tr}(Z) = 1, |Z|_1 \leq k, Z \succeq 0
\]

As shown in [BR13a], thresholding the resulting maximum solves the detection problem as long as $\theta = \tilde{\Omega}(\sqrt{k^2/n})$.

2. **Spectral Algorithm:** Threshold the maximum eigenvalue of $\hat{\Sigma}$. If the data are sampled from $N(0, I_d)$, then the largest eigenvalue is with high probability at most

\[
\lambda_{\max}(\hat{\Sigma}) \leq \frac{d}{n} + \sqrt{\frac{d}{n}} + 1 + o(1)
\]

by standard bounds on the singular values of random Gaussian matrices. Since $d = \Theta(n)$, this algorithm succeeds as long as $\theta = \Omega(1)$. This algorithm was considered in [KNV15].
3. **Sum Test**: Sum the entries of $\hat{\Sigma}$ and threshold the absolute value of the sum. If $v$ has sum exactly zero, then this test will not succeed. However, if we assume that $\hat{\Sigma}$ has even 51% of its nonzero entries of one sign, then this test succeeds if $\theta = \tilde{\Omega}(\sqrt{n}/k)$.

4. **$k$-Sparse Eigenvalue**: Compute and threshold the $k$-sparse unit vector $u$ that maximizes $u^T \hat{\Sigma} u$. This can be found by finding the largest eigenvector of each $k \times k$ principal submatrix of $\hat{\Sigma}$. Note that this takes exponential time. It was shown in [BR13a] that this succeeds as long as $\theta = \tilde{\Omega}(\sqrt{k/n})$.

The boundaries at which these algorithms begin to succeed are shown in Figure 1.2 for the regime $k = \Theta(n^\beta)$ and $\theta = \tilde{O}(n^{-\alpha})$. The computational lower bound mapping to $\theta \approx k^2/n$ in [BR13a] and [GMZ17] is also drawn. As shown, the only point in the parameter diagram for which it matches an algorithmic upper bound is $\alpha = 0$ and $\beta = 1/2$, corresponding to when $\theta = \tilde{\Theta}(1)$ and $k = \tilde{\Theta}(\sqrt{n})$.

For sparse PCA with $d = \Theta(n)$, the SDP is optimal up to $k = \Theta(\sqrt{n})$, at which point the spectral algorithm has stronger guarantees. This algorithmic transition at $k = \Theta(\sqrt{n})$ is characteristic of all of the problems we consider. For the biased variant of sparse PCA where the sum test succeeds, the sum test always does strictly better than the spectral algorithm. Furthermore, the biased variant ceases to have a statistical computational gap around $k = \Theta(n^{2/3})$. While the sum test yields an improved algorithm for detection, unlike the other three algorithms considered above, it does not translate into an algorithm for recovering the support of the sparse component. Given a conjecture about recovery in planted dense subgraph, we show that the best recovery algorithm for biased sparse PCA can only match the guarantees of the spectral algorithm. Thus the biased variant induces a detection-recovery gap when $k \gg \sqrt{n}$. We show that the disappearance of a statistical computation gap at $k = \Theta(n^{2/3})$ and a detection-recovery gap when $k \gg \sqrt{n}$ are features of the problems we consider that admit a sum test. These are biased sparse PCA, planted independent set, planted dense subgraph and biclustering. Distributional lifting gives tight planted clique lower bounds for these problems.

In contrast, rank-1 submatrix, the subgraph stochastic block model and sparse PCA do not admit a sum test. Given the planted clique conjecture, rank-1 submatrix and sparse PCA have no detection-recovery gap and retain their statistical-computational gap for all sparsities $k$. Reflection cloning shows tight lower bounds for these problems in the regime $k \gg \sqrt{n}$, where spectral algorithms become optimal. It is surprising that the planted clique conjecture can tightly capture completely different sets of computational barriers for different problems, illustrating its power as an average-case hardness assumption. Although analogues of the sum test, spectral algorithms and semidefinite programs all have equivalent guarantees up to logarithmic factors for planted clique, reductions from planted clique show tight hardness in problems for which this is not true.
1.3. PRIOR WORK

1.3 Prior Work

This work is part of a growing body of literature giving rigorous evidence for computational-statistical gaps in high-dimensional inference problems. We focus on average-case reductions to directly relate computational-statistical gaps in different problems, as opposed to giving worst-case evidence for hardness in statistical problems [ZWJ14, HMRW14, CPR16]. A survey of prior results on computational-statistical gaps with a focus on predictions from statistical physics can be found in [BPW18] and a general analysis of gaps for algorithms from several convex relaxation hierarchies can be found in [CJ13].

Planted Clique and Independent Set Our computational lower bounds are based on average-case reductions to the problem of finding a planted clique of size $k$ in an Erdős-Rényi graph with $n$ vertices. The planted clique problem was introduced in [AKS98], where a spectral algorithm was shown to recover the planted clique if $k = \Omega(\sqrt{n})$. A number of algorithms for planted clique have since been studied, including approximate message passing, semidefinite programming, nuclear norm minimization and several combinatorial approaches [FK00, McS01, FR10, AV11, DGGP14, DM15a, CX16]. All of these algorithms require that $k = \Omega(\sqrt{n})$, which has led to the planted clique conjecture that no polynomial time algorithm can recover the planted clique if $k = o(\sqrt{n})$. It was also shown in [AAK+07], recovering and detecting the planted clique are equivalent up to $\log n$ factors in $k$. There have been a number of previous average-case reductions from the planted clique conjecture, which we discuss in more detail in the prior work section on average-case reductions.

Several works have considered finding independent sets in sparse Erdős-Rényi graphs, similar to the regime with edge density $q = \tilde{\Theta}(n^{-\alpha})$ where $\alpha \in (0, 2)$ we consider here. In [COE15, GS14, RV+17], the authors examine greedy and local algorithms to find independent sets in the regime $q = \tilde{\Theta}(n^{-1})$ in models related to Erdős-Rényi graphs. In [FO05], a spectral algorithm is given to find a planted independent set in the regime $q = \tilde{\Theta}(n^{-1})$ and in [CO03], the planted independent set recovery problem is shown to be possible in polynomial time in the regime $\alpha \in (0, 1)$ when $q \gg \frac{n}{k^2}$ even in a semirandom model. The algorithms of [CX16] also apply to recovering planted independent sets after taking the complement of the input graph.

Planted Dense Subgraph and Community Detection The planted dense subgraph detection problem was considered in [ACV+14, BI13, VAC+15, HWX15] and generalizations of the recovery problem were considered in [CX16, HWX16b, Mon15, CC18]. In [HWX15], a reduction from planted clique was given for the regime $p = cq$ for some constant $c > 1$ and $q = \tilde{\Theta}(n^{-\alpha})$ and $k$ is binomially distributed, where $k$, $n$, $p$ and $q$ are the size of the community, size of the graph,
community edge density and graph edge density, respectively. Our results strengthen this lower
bound to apply for deterministic $k$ and for all $p > q$ with $p - q = O(q)$ where $q = \Theta(n^{-\alpha})$. When
$p = \omega(q)$, the resulting regime is the planted dense subgraph problem considered in [BCC+10]. The
computational barrier for this problem is conjectured to be the log-density threshold $k = \Theta(n^{\log_3 p})$
when $k \ll \sqrt{n}$ and is achieved by very different algorithms than those that are optimal when
$p = O(q)$ [CDK12, CDM17]. Recently, it was shown in [CM18] that $\Omega(\log n)$ rounds of the Sherali-
Adams hierarchy cannot solve the planted dense subgraph detection problem below the log-density
threshold in the regime $p = \omega(q)$. Hardness below the log-density threshold has been used as an
average-case assumption in several reductions, as outlined in the prior work section on average-case
reductions.

Community detection in the stochastic block model has been the focus of an extensive body
of literature surveyed in [Abb17]. It recently has been shown that the two-community stochastic
block model does not exhibit statistical-computational gaps for partial and exact recovery, which
are possible when the edge density scales like $\Theta(n^{-1})$ [MNS12, MNS13, Masl14] and $\Theta(n^{-1} \log n)$
[MNS14, HWX16a, ABH16], respectively. In contrast, the subgraph variant of the two-community
stochastic block model that we introduce has computational-statistical gaps for partial recovery,
exact recovery and detection, given the planted clique conjecture. The $k$-block stochastic block
model is also conjectured to have statistical-computational gaps starting at $k \geq 4$ [AS15].

Biclustering and the Spiked Wigner Model Gaussian biclustering was considered as a de-
tection problem in [BI13, MW15b, MRZ15] and as a recovery problem in [SWP+09, KBRS11,
BKR+11, CLR15, CX16, HWX16b]. In [MW15b], a reduction from planted clique was given for
a simple vs. composite hypothesis testing variant of the biclustering detection problem, where
the size and mean entries of the planted submatrix were allowed to vary. In [CLR15], submatrix
localization with subgaussian noise was shown to be hard assuming a variant of the planted clique conjecture for regular graphs.

A large body of literature has studied the spectrum of the spiked Wigner model [Pèc06, FP07,
CDMF+09, BGN11]. Spectral algorithms and information-theoretic lower bounds for the spiked
Wigner model detection and recovery problems were considered in [MRZ15, PWB16, PWBM16].
The sparse spiked Wigner model where the sparsity $k$ of the planted vector satisfies $k = \Theta(n)$
was studied in [PWB16, PWBM16, BMV+18]. The sparse spiked Wigner model with $k = \hat{\Theta}(n^3)$
for some $\beta \in (0, 1)$ was considered in [HKP+17], where the authors showed sum of squares lower
bounds matching our planted clique reductions.
1.3. PRIOR WORK

Sparse PCA Since its introduction in [JL04], sparse principal component analysis has been studied broadly in the statistics and computer science communities. A number of algorithms solving sparse PCA under the spiked covariance model have been proposed [AW09, Ma13, CMW+13, BR13b, BR13a, SSM13, KNV15, DM14, WBS16]. The information-theoretic limits for detection and recovery in the spiked covariance model have also been examined extensively [AW09, VL12, BR13b, BJNP13, CMW+13, WBS16, CMW15]. The computational limits of sparse PCA problems have also been considered in the literature. Degree four SOS lower bounds for the spiked covariance model were shown in [MW15a]. In [BR13a], the authors give a reduction from planted clique to a subgaussian composite vs. composite hypothesis testing formulation of sparse PCA as a detection problem, and [BR13b] gives a reduction from planted clique showing hardness for semidefinite programs. In [GMZ17], the authors give a reduction from planted clique to a simple vs. composite hypothesis testing formulation of detection in the spiked covariance model matching our reduction when \( k \ll \sqrt{n} \). In [WBS16], the authors give a reduction from planted clique to a subgaussian variant of the sparse PCA recovery problem. As mentioned in the introduction, these planted clique lower bounds do not match the conjectured algorithmic upper bounds when \( k \) differs in a polynomial factor from \( \sqrt{n} \).

Average-Case Reductions While the theory of worst-case complexity has flourished to the point that many natural problems now known to be NP-hard or even NP-complete, the theory of average-case complexity is far less developed. In the seminal work of [Lev86], it was shown that an average-case complete problem exists. However, no natural problem with a natural distribution on inputs has yet been shown to be average-case complete. As mentioned in this chapter, there are obfuscations to basing average-case complexity on worst-case complexity [BT06a]. For more on the theory of average-case complexity, see Chapter 18 of [AB09] and [BT+06b].

As previously mentioned, there have been a number of average-case reductions from planted clique to average-case problems in both the computer science and statistics literature. These include reductions to testing \( k \)-wise independence [AAK+07], biclustering detection and recovery [MW15b, CLR15, CW18], planted dense subgraph [HWX15], RIP certification [WBP16, KZ14], matrix completion [Che15], minimum circuit size and minimum Kolmogorov time-bounded complexity [HS17a] and sparse PCA [BR13b, BR13a, WBS16, GMZ17]. The planted clique conjecture has also been used as a hardness assumption for average-case reductions in cryptography [JP00, ABW10], as described in Sections 2.1 and 6 of [Bar17]. There have also been a number of average-case reductions from planted clique to show worst-case lower bounds such as hardness of approximation. Planted clique has been used to show worst-case hardness of approximating
densest $k$-subgraph [AAM+11], finding approximate Nash equilibria [MV09, HK11, ABC13], signalling [Dug14, BCKS16], approximating the minmax value of 3-player games [EHV12], aggregating pairwise comparison data [SBW16] and finding endogenously formed communities [BBB+13].

A number of average-case reductions in the literature have started with different average-case assumptions than the planted clique conjecture. Variants of planted dense subgraph have been used to show hardness in a model of financial derivatives under asymmetric information [ABBG11], link prediction [BB18], finding dense common subgraphs [CNW18] and online local learning of the size of a label set [ACLR15]. Hardness conjectures for random constraint satisfaction problems have been used to show hardness in improper learning complexity [DLSS14], learning DNFs [DSS16] and hardness of approximation [Fei02]. There has also been a recent reduction from a hypergraph variant of the planted clique conjecture to tensor PCA [ZX17].

**Lower Bounds for Classes of Algorithms** As described in the introduction, recently there has been a focus on showing unconditional hardness results for restricted models of computation and classes of algorithms. In [Jer92], it was shown that the Metropolis process cannot find large cliques in samples from planted clique. The fundamental limits of spectral algorithms for biclustering and low-rank planted matrix problems were examined in [MRZ15]. Integrality gaps for SDPs solving sparse PCA, planted dense subgraph and submatrix localization were shown in [KNV15] and [CX16]. SOS lower bounds have been shown for a variety of average-case problems, including planted clique [DM15b, RS15, HKP+16, BHK+16], sparse PCA [MW15a], sparse spiked Wigner and tensor PCA [HKP+17], maximizing random tensors on the sphere [BGL17] and random CSPs [KMOW17]. Lower bounds for relaxations of planted clique and maximum independent set in the Lovász-Schrijver hierarchy are shown in [FK03] and lower bounds for Sherali-Adams relaxations of planted dense subgraph in the log-density regime are shown in [CM18]. Tight lower bounds have been shown in the statistical query model for planted clique [FGR+13], random CSPs [FPV15] and robust sparse mean estimation [DKS16]. It also has been recently shown that planted clique with $k \ll \sqrt{n}$ is hard for regular resolution [AR18]. In [HS17b], a meta-algorithm for Bayesian estimation based on low-degree polynomials, SDPs and tensor decompositions is introduced and shown to achieve the best known upper bound for the $k$-block stochastic block model, with a matching lower bound for the meta-algorithm.

### 1.4 Notation

In this paper, we adopt the following notational conventions. Let $\mathcal{L}(X)$ denote the distribution law of a random variable $X$. Given a distribution $\mathbb{P}$, let $\mathbb{P}^{\otimes n}$ denote the distribution of $(X_1, X_2, \ldots, X_n)$.
where the $X_i$ are i.i.d. according to $\mathbb{P}$. Similarly, let $\mathbb{P}^{\otimes m \times n}$ denote the distribution on $\mathbb{R}^{m \times n}$ with i.i.d. entries distributed as $\mathbb{P}$. Given a finite or measurable set $\mathcal{X}$, let $\text{Unif}[\mathcal{X}]$ denote the uniform distribution on $\mathcal{X}$. Let $d_{TV}$ and $d_{KL}$ denote total variation distance and Kullback-Leibler divergence, respectively. Given a measurable set $\mathcal{X}$, let $\Delta(\mathcal{X})$ denote the set of all distributions $\pi$ on $\mathcal{X}$. If $\mathcal{X}$ is itself a set of distributions, we refer to $\Delta(\mathcal{X})$ as the set of priors on $\mathcal{X}$. Throughout the paper, $C$ refers to any constant independent of the parameters of the problem at hand and will be reused for different constants.

Let $N(\mu, \sigma^2)$ denote a normal random variable with mean $\mu$ and variance $\sigma^2$ when $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. Let $N(\mu, \Sigma)$ denote a multivariate normal random vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma$, where $\Sigma$ is a $d \times d$ positive semidefinite matrix. Let $\beta(x, y)$ denote a beta distribution with parameters $x, y > 0$ and let $\chi^2(k)$ denote a $\chi^2$-distribution with $k$ degrees of freedom. Let $\mathcal{B}_0(k)$ denote the set of all unit vectors $v \in \mathbb{R}^d$ with $\|v\|_2 \leq k$. Let $[n] = \{1, 2, \ldots, n\}$ and $\binom{[n]}{k}$ denote the set of all size $k$ subsets of $[n]$. Let $\mathcal{G}_n$ denote the set of all simple graphs on vertex set $[n]$. Let the Orthogonal group on $\mathbb{R}^{d \times d}$ be $O_d$. Let $1_S$ denote the vector $v \in \mathbb{R}^n$ with $v_i = 1$ if $i \in S$ and $v_i = 0$ if $i \notin S$ where $S \subseteq [n]$. For subsets $S \subseteq \mathbb{R}$, let $1_S$ denote the indicator function of the set $S$. Let $\Phi$ denote the cumulative distribution of a standard normal random variable with $\Phi(x) = \int_{-\infty}^{x} e^{-t^2/2} dt$. Given a simple undirected graph $G$, let $V(G)$ and $E(G)$ denote its vertex and edge sets, respectively. The notation $a(n) \gg b(n)$ will denote $a$ growing polynomially faster in $n$ than $b$. In other words, $a \gg b$ if $\liminf_{n \to \infty} a(n)/n > \limsup_{n \to \infty} b(n)/n$. The notation $a = \tilde{\Theta}(b)$ denotes the equality $\lim_{n \to \infty} a(n)/n = \lim_{n \to \infty} b(n)/n$. Here, $a \lesssim b$ denotes $a \leq b$ up to polylogarithmic factors in $n$. 


Chapter 2

Summary of Results

2.1 Detection and Recovery Problems

We consider problems $\mathcal{P}$ with planted sparse structure as both detection and recovery tasks, which we denote by $\mathcal{P}_D$ and $\mathcal{P}_R$, respectively.

Detection  In detection problems $\mathcal{P}_D$, the algorithm is given a set of observations and tasked with distinguishing between two hypotheses:

- a uniform hypothesis $H_0$, under which observations are generated from the natural noise distribution for the problem; and

- a planted hypothesis $H_1$, under which observations are generated from the same noise distribution with a latent planted sparse structure.

In all of the detection problems we consider, $H_0$ is a simple hypothesis consisting of a single distribution and $H_1$ is either also simple or a composite hypothesis consisting of several distributions. When $H_1$ is a composite hypothesis, it consists of a set of distributions of the form $P_\theta$ where $\theta$ is the latent sparse structure of interest. Often $H_1$ is a simple hypothesis consisting of a single distribution which is a mixture of $P_\theta$ with $\theta$ in some sense chosen uniformly at random. In both cases, we will abuse notation and refer to $H_1$ as a set of distributions. Given an observation $X$, an algorithm $A(X) \in \{0, 1\}$ solves the detection problem with nontrivial probability if there is an $\epsilon > 0$ such that its Type I+II error satisfies that

$$\limsup_{n \to \infty} \left( \sup_{P \in H_0} \mathbb{P}_0[A(X) = 1] + \sup_{P \in H} \mathbb{P}_{X \sim P}[A(X) = 0] \right) \leq 1 - \epsilon$$
where $n$ is the parameter indicating the size of $X$. We refer to this quantity as the asymptotic Type I+II error of $A$ for the problem $\mathcal{P}_D$. If the asymptotic Type I+II error of $A$ is zero, then we say $A$ solves the detection problem $\mathcal{P}_D$. Our reductions under total variation all yield exact correspondences between asymptotic Type I+II errors. Specifically, they show that if a polynomial time algorithm has asymptotic Type I+II error of $\epsilon$ on the problem of interest then there is a polynomial time algorithm with asymptotic Type I+II error $\epsilon$ on the problem being reduced from.

**Recovery** In recovery problems $\mathcal{P}_R$, the algorithm is given an observation from $P_\theta$ for some latent $\theta$ from a space $\Theta$ and the task is to recover the support $S(\theta)$ of the sparse structure $\theta$. There are several variants of the recovery task. Given a randomized algorithm with output $A(X) \in \{S(\theta) : \theta \in \Theta\}$ and a distribution $\pi$ on the latent space $\Theta$, the variants of the recovery task are as follows.

- **Partial Recovery:** $A$ solves partial recovery if
  \[
  \mathbb{E}_{\chi \sim \pi, P_\theta} [||A(X) \cap S(\theta)||] = \Omega(||S(\theta)||) \quad \text{as } n \to \infty
  \]

- **Weak Recovery:** $A$ solves weak recovery if
  \[
  \mathbb{E}_{\chi \sim \pi, P_\theta} [||A(X) \Delta S(\theta)||] = o(||S(\theta)||) \quad \text{as } n \to \infty
  \]

- **Exact Recovery:** $A$ solves exact recovery with nontrivial probability $\epsilon > 0$ if for all $\theta \in \Theta$
  \[
  \liminf_{n \to \infty} \mathbb{P}_{\chi \sim \pi, P_\theta} [A(X) = S(\theta)] \geq \epsilon
  \]

Here, $\mathbb{E}_{\pi, P_\theta}$ denotes the mixture of $P_\theta$ induced by $\pi$ and $\Delta$ denotes the symmetric difference between two sets. Whenever the corresponding detection problem $\mathcal{P}_D$ has a simple hypothesis $H_1$, $\pi$ will be the prior on $\Theta$ as in $H_1$, which typically is a uniform prior. When $\mathcal{P}_D$ has a composite hypothesis $H_1$, then an algorithm $A$ solves each of the three variants of the recovery task if the above conditions are met for all distributions $\pi$. We remark that this is equivalent to the above conditions being met only for distributions $\pi$ with all of their mass on a single $\theta \in \Theta$. Given a problem $\mathcal{P}$, the notation $\mathcal{P}_R$ will denote the exact recovery problem, and $\mathcal{P}_{PR}$ and $\mathcal{P}_{WR}$ will denote partial and weak recovery, respectively. All of our recovery reductions will apply to all recovery variants simultaneously. In other words, a partial, weak or exact recovery algorithm for the problem of interest implies the same for the problem being reduced from. The computational and statistical barriers for partial, weak and exact recovery generally differ by sub-polynomial factors for the problems we consider. Therefore the polynomial order of the barriers for $\mathcal{P}_R$ will in general also be those for $\mathcal{P}_{PR}$ and $\mathcal{P}_{WR}$. This is discussed in more detail in Chapter 9.
2.2. PROBLEM FORMULATIONS

An instance of a detection problem $\mathcal{P}_D$ hereby refers to an observation $X$. If $\mathcal{P}_D$ is a simple vs. simple hypothesis testing problem, then the instance $X$ takes one of two distributions – its distribution under $H_0$ and $H_1$, which we respectively denote by $\mathcal{L}_{H_0}(X)$ and $\mathcal{L}_{H_1}(X)$. If $H_1$ is composite, then the distribution of $X$ under $\mathbb{P}$ is denoted as $\mathcal{L}_{\mathbb{P}}(X)$ for each $\mathbb{P} \in H_1$. An instance of a recovery problem $\mathcal{P}_R$ refers to an observation from some $P_\theta$ for some latent $\theta \in \Theta$ if the corresponding detection problem has a composite $H_1$ or from $\mathcal{L}_{H_1}(X) = \mathbb{E}_\pi P_\theta$ if the corresponding detection problem has a simple $H_1$.

Computational Model The algorithms we consider here are either unconstrained or run in randomized polynomial time. An unconstrained algorithm refers to any randomized function or Markov transition kernel from one space to another. These algorithms considered in order to show that information-theoretic lower bounds are asymptotically tight. An algorithm that runs in randomized polynomial time has access to $\text{poly}(n)$ independent random bits and must run in $\text{poly}(n)$ time where $n$ is the size of the input. For clarity of exposition, we assume that explicit expressions can be exactly computed and assume that $N(0, 1)$ and Poisson random variables can be sampled in $O(1)$ time.

2.2 Problem Formulations

In this section, we define the problems that we show computational lower bounds for and the conjectures on which these lower bounds are based. Each problem we consider has a natural parameter $n$, which typically denotes the number of samples or dimension of the data, and sparsity parameter $k$. Every parameter for each problem is implicitly a function of $n$, that grows or decays polynomially in $n$. For example, $k = k(n) = \tilde{\Theta}(n^\beta)$ for some constant $\beta \in (0, 1)$ throughout the paper. For simplicity of notation, we do not write this dependence on $n$. We mostly will be concerned with the polynomial order of growth of each of the parameters and not with subpolynomial factors. We now formally define the problems we consider.

Planted Clique and Independent Set The hypotheses in the planted clique detection problem $\mathcal{P}_{\text{CD}}(n, k, p)$ are

$$H_0 : G \sim G(n, p) \quad \text{and} \quad H_1 : G \sim G(n, k, p)$$

where $G(n, p)$ is an Erdős-Rényi random graph with edge probability and $G(n, k, p)$ is a sample of $G(n, p)$ with a clique of size $k$ planted uniformly at random. As mentioned in the Prior Work section, all known polynomial time algorithms for planted clique fail if $k \ll \sqrt{n}$. This has led to the following hardness conjecture.
Conjecture 2.2.1 (PC Conjecture). Fix some constant $p \in (0, 1)$. Suppose that \( \{ A_n \} \) is a sequence of randomized polynomial time algorithms $A_n : \mathcal{G}_n \to \{0, 1\}$ and $k_n$ is a sequence of positive integers satisfying that $\limsup_{n \to \infty} \log n k_n < \frac{1}{2}$. Then if $G$ is an instance of $\text{PCD}(n, k, p)$, it holds that

$$\liminf_{n \to \infty} (\mathbb{P}_{H_0} [A_n(G) = 1] + \mathbb{P}_{H_1} [A_n(G) = 0]) \geq 1.$$ 

The hardness assumption we use throughout our results is the planted clique conjecture. Other than to show hardness for planted dense subgraph in the sparsest regime, we will only need the planted clique conjecture with edge density $p = 1/2$. An interesting open problem posed in [HWX15] is to show that the PC conjecture at $p = 1/2$ implies it for any fixed constant $p < 1/2$.

The hypotheses in the planted independent set detection problem $\text{PISD}(n, k, p)$ are

$$H_0 : G \sim G(n, p) \quad \text{and} \quad H_1 : G \sim G_I(n, k, p)$$

where $G_I(n, k, p)$ is a sample of $G(n, p)$ where all of the edges of vertex set of size $k$ removed uniformly at random. The recovery $\text{PCR}(n, k, p)$ and $\text{PISR}(n, k, p)$ problems are to estimate the latent clique and independent set supports given samples from $G(n, k, p)$ and $G_I(n, k, p)$, respectively.

**Planted Dense Subgraph** The hypotheses in the planted dense subgraph detection problem $\text{PDS}_D(n, k, p, q)$ are

$$H_0 : G \sim G(n, q) \quad \text{and} \quad H_1 : G \sim G(n, k, p, q)$$

where $G(n, k, p, q)$ is the distribution on $\mathcal{G}_n$ formed by selecting a size $k$ subset $S$ of $[n]$ uniformly at random and joining every two nodes in $S$ with probability $p$ and every other two nodes with probability $q$. The recovery problem $\text{PDSR}_R(n, k, p, q)$ is to estimate the latent planted dense subgraph support $S$ given samples from $G(n, k, p, q)$.

A phenomenon observed in [HWX15] and in [CX16] is that planted dense subgraph appears to have a detection-recovery gap in the regime where $k \gg \sqrt{n}$. The following is a formulation of the conjectured sharper recovery lower bound.

Conjecture 2.2.2 (PDS Recovery Conjecture). Suppose that $G \sim G(n, k, p, q)$ and

$$\liminf_{n \to \infty} \log n k > \frac{1}{2} \quad \text{and} \quad \limsup_{n \to \infty} \log n \left( \frac{k^2 (p - q)^2}{q (1 - q)} \right) < 1$$

then there is no sequence of randomized polynomial-time algorithms $A_n : \mathcal{G}_n \to \binom{n}{k}$ such that $A_n(G)$ achieve partial recovery of the vertices in the latent planted dense subgraph as $n \to \infty$.

This conjecture asserts that the threshold of $n/k^2$ on the signal $\frac{(p-q)^2}{q(1-q)}$ of an instance of PDS is tight for the recovery problem. In contrast, our results show that the tight detection threshold
for PDS given the PC conjecture is lower, at \( n^2/k^4 \). We will use this conjecture to establish similar detection-recovery gaps for biased sparse PCA and biclustering. We note that a related detection-recovery gap for BC was shown in [CLR15]. The same lower bound was established for strong recovery algorithms that solve biclustering for all subgaussian noise distributions assuming hardness of planted clique for a different distribution on random graphs than Erdős-Rényi.

Subgraph Stochastic Block Model We introduce a planted subgraph variant of the two community stochastic block model without the edge-thresholding test that produced the conjectural detection-recovery gap in planted dense subgraph. Let \( G_B(n, k, q, p) \) denote the set of distributions on \( \mathcal{G}_n \) generated a graph \( G \) as follows. Fix any two positive integers \( k_1 \) and \( k_2 \) satisfying that

\[
\frac{k}{2} - k^{1-\delta} \leq k_1, k_2 \leq \frac{k}{2} + k^{1-\delta}
\]

where \( \delta = \delta_{SSBM} > 0 \) is a small constant that will remained fixed throughout the paper. Let \( S = [k_1] \) and \( T = [k_1 + k_2] \setminus [k_1] \). Then generate the edges of \( G \) independently as follows:

1. include edges within \( S \) or within \( T \) with probability at least \( q + p \);
2. include edges between \( S \) and \( T \) with probability at most \( q - p \); and
3. include all other edges with probability \( q \).

Then permute the vertices of \( G \) according to a permutation selected uniformly at random. The communities of the graph are defined to be the images of \( S \) and \( T \) under this permutation. Note that \( G_B(n, k, q, p) \) defines a set of distributions since \( k_1 \) and \( k_2 \) are permitted to vary and the edges between \( S \) and \( T \) are included independently with a probability at least \( q + p \) for each edge. Thus given the random permutation, \( G \) is distributed as an inhomogeneous random graph with independent edges. The subgraph stochastic block model detection problem \( \text{SSBM}_D(n, k, q, p) \) has hypotheses given by

\[
H_0 : G \sim G(n, q) \quad \text{and} \quad H_1 : G \sim \mathcal{P} \quad \text{for some} \quad \mathcal{P} \in G_B(n, k, q, p)
\]

Biclustering Let \( \mathcal{M}_{n,k} \subseteq \mathbb{R}^{n \times n} \) be the set of sparse matrices supported on a \( k \times k \) submatrix with each nonzero entry equal to 1. The biclustering detection problem \( \text{BCD}_D(n, k, \mu) \) has hypotheses

\[
H_0 : M \sim N(0, 1)^{\otimes n \times n} \quad \text{and} \quad H_1 : M \sim \mu \cdot A + N(0, 1)^{\otimes n \times n} \quad \text{where} \quad A \sim \text{Unif}[\mathcal{M}_{n,k}]
\]

The recovery problem \( \text{BCR} \) is to estimate the latent support matrix \( A \) given a sample from \( H_1 \).
CHAPTER 2. SUMMARY OF RESULTS

Rank-1 Submatrix and Sparse Spiked Wigner In rank-1 submatrix, sparse spiked Wigner and sparse PCA, the planted sparse vectors will have sufficiently large entries for support recovery to be possible. We consider the following set of near-uniform magnitude unit vectors

\[ V_{d,k} = \left\{ v \in S^{d-1} : \frac{k}{\log k} \leq \|v\|_0 \leq k \text{ and } \|v_i\| \geq \frac{1}{\sqrt{k}} \text{ for } i \in \text{supp}(v) \right\} \]

The function \( \log k \) can be replaced by any sub-polynomially growing function but is given explicitly for simplicity. The detection problem \( \text{ROS}_D(n, k, \mu) \) has hypotheses

\[ H_0 : M \sim N(0,1)^{\otimes n \times n} \quad \text{and} \quad H_1 : M \sim \mu \cdot rc^\top + N(0,1)^{\otimes n \times n} \quad \text{where } r, c \in V_{n,k} \]

The recovery problem \( \text{ROS}_R \) is to estimate the latent supports \( \text{supp}(r) \) and \( \text{supp}(c) \). An \( n \times n \) GOE matrix \( \text{GOE}(n) \) is a symmetric matrix with i.i.d. \( N(0,1) \) entries below its main diagonal and i.i.d. \( N(0,2) \) entries on its main diagonal. The sparse spiked Wigner detection problem \( \text{SSW}_D \) has hypotheses

\[ H_0 : M \sim \text{GOE}(n) \quad \text{and} \quad H_1 : M \sim \mu \cdot rr^\top + \text{GOE}(n) \quad \text{where } r \in V_{n,k} \]

and the recovery problem \( \text{SSW}_R \) is to estimate the latent supports \( \text{supp}(r) \). A simple intermediate variant that will be useful in our reductions is SROS, which is ROS constrained to have a symmetric spike \( r = c \). Note that if \( M \) is an instance of \( \text{SROS}_D(n, k, \mu) \) then it follows that \( \frac{1}{\sqrt{2}}(M + M^\top) \) is an instance of \( \text{SSW}_D(n, k, \mu/\sqrt{2}) \).

Sparse PCA Detection in the spiked covariance model of \( \text{SPCA}_D(n, k, d, \theta) \) has hypotheses

\[ H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^{\otimes n} \quad \text{and} \quad H_1 : X_1, X_2, \ldots, X_n \sim N \left( 0, I_d + \theta vv^\top \right)^{\otimes n} \quad \text{where } v \in V_{d,k} \]

The recovery task \( \text{SPCA}_R \) is to estimate \( \text{supp}(v) \) given observations \( X_1, X_2, \ldots, X_n \) sampled from \( N \left( 0, I_d + \theta vv^\top \right)^{\otimes n} \) where \( v \in V_{d,k} \). We also consider a simple hypothesis testing variant \( \text{USPCA}_D \) of sparse PCA with hypotheses

\[ H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^{\otimes n} \quad \text{and} \quad H_1 : X_1, X_2, \ldots, X_n \sim N \left( 0, I_d + \theta vv^\top \right)^{\otimes n} \quad \text{where } v \sim \text{Unif}[S_k] \]

where \( S_k \subseteq \mathbb{R}^d \) consists of all \( k \)-sparse unit vectors with nonzero coordinates equal to \( \pm 1/\sqrt{k} \).
2.3. OUR RESULTS

Biased Sparse PCA We introduce a variant of the spiked covariance model with an additional promise. In particular, \(v\) is restricted to the set \(\mathcal{B}V_{d,k}\) of vectors in \(V_{d,k}\) with some overall positive or negative bias. Formally, if \(\|v\|_0^+\) denotes the number of positive entries of \(v\) then

\[
\mathcal{B}V_{d,k} = \left\{ v \in V_{d,k} : \|v\|_0^+ \geq \left( \frac{1}{2} + \delta \right) k \text{ or } \|v\|_0^+ \leq \left( \frac{1}{2} - \delta \right) k \right\}
\]

where \(\delta = \delta_{\text{BSPCA}} > 0\) is an arbitrary constant that will remain fixed throughout the paper. The detection problem \(\text{BSPCA}_D(n,k,d,\theta)\) has hypotheses

\[
H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^\otimes n \quad \text{and} \quad H_1 : X_1, X_2, \ldots, X_n \sim N\left(0, I_d + \theta vv^T\right)^\otimes n \text{ where } v \in \mathcal{B}V_{d,k}
\]

The recovery problem \(\text{BSPCA}_R\) is to estimate \(\text{supp}(v)\) given observations \(X_1, X_2, \ldots, X_n\) sampled from \(N\left(0, I_d + \theta vv^T\right)^\otimes n\) where \(v \in \mathcal{B}V_{d,k}\). We also consider a simple hypothesis testing variant \(\text{UBSPCA}_D\) defined similarly to \(\text{USPCAD}\) with \(v \sim \text{Unif}[BS_k]\) where \(BS_k\) is the set of all \(k\)-sparse unit vectors with nonzero coordinates equal to \(1/\sqrt{k}\).

2.3 Our Results

In this paper, our aim is to establish tight characterizations of the statistical-computational gaps for the problems formulated in the previous section. Each problem has three regimes for each of its detection and recovery variants. In the easy regime, there is a polynomial-time algorithm for the task. In the hard regime, the PC or PDS recovery conjecture implies that there is no polynomial-time algorithm but there is an inefficient algorithm solving the task. In the impossible regime, the task is information-theoretically impossible. Our results complete the hardness picture for each of these problems other than sparse PCA when \(k \ll \sqrt{n}\) and biased sparse PCA. Our results are informally stated below and depicted visually in Figure 2.1.

**Theorem 2.3.1** (Informal Main Theorem). Given the PC and PDS recovery conjectures, the easy, hard and impossible regimes of the problems in Section 2.2 are classified in Figure 2.3 as following one of the configurations of regimes in Figure 2.2.

We remark that all of the computational lower bounds in Theorem 2.3.1 follow from the PC conjecture other than those for \(\text{PISR}, \text{PDSR}, \text{BCR}\) and \(\text{BSPCAWR}\) which follow from the PDS conjecture. The computational lower bounds implicit in the hard regimes in Figures 2.2 and 2.3 are the focus of the present work. Chapter 4 introduces PC-LIFTING to reduce from \(\text{PC}_D\) to \(\text{PIS}_D\). Chapter 5 introduces rejection kernels and general DISTRIBUTIONAL-LIFTING which are
then applied in Chapter 6 to reduce from PC_D to all regimes of PDS_D. Chapter 7 introduces reflection cloning to reduce from BC_D to ROS_D, SSW_D and SSBM_D. Chapter 8 introduces random rotations to reduce from SSW_D to SPCAD and from BC_D to BSPCAD. In Chapter 6, we also give a reduction from PDS_R to BC_R and in Chapter 9, we reduce from PDS_R to BSPCAWR. In Chapter 9, we establish the algorithmic upper bounds and information-theoretic lower bounds needed to complete the proof of Theorem 2.3.1. We remark that in general detection lower bounds imply recovery lower bounds for the problems we consider since, given the support of the latent sparse structure, detection is easy. Thus it suffices to establish the hard regimes in Figures 2.2 and 2.3 only for detection when there is no detection-recovery gap.

Note that this gives a complete characterization of the easy, hard and impossible regions for all of the problems we consider other than sparse PCA and biased sparse PCA. The SDP relaxation of the MLE for sparse PCA was shown in [BR13b] to succeed if \( d = \Theta(n) \) and \( k \ll n^{1/2-\delta} \) down to the signal level of \( \theta \approx k/\sqrt{n} \), which is generally conjectured to be optimal. There is a gap between the lower bounds we prove here, which match those of [BR13b] and [GMZ17] and this conjecturally optimal threshold when \( k \ll \sqrt{n} \). There is also a gap for biased sparse PCA detection when \( k \gg \sqrt{n} \). These gaps are depicted as the black region in Figures 2.1(f) and 2.1(g). Showing tight hardness for sparse PCA for any parameters \( k \ll \sqrt{n} \) remains an interesting open problem. Notably, we do not consider the recovery problem for the subgraph stochastic block model and only consider weak, rather than exact, recovery for sparse PCA and biased sparse PCA. While we obtain computational lower bounds that apply to these recovery variants, obtaining matching information-theoretic lower bounds and algorithms remains an open problem. These open problems are discussed in more detail in Chapter 10.

In Figure 2.1 and Theorem 2.3.1, we depict the implications of our hardness results for SPCA and BSPCA in the case where \( d = \Theta(n) \) for simplicity. Our hardness results extend mildly beyond this regime, but not tightly. However, we remark that the assumptions \( d = \Theta(n) \) or \( d = \Omega(n) \) have become commonplace in the literature on lower bounds for sparse PCA. The reductions from PC to sparse PCA in [BR13a], [WBS16] and [GMZ17] construct hard instances that are only tight when \( d = \Omega(n) \). The sum of squares lower bounds for sparse PCA in [MW15a] and [HLP+17] also are in the regime \( d = \Theta(n) \). The SOS lower bounds in [HLP+17] are actually for the spiked Wigner model rather than the spiked covariance model of sparse PCA that we consider.
2.4 Our Techniques

Distributional Lifting We introduce several new techniques resembling graph lifts to increase the size \( k \) of a sparse structure, while appropriately maintaining the level of signal and independence in the noise distribution. Given a graph \( G \), the main idea behind our techniques is to replace the \{0, 1\}-valued edge indicators with Gaussian and Poisson random variables. We then increase the size of an instance by a factor of two iteratively, while maintaining the signal and independence, through distributional tricks such as Poisson thinning and the rotational invariance of independent Gaussians. In [HWX15], the reduction from planted clique to planted dense subgraph also expands an input graph. Rather than proceed iteratively, their method expands the graph in one step. The main technical issue arising from this is that the diagonal entries of the graph's adjacency matrix are mapped to low-density subgraphs in the hidden community. Showing that these are not detectable requires a subtle argument and that the hidden community is randomly sized according to a Binomial distribution. By proceeding incrementally as in our approach, the diagonal entries become much easier to handle. However for many regimes of interest, incremental approaches that preserve the fact that the instance is a graph seem to unavoidably introduce dependence between edges. Our insight is to map edges to other random variables that can be preserved incrementally while maintaining independence and the desired parameter scaling to produce tight lower bounds. Using Poisson and Gaussian variants of this lifting procedure, we are able to reduce from planted clique in a wide range of parameter regimes. Gaussian lifting also recover a simple vs. simple hypothesis testing variant of the lower bounds for biclustering shown in [MW15b] as an intermediate step towards reducing to planted dense subgraph.

Rejection Kernels We give a simple scheme based on rejection sampling that approximately maps a sample from Bern\((p)\) to a sample from \( P \) and from Bern\((q)\) to \( Q \) where \( p, q \in [0, 1] \) and \( P \) and \( Q \) are two distributions on \( \mathbb{R} \). By truncating samples from a pair of distributions \( P' \) and \( Q' \), and then mapping the resulting Bernoulli random variables to a pair of target distributions \( P \) and \( Q \), this method yields an efficient procedure to simultaneously perform two changes of measure. This method is used in distributional lifting to map from edge indicators to a pair of chosen distributions. This framework extends and uses similar ideas to the approximate sampling methods introduced in [HWX15] and [MW15b].

Reflection Cloning While distributional lifting appears to accurately characterize the hard regimes in many detection problems with an optimal test involving summing the entire input matrix, it is fundamentally lossy. In each iteration, these lifting techniques generate additional randomness
in order to maintain independence in the noise distribution of the instance. In a problem like rank-1 submatrix detection that does not admit a sum test, these cloning techniques do not come close to showing hardness at the computational barrier. We introduce a more sophisticated cloning procedure for cases of Gaussian noise that introduces significantly less randomness in each iteration. Let \( R \) denote the linear operator on \( n \times n \) matrices that reflects the matrix about its vertical axis of symmetry and let \( F \) denote the linear operator that multiplies each entry on the right half of the matrix by \(-1\). Then one step of reflection cloning replaces a matrix \( W \) with

\[
W \leftarrow \frac{1}{\sqrt{2}} (RW^\sigma + FW^\sigma)
\]

where \( \sigma \) is a random permutation. Reflection cloning then repeats this for rows instead of columns.

If \( W = uv^T + G \) has even dimensions and \( G \) has i.i.d. \( N(0,1) \) entries, then reflection cloning effectively doubles the sparsity of \( u \) and \( v \) while mildly decreasing the signal. Importantly, it can be checked that the gaussian noise matrix retains the fact that it has independent entries. It is interesting to note that this cloning procedure precludes the success of a sum test by negating parts of the signal. Using reflection cloning, we establish tight computational lower bounds for rank-1 submatrix detection and for sparse PCA in the regime \( k > \sqrt{n} \). This marks the first tight computational lower bound for sparse PCA over an entire parameter regime.

**Random Rotations and Sparse PCA** We also introduce a simple connection between sparse PCA, biclustering and rank-1 submatrix through random rotations. This yields lower bounds matching those of [GMZ17] and [BR13a]. Although often suboptimal, the random rotations map we introduce tightly gives lower bounds in the regime \( k \gg \sqrt{n} \), using reflection cloning as an intermediate. This illustrates the utility of natural average-case problems as reduction intermediates, suggesting that webs of reductions among problems can be useful beyond worst-case complexity.
2.4. OUR TECHNIQUES

(a) PIS($n, k, q$) and PDS($n, k, p, q$) with $p = cq = \Theta(n^{-\alpha})$

(b) PDS($n, k, p, q$) with $q = \Theta(n^{-\alpha})$ and $p - q = \Theta(n^{-5\alpha/4})$

(c) SSBM($n, k, q, p$) with $q = \Theta(1)$ and $p = \Theta(n^{-\alpha})$

(d) BC($n, k, \mu$) with $\mu = \Theta(n^{-\alpha})$

(e) ROS($n, k, \mu$) and SSW($n, k, \mu$) with $\frac{k}{\mu} = \Theta(n^{-\alpha})$

(f) SPCA($n, k, d, \theta$) with $d = \Theta(n)$ and $\theta = \Theta(n^{-\alpha})$

(g) BSPCA($n, k, d, \theta$) with $d = \Theta(n)$ and $\theta = \Theta(n^{-\alpha})$

Figure 2.1: Parameter regimes by problem plotted as signal vs. sparsity. Sparsity is $k = \Theta(n^\beta)$. First labels characterize detection and second labels characterize exact recovery. Recovery is not considered for SSBM and weak recovery is considered for SPCA and BSPCA. In Easy (E) regimes, there is a polynomial-time algorithm. In Hard (H) regimes, the PC or PDS conjecture implies there is no polynomial-time algorithm. In Impossible (I) regimes, the task is information-theoretically impossible. Hardness in black regions is open.
**Figure 2.2:** Types of hardness regimes by SNR in Theorem 2.3.1. Type I and Type III each have two variants A and B depending on the Easy regime when $k \ll n^{1/2}$. 

### Type I

<table>
<thead>
<tr>
<th>Type</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k \ll n^{2/3}$</th>
<th>$n^{2/3} \ll k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impossible</td>
<td>SNR $\ll \frac{1}{k}$</td>
<td>SNR $\ll \frac{1}{k}$</td>
<td>SNR $\ll \frac{n^2}{k^4}$</td>
</tr>
<tr>
<td>Hard</td>
<td>SNR $\gtrsim \frac{1}{k}$</td>
<td>$\frac{1}{k} \lesssim$ SNR $\ll \frac{n^2}{k^4}$</td>
<td>None</td>
</tr>
<tr>
<td>Easy</td>
<td>(A) None</td>
<td>SNR $\gtrsim \frac{n^2}{k^4}$</td>
<td>SNR $\gtrsim \frac{n^2}{k^4}$</td>
</tr>
<tr>
<td></td>
<td>(B) SNR $\gtrsim 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Type II

<table>
<thead>
<tr>
<th>Type</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k \ll n^{2/3}$</th>
<th>$n^{2/3} \ll k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impossible</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
</tr>
<tr>
<td>Hard</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>None</td>
</tr>
<tr>
<td>Easy</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
</tr>
</tbody>
</table>

### Type III

<table>
<thead>
<tr>
<th>Type</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k$</th>
<th>Type IV</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impossible</td>
<td>SNR $\ll \frac{1}{k}$</td>
<td>SNR $\ll \frac{1}{k}$</td>
<td>Impossible</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
</tr>
<tr>
<td>Hard</td>
<td>SNR $\gtrsim \frac{1}{k}$</td>
<td>$\frac{1}{k} \lesssim$ SNR $\ll \frac{n^2}{k^4}$</td>
<td>Hard</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll 1$</td>
</tr>
<tr>
<td>Easy</td>
<td>(A) None</td>
<td>SNR $\gtrsim \frac{n^2}{k^4}$</td>
<td>Easy</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>SNR $\gtrsim 1$</td>
</tr>
<tr>
<td></td>
<td>(B) SNR $\gtrsim 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Type IV

<table>
<thead>
<tr>
<th>Type</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k$</th>
<th>Type IV</th>
<th>$k \ll n^{1/2}$</th>
<th>$n^{1/2} \lesssim k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impossible</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>Impossible</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
<td>SNR $\ll \sqrt{\frac{k}{n}}$</td>
</tr>
<tr>
<td>Hard</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>Hard</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$</td>
<td>$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll 1$</td>
</tr>
<tr>
<td>Easy</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>Easy</td>
<td>SNR $\gtrsim \sqrt{\frac{k^2}{n}}$</td>
<td>SNR $\gtrsim 1$</td>
</tr>
</tbody>
</table>
2.4. OUR TECHNIQUES

<table>
<thead>
<tr>
<th>Problems</th>
<th>Parameter Regime</th>
<th>SNR</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PIS_D(n, k, q), PDS_D(n, k, cq, q)$</td>
<td>$q = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$ and $c &gt; 1$</td>
<td>$q$</td>
<td>Type IIA</td>
</tr>
<tr>
<td>$PIS_R(n, k, q), PDS_R(n, k, cq, q)$</td>
<td>$q = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$ and $c &gt; 1$</td>
<td>$q$</td>
<td>Type IIIA</td>
</tr>
<tr>
<td>$PDS_D(n, k, p, q)$</td>
<td>$q = \Theta(n^{-\alpha})$ and $p - q = \Theta(n^{-\gamma})$ with $p &gt; q$ for fixed $\alpha, \gamma \in [0, 1)$</td>
<td>$(p-q)^2$</td>
<td>Type IIA</td>
</tr>
<tr>
<td>$PDS_R(n, k, p, q)$</td>
<td>$q = \Theta(n^{-\alpha})$ and $p - q = \Theta(n^{-\gamma})$ with $p &gt; q$ for fixed $\alpha, \gamma \in [0, 1)$</td>
<td>$(p-q)^2$</td>
<td>Type IIIA</td>
</tr>
<tr>
<td>$SSBM_D(n, k, q, \rho)$</td>
<td>$q = \Theta(1)$ and $\rho = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\rho^2$</td>
<td>Type IIIA</td>
</tr>
<tr>
<td>$BC_D(n, k, \mu)$</td>
<td>$\mu = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\mu^2$</td>
<td>Type IB</td>
</tr>
<tr>
<td>$BC_R(n, k, \mu)$</td>
<td>$\mu = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\mu^2$</td>
<td>Type IIIB</td>
</tr>
<tr>
<td>$ROSD(n, k, \mu), ROS_R(n, k, \mu), SSW_D(n, k, \mu), SSW_R(n, k, \mu)$</td>
<td>$\mu = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\mu^2$/$k^2$</td>
<td>Type IIIB</td>
</tr>
<tr>
<td>$SPCA_D(n, k, d, \theta)$,</td>
<td>$d = \Theta(n) \text{ and } \theta = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\theta$</td>
<td>Type II</td>
</tr>
<tr>
<td>$SPCA_WR(n, k, d, \theta)$,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BSPCA_WR(n, k, d, \theta)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BSPCA_D(n, k, d, \theta)$</td>
<td>$d = \Theta(n) \text{ and } \theta = \Theta(n^{-\alpha})$ for fixed $\alpha \in [0, 1)$</td>
<td>$\theta$</td>
<td>Type IV</td>
</tr>
</tbody>
</table>

**Figure 2.3:** Classification of regimes for each problem as in Theorem 2.3.1. For each problem, $k$ is in the regime $k = \Theta(n^\beta)$ where $\beta \in (0, 1)$ is a constant.
Chapter 3

Average-Case Reductions under Total Variation

The typical approach to show computational lower bounds for detection problems is to reduce an instance of one problem to a random object close in total variation distance to an instance of another problem in randomized polynomial time. More precisely, let \( P \) and \( P' \) be detection problems and \( X \) and \( Y \) be instances of \( P \) and \( P' \), respectively. Suppose we are given a polynomial-time computable map \( \phi \) taking an object \( \phi(X) \) with total variation distance to \( Y \) decaying to zero simultaneously under each of \( H_0 \) and \( H_1 \). Then any algorithm that can distinguish \( H_0 \) and \( H_1 \) for \( P' \) in polynomial time when applied to \( \phi(X) \) also distinguishes \( H_0 \) and \( H_1 \) for \( P \). Taking \( P \) to be \( P_C \) and \( P' \) to be the problem of interest then yields a computational hardness result for \( P' \) conditional on the \( P_C \) conjecture. The general idea in this approach is formalized in the following simple lemma.

Lemma 3.0.1. Let \( P \) and \( P' \) be detection problems with hypotheses \( H_0, H_1, H_0', H_1' \) and let \( X \) and \( Y \) be instances of \( P \) and \( P' \), respectively. Suppose there is a polynomial time computable map \( \phi \) satisfying

\[
d_{TV}(\mathcal{L}_{H_0}(\phi(X)), \mathcal{L}_{H_0'}(Y)) + \sup_{P \in H_1} \inf_{\pi \in \Delta(H_1')} \mathbb{E}_{x \sim P} \left[ \int_{H_1'} \mathcal{L}_{P'}(Y) d\pi(P') \right] \leq \delta
\]

If there is a polynomial time algorithm solving \( P' \) with Type I+II error at most \( \epsilon \), then there is a polynomial time algorithm solving \( P \) with Type I+II error at most \( \epsilon + \delta \).

Proof. Let \( \psi \) be a polynomial time computable test function solving \( P' \) with Type I+II error at most \( \epsilon \). Note that for any observation \( X \) of \( P \), the value \( \psi \circ \phi(X) \in \{0, 1\} \) can be computed in polynomial time. This is because the fact that \( \phi(X) \) can be computed in polynomial time implies
that $\phi(X)$ has size polynomial in the size of $X$. We claim that $\psi \circ \phi$ solves the detection problem $P'$. Now fix some distribution $P \in H_1$ and prior $\pi \in \Delta(H'_1)$. By the definition of total variation,

$$\left| P_{H_0}[\psi \circ \phi(X) = 1] - P_{H'_0}[\psi(Y) = 1] \right| \leq d_{TV}(L_{H_0}(\phi(X)), L_{H'_0}(Y))$$

$$\left| P_{X \sim P}[\psi \circ \phi(X) = 0] - \int_{H'_1} P_{Y \sim P'}[\psi(Y) = 0] d\pi(P') \right| \leq d_{TV}(L_P(\phi(X)), \int_{H'_1} L_{P'}(Y)d\pi(P'))$$

Also note that since $\pi$ is a probability distribution,

$$\int_{H'_1} P_{Y \sim P'}[\psi(Y) = 0] d\pi(P') \leq \sup_{P' \in H'_1} P_{Y \sim P'}[\psi(Y) = 0]$$

Combining these inequalities with the triangle inequality yields that

$$P_{H_0}[\psi \circ \phi(X) = 1] + P_{X \sim P}[\psi \circ \phi(X) = 0] \leq \epsilon + d_{TV}(L_{H_0}(\phi(X)), L_{H'_0}(Y))$$

$$+ d_{TV}(L_P(\phi(X)), \int_{H'_1} L_{P'}(Y)d\pi(P'))$$

Fixing $P$ and choosing the prior $\pi$ so that the second total variation above approaches its infimum yields that the right hand side above is upper bounded by $\epsilon + \delta$. The fact that this bound holds for all $P \in H_1$ proves the lemma. \qed

We remark that the second term in the total variation condition of Lemma 1 can be interpreted as ensuring that each distribution $P \in H_1$ is close to a distribution formed by taking a prior $\pi$ over the distributions in hypothesis $H'_1$. In light of this lemma, to reduce one problem to another it suffices to find such a map $\phi$. In the case that $P$ and $P'$ are both simple hypothesis testing problems, the second term is simply $d_{TV}(L_{H_1}(\phi(X)), L_{H'_1}(Y))$.

Throughout the analysis of our average-case reductions, total variation will be the key object of interest. We will make use of several standard results concerning total variation, including the triangle inequality, data processing inequality and tensorization of total variation. The latter two results are stated below.

**Lemma 3.0.2 (Data Processing).** Let $P$ and $Q$ be distributions on a measurable space $(\mathcal{X}, \mathcal{B})$ and let $f : \mathcal{X} \to \mathcal{Y}$ be a Markov transition kernel. If $A \sim P$ and $B \sim Q$ then

$$d_{TV}(L(f(A)), L(f(B))) \leq d_{TV}(P, Q)$$

**Lemma 3.0.3 (Tensorization).** Let $P_1, P_2, \ldots, P_n$ and $Q_1, Q_2, \ldots, Q_n$ be distributions on a measurable space $(\mathcal{X}, \mathcal{B})$. Then

$$d_{TV}\left(\prod_{i=1}^n P_i, \prod_{i=1}^n Q_i\right) \leq \sum_{i=1}^n d_{TV}(P_i, Q_i)$$
A typical analysis of a multi-step algorithm will proceed as follows. Suppose that \( A = A_2 \circ A_1 \) is an algorithm with two steps \( A_1 \) and \( A_2 \). Let \( P_0 \) be the input distribution and \( P_2 \) be the target distribution that we would like to show is close in variation to \( A(P_0) \). Let \( P_1 \) be an intermediate distribution which is close in total variation to \( A_1(P_0) \). By the triangle inequality,

\[
d_{TV}(A(P_0), P_2) \leq d_{TV}(A(P_0), A_2(P_1)) + d_{TV}(A_2(P_1), P_2)
\]

\[
= d_{TV}(A_2 \circ A_1(P_0), A_2(P_1)) + d_{TV}(A_2(P_1), P_2)
\]

\[
\leq d_{TV}(A_1(P_0), P_1) + d_{TV}(A_2(P_1), P_2)
\]

by the data-processing inequality. Thus total variation accumulates over the steps of a multi-step algorithm. This style of analysis will appear frequently in our reductions. Another lemma about total variation that will be useful throughout this work is as follows.

**Lemma 3.0.4.** For any random variable \( Y \) and event \( A \) in the \( \sigma \)-algebra \( \sigma\{Y\} \), it holds that

\[
d_{TV}(\mathcal{L}(Y|A), \mathcal{L}(Y)) = \mathbb{P}[Y \in A^c]
\]

**Proof.** Let \( B \) be any event in \( \sigma\{Y\} \). It follows that

\[
\mathbb{P}[Y \in B|Y \in A] - \mathbb{P}[Y \in B] = \mathbb{P}[Y \in B|Y \in A] \cdot (1 - \mathbb{P}[Y \in A]) - \mathbb{P}[Y \in B \cap A^c]
\]

Since \( \mathbb{P}[Y \in B \cap A^c] \leq \mathbb{P}[Y \in A^c] = 1 - \mathbb{P}[Y \in A] \) and \( \mathbb{P}[Y \in B|Y \in A] \in [0, 1] \), we have that the quantity above is between \(-\mathbb{P}[Y \in A^c]\) and \(\mathbb{P}[Y \in A^c]\). Therefore, by the definition of total variation distance

\[
d_{TV}(\mathcal{L}(Y|A), \mathcal{L}(Y)) = \sup_{B \in \sigma\{Y\}} \|\mathbb{P}[Y \in B|Y \in A] - \mathbb{P}[Y \in B]\| = \mathbb{P}[Y \in A^c]
\]

where the equality case is achieved by setting \( B = A \). This proves the lemma. \( \square \)
CHAPTER 3. AVERAGE-CASE REDUCTIONS UNDER TOTAL VARIATION
Chapter 4

Densifying Planted Clique and Planted Independent Set

In this section, we give a reduction increasing the ambient edge density in planted clique and showing tight hardness for the planted independent set problem. This reduction, which we term PC-LIFTING, serves as an introduction to the general distributional lifting procedure in the next chapter. Distributional lifting will subsequently be specialized to produce Poisson and Gaussian variants of the procedure, which will be used to prove hardness for biclustering and different regimes of planted dense subgraph.

4.1 Detecting Planted Generalized Diagonals

We first prove a technical lemma that will be used in all of our cloning procedures. Given a matrix $M$, let $M_{\sigma_1,\sigma_2}$ denote the matrix formed by permuting rows according to $\sigma_1$ and columns according to $\sigma_2$. Let $id$ denote the identity permutation.

Lemma 4.1.1. Let $P$ and $Q$ be two distributions such that $Q$ dominates $P$ and $\chi^2(P, Q) \leq 1$. Suppose that $M$ is an $n \times n$ matrix with all of its non-diagonal entries i.i.d. sampled from $Q$ and all of its diagonal entries i.i.d. sampled from $P$. Suppose that $\sigma$ is a permutation on $[n]$ chosen uniformly at random. Then

$$d_{TV}(\mathcal{L}(M_{id,\sigma}), Q^{\otimes n \times n}) \leq \sqrt{\frac{\chi^2(P, Q)}{2}}$$

Proof. Let $\sigma'$ be a permutation of $[n]$ chosen uniformly at random and independent of $\sigma$. Note that
by Fubini’s theorem we have that
\[ \chi^2 \left( \mathcal{L}(M^{\text{id},\sigma}), Q^\otimes_n \right) + 1 = \int \frac{\mathbb{E}_\sigma [P_{M^{\text{id},\sigma}}(X|\sigma)]^2}{P_{Q^\otimes_n}(X)} dX = \mathbb{E}_{\sigma,\sigma'} \int \frac{P_{M^{\text{id},\sigma}}(X|\sigma)P_{M^{\text{id},\sigma'}}(X|\sigma')}{P_{Q^\otimes_n}(X)} dX \]

Now note that conditioned on \( \sigma \), the entries of \( M^{\text{id},\sigma} \) are independent with distribution
\[ P_{M^{\text{id},\sigma}}(X|\sigma) = \prod_{i=1}^n P(X_{i\sigma(i)}) \prod_{j \neq \sigma(i)} Q(X_{ij}) \]

Therefore we have that
\[
\int \frac{P_{M^{\text{id},\sigma}}(X|\sigma)P_{M^{\text{id},\sigma'}}(X|\sigma')}{P_{Q^\otimes_n}(X)} dX = \int \left( \prod_{i: \sigma(i) = \sigma'(i)} \frac{P(X_{i\sigma(i)})}{Q(X_{i\sigma(i)})} \right) \left( \prod_{i: \sigma(i) \neq \sigma'(i)} P(X_{i\sigma(i)}) \right) \times \left( \prod_{i: \sigma(i) \neq \sigma'(i)} P(X_{i\sigma'(i)}) \right) \left( \prod_{j \neq \sigma(i), j \neq \sigma'(i)} Q(X_{ij}) \right) dX
\]
\[
= \prod_{i: \sigma(i) = \sigma'(i)} \left( \int \frac{P(X_{i\sigma(i)})}{Q(X_{i\sigma(i)})} dX_{i\sigma(i)} \right) - 1 \leq \left( 1 + \chi^2(P,Q) \right)^{|\{i: \sigma(i) = \sigma'(i)\}|}
\]

If \( \tau = \sigma' \circ \sigma^{-1} \), then \( \tau \) is a uniformly at random chosen permutation and \( Y = |\{i: \sigma(i) = \sigma'(i)\}| \) is the number of fixed points of \( \tau \). As in [Pit97], the \( i \)th moment of \( Y \) is the \( i \)th Bell number for \( i \leq n \) and for \( i > n \), the \( i \)th moment of \( Y \) is at most the \( i \)th Bell number. Since a Poisson distribution with rate 1 has its \( i \)th moment given by the \( i \)th Bell number for all \( i \), it follows that for each \( t \geq 0 \) the MGF \( \mathbb{E}[e^{tY}] \) is at most that of a Poisson with rate 1, which is \( \exp(e^t - 1) \). Setting \( t = \log(1 + \chi^2(P,Q)) > 0 \) yields that
\[ \chi^2 \left( \mathcal{L}(M^{\text{id},\sigma}), Q^\otimes_n \right) = \mathbb{E} \left[ (1 + \chi^2(P,Q))^Y \right] - 1 \leq \exp(\chi^2(P,Q) - 1) \leq 2 \cdot \chi^2(P,Q) \]
since \( e^x \leq 1 + 2x \) for \( x \in [0, 1] \). Now by Cauchy-Schwarz we have that
\[ d_{TV} \left( \mathcal{L}(M^{\text{id},\sigma}), Q^\otimes_n \right) \leq \frac{1}{2} \sqrt{\chi^2(M^{\text{id},\sigma}), Q^\otimes_n} \leq \sqrt{\frac{\chi^2(P,Q)}{2}} \]
which completes the proof of the lemma. \( \square \)

### 4.2 Planted Clique Lifting

In this section, we analyze the reduction PC-LIFTING, which is given in Figure 4.1. This reduction will be shown to approximately take an instance of PC\((n, n^{1/2-\epsilon}, 1/2)\) to PC\((N, k, 1 - q)\) where
4.2. PLANTED CLIQUE LIFTING

\( N = \Theta(n^{1+\alpha}), \ k = \Theta(n^{1/2+\alpha/2-\epsilon}) \) and \( q = \Theta(n^{-\alpha}) \). By taking the complement of the resulting graph, this shows planted clique lower bounds for \( PISD(N, k, q) \) up to the boundary \( \frac{N^2}{4k} \gg q \), exactly matching the computational boundary stated in Theorem 1. The reduction PC-LIFTING proceeds iteratively, with PC\((n, k, p)\) approximately mapped at each step to PC\((2n, 2k, p^{1/4})\).

Given a labelled graph \( G \) on \( n \) vertices and a permutation \( \sigma \) on \([n]\), let \( G^\sigma \) denote the labelled graph formed by permuting the vertex labels of \( G \) according to \( \sigma \). Given disjoint subsets \( S, T \subseteq [n] \), let \( G[S] \) denote the induced subgraph on the set \( S \) and \( G[S \times T] \) denote the induced bipartite subgraph between \( S \) and \( T \). Also let \( B(m, n, p) \) denote the random bipartite graph with parts of sizes \( m \) and \( n \), respectively, where each edge is included independently with probability \( p \). Let \( G(n, p, S) \) where \( S \) is a \( k \)-subset of \([n]\) denote an instance of \( G(n, k, p) \) where the planted clique is conditioned to be on \( S \).

Lemma 4.2.1 (Planted Clique Lifting). Suppose that \( n \) and \( \ell \) are such that \( \ell = O(\log n) \) and are sufficiently large. Let \( w(n) > 2 \) be an increasing function with \( w(n) \to \infty \) as \( n \to \infty \). Then \( \phi = PC-LIFTING \) is a randomized polynomial time computable map \( \phi : G_n \to G_{2^\ell n} \) such that under both \( H_0 \) and \( H_1 \), it holds that

\[
d_{TV}(\phi(PC(n, k, 1/2)), PC(2^\ell n, 2^{\ell} k, (1 - w(n))^{-\frac{1}{4}})) \leq \frac{2}{\sqrt{w(n)}}
\]

Proof. If \( \ell = O(\log n) \), this algorithm runs in randomized polynomial time. Let \( \phi_\ell \) be the algorithm that outputs the value of \( H \) in \( \phi \) after \( \ell \) iterations of Step 3. Note that \( \phi_0 \) outputs \( H \) after Steps 1 and 2.

We first consider a single iteration of Step 3 applied to \( G \sim G(n, p, S) \), where \( G(n, p, S) \) is the distribution of Erdős-Renyi graphs with a planted clique on a fixed vertex set \( S \subseteq [n] \) of size \( |S| = k \) and \( p \geq 1/2 \). For each pair of distinct \( \{i, j\} \notin \binom{S}{2} \), it holds that \( 1_{\{i, j\} \in E(G)} \sim \text{Bern}(p) \) and by the probability in Step 3a that \( x_{ij} \sim \text{Bern}(p^{1/4}) \otimes \mathcal{A} \). Therefore the graph \( H' \) constructed in Step 3b satisfies that:

- \( S' = S \cup \{2n + 1 - i : i \in S\} \) forms a clique of size \( 2k \);
- \( \{2n + 1 - i, i\} \in E(H') \) for each \( i \in [n] \); and
- each other edge is in \( E(H') \) independently with probability \( p^{1/4} \).

Now consider the graph \( \phi_1(G) = H = (H')^\sigma \) conditioned on the set \( \sigma(S') \). We will show that this graph is close in total variation to \( G(2n, p^{1/4}, \sigma(S')) \). Let \( T_1 = [n] \setminus S \) and \( T_2 = [2n] \setminus \{2n + 1 - i : i \in S\} \). Note that every pair of vertices of the form \( \{2n + 1 - i, i\} \) in \( H' \) are either both in \( S' \) or between \( T_1 \) and \( T_2 \). This implies that every pair of distinct vertices not in \( \sigma(S')^2 \) or \( \sigma(T_1) \times \sigma(T_2) \) is in \( E(H) \)
Algorithm PC-Lifting

Inputs: Planted clique instance $G \in G_n$, number of iterations $\ell$, function $w$ with $w(n) \to \infty$

1. For each pair of vertices $\{i, j\} \notin E(G)$, add the edge $\{i, j\}$ to $E(G)$ independently with probability $1 - 2 \cdot w(n)^{-1}$

2. Initialize $H \leftarrow G$, $m \leftarrow n$ and $p \leftarrow 1 - w(n)^{-1}$

3. For $i = 0, 1, \ldots, \ell - 1$ do:
   a. For each pair $\{i, j\}$ of distinct vertices in $[m]$, sample $x^{ij} \in \{0, 1\}^4$ such that
      - If $\{i, j\} \in E(H)$, then $x^{ij} = (1, 1, 1, 1)$
      - If $\{i, j\} \notin E(H)$, then $x^{ij} = v$ with probability
        $\mathbb{P}[x^{ij} = v] = \frac{p^{v_{1/4}} (1 - p^{1/4})^{4 - |v_1|}}{1 - p}$
        for each $v \in \{0, 1\}^4$ with $v \neq (1, 1, 1, 1)$
   b. Construct the graph $H'$ on the vertex set $[2m]$ such that for distinct $i, j \in [m]$
      - $\{i, j\} \in E(H')$ if $x^{ij}_1 = 1$
      - $\{2m + 1 - i, j\} \in E(H')$ if $x^{ij}_2 = 1$
      - $\{i, 2m + 1 - j\} \in E(H')$ if $x^{ij}_3 = 1$
      - $\{2m + 1 - i, 2m + 1 - j\} \in E(H')$ if $x^{ij}_4 = 1$
      and for each $i \in [m]$, add the edge $\{i, 2m + 1 - i\} \in E(H')$
   c. Generate a permutation $\sigma$ on $[2m]$ uniformly at random
   d. Update $H \leftarrow (H')^\sigma$, $p \leftarrow p^{1/4}$ and $m \leftarrow 2m$

4. Output $H$

**Figure 4.1:** Planted clique lifting procedure in Lemma 4.2.1.

independent with probability $p^{1/4}$, exactly matching the corresponding edges in $G(2n, p^{1/4}, \sigma(S'))$. Coupling these corresponding edges yields only the edges between $\sigma(T_1)$ and $\sigma(T_2)$ uncoupled.
4.2. PLANTED CLIQUE LIFTING

Therefore we have that
\[ d_{TV} \left( \mathcal{L}(H|\sigma(S')), G\left(2n, p^{1/4}, \sigma(S')\right) \right) = d_{TV} \left( \mathcal{L}(H[\sigma(T_1) \times \sigma(T_2)]), B\left(n - k, n - k, p^{1/4}\right) \right) \]

Now let the \((n-k) \times (n-k)\) matrix \(M\) have 1's on its main diagonal and each other entry distributed sampled i.i.d. from \(\text{Bern}(p^{1/4})\). If \(\tau\) is a random permutation on \([n-k]\), then the adjacency matrix of \(H[\sigma(T_1) \times \sigma(T_2)]\) conditioned on \(\sigma(S')\) is distributed as \(\mathcal{L}(M^{\text{id},\tau})\), since \(T_1\) and \(T_2\) are disjoint. Therefore it follows that
\[ d_{TV} \left( \mathcal{L}(H[\sigma(T_1) \times \sigma(T_2)]), B\left(n - k, n - k, p^{1/4}\right) \right) = d_{TV} \left( \mathcal{L}(M^{\text{id},\tau}), \text{Bern}(p^{1/4}) \otimes (n-k) \times (n-k) \right) \leq \sqrt{\frac{\chi^2(\text{Bern}(1), \text{Bern}(p^{1/4}))}{2}} \leq \sqrt{1 - p^{1/4}} \]
by Lemma 4.1.1 and since \(p \geq 1/2\). It follows by the triangle inequality that
\[ d_{TV} \left( \phi_1(G(n, p^{1/4}, S)), G(2n, 2k, p^{1/4}) \right) \leq \mathbb{E}_{\sigma(S')} \left[ d_{TV} \left( \mathcal{L}(H|\sigma(S')), G\left(2n, p^{1/4}, \sigma(S')\right) \right) \right] \]

Letting \(S\) be chosen uniformly at random over all subsets of \([n]\) of size \(k\), applying the triangle inequality again and combining the inequalities above yields that
\[ d_{TV} \left( \phi_1(G(n, k, p)), G(2n, 2k, p^{1/4}) \right) \leq \mathbb{E}_S \left[ d_{TV} \left( \phi_1(G(n, p, S)), G(2n, 2k, p^{1/4}) \right) \right] \leq \sqrt{1 - p^{1/4}} \]
A nearly identical but slightly simpler argument shows that
\[ d_{TV} \left( \phi_1(G(n, p)), G(2n, p^{1/4}) \right) \leq \sqrt{1 - p^{1/4}} \]
For each \(\ell \geq 0\), let \(p_\ell = (1 - w(n)^{-1})^{1/\ell}\) be the value of \(p\) after \(\ell\) iterations of Step \(2\). Now note that for each \(\ell \geq 0\), we have by triangle inequality and data processing inequality that
\[ d_{TV} \left( \phi_{\ell+1}(G(n, k, 1/2)), G\left(2^{\ell+1}n, 2^{\ell+1}k, p_{\ell+1}\right) \right) \leq d_{TV} \left( \phi_1 \left( \phi_\ell \left( G(n, k, 1/2) \right) \right), \phi_1 \left( G\left(2^\ell n, 2^\ell k, p_\ell\right) \right) \right) + d_{TV} \left( \phi_1 \left( G\left(2^\ell n, 2^\ell k, p_\ell\right) \right), G\left(2^{\ell+1}n, 2^{\ell+1}k, p_{\ell+1}\right) \right) \leq d_{TV} \left( \phi_\ell \left( G(n, k, 1/2) \right), G\left(2^\ell n, 2^\ell k, p_\ell\right) \right) + \sqrt{1 - p_{\ell+1}} \]
and an identical inequality for \(\phi_\ell(G(n, 1/2))\). Noting that this total variation is zero when \(\ell = 0\) and applying these inequalities inductively yields that
\[ d_{TV} \left( \phi_\ell \left( G(n, k, 1/2) \right), G\left(2^\ell n, 2^\ell k, p_\ell\right) \right) \leq \sum_{i=1}^{\ell} \sqrt{1 - p_i} \]
48 CHAPTER 4. DENSIFYING PLANTED CLIQUE AND PLANTED INDEPENDENT SET

and an identical inequality for \( \phi_d(G(n, 1/2)) \). Now note that if \( x \leq 1/2 \) then \((1 - x)^{1/4} \geq 1 - x/3\). Iterating this inequality yields that \( 1 - p_i \leq 3^{-i} w(n)^{-1} \). Therefore

\[
\sum_{i=1}^{\ell} \sqrt{1 - p_i} \leq \frac{1}{\sqrt{w(n)}} \sum_{i=1}^{\ell} 3^{-i/2} < \frac{2}{\sqrt{w(n)}}
\]

This completes the proof of the lemma. \( \square \)

The next theorem formally gives the hardness result guaranteed by the reduction analyzed above together with the PC conjecture. There will be many theorems of this form throughout the paper, which will typically resolve to applying a total variation bound guaranteed in a previous lemma with Lemma 3.0.1, and analyzing the asymptotic regime of several parameters.

**Theorem 4.2.1.** Let \( \alpha \in [0, 2) \) and \( \beta \in (0, 1) \) be such that \( \beta < \frac{1}{2} + \frac{\alpha}{4} \). There is a sequence \( \{(N_n, K_n, q_n)\}_{n \in \mathbb{N}} \) of parameters such that:

1. The parameters are in the regime \( q = \tilde{\Theta}(N^{-\alpha}) \) and \( K = \tilde{\Theta}(N^\beta) \) or equivalently,

\[
\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
\]

2. For any sequence of randomized polynomial-time tests \( \phi_n : \mathcal{G}_{N_n} \to \{0, 1\} \), the asymptotic Type I+II error of \( \phi_n \) on the problem \( \text{PISD}(N_n, K_n, q_n) \) is at least 1 assuming the PC conjecture holds with density \( p = 1/2 \).

Therefore the computational boundary for \( \text{PISD}(n, k, q) \) in the parameter regime \( q = \tilde{\Theta}(n^{-\alpha}) \) and \( k = \tilde{\Theta}(n^\beta) \) is \( \beta^* = \frac{1}{2} + \frac{\alpha}{4} \).

**Proof.** If \( \beta < \alpha \) then PIS is information-theoretically impossible. Thus we may assume that \( \beta \geq \alpha \). Let \( \gamma = \frac{2\beta - \alpha}{2 - \alpha} \) and note that \( \gamma \in (0, 1/2) \). Now set

\[
\ell_n = \left[ \frac{\alpha \log_2 n}{2 - \alpha} \right], \quad k_n = \lceil n^\gamma \rceil, \quad N_n = 2^{\ell_n} n, \quad K_n = 2^{\ell_n} k_n, \quad q_n = 1 - (1 - w(n)^{-1})^{1/4n}
\]

where \( w(n) \) is any sub-polynomial increasing function tending to infinity. By Lemma 4.2.1, there is a randomized polynomial time algorithm mapping \( \text{PCD}(n, k_n, 1/2) \) to \( \text{PCD}(N_n, K_n, 1 - q_n) \) with total variation converging to zero as \( n \to \infty \). Now note that flipping every edge to a non-edge and non-edge to an edge maps \( \text{PCD}(N_n, K_n, 1 - q_n) \) to \( \text{PISD}(N_n, K_n, q_n) \). This map with Lemma 1 now implies that property 2 above holds. We now verify property 1. Note that

\[
\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \left[ \frac{\alpha \log_2 n}{2 - \alpha} \right] \cdot \log 2 + \left( \frac{2\beta - \alpha}{2 - \alpha} \right) \log n = \frac{\alpha}{2 - \alpha} + \frac{2\beta - \alpha}{2 - \alpha} + 1 = \beta
\]
4.2. PLANTED CLIQUE LIFTING

Note that as \( n \to \infty \), it follows that since \( 4^{-\ell_n} \log(1 - w(n)^{-1}) \to 0 \),

\[
q_n = 1 - (1 - w(n)^{-1})^{1/4} = 1 - e^{4^{-\ell_n} \log(1-w(n)^{-1})} \sim 4^{-\ell_n} \log(1 - w(n)^{-1})
\]

Now it follows that

\[
\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2 \left[ \frac{\alpha \log_2 n}{2-\alpha} \right] \log 2 - \log(1 - w(n)^{-1})}{\left[ \frac{\alpha \log_2 n}{2-\alpha} \right] \cdot \log 2 + \log n} = \frac{2\alpha}{2-\alpha} + 1 = \alpha
\]

which completes the proof. \( \square \)
Chapter 5

Rejection Kernels and Distributional Lifting

In this chapter, we generalize the idea in PC-LIFTING to apply to any distribution with a natural cloning operation, analogous to Step 3a in PC-LIFTING. Before describing this general distributional lifting procedure, we first will establish several results on applying rejection kernels, a general method for changes of measure such as from Bernoulli edge indicators to Gaussians, that we will need throughout our reductions.

5.1 Rejection Kernels

All of our remaining reductions will involve approximately mapping from a pair of Bernoulli random variables, typically edge indicators in random graphs, to a given pair of random variables. Similar entry-wise transformations of measure were used in [MW15b] and [GMZ17] for mapping from Bernoulli random variables to Gaussian random variables. We generalize these maps to arbitrary distributions and give a simple algorithm using rejection sampling to implement them. The general objective is to construct a single randomized function $RK : \{0, 1\} \to \mathbb{R}$ that simultaneously maps $\text{Bern}(p)$ to the distribution $f_X$ and $\text{Bern}(q)$ to $g_X$, approximately in total variation distance. For maps from instances $G$ of planted clique, such a map with $p = 1$ and $q = 1/2$ approximately sends the edge indicators $1_{\{i,j\} \in E(G)}$ to $f_X$ if $i$ and $j$ are in the planted clique and to $g_X$ otherwise.

We first describe the general structure of the maps $RK$ and their precise total variation guarantees in the following lemma. Then we give particular rejection kernels that we will use in our reductions.
Algorithm \texttt{RK}(x)

\textbf{Parameters:} Input \( x \in \{0, 1\} \), a pair of PMFs or PDFs \( f_X \) and \( g_X \) that can be efficiently computed and sampled, Bernoulli probabilities \( p, q \in [0, 1] \), number of iterations \( N \)

1. Initialize \( Y \leftarrow 0 \)

2. For \( N \) iterations do:
   a. If \( x = 0 \), sample \( Z \sim g_X \) and if
      \[
      p \cdot g_X(Z) \geq q \cdot f_X(Z)
      \]
      then with probability \( 1 - \frac{q \cdot f_X(Z)}{p \cdot g_X(Z)} \), update \( Y \leftarrow Z \) and break
   b. If \( x = 1 \), sample \( Z \sim f_X \) and if
      \[
      (1 - q) \cdot f_X(Z) \geq (1 - p) \cdot g_X(Z)
      \]
      then with probability \( 1 - \frac{(1 - p) \cdot g_X(Z)}{(1 - q) \cdot f_X(Z)} \), update \( Y \leftarrow Z \) and break

3. Output \( Y \)

\textbf{Figure 5.1:} Rejection kernel in Lemma 5.1.1

\textbf{Lemma 5.1.1.} Let \( f_X \) and \( g_X \) be probability mass or density functions supported on subsets of \( \mathbb{R} \) such that \( g_X \) dominates \( f_X \). Let \( p, q \in [0, 1] \) be such that \( p > q \) and let

\[
S = \left\{ x \in \mathbb{R} : \frac{1 - p}{1 - q} \leq \frac{f_X(x)}{g_X(x)} \leq \frac{p}{q} \right\}
\]

Suppose that \( f_X(x) \) and \( g_X(x) \) can be computed in \( O(T_1) \) time and samples from \( f_X \) and \( g_X \) can be generated in randomized \( O(T_2) \) time. Then there is a randomized \( O(N(T_1 + T_2)) \) time computable map \( \texttt{RK} : \{0, 1\} \to \mathbb{R} \) such that \( d_{TV}(\texttt{RK}(\text{Bern}(p)), f_X) \leq \Delta \) and \( d_{TV}(\texttt{RK}(\text{Bern}(q)), g_X) \leq \Delta \) where

\[
\Delta = \max \left\{ \frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p - q} + \left( \frac{\mathbb{P}_{X \sim g_X}[X \notin S]}{q/p} + \frac{p}{q} \right)^N, \right. \\
\left. \frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q} + \left( \frac{\mathbb{P}_{X \sim f_X}[X \notin S]}{1 - q} + \frac{1 - p}{1 - q} \right)^N \right\}
\]
5.1. REJECTION KERNELS

Proof. Let RK be implemented as shown in Figure 5.1.1 and note that RK runs in randomized $O(N(T_1 + T_2))$ time. Define $S_0$ and $S_1$ by

$$S_0 = \left\{ x \in \mathbb{R} : \frac{f_X(x)}{g_X(x)} \leq \frac{p}{q} \right\} \quad \text{and} \quad S_1 = \left\{ x \in \mathbb{R} : \frac{1-p}{1-q} \leq \frac{f_X(x)}{g_X(x)} \right\}$$

Now define the distributions by the densities or mass functions

$$\varphi_0(x) = \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p \cdot \mathbb{P}_{X \sim g_X}[X \in S_0] - q \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]} \quad \text{for } x \in S_0$$

$$\varphi_1(x) = \frac{(1-q) \cdot f_X(x) - (1-p) \cdot g_X(x)}{(1-q) \cdot \mathbb{P}_{X \sim f_X}[X \in S_1] - (1-p) \cdot \mathbb{P}_{X \sim g_X}[X \in S_1]} \quad \text{for } x \in S_1$$

Note that these are both well-defined PDFs or PMFs since they are nonnegative by the definitions of $S_0$ and $S_1$ and normalized. First consider the case when $x = 0$. Let $A_i$ be the event that the update $Y \leftarrow Z$ occurs in the $i$th iteration of Step 2a. The probability of $A_i$ is

$$\mathbb{P}[A_i] = \int_{S_0} g_X(x) \left(1 - \frac{q \cdot f_X(x)}{p \cdot g_X(x)}\right) dx = \mathbb{P}_{X \sim g_X}[X \in S_0] - \frac{q}{p} \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]$$

The density of $Y$ given the event $A_i$ is therefore given by

$$f_{Y|A_i}(x) = \mathbb{P}[A_i]^{-1} \cdot f_X|_{X \in S_0}(x) \cdot \left(1 - \frac{q \cdot f_X(x)}{p \cdot g_X(x)}\right) = \varphi_0(x)$$

If $A = A_1 \cup A_2 \cup \cdots \cup A_N$ is the event that the update $Y \leftarrow Z$ occurs in an iteration of Step 2a, then it follows by independence that

$$\mathbb{P}[A^c] = \prod_{i=1}^{N} (1 - \mathbb{P}[A_i]) = \left(1 - \mathbb{P}_{X \sim g_X}[X \in S_0] + \frac{q}{p} \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]\right)^N \leq \left(\mathbb{P}_{X \sim g_X}[X \not\in S] + \frac{q}{p}\right)^N$$

since $S \subseteq S_0$. Note that $f_{Y|A}(x) = \varphi_0(x)$ and RK(0) is $Y$ if $x = 0$. Therefore it follows by Lemma 3.0.4 that

$$d_{TV}(RK(0), \varphi_0) = \mathbb{P}[A^c] \leq \left(\mathbb{P}_{X \sim g_X}[X \not\in S] + \frac{q}{p}\right)^N$$

A symmetric argument shows that when $x = 1$,

$$d_{TV}(RK(1), \varphi_1) \leq \left(\mathbb{P}_{X \sim f_X}[X \not\in S] + \frac{1-p}{1-q}\right)^N$$
Now note that
\[
\left\| \varphi_0 - \frac{p \cdot g_X - q \cdot f_X}{p - q} \right\|_1 = \int_{S_0} \left| \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p - q} - \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p - q} \right| dx \\
+ \int_{S_0^c} \frac{q \cdot f_X(x) - p \cdot g_X(x)}{p - q} dx \\
= \left| 1 - \frac{p \cdot \mathbb{P}_{X \sim g_X} [X \in S_0] - q \cdot \mathbb{P}_{X \sim f_X} [X \in S_0]}{p - q} \right| \mathbb{P}_{X \sim f_X} [X \notin S_0] \\
+ \frac{q \cdot \mathbb{P}_{X \sim f_X} [X \notin S_0] - p \cdot \mathbb{P}_{X \sim g_X} [X \notin S_0]}{p - q} \\
= \frac{2(q \cdot \mathbb{P}_{X \sim f_X} [X \notin S_0] - p \cdot \mathbb{P}_{X \sim g_X} [X \notin S_0])}{p - q} \\
\leq \frac{2 \cdot \mathbb{P}_{X \sim f_X} [X \notin S]}{p - q}
\]
since \(S_0 \subseteq S\). A similar computation shows that
\[
\left\| \varphi_1 - \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} \right\|_1 = \frac{2((1 - p) \cdot \mathbb{P}_{X \sim g_X} [X \notin S_1] - (1 - q) \cdot \mathbb{P}_{X \sim f_X} [X \notin S_1])}{p - q} \\
\leq \frac{2 \cdot \mathbb{P}_{X \sim g_X} [X \notin S]}{p - q}
\]

Now note that
\[
f_X = p \cdot \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} + (1 - p) \cdot \frac{p \cdot g_X - q \cdot f_X}{p - q}
\]

Therefore by the triangle inequality, we have that
\[
d_{TV}(\text{RK}(\text{Bern}(p)), f_X) \leq d_{TV}(\text{RK}(\text{Bern}(p)), p \cdot \varphi_1 + (1 - p) \cdot \varphi_0) + d_{TV}(p \cdot \varphi_1 + (1 - p) \cdot \varphi_0, f_X) \\
\leq p \cdot d_{TV}(\text{RK}(1), \varphi_1) + p \cdot \left\| \varphi_1 - \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} \right\|_1 \\
+ (1 - p) \cdot d_{TV}(\text{RK}(0), \varphi_0) + (1 - p) \cdot \left\| \varphi_0 - \frac{p \cdot g_X - q \cdot f_X}{p - q} \right\|_1 \\
\leq p \cdot \left( \frac{2 \cdot \mathbb{P}_{X \sim f_X} [X \notin S]}{p - q} + \left( \mathbb{P}_{X \sim g_X} [X \notin S] + \frac{q}{p} \right)^N \right) \\
+ (1 - p) \cdot \left( \frac{2 \cdot \mathbb{P}_{X \sim g_X} [X \notin S]}{p - q} + \left( \mathbb{P}_{X \sim f_X} [X \notin S] + \frac{1 - p}{1 - q} \right)^N \right) \\
\leq \Delta
\]

Similarly, note that
\[
g_X = q \cdot \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} + (1 - q) \cdot \frac{p \cdot g_X - q \cdot f_X}{p - q}
\]
5.1. REJECTION KERNELS

The same triangle inequality applications as above show that

\[
d_{TV}(RK(Bern(q)), gx) \leq q \cdot \left( \frac{2 \cdot \mathbb{P}_{X \sim g_x \uparrow} \{X \not\in S\}}{p - q} + \left( \mathbb{P}_{X \sim g_x \uparrow} \{X \not\in S\} + \frac{q}{p} \right)^N \right) + (1 - q) \cdot \left( \frac{2 \cdot \mathbb{P}_{X \sim g_x \uparrow} \{X \not\in S\}}{p - q} + \left( \mathbb{P}_{X \sim f_x \uparrow} \{X \not\in S\} + \frac{1 - p}{1 - q} \right)^N \right)
\]

\leq \Delta

completing the proof of the lemma. \(\square\)

We will denote RK as defined with the parameters in the lemma above as \(RK(p \rightarrow f_x, q \rightarrow g_x, N)\) from this point forward. We now give the particular rejection kernels we will need in our reductions and their total variation guarantees. The first rejection kernel maps from the edge indicators in planted clique to Poisson random variables and will be essential in Poisson lifting.

**Lemma 5.1.2.** Let \(n\) be a parameter and let \(\epsilon > 0, c > 1\) and \(p \in (0, 1)\) be fixed constants satisfying that \(3\epsilon^{-1} \leq \log_c p^{-1}\). If \(\lambda = \lambda(n)\) satisfies that \(0 < \lambda \leq n^{-\epsilon}\), then the map

\[RK_{P1} = RK(1 \rightarrow \text{Pois}(c\lambda), p \rightarrow \text{Pois}(\lambda), N)\]

where \(N = \lceil 6 \log_{p^{-1}} n \rceil\) can be computed in \(O(\log n)\) time and satisfies that

\[d_{TV}(RK_{P1}(1), \text{Pois}(c\lambda)) = O(n^{-3}) \quad \text{and} \quad d_{TV}(RK_{P1}(Bern(p)), \text{Pois}(\lambda)) = O(n^{-3})\]

**Proof.** Let \(f_x(m)\) and \(g_x(m)\) be the PMFs of \(\text{Pois}(c\lambda)\) and \(\text{Pois}(\lambda)\), respectively. Note that

\[f_x(m) = \frac{e^{-c\lambda}(c\lambda)^m}{m!} \quad \text{and} \quad g_x(m) = \frac{e^{-\lambda}(\lambda)^m}{m!}\]

can be computed and sampled in \(O(1)\) operations. Therefore Lemma 5.1.1 implies that \(RK_{P1}\) can be computed in \(O(N) = O(\log n)\) time. Let the set \(S\) be as defined in Lemma 5.1.1, let \(M = \log_c(2p)^{-1} \geq 3\epsilon^{-1}\) and define the set \(S' = \{m \in \mathbb{Z}_{\geq 0} : m \leq M\}\). Now note that if \(m \in S'\) then

\[
\frac{f_x(m)}{g_x(m)} = \frac{e^{-c\lambda}(c\lambda)^m}{m!} \cdot \frac{e^{-\lambda}(\lambda)^m}{m!} = e^{-(c-1)\lambda} \leq e^M \leq p^{-1}
\]

and therefore it follows that \(S' \subseteq S\). For sufficiently large \(n\), we have that \(M = \log_c p^{-1} \geq cn^{-\epsilon} \geq c\lambda > \lambda\) and therefore a standard Poisson tail bound yields that

\[\mathbb{P}_{X \sim g_x} \{X \not\in S\} \leq \mathbb{P}_{X \sim g_x} \{X \geq M\} \leq e^{-\lambda} \left( \frac{e\lambda}{M} \right)^M \leq \left( \frac{e}{M} \right)^M \lambda \leq \left( \frac{e}{M} \right)^M n^{-3}\]
since $\lambda \leq n^{-\epsilon}$. Similarly, we have that $\mathbb{P}_{X \sim f_X}[X \not\in S] \leq \left(\frac{c}{M}\right)^M n^{-3}$. Now note that for sufficiently large $n$, we have that

$$\frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \not\in S]}{1 - p} + (\mathbb{P}_{X \sim g_X}[X \not\in S] + p)^N \leq 2 \left(\frac{c}{M}\right)^M n^{-3} + \left(\left(\frac{c}{M}\right)^M n^{-3} + p\right)^N$$

$$\leq \left(\frac{2(1 - p)^{-1} \left(\frac{c}{M}\right)^M + 1}{1 - p}\right)n^{-3}$$

By similar reasoning, we have that for sufficiently large $n$

$$\frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \not\in S]}{1 - p} + (\mathbb{P}_{X \sim f_X}[X \not\in S])^N \leq \left(2(1 - p)^{-1} \left(\frac{c}{M}\right)^M + 1\right)n^{-3}$$

Therefore $\Delta \leq \left(2(1 - p)^{-1} \left(\frac{c}{M}\right)^M + 1\right)n^{-3}$ for sufficiently large $n$ and applying Lemma 5.1.1 proves the lemma.

The next lemma gives another approximate map to Poisson random variables from Bernoulli random variables corresponding to the edge indicators in the edge-dense regime of planted dense subgraph. We use the following lemma to apply Poisson lifting after Gaussian lifting in order to deduce hardness in the general regime of PDS. The proof is very similar to that of Lemma 5.1.2.

**Lemma 5.1.3.** Let $\epsilon \in (0, 1)$ be a fixed constant and let $n$ be a parameter. Suppose that:

- $\lambda = \lambda(n)$ satisfies that $0 < \lambda \leq n^{-\epsilon}$;
- $c = c(n) > 1$ satisfies that $c = O(n)$; and
- $\rho = \rho(n) \in (0, 1/2)$ satisfies that $\rho \geq n^{-K}$ for sufficiently large $n$ where $K = \Theta(n)$ is positive and

$$K + 3 \epsilon^{-1} \leq \log_e(1 + 2\rho) = O(n)$$

Then the map

$$\text{RK}_P^2 = \text{RK} \left(\frac{1}{2} + \rho \rightarrow \text{Pois}(c\lambda), \frac{1}{2} \rightarrow \text{Pois}(\lambda), N\right)$$

where $N = [6\rho^{-1} \log n]$ can be computed in poly($n$) time and satisfies

$$d_{TV}(\text{RK}_P^2(\text{Bern}(1/2 + \rho)), \text{Pois}(c\lambda)) = O(n^{-3})$$

and

$$d_{TV}(\text{RK}_P^2(\text{Bern}(1/2)), \text{Pois}(\lambda)) = O(n^{-3})$$
5.1. REJECTION KERNELS

Proof. As in Lemma 5.1.2, let \( f_X(m) \) and \( g_X(m) \) be the PMFs of \( \text{Pois}(c \lambda) \) and \( \text{Pois}(\lambda) \) and note that they can be computed and sampled in \( O(1) \) operations. Lemma 5.1.1 implies that \( \text{RK}_{P2} \) can be computed in \( O(N) = O(n^K \log n) = \text{poly}(n) \) time. Let the set \( S \) be as defined in Lemma 5.1.1, let \( M = \log_e(1 + 2 \rho) \geq (K + 3)\epsilon^{-1} \) and define the set \( S' = \{ m \in \mathbb{Z}_{\geq 0} : m \leq M \} \). Now note that if \( n > 1 \), then \( \lambda \leq n^{-\epsilon} < 1 \) and it follows that

\[
e^{-(c-1)\lambda} \geq 1 - (c - 1)\lambda > 2 - c \geq 2 - (1 + 2\rho)\epsilon/(K+3) > 1 - 2\rho
\]
since \( \epsilon \in (0, 1) \). Therefore if \( m \in S' \), then

\[
1 - 2\rho < e^{-(c-1)\lambda} \leq \frac{f_X(m)}{g_X(m)} = e^{-(c-1)\lambda} e^m \leq e^M \leq 1 + 2\rho
\]
and it follows that \( S' \subseteq S \). By the same Poisson tail bounds as in Lemma 5.1.2, we have that for sufficiently large \( n \)

\[
\mathbb{P}_{X \sim g_X} [X \notin S] \leq \left( \frac{e}{M} \right)^M n^{-K-3} \quad \text{and} \quad \mathbb{P}_{X \sim f_X} [X \notin S] \leq \left( \frac{ce}{M} \right)^M n^{-K-3}
\]
Now note that for sufficiently large \( n \), we have that \( \rho^{-1} \leq n^K \) and \( \left( \frac{e}{M} \right)^M n^{-K-3} \leq \frac{1}{2} n^{-K} \leq \frac{1}{2} \rho \) since \( M = O(n(1)) \). Therefore

\[
2\rho^{-1} \cdot \mathbb{P}_{X \sim f_X} [X \notin S] + \left( \mathbb{P}_{X \sim g_X} [X \notin S] + \frac{1}{1 + 2\rho} \right)^N \leq 2\rho^{-1} \left( \frac{ce}{M} \right)^M n^{-K-3} + \left( \frac{ce}{M} \right)^M n^{-K-3} + 1 - \rho
\]
\[
\leq 2 \left( \frac{ce}{M} \right)^M n^{-3} + \left( 1 - \frac{\rho}{2} \right)^N
\]
\[
\leq 2 \left( \frac{ce}{M} \right)^M n^{-3} + e^{-\rho N/2}
\]
\[
\leq \left( 2 \left( \frac{ce}{M} \right)^M + 1 \right) n^{-3}
\]
By similar reasoning, we have that for sufficiently large \( n \),

\[
2\rho^{-1} \cdot \mathbb{P}_{X \sim g_X} [X \notin S] + \left( \mathbb{P}_{X \sim f_X} [X \notin S] + 1 - 2\rho \right)^N \leq \left( 2 \left( \frac{e}{M} \right)^M + 1 \right) n^{-3}
\]
Therefore \( \Delta \leq \left( 2 \left( \frac{ce}{M} \right)^M + 1 \right) n^{-3} \) for sufficiently large \( n \) and applying Lemma 5.1.1 proves the lemma.

The next lemma of this section approximately maps from Bernoulli to discretized Gaussian random variables. Gaussian random variables appear in two different contexts in our reductions:
the problems ROS, SPCA and BC have observations sampled from multivariate Gaussians; and (2) random matrices with Gaussian entries are used as intermediate in our reductions to PDS in the general regime and SSBM. In both cases, we will need the map in the following lemma. As in the proofs of the previous two lemmas, this next lemma also verifies the conditions of Lemma 5.1.1 for Gaussians and derives an upper bound on $\Delta$.

**Lemma 5.1.4.** Let $n$ be a parameter and suppose that $p = p(n)$ and $q = q(n)$ satisfy that $p > q$, $p, q \in [0, 1]$, $\max(q, 1-q) = \Omega_n(1)$ and $p-q \geq n^{-O_n(1)}$. Let $\delta = \min \left\{ \log \left( \frac{p}{q} \right), \log \left( \frac{q}{1-p} \right) \right\}$. Suppose that $\mu = \mu(n) \in (0, 1)$ is such that

$$\mu \leq \frac{\delta}{2\sqrt{6\log n + 2\log(p-q)-1}}$$

Then the map

$$\text{RK}_G = \text{RK} (p \rightarrow N(\mu, 1), q \rightarrow N(0, 1), N)$$

where $N = \lfloor 6^{\delta^{-1}} \log n \rfloor$ can be computed in $\text{poly}(n)$ time and satisfies

$$d_{TV}(\text{RK}_G(\text{Bern}(p)), N(\mu, 1)) = O_n(n^{-3}) \quad \text{and} \quad d_{TV}(\text{RK}_G(\text{Bern}(q)), N(0, 1)) = O_n(n^{-3})$$

**Proof.** Let $f(x)$ and $g(x)$ be the PDFs of $N(\mu, 1)$ and $N(0, 1)$, respectively, where

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \quad \text{and} \quad g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

which can be computed and sampled in $O(1)$ operations in the given computational model. Now note that since $\log(1+x) \geq x/2$ for $x \in (0, 1)$, we have that

$$\log \left( \frac{p}{q} \right) \geq \frac{p-q}{2q} \geq \frac{1}{2} (p-q) \geq \frac{1}{2} n^{-O_n(1)}$$

and similarly that $\log \left( \frac{1-q}{1-p} \right) \geq \frac{p-q}{2(1-p)} \geq \frac{1}{2} (p-q) \geq \frac{1}{2} n^{-O_n(1)}$. Therefore $N = \text{poly}(n)$ and Lemma 5.1.1 implies that $\text{RK}_G$ can be computed in $\text{poly}(n)$ time. Let the set $S$ be as defined in Lemma 5.1.1, let $M = \sqrt{6\log n + 2\log(p-q)-1}$ and define the set $S' = \{ x \in \mathbb{R} : |x| \leq M \}$. Note that if $x \in S'$ then we have that since $M\mu = \delta/2$,

$$1-p \leq \exp(-2M\mu) \leq \exp \left( -M\mu - \frac{\mu^2}{2} \right) \leq \frac{f(x)}{g(x)} = \exp \left( x\mu - \frac{\mu^2}{2} \right) \leq \exp(M\mu) \leq \frac{p}{q}$$

for sufficiently large $n$ since $M \rightarrow \infty$ as $n \rightarrow \infty$ and $\mu \in (0, 1)$. This implies that $S' \subseteq S$. Using the bound $1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} \cdot t^{-1} e^{-t^2/2}$ for $t \geq 1$, we have that

$$P_{X \sim g_X}[X \not\in S] \leq P_{X \sim g_X}[X \not\in S'] = 2(1 - \Phi(M)) \leq \frac{2}{\sqrt{2\pi}} \cdot M^{-1} e^{-M^2/2}$$
Similarly, it follows that for sufficiently large $n$ we have that $M/(M - \mu) \leq 2$ and thus

$$\mathbb{P}_{X \sim f_X} [X \not\in S] \leq \mathbb{P}_{X \sim f_X} [X \not\in S'] = (1 - \Phi(M - \mu)) + (1 - \Phi(M + \mu))$$

$$\leq \frac{1}{\sqrt{2\pi}} \cdot (M - \mu)^{-1} e^{-(M - \mu)^2/2} + \frac{1}{\sqrt{2\pi}} \cdot (M + \mu)^{-1} e^{-(M + \mu)^2/2}$$

$$\leq \frac{1}{\sqrt{2\pi}} \cdot M^{-1} e^{-(\mu^2 + M^2)/2} \left[ \frac{M}{M - \mu} \cdot e^{M\mu} + 1 \right]$$

$$\leq \frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) \cdot M^{-1} e^{-M^2/2}$$

Now note that $M^{-1} e^{-M^2/2} \leq e^{-M^2/2} \leq (p - q)n^{-3}$ for sufficiently large $n$. If $n$ is large enough then

$$\frac{2}{\sqrt{2\pi}} \cdot n^{-3} \leq \frac{p^{1/2}}{p(p^{1/2} + q^{1/2})} = \Omega_n(1)$$

Rearranging yields that $\frac{2}{\sqrt{2\pi}} (p - q)n^{-3} \leq \sqrt{\frac{p - q}{p}}$. This implies that

$$\frac{2}{p - q} \cdot \mathbb{P}_{X \sim f_X} [X \not\in S] + \left( \mathbb{P}_{X \sim g_X} [X \not\in S] + \frac{p}{q} \right)^N \leq \frac{2}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} + \left( \frac{2}{\sqrt{2\pi}} (p - q)n^{-3} + \frac{q}{p} \right)^N$$

$$\leq \frac{2}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} + \left( \frac{q}{p} \right)^{N/2}$$

$$\leq \left( 1 + \frac{2}{\sqrt{2\pi}} + \frac{4}{\sqrt{2\pi}} \sqrt{\frac{p}{q}} \right) n^{-3} = O_n(n^{-3})$$

Now note that if $\frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} > \frac{(1 - p)^{1/3}}{(1 - q)((1 - p)^{1/3} + (1 - p)^{1/3})}$ then it follows that

$$\frac{1 - p}{1 - q} \leq \left( \frac{1 - q}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) \cdot n^{-3} - 1 \right)^{-1} \leq Cn^{-3}$$

for some constant $C > 0$ if $n$ is sufficiently large, since $1 - q = \Omega_n(1)$. Otherwise, the same manipulation as above implies that $\frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} \leq \sqrt{\frac{1 - p}{1 - q} - \frac{1 - p}{1 - q}}$. Therefore we have in either case that

$$\frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} + \frac{1 - p}{1 - q} \leq \max \left\{ \sqrt{\frac{1 - p}{1 - q}} \cdot \frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} + Cn^{-3} \right\}$$

For sufficiently large $n$, the second term in the maximum above is at most $n^{-2}$. Therefore

$$\frac{2 \cdot \mathbb{P}_{X \sim g_X} [X \not\in S]}{p - q} + \left( \mathbb{P}_{X \sim f_X} [X \not\in S] + \frac{1 - p}{1 - q} \right)^N \leq \left( \frac{4}{\sqrt{2\pi}} \right) n^{-3} + \left( \frac{1}{\sqrt{2\pi}} \left( 1 + 2 \sqrt{\frac{p}{q}} \right) n^{-3} + \frac{1 - p}{1 - q} \right)^N$$

$$\leq \left( \frac{4}{\sqrt{2\pi}} \right) n^{-3} + \max \left\{ \left( \frac{1 - p}{1 - q} \right)^{N/2} , n^{-2N} \right\}$$

$$= O_n(n^{-3})$$

Therefore $\Delta = O_n(n^{-3})$ for sufficiently large $n$ since $q = \Omega_n(1)$. Now applying Lemma 5.1.1 proves the lemma. \qed
5.2 Distributional Lifting

The general distributional lifting procedure begins with an instance \( G \in \mathcal{G}_n \) of a planted dense subgraph problem such as planted clique and applies a rejection kernel element-wise to its adjacency matrix. This yields a symmetric matrix \( M \) with zeros on its main diagonal, i.i.d. entries sampled from \( P_{\lambda_0} \) on entries corresponding to clique edges and i.i.d. entries sampled from \( Q_{\lambda_0} \) elsewhere. As an input to the procedure, we assume a random cloning map \( f_{\text{cl}} \) that exactly satisfies

\[
\begin{align*}
    f_{\text{cl}}(P_{\lambda}) &\sim P^{\otimes 4}_{g_{\text{cl}}(\lambda)} \quad \text{and} \quad f_{\text{cl}}(Q_{\lambda}) \sim Q^{\otimes 4}_{g_{\text{cl}}(\lambda)}.
\end{align*}
\]

for some parameter update function \( g_{\text{cl}} \). Applying this cloning map entry-wise to \( M \) and arranging the resulting entries correctly yields a matrix \( M' \) of size \( 2n \times 2n \) with a planted submatrix of size \( 2k \times 2k \). The only distributional issue that arises is that now entries from \( Q_{g_{\text{cl}}(\lambda)} \) are all i.i.d. entries sampled from \( P_{g_{\text{cl}}(\lambda)} \) instead of from \( P_{\lambda} \), although some should be from \( P_{g_{\text{cl}}(\lambda)} \). We handle these approximately in total variation by randomly permuting the rows and columns and applying Lemma 4.1.1. Iterating this procedure \( \ell \) times yields a matrix \( M'' \) of size \( 2^\ell n \times 2^\ell n \) with a planted submatrix of size \( 2^\ell k \times 2^\ell k \). If \( \lambda_{i+1} = g_{\text{cl}}(\lambda_i) \), then \( M'' \) has all i.i.d. entries from \( Q_{\lambda_{\ell}} \) under \( H_0 \) and a planted submatrix with i.i.d. entries from \( P_{\lambda_{\ell}} \) under \( H_1 \). We then truncate the entries of \( M'' \) to produce the adjacency matrix of a graph with i.i.d. edge indicators, conditioned on the vertices in the planted subgraph. This yields a general procedure to reduce from an instance of planted clique to subgraph problems with larger planted subgraphs.

A natural question is: what is the purpose of the distributions \( P_{\lambda} \) and \( Q_{\lambda} \)? If the initial and final distributions are both graph distributions with Bernoulli edge indicators, it a priori seems unnecessary to use matrix distributions without Bernoulli entries as intermediates. Our main reason for introducing these intermediate distributions is that they achieve the right parameter tradeoffs to match the best known algorithms for PDS where cloning procedures that stay within the set of graph distributions do not. Consider the target planted dense subgraph instance of PDS(\( n, k, p, q \)) where \( p = 2q \) and \( q = \Theta(n^{-\alpha}) \). To produce lower bounds tight with the computational barrier in Theorem 2.3.1, a lifting procedure mapping \( n \rightarrow 2n \) and \( k \rightarrow 2k \) at each step would need its cloning map to satisfy

\[
\begin{align*}
    f_{\text{cl}}(\text{Bern}(q)) &\sim Q = \text{Bern}(q/4)^{\otimes 4} \quad \text{and} \quad f_{\text{cl}}(\text{Bern}(p)) \sim P = \text{Bern}(p/4)^{\otimes 4},
\end{align*}
\]

where \( p = 2q \). It is not difficult to verify that for any random map \( f_{\text{cl}} : \{0, 1\} \rightarrow \{0, 1\}^4 \), it would need to hold that

\[
\frac{1 - p}{1 - q} \leq \frac{P(x)}{Q(x)} \leq \frac{p}{q} \quad \text{for all } x \in \{0, 1\}^4.
\]
5.2. DISTRIBUTIONAL LIFTING

However, \( P(1, 1, 1, 1)/Q(1, 1, 1, 1) = 16 > 2 = p/q \), so no such map can exist. Another approach is to relax \( f_d \) to be an approximate map like a rejection kernel. However, this seems to induce a large loss in total variation from the target distribution on \( M' \). Our solution is to use a rejection kernel to map to distributions with natural cloning maps \( f_d \), such as a Poisson or Gaussian distribution, to front-load the total variation loss to this approximate mapping step and induce no entry-wise total variation loss later in the cloning procedure. We choose the precise distributions \( P_\lambda \) and \( Q_\lambda \) to match the parameter tradeoff along the computational barrier. Note that this general distributional lifting procedure can also be used to map to problems other than variants of subgraph detection, such as biclustering, by not truncating in Step 4.

We remark that the PC-LIFTING reduction presented in the previous chapter is almost an instance of distributional lifting with \( P_\lambda = \text{Bern}(1) \) for all \( \lambda \) and \( Q_\lambda = \text{Bern}(\lambda) \), with the parameter update \( g_{\text{el}}(\lambda) = \lambda^{1/4} \). However in PC-LIFTING, the planted anti-diagonal entries between the vertices \( i \) and \( 2m + 1 - i \) in Step 3b are from the planted distribution \( P_\lambda \), rather than \( Q_\lambda \) as in distributional lifting. This requires a slightly different analysis of the planted anti-diagonal entries with Lemma 4.1.1.

We now proceed to describe distributional lifting and prove its guarantees. Given two distributions \( P \) and \( Q \), let \( M_n(Q) \) denote the distribution on \( n \times n \) symmetric matrices with zero diagonal entries and every entry below the diagonal sampled independently from \( Q \). Similarly, let \( M_n(S, P, Q) \) denote the distribution on random \( n \times n \) symmetric matrices formed by:

1. sampling the entries of the principal submatrix with indices in \( S \) below its main diagonal independently from \( P \);

2. sampling all other entries below the main diagonal independently from \( Q \); and

3. placing zeros on the diagonal.

Let \( M_n(k, P, Q) \) denote the distribution of matrices \( M_n(S, P, Q) \) where \( S \) is a size \( k \) subset of \( [n] \) selected uniformly at random. Given a matrix \( M \in \mathbb{R}^{n \times n} \) and index sets \( S, T \subseteq [n] \), let \( M[S \times T] \) denote the \( |S| \times |T| \) submatrix of \( M \) with row indices in \( S \) and column indices in \( T \). Also let \( G(n, p, q, S) \) where \( S \) is a \( k \)-subset of \( [n] \) denote an instance of \( G(n, k, p, q) \) where the planted dense subgraph is conditioned to be on \( S \). The guarantees of distributional lifting are as follows.

**Theorem 5.2.1 (Distributional Lifting).** Suppose that \( n \) and \( \ell \) are such that \( \ell = O(\log n) \) and are sufficiently large. Let \( p', q' \in [0, 1] \) and define the parameters:

- target planted and noise distribution families \( P_\lambda \) and \( Q_\lambda \) parameterized by \( \lambda \);
Algorithm DISTRIBUTIONAL-LIFTING

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations $\ell$, parameterized families of target planted and noise distributions $P_\lambda$ and $Q_\lambda$, a TV-approximation $Q'_\lambda$ to $Q_\lambda$ that can be efficiently sampled, rejection kernel $\text{RK}$ approximately mapping $\text{Bern}(p') \rightarrow P_{\lambda_0}$ and $\text{Bern}(q') \rightarrow Q_{\lambda_0}$, threshold $t$, cloning map $f_{\text{cl}}$ and corresponding parameter map $g_{\text{cl}}$

1. Form the symmetric matrix $M \in \mathbb{R}^{n \times n}$ with $M_{ii} = 0$ and off-diagonal terms 

$$M_{ij} = \text{RK}(\mathbf{1}_{\{i,j\} \in E(G)})$$

2. Initialize $W \leftarrow M$ and $m \leftarrow n$

3. For $i = 0, 1, \ldots, \ell - 1$ do:
   a. For each pair of distinct $i, j \in [m]$, let $(x^1_{ij}, x^2_{ij}, x^3_{ij}, x^4_{ij}) = f_{\text{cl}}(W_{ij})$
   b. Let $W' \in \mathbb{R}^{2m \times 2m}$ be the symmetric matrix with $W'_{ii} = 0$ and

$$W'_{ij} = x^1_{ij}$$
$$W'_{(2m+1-i)j} = x^2_{ij}$$
$$W'_{i(2m+1-j)} = x^3_{ij}$$
$$W'_{(2m+1-i)(2m+1-j)} = x^4_{ij}$$

for all distinct $i, j \in [m]$ and

$$W'_{i,2m+1-i} \sim \text{i.i.d. } Q'_{\lambda}$$

for all $i \in [m]$

c. Generate a permutation $\sigma$ on $[2m]$ uniformly at random

d. Update $W \leftarrow (W')^{\sigma \sigma}$, $m \leftarrow 2m$ and $\lambda \leftarrow g_{\text{cl}}(\lambda)$

4. Output the graph $H$ with $\{i, j\} \in E(H)$ if $W_{ij} > t$

Figure 5.2: Distributional lifting procedure in Theorem 5.2.1.
5.2. DISTRIBUTIONAL LIFTING

- a rejection kernel $\text{rk}$ that can be computed in randomized $\text{poly}(n)$ time and parameter $\lambda_0$ such that $\text{rk} (\text{Bern}(p')) \sim \tilde{P}_{\lambda_0}$ and $\text{rk} (\text{Bern}(q')) \sim \tilde{Q}_{\lambda_0}$;

- a cloning map $\text{fcl}$ that can be computed in randomized $\text{poly}(n)$ time and parameter map $\text{gcl}$ such that

$$
\text{fcl}(P_\lambda) \sim P_{g_{\text{cl}(\lambda)}}^{\otimes 4} \quad \text{and} \quad \text{fcl}(Q_\lambda) \sim Q_{g_{\text{cl}(\lambda)}}^{\otimes 4}
$$

for each parameter $\lambda$;

- a randomized $\text{poly}(n)$ time algorithm for sampling from $Q_{\lambda_i}'$ for each $1 \leq i \leq \ell$ where the sequence of parameters $\lambda_i$ are such that $\lambda_{i+1} = g_{\text{cl}}(\lambda_i)$ for each $i$; and

- a threshold $t \in \mathbb{R}$.

Then $\phi =$ DISTRIBUTIONAL-LIFTING with these parameters is a randomized polynomial time computable map $\phi : \mathcal{G}_n \rightarrow \mathcal{G}_{2n}$ such that under both $H_0$ and $H_1$, it holds that

$$
d_{TV} \left( \phi (\text{PDS}(n,k,p',q')), \text{PDS} \left( 2^\ell n, 2^\ell k, p, q \right) \right) \leq \left( \frac{n}{2} \right) \cdot \max \left\{ d_{TV} \left( \tilde{P}_{\lambda_0}, P_{\lambda_0} \right), d_{TV} \left( \tilde{Q}_{\lambda_0}, Q_{\lambda_0} \right) \right\} + \sum_{i=1}^{\ell} \left( 2^{i} \cdot d_{TV} (Q_{\lambda_i}, Q_{\lambda_i}') + \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}} \right)
$$

where $p = \mathbb{P}_{X \sim P_{\lambda_1}} [X > t]$ and $q = \mathbb{P}_{X \sim Q_{\lambda_1}} [X > t]$.

Proof. If $\ell = O(\log n)$, the algorithm DISTRIBUTIONAL-LIFTING runs in randomized polynomial time. Let $\phi_i(W)$ be the algorithm that outputs the value of $W$ after $i$ iterations of Step 4 given the original value of $W$ and Let $\phi'_i(G)$ be the algorithm that outputs the value of $W$ after $i$ iterations of Step 4 given the original graph $G$. Note that $\phi'_0$ outputs the value of $M$ after Step 1.

We first consider an iteration of Step 3 applied to $M \sim M_m(S, P_\lambda, Q_\lambda)$ where $|S| = k$. By the definition of $f_{\text{cl}}$, if $i, j$ are distinct and both in $S$, then $(x_{ij}^1, x_{ij}^2, x_{ij}^3, x_{ij}^4) \sim P_{g_{\text{cl}(\lambda)}}^{\otimes 4}$. Similarly, if at least one of $i$ or $j$ is not in $S$, then $(x_{ij}^1, x_{ij}^2, x_{ij}^3, x_{ij}^4) \sim Q_{g_{\text{cl}(\lambda)}}^{\otimes 4}$. Therefore the symmetric matrix $W'$ constructed in Step 4b has independent entries below its main diagonal and satisfies that:

- $W'_{ij} \sim P_{g_{\text{cl}(\lambda)}}$ for all distinct $i, j \in S' = S \cup \{ 2m + 1 - i : i \in S \}$ with $i + j \neq 2m + 1$;

- $W'_{ij} \sim Q_{g_{\text{cl}(\lambda)}}$ for all distinct $(i, j) \notin S' \times S'$;

- $W'_{ij} \sim Q_{g_{\text{cl}(\lambda)}}'$ with $i + j = 2m + 1$; and

- $W'_{ii} = 0$. 
Let $W'_r$ be the matrix with each of its entries identically distributed to those of $W'$ except $(W'_r)_{ij} \sim Q_{g_{cl}(\lambda)}$ if $i + j = 2m + 1$. Coupling entries individually yields that

$$d_{TV}(\mathcal{L}(W'), \mathcal{L}(W'_r)) \leq m \cdot d_{TV}(Q_{g_{cl}(\lambda)}, Q'_{g_{cl}(\lambda)})$$

Now consider the matrix $W_r = (W'_r)^{\sigma \sigma}$ conditioned on the two sets $\sigma(S)$ and $\sigma(S' \setminus S)$ where $\sigma$ is a uniformly at random chosen permutation on $[2m]$. We will show that this matrix is close in total variation to $M_{2m}(\sigma(S'), P_{g_{cl}(\lambda)}, Q_{g_{cl}(\lambda)})$. Note that fully conditioned on $\sigma$, the entries of $W_r$ below the main diagonal are independent and identically distributed to $M_{2m}(\sigma(S'), P_{g_{cl}(\lambda)}, Q_{g_{cl}(\lambda)})$ other than the entries with indices $(\sigma(i), \sigma(2m + 1 - i))$ where $i \in S'$. These entries are distributed as $Q_{g_{cl}(\lambda)}$ in $W_r|\sigma$ and as $P_{g_{cl}(\lambda)}$ in the target distribution $M_{2m}(\sigma(S'), P_{g_{cl}(\lambda)}, Q_{g_{cl}(\lambda)})$. Marginalizing to only condition on the sets $\sigma(S)$ and $\sigma(S', S)$, yields that all entries $(W_r)_{ij}$ with $(i, j) \notin \sigma(S) \times \sigma(S', S) \cup \sigma(S', S) \times \sigma(S)$ are identically distributed in $W_r|\{\sigma(S), \sigma(S' \setminus S)\}$ and the target distribution. Coupling these corresponding entries yields that the total variation between $W_r|\{\sigma(S), \sigma(S' \setminus S)\}$ and the target distribution satisfies that

$$d_{TV}(\mathcal{L}(W_r|\sigma(S), \sigma(S' \setminus S)), M_{2m}(\sigma(S'), P_{g_{cl}(\lambda)}, Q_{g_{cl}(\lambda)})) = d_{TV}(\mathcal{L}(W_r|\sigma(S) \times \sigma(S' \setminus S)), M_k(P_{g_{cl}(\lambda)}))$$

Now observe that $W_r|\sigma(S) \times \sigma(S' \setminus S)$ is distributed as $\mathcal{L}(A^{\text{id}, \tau})$ where $\tau$ is permutation of $[k]$ selected uniformly at random and $A$ is a matrix with its diagonal entries i.i.d. $Q_{g_{cl}(\lambda)}$ and its other entries i.i.d. $P_{g_{cl}(\lambda)}$. By Lemma 4.1.1, we therefore have that

$$d_{TV}(\mathcal{L}(W_r|\sigma(S) \times \sigma(S' \setminus S)), M_k(P_{g_{cl}(\lambda)})) = d_{TV}(\mathcal{L}(A^{\text{id}, \tau}), M_k(P_{g_{cl}(\lambda)})) \leq \sqrt{\frac{1}{2} \cdot \chi^2(Q_{g_{cl}(\lambda)}, P_{g_{cl}(\lambda)})}$$

Now consider the matrix $\phi_1(M) = W = (W')^{\sigma \sigma}$. By the data processing inequality, we have that

$$d_{TV}(\mathcal{L}(W_r|\sigma(S), \sigma(S' \setminus S)), \mathcal{L}(W|\sigma(S), \sigma(S' \setminus S)) \leq d_{TV}(\mathcal{L}(W'), \mathcal{L}(W'_r)) \leq m \cdot d_{TV}(Q_{g_{cl}(\lambda)}, Q'_{g_{cl}(\lambda)})$$

The triangle inequality now implies that

$$d_{TV}(\mathcal{L}(W|\sigma(S), \sigma(S' \setminus S)), \ M_{2m}(\sigma(S'), P_{g_{cl}(\lambda)}, Q_{g_{cl}(\lambda)})) \leq m \cdot d_{TV}(Q_{g_{cl}(\lambda)}, Q'_{g_{cl}(\lambda)}) + \sqrt{\frac{1}{2} \cdot \chi^2(Q_{g_{cl}(\lambda)}, P_{g_{cl}(\lambda)})}$$
Letting $S$ be chosen uniformly at random over all subsets of $[n]$ of size $k$ and the triangle inequality now imply that

\[
d_{TV}(\phi_1(M_m(k, P_\lambda, Q_\lambda)), M_{2m}(2k, P_{gcd(\lambda)}, Q_{gcd(\lambda)}))
\leq \mathbb{E}_S \mathbb{E}_{\sigma(S), \sigma(S'\setminus S)} [d_{TV}(\mathcal{L}(W|\sigma(S), \sigma(S'\setminus S)), M_{2m}(\sigma(S'), P_{gcd(\lambda)}, Q_{gcd(\lambda)}))]
\leq m \cdot d_{TV}(Q_{gcd(\lambda)}, Q'_{gcd(\lambda)}) + \sqrt{\frac{1}{2} \cdot \chi^2(Q_{gcd(\lambda)}, P_{gcd(\lambda)})}
\]

For each $i \geq 0$, combining this inequality with the triangle inequality and data processing inequality yields that

\[
d_{TV}(\phi_{i+1}(M_m(k, P_\lambda, Q_\lambda)), M_{2i+1m}(2^{i+1}k, P_{\lambda_{i+1}}, Q_{\lambda_{i+1}}))
\leq d_{TV}(\phi_1 \circ \phi_i(M_m(k, P_\lambda, Q_\lambda)), \phi_1(M_{2i}m(2^i k, P_{\lambda_i}, Q_{\lambda_i})))
+ d_{TV}(\phi_1(M_{2i}m(2^i k, P_{\lambda_i}, Q_{\lambda_i})), M_{2i+1m}(2^{i+1}k, P_{\lambda_{i+1}}, Q_{\lambda_{i+1}}))
\leq d_{TV}(\phi_i(M_m(k, P_\lambda, Q_\lambda)), M_{2im}(2^i k, P_{\lambda_i}, Q_{\lambda_i}))
+ 2^m \cdot d_{TV}(Q_{\lambda_{i+1}}, Q'_{\lambda_{i+1}}) + \sqrt{\frac{1}{2} \cdot \chi^2(Q_{\lambda_{i+1}}, P_{\lambda_{i+1}})}
\]

Now note that the adjacency matrix $A_{ij}(G) = 1_{\{i,j\} \in E(G)}$ of $G \sim G(n, p', q', S)$ is distributed as $M_n(S, \text{Bern}(p'), \text{Bern}(q'))$. Note that $\phi'_0$ applies RK element-wise to the entries below the main diagonal of $A_{ij}(G)$. Coupling each of the independent entries below the diagonals of $\phi'_0(G(n, p', q', S))$ and $M_n(S, P_{\lambda_{0}}, Q_{\lambda_{0}})$ separately, we have that if $|S| = k$ then

\[
d_{TV}(\phi'_0(G(n, p', q', S)), M_n(S, P_{\lambda_{0}}, Q_{\lambda_{0}})) \leq \binom{k}{2} \cdot d_{TV}(\bar{P}_{\lambda_{0}}, P_{\lambda_{0}}) + \binom{n}{2} - \binom{k}{2} \cdot d_{TV}(\bar{Q}_{\lambda_{0}}, Q_{\lambda_{0}})
\]

Taking $S$ to be uniformly distributed over all $k$ element subsets of $[n]$ yields by triangle inequality,

\[
d_{TV}(\phi'_0(G(n, k, p', q'))), M_n(k, P_{\lambda_{0}}, Q_{\lambda_{0}}))
\leq \mathbb{E}_S [d_{TV}(\phi'_0(G(n, p', q', S)), M_n(S, P_{\lambda_{0}}, Q_{\lambda_{0}}))]
\leq \binom{n}{2} \cdot \max \{d_{TV}(\bar{P}_{\lambda_{0}}, P_{\lambda_{0}}), d_{TV}(\bar{Q}_{\lambda_{0}}, Q_{\lambda_{0}})\}
\]

Applying the bounds above iteratively, the triangle inequality and the data processing inequality
now yields that
\[
\begin{align*}
d_{TV} & \left( \phi'_\ell(G(n, k, q', q')) , \mathcal{M}_{2^\ell n} \left( 2^{\ell} k, P_{\lambda_\ell} , Q_{\lambda_\ell} \right) \right) \\
& \leq d_{TV} \left( \phi_\ell \circ \phi_0(G(n, k, p), \phi_\ell(M_n(k, P_{\lambda_0}, Q_{\lambda_0})) \right) + d_{TV} \left( \phi_\ell(M_n(k, P_{\lambda_0}, Q_{\lambda_0}) , \mathcal{M}_{2^\ell n} \left( 2^{\ell} k, P_{\lambda_\ell} , Q_{\lambda_\ell} \right) \right) \\
& \leq d_{TV} \left( \phi'_0(G(n, k, p)), M_n(k, P_{\lambda_0}, Q_{\lambda_0}) \right) + \sum_{i=1}^{\ell} \left( 2^{i-1} n \cdot d_{TV} \left( Q_{\lambda_i} , Q'_{\lambda_i} \right) + \frac{\chi^2(Q_{\lambda_i}, P_{\lambda_\ell})}{2} \right) \\
& \leq \left( \frac{n}{2} \right) \cdot \max \left\{ d_{TV} \left( \tilde{P}_{\lambda_0}, P_{\lambda_0} \right) , d_{TV} \left( \tilde{Q}_{\lambda_0}, Q_{\lambda_0} \right) \right\} + \sum_{i=1}^{\ell} \left( 2^{i-1} n \cdot d_{TV} \left( Q_{\lambda_i} , Q'_{\lambda_i} \right) + \frac{\chi^2(Q_{\lambda_i}, P_{\lambda_\ell})}{2} \right) 
\end{align*}
\]

By the same reasoning, we have that
\[
\begin{align*}
d_{TV} \left( \phi'_0(G(n, q')), M_n(Q_{\lambda_0}) \right) & \leq \left( \frac{n}{2} \right) \cdot d_{TV} \left( \tilde{Q}_{\lambda_0}, Q_{\lambda_0} \right) 
\end{align*}
\]

Now note that if \( M \sim M_m(Q_{\lambda_0}) \), every entry of \( W'_{ij} \) in Step 3b below the main diagonal is i.i.d. sampled from \( Q_{\lambda_1} \) other than those with \( i + j = 2m + 1 \), which are sampled from \( Q'_{\lambda_1} \). Coupling entries individually implies that
\[
\begin{align*}
d_{TV} \left( \phi_1(M), M_{2m}(Q_{\lambda_1}) \right) & \leq m \cdot d_{TV} \left( Q_{\lambda_1} , Q'_{\lambda_1} \right) 
\end{align*}
\]

By induction we have that the data processing and triangle inequalities imply that
\[
\begin{align*}
d_{TV} \left( \phi_\ell(M), M_{2^\ell n}(Q_{\lambda_\ell}) \right) & \leq \sum_{i=1}^{\ell} 2^{i-1} m \cdot d_{TV} \left( Q_{\lambda_i} , Q'_{\lambda_i} \right) 
\end{align*}
\]

Therefore it follows that
\[
\begin{align*}
\begin{align*}
d_{TV} \left( \phi'_\ell(G(n, q')), M_{2^\ell n}(Q_{\lambda_\ell}) \right) & \leq d_{TV} \left( \phi_\ell \circ \phi_0(G(n, q')), \phi_\ell(M_n(Q_{\lambda_0})) \right) \\
& + d_{TV} \left( \phi_\ell(M_n(Q_{\lambda_0}) , M_{2^\ell n}(Q_{\lambda_\ell}) \right) \\
& \leq d_{TV} \left( \phi'_0(G(n, q')), M_n(Q_{\lambda_0}) \right) + \sum_{i=1}^{\ell} 2^{i-1} m \cdot d_{TV} \left( Q_{\lambda_i} , Q'_{\lambda_i} \right) \\
& \leq \left( \frac{n}{2} \right) \cdot d_{TV} \left( \tilde{Q}_{\lambda_0}, Q_{\lambda_0} \right) + \sum_{i=1}^{\ell} 2^{i-1} m \cdot d_{TV} \left( Q_{\lambda_i} , Q'_{\lambda_i} \right) 
\end{align*}
\]

If \( W \sim M_{2^\ell n}(2^{\ell} k, P_{\lambda_\ell} , Q_{\lambda_\ell}) \), then the graph with adjacency matrix \( A_{ij} = 1_{\{W_{ij} > t\}} \) is distributed as \( G(2^{\ell} n, 2^{\ell} k, p, q) \) where \( p = \mathbb{P}_{X \sim P_{\lambda_\ell}} [X > t] \) and \( q = \mathbb{P}_{X \sim Q_{\lambda_\ell}} [X > t] \). Similarly if \( W \sim M_{2^\ell n}(Q_{\lambda_\ell}) \) then the graph with adjacency matrix \( A_{ij} = 1_{\{W_{ij} > t\}} \) is distributed as \( G(2^{\ell} n, q) \). Now combining the total variation bounds above with the data processing inequality proves the theorem. \( \square \)
Chapter 6

Planted Dense Subgraph and Biclustering

6.1 Poisson Lifting and Lower Bounds for Low-Density PDS

In this section, we introduce Poisson lifting to give a reduction from planted clique to \( \text{PDS}(n, k, p, q) \) in the regime where \( \frac{p}{q} \to c \) as \( n \to \infty \) for some fixed \( c > 1 \) and \( q = \Theta(n^{-\alpha}) \) for some fixed \( \alpha > 0 \). Poisson lifting is a specific instance of DISTRIBUTIONAL-LIFTING with Poisson target distributions. The guarantees of Poisson lifting are captured in the following lemma.

Lemma 6.1.1 (Poisson Lifting). Suppose that \( n \) and \( \ell \) are such that \( \ell = O(\log n) \) and are sufficiently large. Fix arbitrary constants \( \epsilon \in (0, 1) \) and \( c > 1 \) and let \( \lambda_0 = n^{-\epsilon} \). Suppose that \( \gamma \) is a small enough constant satisfying that \( \log_c \gamma^{-1} \geq 3\epsilon^{-1} \). Then \( \phi = \text{POISSON-LIFTING} \) is a randomized polynomial time computable map \( \phi: \mathcal{G}_n \to \mathcal{G}_{2^\ell n} \) such that under both \( H_0 \) and \( H_1 \), it holds that

\[
\text{d}_{TV}(\phi(\text{PC}(n, k, \gamma)), \text{PDS}(2^\ell n, 2^\ell k, p, q)) = O\left(n^{-\epsilon/2}\right)
\]

where \( p = 1 - e^{4^{-\ell}c\lambda_0} \) and \( q = 1 - e^{4^{-\ell}\lambda_0} \).

Proof. Let \( \text{POISSON-LIFTING} \) be the algorithm DISTRIBUTIONAL-LIFTING applied with the parameters in Figure 6.1. Let \( \lambda_{i+1} = g_{cl}(\lambda_i) = \lambda_i/4 \) for each \( 0 \leq i \leq \ell - 1 \). Note that the cloning map \( f_{cl}(x) \) can be computed in \( O(1) \) operations. Furthermore, if \( x \sim \text{Pois}(\lambda) \) then Poisson thinning implies that if \( f_{cl}(x) = (x_1, x_2, x_3, x_4) \) then the \( x_i \) are independent and satisfy that \( x_i \sim \text{Pois}(\lambda/4) \).
Therefore it follows that

\[ f_{cl}(P_\lambda) = f_{cl}(\text{Pois}(c \lambda)) \sim \text{Pois}(c \lambda/4)^\otimes 4 = P_{g_{cl}(\lambda)}^{\otimes 4} \quad \text{and} \]

\[ f_{cl}(Q_\lambda) = f_{cl}(\text{Pois}(\lambda)) \sim \text{Pois}(\lambda/4)^\otimes 4 = Q_{g_{cl}(\lambda)}^{\otimes 4} \]

Furthermore, note that \( P_\lambda = \text{Pois}(c \lambda) \) and \( Q_\lambda = Q'_{\lambda} = \text{Pois}(\lambda) \) can be sampled in \( O(1) \) time. Note that the \( \chi^2 \) divergence between these distributions is

\[
\chi^2(Q_\lambda, P_\lambda) = -1 + \sum_{t=0}^{\infty} \left( \frac{1}{t!} e^{-c \lambda} (c \lambda)^t \right)^2 = -1 + \exp \left( (c^{-1} - 1)^2 \lambda \right) \cdot \sum_{t=0}^{\infty} \frac{e^{-\lambda/c} (\lambda/c)^t}{t!}
\]

\[
= \exp \left( (c^{-1} - 1)^2 \lambda \right) - 1 \leq 2c^{-1}(c - 1)^2 \lambda
\]

as long as \( c^{-1}(c - 1)^2 \lambda \leq 1 \) since \( e^x \leq 1 + 2x \) for \( x \in [0, 1] \). By Lemma 5.1.2, the rejection kernel \( \text{RK}_{P_1} \) can be computed in \( O(\log n) \) time and satisfies that

\[ d_{TV}(\text{RK}_{P_1}(1), P_{\lambda_0}) = O(n^{-3}) \quad \text{and} \quad d_{TV}(\text{RK}_{P_1}(\text{Bern}(\gamma)), Q_{\lambda_0}) = O(n^{-3}) \]
Now note that $P_{\lambda_1} = \text{Pois}(4^{-\epsilon}c\lambda_0)$ and $Q_{\lambda_1} = \text{Pois}(4^{-\epsilon}\lambda_0)$ which implies that $p = \mathbb{P}_{X \sim P_{\lambda_1}}[X > 0] = 1 - e^{4^{-\epsilon}\lambda_0}$ and $q = \mathbb{P}_{X \sim Q_{\lambda_1}}[X > 0] = 1 - e^{4^{-\epsilon}\lambda_0}$. Since $\text{PDS}(n, k, 1, \gamma)$ is the same problem as $\text{PC}(n, k, \gamma)$, applying Theorem 5.2.1 yields that under both $H_0$ and $H_1$, we have

$$
\begin{align*}
&d_{\text{TV}}\left(\phi(\text{PC}(n, k, \gamma)), \text{PDS}\left(\ell n, 2k, p, q\right)\right) \\
&\leq \binom{n}{2} \cdot \max\{d_{\text{TV}}(\text{RK}_P(1), P_{\lambda_1}), d_{\text{TV}}(\text{RK}_P(\text{Bern}(\gamma)), Q_{\lambda_1})\} + \sum_{i=1}^{\ell} \sqrt{\frac{X_i^2(Q_{\lambda_1}, P_{\lambda_1})}{2}} \\
&\leq \binom{n}{2} \cdot O(n^{-3}) + c^{-1/2}(c-1) \sum_{i=1}^{\ell} \sqrt{\lambda_i} \\
&= O(n^{-1}) + c^{-1/2}(c-1)n^{-\ell/2} \sum_{i=1}^{\ell} 2^{-i} = O\left(n^{-1} + n^{-\ell/2}\right)
\end{align*}
$$

which completes the proof of the lemma.

We now use the reduction based on Poisson lifting analyzed above to prove hardness for the sparsest regime of $\text{PDS}$.

**Theorem 6.1.1.** Fix some $c > 1$. Let $\alpha \in [0, 2)$ and $\beta \in (0, 1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{4}$. There is a sequence $\{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $q = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^{\beta})$ or equivalently,

$$
\lim_{n \to \infty} \frac{\log q_n}{\log N_n} = \alpha, \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta \quad \text{and} \quad \lim_{n \to \infty} \frac{p_n}{q_n} = c
$$

2. For any sequence of randomized polynomial-time tests $\phi_n : G_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems $\text{PDS}_D(N_n, K_n, p_n, q_n)$ is at least 1 assuming the $\text{PC}$ conjecture holds for each fixed density $p \leq 1/2$.

Therefore the computational boundary for $\text{PDS}_D(n, k, p, q)$ in the parameter regime $q = \tilde{\Theta}(n^{-\alpha})$, $p \rightarrow c$ and $K = \tilde{\Theta}(n^{\beta})$ is $\beta^* = \frac{1}{2} + \frac{\alpha}{4}$.

**Proof.** If $\beta < \alpha$ then $\text{PDS}$ in this regime is information-theoretically impossible. Thus we may assume that $\beta \geq \alpha$. Take $\epsilon > 0$ to be a small enough constant so that

$$
\beta + \epsilon(1 - \beta) < \frac{1}{2} + \frac{\alpha}{4}
$$

Now let $\gamma = \frac{2\beta - \alpha + \epsilon(1 - \beta)}{2 - \alpha}$. Rearranging the inequality above yields that $\gamma \in (0, 1/2)$. Now set

$$
\ell_n = \left\lceil \frac{(\alpha - \epsilon) \log_2 n}{2 - \alpha} \right\rceil, \quad k_n = \lceil n^\gamma \rceil, \quad N_n = 2^\ell_n n, \quad K_n = 2^\ell_n k_n,
$$
\[ p_n = 1 - e^{4-\epsilon n} c \alpha - \epsilon, \quad q_n = 1 - e^{4-\epsilon n} c \alpha - \epsilon \]

Take \( p \) to be a small enough constant so that \( p < \frac{1}{2} c^{-3} \alpha^{-1} \). By Lemma 6.1.1, there is a randomized polynomial time algorithm mapping \( PCD(n, k, n, p) \) to \( PDSD(N, k, p, q) \) with total variation converging to zero as \( n \to \infty \). This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that
\[
\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{(\alpha - \epsilon) \log_2 n}{2 - \alpha} \cdot \log 2 + \frac{2(2 - \alpha + (1 - \beta))}{2 - \alpha} \log n = \frac{\alpha - \epsilon}{2 - \alpha} + \frac{2(2 - \alpha + (1 - \beta))}{2 - \alpha} = \beta
\]

Note that as \( n \to \infty \), it follows that since \( 4^{-\epsilon n} n^{-\epsilon} \to 0 \),
\[ q_n = 1 - e^{4-\epsilon n} n^{-\epsilon} \sim 4^{-\epsilon n} n^{-\epsilon} \]

Similarly \( p_n \sim 4^{-\epsilon n} c \alpha - \epsilon \) and thus \( \frac{p_n}{q_n} \to c \). Note that
\[
\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2(\alpha - \epsilon) \log_2 n}{2 - \alpha} \cdot \log 2 + \frac{\epsilon \log n}{2 - \alpha} = \frac{2\alpha - \epsilon}{2 - \alpha} + \frac{\epsilon}{2 - \alpha} = \alpha
\]

which completes the proof. \( \square \)

In this section, we gave a planted clique lower bound for \( PDSD(n, k, p, q) \) with \( \frac{p}{q} \to c \) as opposed to \( p = cq \) exactly. We now will describe a simple reduction from \( PDSD(n, k, p, q) \) with \( \frac{p}{q} \to c_1 \) where \( c_1 > c \) to \( PDSD(n, k, p_1, q_1) \) where \( p_1 = cq_1 \) and \( q_1 = \Theta(q) \). Given an instance of \( PDSD(n, k, p, q) \), add in every non-edge independently with probability \( \rho = \frac{p - cq}{c_1 - cq} \) which is in \((0, 1)\) since \( c_1 > c \) implies that \( p > cq \) for large enough \( n \). This yields an instance of \( PDSD \) with \( p_1 = 1 - (1 - \rho)(1 - p) = p + \rho - \rho p \) and \( q_1 = 1 - (1 - \rho)(1 - q) = q + \rho - \rho q \). The choice of \( \rho \) implies that \( p_1 = cq_1 \) exactly and \( \rho = \Theta(q) \) since \( \frac{p}{q} \to c_1 > c \). Applying this reduction after POISSON-LIFTING yields that \( PDSD(n, k, cq, q) \) has the same planted clique lower bound as in the previous theorem.

### 6.2 Gaussian Lifting and Lower Bounds for High-Density PDS and BC

In parallel to the previous section, here we introduce Gaussian lifting to give a reduction from planted clique to the dense regime of \( PD(n, k, p, q) \) where \( q = \Theta(1) \) and \( p - q = \Theta(n^{-\alpha}) \) for some fixed \( \alpha > 0 \).

The next lemma we prove is an analogue of Lemma 6.1.1 for GAUSSIAN-LIFTING and follows the same structure of verifying the preconditions for and applying Lemma 5.2.1.
Algorithm \textsc{Gaussian-Lifting}

\textit{Inputs:} Graph $G \in \mathcal{G}_n$, iterations $\ell$

Return the output of DISTRIBUTIONAL-LIFTING applied to $G$ with $\ell$ iterations and parameters:

- initial densities $p' = 1$ and $q' = 1/2$
- target families $P_A = N(\lambda, 1)$ and $Q_A = Q'_A = N(0, 1)$
- rejection kernel $R_{K_G} = R_K (1 \rightarrow N(\lambda_0, 1), 1/2 \rightarrow N(0, 1), N)$ where $N = \left[6 \log_2 n\right]$ and
  $\lambda_0 = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$
- cloning map $f_{cl}(x) = (x_1, x_2, x_3, x_4)$ computed as follows:
  1. Generate $G_1, G_2, G_3 \sim_{\text{i.i.d.}} N(0, 1)$
  2. Compute $(x_1, x_2, x_3, x_4)$ as
    \begin{align*}
    x_1 &= \frac{1}{2} (x + G_1 + G_2 + G_3) \\
    x_2 &= \frac{1}{2} (x - G_1 + G_2 - G_4) \\
    x_3 &= \frac{1}{2} (x + G_1 - G_2 - G_3) \\
    x_4 &= \frac{1}{2} (x - G_1 - G_2 + G_3)
    \end{align*}
- parameter map $g_{cl}(\lambda) = \lambda/2$
- threshold $t = 0$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{gaussian_lifting.png}
\caption{Gaussian lifting procedure in Lemma 6.2.1.}
\end{figure}

\textbf{Lemma 6.2.1} (Gaussian Lifting). Suppose that $n$ and $\ell$ are such that $\ell = O(\log n)$ and are sufficiently large and let

$$\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$$

Then $\phi = \textsc{Gaussian-Lifting}$ is a randomized polynomial time computable map $\phi : \mathcal{G}_n \rightarrow \mathcal{G}_{2 \ell n}$ such that under both $H_0$ and $H_1$, it holds that

$$d_{TV}\left(\phi(\text{PC}(n, k, 1/2)), \text{PDS}\left(2^{\ell n}, 2^{\ell k}, \Phi\left(2^{-\ell} \mu\right), 1/2\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$
Proof. Let \textsc{Gaussian-Lifting} be the algorithm \textsc{Distributional-Lifting} applied with the parameters in Figure 6.2. Let \( \lambda_{i+1} = g_{cl}(\lambda_i) = \lambda_i/2 \) for each \( 0 \leq i \leq \ell - 1 \). Note that the cloning map \( f_{cl}(x) \) can be computed in \( O(1) \) operations. Now suppose that \( x \sim N(\lambda, 1) \). If \( f_{cl}(x) = (x_1, x_2, x_3, x_4) \), then it follows that

\[
\begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3 \\
    x_4
\end{bmatrix} = \frac{\lambda}{2} \begin{bmatrix}
    1 \\
    1 \\
    1 \\
    1
\end{bmatrix} + \frac{1}{2} \begin{bmatrix}
    1 & 1 & 1 & 1 \\
    1 & -1 & 1 & -1 \\
    1 & 1 & -1 & -1 \\
    1 & -1 & -1 & 1
\end{bmatrix} \begin{bmatrix}
    x - \lambda
\end{bmatrix}
\]

Since \( x - \mu, G_{ij}^1, G_{ij}^2 \) and \( G_{ij}^3 \) are zero-mean and jointly Gaussian with covariance matrix \( I_4 \), it follows that the entries of \((x_1, x_2, x_3, x_4)\) are also jointly Gaussian. Furthermore, the coefficient matrix above is orthonormal, implying that the covariance matrix of \((x_1, x_2, x_3, x_4)\) remains \( I_4 \). Therefore it follows that \( f_{cl}(N(\lambda, 1)) \sim N(\lambda/2, 1)^{\otimes 4} \). Applying this identity with \( \lambda = 0 \) yields that \( f_{cl}(N(0, 1)) \sim N(0, 1)^{\otimes 4} \). Thus \( f_{cl} \) is a valid cloning map for \( P_{\lambda} \) and \( Q_{\lambda} \) with parameter map \( g_{cl}(\lambda) = \lambda/2 \).

Observe that \( P_{\lambda} = N(\lambda, 1) \) and \( Q_{\lambda} = Q'_{\lambda} = N(0, 1) \) can be sampled in \( O(1) \) time in the given computational model. Note that the \( \chi^2 \) divergence between these distributions is

\[
\chi^2(Q_{\lambda}, P_{\lambda}) = -1 + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} d\lambda = -1 + \frac{e^{2\lambda^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x+\lambda)^2/2} dx = e^{2\lambda^2} - 1 \leq 4\lambda^2
\]
as long as \( 4\lambda^2 \leq 1 \) since \( e^x \leq 1 + 2x \) for \( x \in [0, 1] \), which is the case for all \( \lambda = \lambda_i \). By Lemma 5.1.4, the rejection kernel \( \text{RKG} \) can be computed in \( \text{poly}(n) \) time and satisfies that

\[
d_{TV}(\text{RKG}(1), P_{\lambda_0}) = O(n^{-3}) \quad \text{and} \quad d_{TV}(\text{RKG}(\text{Bern}(1/2)), Q_{\lambda_0}) = O(n^{-3})
\]

Now note that \( P_{\lambda} = N(2^{-\ell} \mu, 1) \) and \( Q_{\lambda} = N(0, 1) \) which implies that \( p = \mathbb{P}_{X \sim P_{\lambda}}[X > 0] = \Phi(2^{\ell} \mu) \) and \( q = \mathbb{P}_{X \sim Q_{\lambda}}[X > 0] = 1/2 \). Since \( \text{PDS}(n, k, 1, 1/2) \) is the same problem as \( \text{PC}(n, k, 1/2) \), applying Theorem 5.2.1 yields that under both \( H_0 \) and \( H_1 \), we have

\[
d_{TV}(\phi(\text{PC}(n, k, 1/2)), \text{PDS}(2^\ell n, 2^{\ell} k, p, q)) \leq \left(\frac{n}{2}\right) \cdot \max\{d_{TV}(\text{RKG}(1), P_{\lambda_0}), d_{TV}(\text{RKG}(\text{Bern}(1/2)), Q_{\lambda_0})\} + \sum_{i=1}^{\ell} \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}
\]

\[
\leq \left(\frac{n}{2}\right) \cdot O(n^{-3}) + \sqrt{2} \cdot \sum_{i=1}^{\ell} \lambda_i
\]

\[
= O(n^{-1}) + \mu \sqrt{2} \cdot \sum_{i=1}^{\ell} 2^{-i} = O\left(n^{-1} + \frac{1}{\sqrt{\log n}}\right)
\]
which completes the proof of the lemma.

We now use this Gaussian-lifting reduction to deduce hardness for the dense variant of planted dense subgraph, which has a slightly different computational boundary than the sparsest variant.

**Theorem 6.2.1.** Let \( \alpha \in [0, 2) \) and \( \beta \in (0, 1) \) be such that \( \beta < \frac{1}{2} + \frac{a}{2} \). There is a sequence \( \{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}} \) of parameters such that:

1. The parameters are in the regime \( q = \Theta(1) \), \( p - q = \Theta(N^{-\alpha}) \) and \( K = \Theta(N^\beta) \) or equivalently,

\[
\lim_{n \to \infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
\]

2. For any sequence of randomized polynomial-time tests \( \phi_n : \mathcal{G}_{N_n} \to \{0, 1\} \), the asymptotic Type I+II error of \( \phi_n \) on the problems \( \text{PDS}_D(N_n, K_n, p_n, q_n) \) is at least 1 assuming the PC conjecture holds with density \( p = 1/2 \).

Therefore the computational boundary for \( \text{PDS}_D(n, k, p, q) \) in the parameter regime \( q = \Theta(1) \), \( p - q = \Theta(n^{-\alpha}) \) and \( k = \Theta(n^\beta) \) is \( \beta^* = \frac{1}{2} + \frac{a}{2} \).

**Proof.** If \( \beta < 2\alpha \) then PDS is information-theoretically impossible. Thus we may assume that \( \beta \geq 2\alpha \). Let \( \gamma = \frac{\beta - \alpha}{1 - \alpha} \) and note that \( \gamma \in (0, 1/2) \). Now set

\[
\ell_n = \left\lfloor \frac{\alpha \log_2 n}{1 - \alpha} \right\rfloor, \quad k_n = \lceil n^\gamma \rceil, \quad N_n = 2^{\ell_n} n, \quad K_n = 2^{\ell_n} k_n,
\]

\[
p_n = \Phi \left( 2^{-\ell_n} \cdot \frac{\log 2}{2\sqrt{6} \log n + 2 \log 2} \right)
\]

By Lemma 6.1.1, there is a randomized polynomial time algorithm mapping \( \text{PC}_D(n, k_n, 1/2) \) to the detection problem \( \text{PDS}_D(N_n, K_n, p_n, q_n) \) with total variation converging to zero as \( n \to \infty \). This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that

\[
\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{\left\lfloor \frac{\alpha \log_2 n}{1 - \alpha} \right\rfloor \cdot \log 2 + \left( \beta - \alpha \right) \log n}{\left\lfloor \frac{\alpha \log_2 n}{1 - \alpha} \right\rfloor \cdot \log 2 + \log n} = \frac{\alpha}{1 - \alpha} + \frac{\beta - \alpha}{1 - \alpha} = \beta
\]

Let \( \mu = \frac{\log 2}{2\sqrt{6} \log n + 2 \log 2} \) and note that as \( n \to \infty \), we have that

\[
\lim_{n \to \infty} \frac{\Phi \left( 2^{-\ell_n} \mu \right) - \frac{1}{2}}{2^{-\ell_n} \mu} = \lim_{\tau \to 0} \left( \frac{1}{\tau \sqrt{2\pi}} \int_0^\tau e^{-x^2/2} dx \right) = \frac{1}{\sqrt{2\pi}}
\]
Therefore $p_n - q_n \sim \frac{2^{-\ell n} \mu}{\sqrt{2\pi}}$ as $n \to \infty$. This implies

$$\lim_{n \to \infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2 \left[ \frac{\alpha \log_2 n}{1-\alpha} \right] \cdot \log 2 - \log \mu}{\left[ \frac{\alpha \log_2 n}{1-\alpha} \right] \cdot \log 2 + \log n} = \frac{2\alpha}{1-\alpha} + 1 = \alpha$$

which completes the proof.

Note that to prove this theorem, it was only necessary to map to instances with ambient density $q = 1/2$. We remark that it is possible to map from $q = 1/2$ to any constant $q$ by removing edges with a constant probability $p < 1$ or removing non-edges with probability $p$. Note that this still preserves the asymptotic regime $p - q = \tilde{\Theta}(n^{-\alpha})$. We now use GAUSSIAN-LIFTING to give a reduction from planted clique to biclustering.

**Lemma 6.2.2.** Suppose that $n$ and $\ell$ are such that $\ell = O(\log n)$ and are sufficiently large and

$$\mu \leq \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$$

Then there is a randomized polynomial time computable map $\phi = \text{BC-REDUCTION}$ with $\phi : \mathcal{G}_n \to \mathbb{R}^{2^\ell n \times 2^\ell n}$ such that under $H_0$ and $H_1$, it holds that

$$d_{TV} \left( \phi(\text{PC}(n, k, 1/2)), \text{BC} \left( 2^\ell n, 2^\ell k, 2^{-\ell - 1/2} \mu \right) \right) = O \left( \frac{1}{\sqrt{\log n}} \right)$$


**Algorithm BC-REDUCTION**

*Inputs:* Graph $G \in \mathcal{G}_n$, iterations $\ell$

1. Set $W$ to be the output of GAUSSIAN-LIFTING applied to $G$ with $\ell$ iterations without the thresholding in Step 4 of DISTRIBUTIONAL-LIFTING

2. Replace the diagonal entries with $W_{ii} \sim \text{i.i.d. } N(0, 2)$

3. Generate an antisymmetric $2^\ell n \times 2^\ell n$ matrix $A$ of with i.i.d. $N(0, 1)$ random variables below its main diagonal and set

$$W \leftarrow \frac{1}{\sqrt{2}} (W + A)$$

4. Generate a permutation $\sigma$ of $[2^\ell n]$ uniformly at random and output $W^{\text{id}, \sigma}$

**Figure 6.3:** Reduction to biclustering in Lemma 6.2.2.
6.2. GAUSSIAN LIFTING AND LOWER BOUNDS FOR HIGH-DENSITY PDS AND BC

Proof. Let \( \phi = BC\text{-REDUCTION} \) be as in Figure 6.5. Let \( \phi'_\ell \) denote GAUSSIAN-LIFTING applied to \( G \) that outputs \( W \) after \( \ell \) iterations without the thresholding in Step 4 of DISTRIBUTIONAL-LIFTING. Lemmas 5.2.1 and 6.2.1 imply GAUSSIAN-LIFTING ensures that

\[
d_{TV}(\phi'_\ell(G(n,1/2)), M_{2\varepsilon n}(N(0,1))) = O\left(\frac{1}{\sqrt{\log n}}\right)
\]

Now suppose that \( W \sim M_{2\varepsilon n}(N(0,1)) \) and let \( W' = \phi_{2\cdot3}(W) \) denote the value of \( W \) after applying Steps 2 and 3 in Figure 6.5 to \( W \). Note that the diagonal entries of \( W \) are i.i.d. \( N(0,1) \) since the diagonal entries of \( A \) are zero. If \( i < j \), then it follows that

\[
W_{ij} = \frac{1}{\sqrt{2}}(W_{ij} + G_{ij}) \quad \text{and} \quad W_{ji} = \frac{1}{\sqrt{2}}(W_{ij} - G_{ij})
\]

Since \( W_{ij} \) and \( G_{ij} \) are independent and distributed as \( N(0,1) \), it follows that \( W_{ij} \) and \( W_{ji} \) are jointly Gaussian and uncorrelated, which implies that they are independent. Furthermore, \( W_{ij} \) and \( W_{ji} \) are both in the \( \sigma \)-algebra \( \sigma\{W_{ij}, G_{ij}\} \) and collection of \( \sigma \)-algebras \( \sigma\{W_{ij}, G_{ij}\} \) with \( i < j \) is independent. Thus it follows that both \( W' \) and \( (W')_{id,\sigma} \) are distributed as \( N(0,1)^{\otimes 2^n \times 2^n} \). It follows by the data processing inequality that

\[
d_{TV}(\phi(G(n,1/2)), N(0,1)^{\otimes 2^n \times 2^n}) \leq d_{TV}(\phi'_{\ell}(G(n,1/2)), \mathcal{L}(W)) = O\left(\frac{1}{\sqrt{\log n}}\right)
\]

Now consider \( G \sim G(n,k,1/2) \) and note that GAUSSIAN-LIFTING ensures that

\[
d_{TV}(\phi(G), M_{2\varepsilon n}(2^\ell k, N(2^{-\ell} \mu, 1), N(0,1))) = O\left(\frac{1}{\sqrt{\log n}}\right)
\]

Now let \( W' \sim M_{2\varepsilon n}(S, N(2^{-\ell} \mu, 1), N(0,1)) \) where \( S \) is a subset of \([2^n]\) of size \( 2^\ell k \) and let \( W \) be the matrix formed by applying Steps 1 and 2 above to \( W' \) in place of \( \phi'_{\ell}(G) \). By the same jointly Gaussian independence argument above, it follows that the entries of \( W \) are independent and distributed as:

- \( W_{ij} \sim N(2^{-\ell-1/2} \mu, 1) \) if \((i,j) \in S \times S \) and \( i \neq j \); and
- \( W_{ij} \sim N(0,1) \) if \((i,j) \notin S \times S \) or \( i = j \).

Now consider the matrix \((W')_{id,\sigma} \) conditioned on the permutation \( \sigma \). Its entries are independent and identically distributed to the corresponding entries of \( 2^{-\ell-1/2} \mu \cdot 1_S^T \cdot 1 + N(0,1)^{\otimes 2^n \times 2^n} \) where \( T = \sigma(S) \) other than at the indices \((i, \sigma(i)) \) for \( i \in S \). Marginalizing to condition only on \( \sigma(S) = T \)
and coupling all entries with indices outside of $S \times T$ yields that
\[
d_{TV} \left( \mathcal{L}((W')^{id,\sigma} | \sigma(S) = T), \mathcal{L} \left( 2^{-\ell - 1/2} \mu \cdot 1_S 1_T^T + N(0,1)^{\otimes 2^n \times 2^n} \right) \right)
d_{TV} \left( \mathcal{L} \left( (W')^{id,\sigma}[S \times T] | \sigma(S) = T \right), N(2^{-\ell - 1/2} \mu, 1)^{\otimes 2^k \times 2^k} \right)
\leq \sqrt{\frac{1}{2}} \cdot \chi^2 \left( (N(0,1), N(2^{-\ell - 1/2} \mu, 1)) \right)
\leq 2^{-\ell} \mu \leq \frac{\log 2}{2^{\ell + 1} \sqrt{\log n + 2 \log 2}} = O \left( \frac{1}{2^{\ell} \sqrt{\log n}} \right)
\]
by applying Lemma 4.1.1 and the $\chi^2$ upper bound shown in Lemma 6.2.1. Note that Lemma 4.1.1 applies because $(W')^{id,\sigma}[S \times T]$ conditioned only on $\sigma(S) = T$ is distributed as a $2^k \times 2^k$ matrix with i.i.d. entries $N(2^{-\ell - 1/2} \mu, 1)$, other than its diagonal entries which are i.i.d. $N(0, 1)$, and with its columns randomly permuted. Letting $\sigma(S) = T$ be chosen uniformly at random over all ordered pairs of size $2^k$ subsets of $[2^n]$ yields by the triangle inequality that
\[
d_{TV} \left( \mathcal{L}((W')^{id,\sigma}), \int \mathcal{L} \left( 2^{-\ell - 1/2} \mu \cdot 1_S 1_T^T + N(0,1)^{\otimes 2^n \times 2^n} \right) d\pi'(T) \right) = O \left( \frac{1}{2^{\ell} \sqrt{\log n}} \right)
\]
where $\pi'$ is the uniform distribution on all size $2^k$ subsets of $[2^n]$. Let $\pi(S,T)$ be the uniform distribution on all pairs of subsets of size $2^k$ of $[2^n]$. Taking $S$ to be also chosen uniformly at random yields by the data processing and triangle inequalities that
\[
d_{TV} \left( \mathcal{L}(\phi(G)), \int \mathcal{L} \left( 2^{-\ell - 1/2} \mu \cdot 1_S 1_T^T + N(0,1)^{\otimes 2^n \times 2^n} \right) d\pi(S,T) \right)
\leq d_{TV} \left( \phi'(G), M_{2^n}(2^k, N(2^{-\ell} \mu, 1), N(0, 1)) \right)
+ E_S \left[ d_{TV} \left( \mathcal{L}((W')^{id,\sigma}), \int \mathcal{L} \left( 2^{-\ell - 1/2} \mu \cdot 1_S 1_T^T + N(0,1)^{\otimes 2^n \times 2^n} \right) d\pi'(T) \right) \right]
= O \left( \frac{1}{\sqrt{\log n}} \right)
\]
which completes the proof of the lemma.

Note that Lemma 6.2.2 provides a randomized polynomial time map that exactly reduces from $PC_D(n, k, 1/2)$ to $BC_D(2^n, 2^k, 2^{-\ell-1/2} \mu)$. This reduction yields tight computational lower bounds for a simple vs. simple hypothesis testing variant of biclustering as stated in Theorem 6.2.2. This follows from setting $\ell_n, k_n, N_n$ and $K_n$ as in Theorem 6.2.1 and $\mu_n = 2^{-\ell_n - 1/2} \mu$, then applying an identical analysis as in Theorem 6.2.1. Note that when $\beta < \frac{1}{2}$, this choice sets $\ell_n = 0$ and deduces that $BC_D$ is hard when $\alpha > 0$.

**Theorem 6.2.2.** Let $\alpha > 0$ and $\beta \in (0,1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{2}$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:
Algorithm BC-RECOVERY

Inputs: Graph $G \in \mathcal{G}_n$, density bias $\rho$

1. Let $\mathbf{R}_G = \mathbf{R}(\frac{1}{2} + \rho \rightarrow N(\mu, 1), \frac{1}{2} \rightarrow N(0, 1), N)$ where $\mu = \frac{\log(1 + 2\rho)}{2\sqrt{6 \log n + 2 \log 2}}$ and $N = \lceil 6 \log_{1+2\rho} n \rceil$ and compute the symmetric matrix $W \in \mathbb{R}^{n \times n}$ with

$$W_{ij} = \mathbf{R}_G(1_{\{i,j\} \in E(G)})$$

for all $i \neq j$ and $W_{ii} \sim \text{i.i.d.} N(0, 2)$

2. Generate an antisymmetric matrix $A \in \mathbb{R}^{n \times n}$ of i.i.d. $N(0, 1)$ random variables below its main diagonal and set

$$W \leftarrow \frac{1}{\sqrt{2}} (W + A)$$

3. Generate a permutation $\sigma$ of $[n]$ uniformly at random and output $W^{id, \sigma}$

Figure 6.4: Reduction to biclustering recovery in Lemma 6.2.3.

1. The parameters are in the regime $\mu = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

$$\lim_{n \to \infty} \frac{\log \mu_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. For any sequence of randomized polynomial-time tests $\phi_n : \mathbb{R}^{N_n \times N_n} \to \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems $\text{BCD}(N_n, K_n, \mu_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.

Therefore the computational boundary for $\text{BCD}(n, k, \mu)$ in the parameter regime $\mu = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

We now deduce the computational barrier for the biclustering recovery problem from the PDS recovery conjecture. The obtained boundary of $\beta^* = \frac{1}{2} + \alpha$ is stronger than the detection boundary $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$ in the previous theorem. First we will need the following lemma, which gives the necessary total variation guarantees for our reduction. We omit details that are identical to the proof of Lemma 6.2.2.

Lemma 6.2.3. Suppose that $n, \mu$ and $\rho \geq n^{-1}$ are such that

$$\mu = \frac{\log(1 + 2\rho)}{2\sqrt{6 \log n + 2 \log 2}}$$
Then there is a randomized polynomial time computable map \( \phi = \text{BC-RECOVERY} \) with \( \phi : \mathcal{G}_n \to \mathbb{R}^{n \times n} \) such that for any subset \( S \subseteq [n] \) with \( |S| = k \), it holds that

\[
d_{TV}\left( \phi(G(n, 1/2 + \rho, 1/2, S)), \int L \left( \mu \cdot 1_S 1_T^\top + N(0, 1)^{\otimes n \times n} \right) d\pi(T) \right) = O\left( \frac{1}{\sqrt{\log n}} \right)
\]

where \( \pi \) is the uniform distribution on subsets of \([n]\) of size \( k \).

**Proof.** Let \( \phi = \text{BC-RECOVERY} \) be as in Figure 6.4. Applying Lemma 5.1.4, it holds that \( \text{RK}_G \) can be computed in \( \text{poly}(n) \) time and that

\[
d_{TV}(\text{RK}_G(\text{Bern}(1/2 + \rho)), N(\mu, 1)) = O(n^{-3}) \quad \text{and} \quad d_{TV}(\text{RK}_G(\text{Bern}(1/2)), N(0, 1)) = O(n^{-3})
\]

Let \( W_1 \) and \( W_2 \) be the values of \( W \) after Steps 1 and 2, respectively, applied to an input graph \( G \sim G(n, 1/2 + \rho, 1/2, S) \). Let \( M \) be a sample from \( M_n(S, N(\mu, 1), N(0, 1)) \) with i.i.d. \( N(0, 2) \) random variables on its diagonal. Coupling entries individually yields that

\[
d_{TV}(\mathcal{L}(W_1), \mathcal{L}(M)) \leq \binom{k}{2} \cdot d_{TV}(\text{RK}_G(\text{Bern}(1/2 + \rho)), N(\mu, 1))
\]

\[
\quad + \left( \binom{n}{2} - \binom{k}{2} \right) \cdot d_{TV}(\text{RK}_G(\text{Bern}(1/2)), N(0, 1))
\]

\[
= \binom{n}{2} \cdot O(n^{-3}) = O(n^{-1})
\]

An identical argument as in Lemma 6.2.2 now shows that \( W_2 \) is at total variation distance \( O(n^{-1}) \) from \( \mu \cdot 1_S 1_T^\top + N(0, 1)^{\otimes n \times n} \) with all of its diagonal entries replaced with i.i.d. samples from \( N(0, 1) \). The same permutation argument applying Lemma 4.1.1 now yields that

\[
d_{TV}\left( \mathcal{L}(W_2^{id, \sigma}), \int L \left( \mu \cdot 1_S 1_T^\top + N(0, 1)^{\otimes n \times n} \right) d\pi(T) \right) = O\left( \frac{1}{\sqrt{\log n}} \right)
\]

Applying the triangle and data processing inequalities as in the conclusion of Lemma 6.2.2 completes the proof of the lemma. \( \square \)

With this lemma, we now deduce the recovery barrier for biclustering from the PDS and PC conjectures. Note that the recovery barrier of \( \beta^* = \frac{1}{2} + \alpha \) and detection barrier of \( \beta^* = \frac{1}{2} + \frac{\alpha}{2} \) indicates that recovery is conjectured to be strictly harder than detection for the formulations we consider in the regime \( \beta > \frac{1}{2} \).

**Theorem 6.2.3.** Let \( \alpha > 0 \) and \( \beta \in (0, 1) \). There is a sequence \( \{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}} \) of parameters such that:
6.2. GAUSSIAN LIFTING AND LOWER BOUNDS FOR HIGH-DENSITY PDS AND BC

1. The parameters are in the regime \( \mu = \tilde{O}(N^{-\alpha}) \) and \( K = \tilde{O}(N^\beta) \) or equivalently,

\[
\lim_{n \to \infty} \frac{\log \mu_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
\]

2. If \( \beta \geq \frac{1}{2} \) and \( \beta < \frac{1}{2} + \alpha \), then the following holds. Let \( \epsilon > 0 \) be fixed and let \( M_n \) be an instance of \( \text{BC}_R(N_n, K_n, \mu_n) \). There is no sequence of randomized polynomial-time computable functions \( \phi_n : \mathbb{R}^{N_n \times N_n} \to \binom{[N_n]}{k}^2 \) such that for all sufficiently large \( n \) the probability that \( \phi_n(M_n) \) is exactly the pair of latent row and column supports of \( M_n \) is at least \( \epsilon \), assuming the PDS recovery conjecture.

3. If \( \beta < \frac{1}{2} \) and \( \alpha > 0 \), then the following holds. There is no sequence of randomized polynomial-time computable functions \( \phi_n : \mathbb{R}^{N_n \times N_n} \to \binom{[N_n]}{k}^2 \) such that for all sufficiently large \( n \) the probability that \( \phi_n(M_n) \) is exactly the pair of latent row and column supports of \( M_n \) is at least \( \epsilon \), assuming the PC conjecture.

Therefore, given the PDS recovery conjecture, the computational boundary for \( \text{BC}_R(n, k, \mu) \) in the parameter regime \( \mu = \tilde{O}(n^{-\alpha}) \) and \( k = \tilde{O}(n^\beta) \) is \( \beta^* = \frac{1}{2} + \alpha \) when \( \beta \geq \frac{1}{2} \) and \( \alpha^* = 0 \) when \( \beta < \frac{1}{2} \).

**Proof.** First we consider the case when \( \beta \geq 1/2 \) and \( \beta < \frac{1}{2} + \alpha \). Now set

\[
k_n = \left[ n^{\beta} \right], \quad \rho_n = n^{-\alpha}, \quad N_n = n, \quad K_n = k_n, \quad \mu_n = \frac{\log(1 + 2\rho_n)}{2\sqrt{6} \log n + 2 \log 2}
\]

Assume for contradiction that there is a sequence of randomized polynomial-time computable functions \( \phi_n \) as described above and let \( \phi_n^\tau \) denote the restriction of \( \phi_n \) to output latent row support only. Let \( \varphi_n = \text{BC-RECOVERY} \) be the reduction in Lemma 6.2.3, let \( G_n \sim G(n, k_n, 1/2 + \rho_n, 1/2) \) and let \( M_n = \varphi_n(G_n) \). Let \( \mathcal{L}_{n,S,T} = \mathcal{L} \left( \mu_n 1_S 1_T^T + N(0, 1)^{\otimes n \times n} \right) \) and let \( \mathcal{L}_{n,S} = \int \mathcal{L}_{n,S,T} d\pi(T) \) where \( \pi \) is the uniform distribution over \( k\)-subsets of \([n]\) be the distribution of an instance of \( \text{BC}_R(N_n, K_n, \mu_n) \) conditioned on the event that the row support of its planted submatrix is \( S \).

Now observe that

\[
|\mathbb{P}_{M \sim \mathcal{L}(M_n)} [\phi_n^\tau(M) = S] - \mathbb{P}_{M \sim \mathcal{L}_{n,S}} [\phi_n^\tau(M) = S]| \leq d_{TV} \left( \mathcal{L}(M_n), \mathcal{L}_{n,S} \right) = O \left( \frac{1}{\sqrt{\log n}} \right)
\]

because of Lemma 6.2.3. Now note that \( \mathbb{P}_{M \sim \mathcal{L}_{n,S}} [\phi_n^\tau(M) = S] = \mathbb{E}_{T \sim \pi} \mathbb{P}_{M \sim \mathcal{L}_{n,S,T}} [\phi_n^\tau(M) = S] \geq \epsilon \)

for sufficiently large \( n \) by assumption. Therefore it follows that

\[
\mathbb{P}[\phi_n^\tau \circ \varphi_n(G_n) = S] \geq \epsilon - O \left( \frac{1}{\sqrt{\log n}} \right)
\]

which is at least \( \epsilon/2 \) for sufficiently large \( n \). Now observe that

\[
\lim_{n \to \infty} \frac{\log k_n}{\log n} = \beta \quad \text{and} \quad \lim_{n \to \infty} \frac{\log \left( \frac{k_n^2 \rho_n^2}{\left( \frac{1}{4} - \rho^2 \right)} \right)}{\log n} = 2\beta - 2\alpha < 1
\]
Since the sequence of functions \( \phi_n \circ \phi_n \) can be computed in randomized polynomial time, this contradicts the PDS recovery conjecture. Therefore no such sequence of functions \( \phi_n \) exists for the parameter sequence \( \{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}} \) defined above. Now note that as \( n \to \infty \),

\[
\mu_n = \frac{\log(1 + 2\rho_n)}{2\sqrt{6} \log n + 2 \log 2} \sim \frac{\rho_n}{\sqrt{6} \log n} = \frac{n^{-\alpha}}{\sqrt{6} \log n}
\]

Therefore it follows that

\[
\lim_{n \to \infty} \frac{\log \mu_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{\alpha \log n + \frac{1}{2} \log(6 \log n)}{\log n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
\]

This completes the proof in the case that \( \beta \geq 1/2 \). Now consider the case where \( \beta < 1/2 \). Set \( \rho_n = 1/2 \) and all other parameters as above. Let \( G_n \sim G(n, k_n, S) \) and repeat the same argument as above to obtain that \( \mathbb{P}[\phi_n(\phi_n(G_n)) = S] \geq \epsilon - o(1) \). Now consider the algorithm \( \phi'_n : G_n \to \{0, 1\} \) that computes \( S' = \phi'_n(\phi_n(G_n)) \) and checks if \( S' \) is a clique, outputting a 1 if it is and 0 or 1 uniformly at random otherwise. If \( G_n \sim G(n, 1/2) \), then with probability \( 1 - o(1) \) the largest clique of \( G_n \) is less than \((2 + \epsilon) \log 2 \) for any fixed \( \epsilon > 0 \). It follows by the definition of \( \phi_n \) that \( |S'| = k_n = \Theta(n^\beta) = o(\log n) \) and thus with probability \( 1 - o(1) \), \( \phi'_n \) outputs a random bit. Therefore \( \mathbb{P}_{G_n \sim G(n, 1/2)}[\phi'_n(G_n) = 1] = 1/2 + o(1) \). If \( G_n \sim G(n, k, 1/2) \), then with probability at least \( \epsilon - o(1) \), \( S' \) is the support of the planted clique and \( \phi'_n \) outputs a 1. Otherwise, \( \phi'_n \) outputs a random bit. Therefore \( \mathbb{P}_{G_n \sim G(n, k, 1/2)}[\phi'_n(G_n) = 0] = (1 - \epsilon)/2 + o(1) \). Therefore it follows that the Type I+II error of \( \phi'_n \) is

\[
\mathbb{P}_{G_n \sim G(n, 1/2)}[\phi'_n(G_n) = 1] + \mathbb{P}_{G_n \sim G(n, k, 1/2)}[\phi'_n(G_n) = 0] = 1 - \frac{\epsilon}{2} + o(1)
\]

which contradicts the PC conjecture. This completes the proof of the theorem. \( \square \)

### 6.3 Lower Bounds for General PDS

In this section we give a reduction to the general regime of PDS where \( q = \tilde{\Theta}(n^{-\alpha}) \) and \( p - q = \tilde{\Theta}(n^{-\beta}) \) where \( \beta > \alpha \). Note that in order to completely characterize PDS when \( p - q = O(q) \), we also need the computational lower bound shown in Section 6.1 when \( \alpha = \beta \). We now give this reduction, which applies GAUSSIAN-LIFTING and POISSON-LIFTING in sequence.

**Lemma 6.3.1.** Fix some arbitrary \( \epsilon \in (0, 1) \). Suppose that \( n, \ell_1 \) and \( \ell_2 \) are such that \( \ell_1, \ell_2 = O(\log n) \) and are sufficiently large. Let \( \ell = \ell_1 + \ell_2 \) and

\[
\mu = \frac{\log 2}{2\sqrt{6} \log n + 2 \log 2}
\]
Algorithm \textsc{General-PDS-Reduction}

\textbf{Inputs:} Graph \(G \in \mathcal{G}_n\), iterations \(\ell_1, \ell_2\)

1. Let \(H\) be the output of \textsc{Gaussian-Lifting} applied to \(G\) with \(\ell_1\) iterations

2. Update \(H\) to be the output of \textsc{Poisson-Lifting} applied to \(H\) with \(\ell_2\) iterations where
   the rejection kernel is replaced with
   \[
   \text{RK}_{F_2} = \text{RK} \left( \frac{1}{2} + \rho \to \text{Pois}(c\lambda), \frac{1}{2} \to \text{Pois}(\lambda), N \right)
   \]
   where the rejection kernel has natural parameter \(2^{\ell_1} \mu\) and satisfies
   \[
   \lambda = (2^{\ell_1} \mu)^{-\epsilon}, \quad N = \left\lceil 6 \rho^{-1} \log(2^{\ell_1} \mu) \right\rceil, \quad \rho = \Phi(2^{-\ell_1} \mu) - 1/2 \quad \text{and} \quad c = (2\Phi(2^{-\ell_1} \mu))^{\epsilon/4}
   \]

3. Output \(H\)

\textbf{Figure 6.5:} Reduction to the general regime of planted dense subgraph in Lemma 6.3.1.

Then there is a randomized polynomial time computable map \(\phi = \textsc{General-PDS-Reduction}\) with \(\phi : \mathcal{G}_n \to \mathcal{G}_n\) such that under both \(H_0\) and \(H_1\), it holds that

\[
\text{d}_{TV} \left( \phi(\text{PC}(n, k, 1/2)), \text{PDS} \left( 2^{\ell_1} n, 2^{\ell_1} k, p_{\ell_1, \ell_2}, q_{\ell_1, \ell_2} \right) \right) = O \left( \frac{1}{\sqrt{\log n}} \right)
\]

where \(p_{\ell_1, \ell_2}\) and \(q_{\ell_1, \ell_2}\) are defined to be

\[
p_{\ell_1, \ell_2} = 1 - \exp \left( 4^{-\ell_2} \left( 2^{\ell_1} \mu \right)^{-\epsilon} \cdot (2\Phi(2^{-\ell_1} \mu))^{\epsilon/4} \right) \quad \text{and} \quad q_{\ell_1, \ell_2} = 1 - \exp \left( 4^{-\ell_2} \left( 2^{\ell_1} \mu \right)^{-\epsilon} \right)
\]

\textbf{Proof.} Let \(\phi = \textsc{General-PDS-Reduction}\) be as in Figure 6.5. Let \(\phi_1(G)\) denote the map in Step 1 applying \textsc{Gaussian-Lifting} for \(\ell_1\) iterations and let \(\phi_2(G)\) denote the map in Step 2 applying the modified version of \textsc{Poisson-Lifting} for \(\ell_2\) iterations. If \(G \sim G(n, k, 1/2)\) then Lemma 6.2.1 implies that

\[
\text{d}_{TV} \left( \phi_1(G), G \left( 2^{\ell_1} n, 2^{\ell_1} k, \Phi(2^{-\ell_1} \mu), 1/2 \right) \right) = O \left( \frac{1}{\sqrt{\log n}} \right)
\]

For the values of \(\lambda, \rho\) and \(c\) in Step 2, we have that \(c < 2^{1/4}\) and \(\log_c(1 + 2\rho) = 4^{-1} = O(1)\). Also observe that as \(n \to \infty\),

\[
\rho = \Phi(2^{-\ell_1} \mu) - 1/2 \sim \frac{1}{\sqrt{2\pi}} \cdot 2^{-\ell_1} \mu = \Omega \left( \frac{1}{2^{\ell_1} \sqrt{\log n}} \right) = \omega \left( \frac{1}{2^{\ell_1} n} \right)
\]
Therefore the values $K = 1, \epsilon, \lambda, \rho, c$ and natural parameter $2^{\ell_1}n$ satisfy the preconditions to apply Lemma 5.1.4. It follows that

$$d_{TV}(\text{RKP}_2(\text{Bern}(1/2 + \rho)), \text{Pois}(c\lambda)) \leq O\left(2^{-3\ell_1}n^{-3}\right),$$

and

$$d_{TV}(\text{RKP}_2(\text{Bern}(1/2)), \text{Pois}(\lambda)) \leq O\left(2^{-3\ell_1}n^{-3}\right).$$

Now let $H'$ be a sample from $G\left(2^{\ell_1}n, 2^{\ell_1}k, \Phi\left(2^{-\ell_1}\mu\right), 1/2\right)$. The argument in Lemma 6.1.1 applied with these total variation bounds yields that

$$d_{TV}\left(\phi_2(H'), G\left(2^{\ell_1}n, 2^{\ell_1}k, p_{\ell_1, \ell_2}, q_{\ell_1, \ell_2}\right)\right) = O\left(2^{-\ell_1\epsilon/2}n^{-\epsilon/2}\right)$$

Applying the triangle inequality and data processing inequality yields that

$$d_{TV}\left(\phi(G(n, k, 1/2)), G\left(2^{\ell_1}n, 2^{\ell_1}k, p_{\ell_1, \ell_2}, q_{\ell_1, \ell_2}\right)\right)$$

$$\leq d_{TV}\left(\phi_2 \circ \phi_1(G), \phi_2(H')\right) + d_{TV}\left(\phi_2(H'), G\left(2^{\ell_1}n, 2^{\ell_1}k, p_{\ell_1, \ell_2}, q_{\ell_1, \ell_2}\right)\right)$$

$$= O\left(\frac{1}{\log n} + 2^{-\ell_1\epsilon/2}n^{-\epsilon/2}\right) = O\left(\frac{1}{\sqrt{\log n}}\right).$$

By the same argument, if $G \sim G(n, 1/2)$ then

$$d_{TV}\left(\phi(G(n, 1/2)), G\left(2^{\ell_1}n, q_{\ell_1, \ell_2}\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

which completes the proof of the lemma. \qed

We now use this reduction combining GAUSSIAN-LIFTING and POISSON-LIFTING to identify the computational barrier in the general PDS problem.

**Theorem 6.3.1.** Let $\alpha, \gamma \in [0, 2)$ and $\beta \in (0, 1)$ be such that $\gamma \geq \alpha$ and $\beta < \frac{1}{2} + \frac{\gamma}{2} - \frac{\alpha}{4}$. There is a sequence $\{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $p - q = \tilde{O}(N^{-\gamma})$, $q = \tilde{O}(N^{-\alpha})$ and $K = \tilde{O}(N^\beta)$ or equivalently,

$$\lim_{n \to \infty} \frac{\log q_n}{\log N_n} = \alpha, \quad \lim_{n \to \infty} \frac{\log (p_n - q_n)^{-1}}{\log N_n} = \gamma \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. For any sequence of randomized polynomial-time tests $\phi_n : G_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems PDS$_D(N_n, K_n, c q_n, q_n)$ is at least 1 assuming the PC conjecture holds for $p = 1/2$.

Therefore the computational boundary for PDS$_D(n, k, p, q)$ in the parameter regime $p - q = \tilde{O}(n^{-\gamma})$, $q = \tilde{O}(n^{-\alpha})$ and $k = \tilde{O}(n^\beta)$ where $\gamma \geq \alpha$ is $\beta^* = \frac{1}{2} + \frac{\gamma}{2} - \frac{\alpha}{4}$. 

6.3. LOWER BOUNDS FOR GENERAL PDS

Proof. If $\beta < 2\gamma - \alpha$ then PDS in this regime is information-theoretically impossible. Thus we may assume that $1 > \beta \geq 2\gamma - \alpha$. Now let

$$\eta = 1 - (1 - \beta) \cdot \frac{2 - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)}$$

Note that the given condition on $\alpha$, $\beta$, and $\gamma$ rearranges to $\frac{1 - \beta}{2 - \alpha - 2(\gamma - \alpha)} > \frac{1}{4}$. Therefore taking $\epsilon > 0$ to be small enough ensures that $2 - \alpha - (2 - \epsilon)(\gamma - \alpha) > 0$, $\eta \in (0, 1/2)$ and $\alpha > \epsilon$. Now set

$$\ell_1^n = \left[ \frac{(\gamma - \alpha)(2 - \epsilon) \log_2 n}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} \right], \quad \ell_2^n = \left[ \frac{(\alpha - \epsilon) \log_2 n}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} \right],$$

$$k_n = \lceil n^\eta \rceil, \quad N_n = 2^{\ell_1^n + \ell_2^n} n, \quad K_n = 2^{\ell_1^n + \ell_2^n} n_n,$$

$$p_n = 1 - \exp \left( 4^{-\ell_2^n} \left( 2^{\ell_1^n} n \right)^{-\epsilon} \cdot \left( 2\Phi \left( 2^{-\ell_1^n} \mu \right) \right)^{\epsilon/4} \right), \quad q_n = 1 - \exp \left( 4^{-\ell_2^n} \left( 2^{\ell_1^n} n \right)^{-\epsilon} \right)$$

where $\mu = \frac{\log 2}{2\sqrt{\varepsilon \log n + 2\log 2}}$. By Lemma 6.3.1, there is a randomized polynomial time algorithm mapping $PCD(n, k_n, 1/2)$ to $PDS_D(N_n, K_n, p_n, q_n)$ with total variation converging to zero as $n \to \infty$. This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n \to \infty} \log \frac{K_n}{N_n} = \frac{(\gamma - \alpha)(2 - \epsilon)(\gamma - \alpha)}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + \frac{\alpha - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + 1 - \frac{(1 - \beta)(2 - \epsilon)}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)}$$

$$= \frac{2 - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} - \frac{2 - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + 1 = \beta$$

Using the approximations in the proof of Theorem 6.2.1, we obtain as $n \to \infty$

$$q_n \sim 4^{-\ell_2^n} \left( 2^{\ell_1^n} n \right)^{-\epsilon},$$

$$p_n - q_n \sim 4^{-\ell_2^n} \left( 2^{\ell_1^n} n \right)^{-\epsilon} \left[ \left( 2\Phi \left( 2^{-\ell_1^n} \mu \right) \right)^{\epsilon/4} - 1 \right] \sim 4^{-\ell_2^n} \left( 2^{\ell_1^n} n \right)^{-\epsilon} \cdot \frac{\epsilon}{2\sqrt{2\pi} \cdot 2^{\ell_1^n} \mu}$$

Now it follows that

$$\lim_{n \to \infty} \frac{\log \left( q_n^{-1} \right)}{\log N_n} = \frac{2 \cdot \frac{\alpha - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + \epsilon \cdot \frac{\gamma - \alpha - (2 - \epsilon)}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} = \alpha$$

$$\lim_{n \to \infty} \frac{\log \left( p_n - q_n \right)^{-1}}{\log N_n} = \frac{2 \cdot \frac{\alpha - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + (1 + \epsilon) \cdot \frac{\gamma - \alpha - (2 - \epsilon)}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} + \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} = \gamma$$

which completes the proof. \qed
Chapter 7

Reflection Cloning and Subgraph SBM

7.1 Reflecting Cloning and Rank-1 Submatrix

Suppose that $n$ is even and fixed. Let $R$ denote the linear operator on $\mathbb{R}^{n \times n}$ matrices that reflects a matrix horizontally about its vertical axis of symmetry. Let $F$ denote the linear operator that multiplies each entry on the right half of a matrix by $-1$. The reflection cloning reduction is to iteratively update the matrix $W$ to

$$W \leftarrow \frac{1}{\sqrt{2}} (RW^\sigma + FW^\sigma)$$

where $\sigma$ is chosen uniformly at random and then to update $W$ similarly with vertical analogues of $R$ and $F$. This achieves the same scaling of $\mu$ as GAUSSIAN-LIFTING but does not increase $n$. This ends up tightly achieving the right parameter scaling to deduce the sharper hardness of ROS$_D$ and, indirectly, SPCA$_D$ over problems that admit sum-tests such as PIS$_D$, PDS$_D$, BC$_D$ and BSPCA$_D$. The parameter scaling in these problems is captured exactly by the cloning methods in the previous two sections. Note that even without Step 3, REFLECTION-CLONING necessarily causes $r'$ and $c'$ to have negative entries and hence cannot show hardness of BC$_D$. We remark that all previous cloning methods are in some sense lossy, introducing independent sources of randomness at each entry of the input matrix. In contrast, the only randomness introduced in REFLECTION-CLONING are random permutations of rows and columns, and ends up achieving a much sharper scaling in $\mu$.

We remark that if $r, c \in \{-1, 0, 1\}$ then $r'$ and $c'$ have most of their entries in $\{-1, 0, 1\}$. However, those that are not are information-theoretically detectable. We note that if it were possible to reduce from $r, c \in \{-1, 0, 1\}$ to $r'$ and $c'$ with this property, then this would strengthen our
Algorithm REFLECTION-CLONING

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$ where $n$ is even, number of iterations $\ell$

1. Initialize $W \leftarrow M$

2. For $i = 0, 1, \ldots, \ell - 1$ do:
   a. Generate a permutation $\sigma$ of $[n]$ uniformly at random
   a. Let $W' \in \mathbb{R}^{n \times n}$ have entries
      $$W'_{ij} = \frac{1}{2} \left( W_{ij}^{\sigma, \sigma} + W_{(n+1-i)j}^{\sigma, \sigma} + W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} + W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} \right)$$
      $$W'_{(n+1-i)j} = \frac{1}{2} \left( W_{ij}^{\sigma, \sigma} - W_{(n+1-i)j}^{\sigma, \sigma} + W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} - W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} \right)$$
      $$W'_{i(n+1-j)} = \frac{1}{2} \left( W_{ij}^{\sigma, \sigma} - W_{(n+1-i)j}^{\sigma, \sigma} - W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} + W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} \right)$$
      $$W'_{(n+1-i)(n+1-j)} = \frac{1}{2} \left( W_{ij}^{\sigma, \sigma} + W_{(n+1-i)j}^{\sigma, \sigma} + W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} - W_{(n+1-i)(n+1-j)}^{\sigma, \sigma} \right)$$
      for each $1 \leq i, j \leq n/2$
   b. Set $W \leftarrow W'$

3. Output $W$

Figure 7.1: Reflection cloning procedure in Lemma 7.1.1.

hardness results for SSBM$_D$, SSW$_D$, ROS$_D$ and SPCAD$_D$. Given a vector $v \in \mathbb{R}^n$ and permutation $\sigma$ of $[n]$, let $r^\sigma$ denote the vector formed by permuting the indices of $r$ according to $\sigma$.

Lemma 7.1.1 (Reflection Cloning). Suppose $n$ is even and $\ell = O(\log n)$. There is a randomized polynomial-time computable map $\phi = \text{REFLECTION-CLONING}$ with $\phi : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ and

1. It holds that
   $$\phi \left( N(0, 1)^{\otimes n \times n} \right) \sim N(0, 1)^{\otimes n \times n}$$

2. Consider any $\lambda > 0$ and any pair of vectors $r, c \in \mathbb{Z}^n$. Then there is a distribution $\pi$ over vectors $r', c' \in \mathbb{Z}^n$ with $\|r'\|_2^2 = 2^\ell \|r\|_2^2$ and $\|c'\|_2^2 = 2^\ell \|c\|_2^2$ such that
   $$\phi \left( \lambda \cdot r c^T + N(0, 1)^{\otimes n \times n} \right) \sim \int \mathcal{L} \left( \frac{\lambda}{2^\ell} \cdot r' c'^T + N(0, 1)^{\otimes n \times n} \right) d\pi(r', c')$$
where it holds with probability at least $1 - 4\|r\|_0^{-1} - 4\|c\|_0^{-1}$ over $\pi$ that

$$2^\ell \|r\|_0 \geq \|r'\|_0 \geq 2^\ell \|r\|_0 \left(1 - \max \left(\frac{2Cl \cdot \log(2^\ell \|r\|_0)\ell \|r\|_0}{\ell \|r\|_0}, \frac{2\|c\|_0}{\ell \|c\|_0}\right)\right)$$

$$2^\ell \|c\|_0 \geq \|c'\|_0 \geq 2^\ell \|c\|_0 \left(1 - \max \left(\frac{2Cl \cdot \log(2^\ell \|c\|_0)\ell \|c\|_0}{\ell \|c\|_0}, \frac{2\|c\|_0}{\ell \|c\|_0}\right)\right)$$

for some constant $C$ if $\|r\|_0$ and $\|c\|_0$ are sufficiently large and at most $2^{\ell - 1} n$. Furthermore, if $r = c$ then $r' = c'$ holds almost surely.

Proof. Let $\phi(M)$ be implemented by the procedure $\text{REFLECTION-CLONING}(M, \ell)$ as in Figure 6. If $\ell = O(\log n)$, this algorithm runs in randomized polynomial time. Let $\phi_1(W)$ denote the map that takes the value $W$ prior to an iteration of Step 2 as its input and outputs the value of $W$ after this iteration.

If $W \sim N(0, 1)^{\otimes n \times n}$, then it follows by a similar argument as in Lemma 7 that $\phi_1(W) \sim N(0, 1)^{\otimes n \times n}$. Specifically, for each $1 \leq i, j \leq n/2$ we have that

$$\begin{bmatrix}
W'_{ij} \\
W'_{(n+i-j)j} \\
W'_{(n+i-j)(n+j)}
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix} \begin{bmatrix}
W_{ij}^{\sigma,\sigma} \\
W_{(n+i-j)j}^{\sigma,\sigma} \\
W_{(n+i-j)(n+j)}^{\sigma,\sigma}
\end{bmatrix}$$

where $\sigma$ is the random permutation generated in Step 2a. It holds that $W^{\sigma,\sigma} \sim N(0, 1)^{\otimes n \times n}$ and therefore that the vector of entries on the right hand side above is distributed as $N(0, 1)^{\otimes 4}$. Since the coefficient matrix is orthogonal, it follows that the vector on the left hand side is also distributed as $N(0, 1)^{\otimes 4}$. Since the $\sigma$-algebras $\sigma\{W_{ij}^{\sigma,\sigma}, W_{(n+i-j)j}^{\sigma,\sigma}, W_{(n+i-j)(n+j)}^{\sigma,\sigma}\}$ are independent as $(i, j)$ ranges over $[n/2]^2$, it follows that $W' = \phi_1(W) \sim N(0, 1)^{\otimes n \times n}$. Iterating, it now follows that $\phi(N(0, 1)^{\otimes n \times n}) \sim N(0, 1)^{\otimes n \times n}$, establishing Property 1.

Now consider the case when $W = \lambda \cdot Rc^T + U$ where $U \sim N(0, 1)^{\otimes n \times n}$. Note that $W'$ can be expressed in terms of $W^{\sigma,\sigma}$ as

$$W' = \frac{1}{2} (A + B)^T W^{\sigma,\sigma} (A + B)^T = \frac{\lambda}{2} (Ar^\sigma + Br^\sigma) (Ac^\sigma + Bc^\sigma)^T + \frac{1}{2} (A + B)^T U^{\sigma,\sigma} (A + B)$$

where $B$ is the $n \times n$ matrix with ones on its anti-diagonal and zeros elsewhere, and $A$ is given by

$$A = \begin{bmatrix} I_{n/2} & 0 \\ 0 & -I_{n/2} \end{bmatrix}$$

Note that $\frac{1}{2} (A + B)^T U^{\sigma,\sigma} (A + B)$ is distributed as $\phi_1(U) \sim N(0, 1)^{\otimes n \times n}$. Since $A + B$ is symmetric and satisfies that $(A + B)^2 = 2 \cdot I_n$, we have that

$$\|Ar^\sigma + Br^\sigma\|_2^2 = 2\|r^\sigma\|_2^2 = 2\|r\|_2^2 \quad\text{and}\quad \|Ac^\sigma + Bc^\sigma\|_2^2 = 2\|c^\sigma\|_2^2 = 2\|c\|_2^2$$
Let $r_0 = r$ and $r_{i+1} = Ar_{i+1} + Br_{i+1}$ where $\sigma_i$ is the permutation generated in the $i$th iteration of Step 2. It follows by induction that $r_\ell \in \mathbb{Z}^n$ and $\|r_\ell\|_2^2 = 2^\ell \|r\|_2^2$ hold almost surely. Analogously define $c_\ell$ for each $0 \leq i \leq \ell$ and note that

$$\phi \left( \lambda \cdot r_\ell^T + N(0,1)^{\otimes n \times n} \right) \sim \mathcal{L} \left( \frac{\lambda}{2^\ell} \cdot r_\ell^T + N(0,1)^{\otimes n \times n} \right)$$

Furthermore note that if $r_0 = r = c = c_0$, then $r_i = c_i$ for all $i$ holds almost surely. Thus it suffices to show that the desired bounds on $\|r_\ell\|_0$ and $\|c_\ell\|_0$ hold with high probability in order to establish Property 2.

Note that since $A + B$ has two nonzero entries per row and column, it follows that $2\|r_i\|_0 \geq \|r_{i+1}\|_0$ for all $i$. Now consider the set

$$S_i = \{ \{j, n + 1 - j\} : j, n + 1 - j \in \text{supp}(r_i^2) \}$$

Note that if $j \in \text{supp}(r_i^2)$, then $j$ is only not in the support of $r_{i+1}$ if it is in some unordered pair in $S_i$. Also note that $(r_{i+1})_j + (r_{i+1})_{n+1-j} = 2r_i^2 \neq 0$ if $j$ is in some unordered pair in $S_i$. Therefore at most one element per unordered pair of $S_i$ can be absent from the support of $r_{i+1}$. This implies that $2\|r_i\|_0 - \|r_{i+1}\|_0 \leq |S_i|$ for each $i$. For each $1 \leq j \leq n/2$, let $X_j$ be the indicator for the event that $\{j, n + 1 - j\} \in S_i$. Let $t = \|r_i\|_0$ and note that $|S_i| = X_1 + X_2 + \cdots + X_{n/2}$. For any subset $T \subseteq [n/2]$, it follows that if $|T| \leq n/2$ then

$$E \left[ \prod_{j \in T} X_j \right] = \frac{t(t-1) \cdots (t-2|T|+1)}{n(n-1) \cdots (n-2|T|+1)} \leq \left( \frac{t}{n} \right)^{2|T|}$$

and if $|T| > n/2$, then this expectation is zero. Let $Y \sim \text{Bin}(n/2, t^2/n^2)$ and note that the above inequality implies that $E[|S_i|^k] \leq E[Y^k]$ for all $j \geq 0$. This implies that if $\theta \geq 0$, then

$$E[\exp(\theta |S_i|)] \leq E[\exp(\theta Y)] = \left( 1 + (e^\theta - 1) \cdot \frac{t^2}{n^2} \right)^{n/2} \leq \exp \left( (e^\theta - 1) \cdot \frac{t^2}{2n} \right)$$

A Chernoff bound now yields that

$$P[|S_i| \geq k] \leq \exp \left( (e^\theta - 1) \cdot \frac{t^2}{2n} - \theta k \right)$$

Setting $k = t^2/n$ and $\theta = \ln 2$ yields that

$$P \left[ |S_i| \geq \frac{t^2}{n} \right] \leq \left( \frac{e}{4} \right)^{\frac{t^2}{2n}} \leq \frac{1}{t}$$

if $t^2/2n \geq \log_{4/e} t$. If $t^2/2n < \log_{4/e} t$, setting $\theta = \ln 2$ and $k = \frac{1}{\ln 2} \left( \log_{4/e} t + \ln t \right) = C \log t$ yields

$$P \left[ |S_i| \geq C \log t \right] \leq \exp \left( \frac{t^2}{2n} - (\ln 2)k \right) = \frac{1}{t}$$
Therefore with probability at least $1 - 1/\|r_i\|_0$, it follows that $|S_i| < \max(C \log \|r_i\|_0, \|r_i\|_0^2/n)$. Note that this inequality implies that

$$\|r_{i+1}\|_0 \geq 2\|r_i\|_0 - |S_i| \geq 2\|r_i\|_0 \left(1 - \max \left(\frac{C \log \|r_i\|_0}{\|r_i\|_0}, \frac{\|r_i\|_0}{n}\right)\right)$$

We now will show by induction on $0 \leq j \leq \ell$ that as long as $\|r\|_0$ is sufficiently large, we have

$$\|r_j\|_0 \geq 2^j \|r\|_0 \left(1 - \max \left(\frac{2Cj \cdot \log(2^j \|r\|_0)}{\|r\|_0}, \frac{2^j \|r\|_0}{n}\right)\right)$$

holds with probability at least

$$1 - \frac{4(1 - 2^{-j})}{\|r\|_0}$$

The claim is vacuously true when $j = 0$. Now assume that this holds for a particular $j$. Now note that since $j \leq \ell$ and $\|r\|_0$ is sufficiently large but at most $2^{\ell-1}n$, we have that

$$\max \left(\frac{2Cj \cdot \log(2^j \|r\|_0)}{\|r\|_0}, \frac{2^j \|r\|_0}{n}\right) \leq \frac{1}{2}$$

Therefore $\|r_j\|_0 \geq 2^{j-1} \|r\|_0$. We now condition on the value of $\|r_j\|_0$ and the event (7.1) holds. Under this conditioning, it follows by the induction hypothesis that with probability at least $1 - 1/\|r_j\|_0 \geq 1 - 2^{1-j} \|r\|^{-1}_0$, we have that

$$\|r_{j+1}\|_0 \geq 2^{j+1} \|r\|_0 \left(1 - \max \left(\frac{2C(j + 1) \cdot \log(2^{j+1} \|r\|_0)}{\|r\|_0}, \frac{2^{j+1} \|r\|_0}{n}\right)\right) \left(1 - \max \left(\frac{C \log \|r_j\|_0}{\|r_j\|_0}, \frac{\|r_j\|_0}{n}\right)\right)$$

$$2^{j+1} \|r\|_0 \left(1 - \max \left(\frac{2C(j + 1) \cdot \log(2^{j+1} \|r\|_0)}{\|r\|_0}, \frac{2^{j+1} \|r\|_0}{n}\right)\right)$$

since $2^{j-1} \|r\|_0 \leq \|r_j\|_0 \leq 2^j \|r\|_0$. Marginalizing over $\|r_j\|_0$ and whether or not (7.1) holds yields that the above inequality holds with probability at least

$$\left(1 - \frac{4(1 - 2^{-j})}{\|r\|_0}\right) \left(1 - \frac{2^{1-j} \|r\|_0}{\|r\|_0}\right) \geq 1 - \frac{4(1 - 2^{-j})}{\|r\|_0}$$

which completes the induction. Obtaining symmetric results for $c$ and taking $j = \ell$ completes the proof of the lemma.

Combining this reflection cloning procedure with the reduction from planted clique to biclustering yields a reduction from planted clique to rank-1 submatrix.

**Lemma 7.1.2.** Suppose that $n$ is even and $2^k k < \frac{n}{\log k}$ where $n$ is sufficiently large. Let

$$\mu \leq \frac{\log 2}{2\sqrt{6 \log n} + 2 \log 2}$$
Algorithm ROS-REDUCTION

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations $\ell$

1. Compute the output $W$ of BC-REDUCTION applied to $G$ with zero iterations

2. Return the output of REFLECTION-CLONING applied to $W$ with $\ell$ iterations

There is a randomized polynomial time computable map $\phi = \text{ROS-REDUCTION}$ with $\phi : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ such that if $G$ is an instance of $\text{PC}(n, k, 1/2)$ then under $H_0$, it holds that

$$d_{TV}(\mathcal{L}_{H_0}(\phi(G)), N(0, 1)^{\otimes n \times n}) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

and, under $H_1$, there is a prior $\pi$ on pairs of unit vectors in $\mathcal{V}_{n, 2 \ell k}$ such that

$$d_{TV}(\mathcal{L}_{H_1}(\phi(G)), \int \mathcal{L} \left(\frac{\mu}{\sqrt{2}} \cdot uv^\top + N(0, 1)^{\otimes n \times n}\right) \, d\pi(u, v)) = O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

Proof. Let $\phi = \text{ROS-REDUCTION}$ be as in Figure 7.2. Let $\phi_1 : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ and $\phi_2 : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ denote the maps in Steps 1 and 2, respectively. By Lemma 6.2.2, it holds that

$$d_{TV}(\phi_1(G(n, 1/2)), N(0, 1)^{\otimes n \times n}) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$$d_{TV}(\phi_1(G(n, k, 1/2)), \int \mathcal{L} \left(\frac{\mu}{\sqrt{2}} \cdot 1_S 1_T^\top + N(0, 1)^{\otimes n \times n}\right) \, d\pi'(S, T)) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

where $\pi'$ is the uniform distribution over pairs of $k$-subsets $S, T \subseteq [n]$. By Lemma 7.1.1, it holds that $\phi_2(N(0, 1)^{\otimes n \times n}) \sim N(0, 1)^{\otimes n \times n}$. By the data processing inequality, we have that

$$d_{TV}(\mathcal{L}_{H_0}(\phi(G)), N(0, 1)^{\otimes n \times n}) = d_{TV}(\phi_2 \circ \phi_1(G(n, 1/2)), \phi_2(N(0, 1)^{\otimes n \times n}))$$

$$\leq d_{TV}(\phi_1(G(n, 1/2)), N(0, 1)^{\otimes n \times n}) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Let $M$ be a the matrix distributed as $\frac{\mu}{\sqrt{2}} \cdot 1_S 1_T^\top + N(0, 1)^{\otimes n \times n}$ where $S$ and $T$ are $k$-element subsets of $[n]$ chosen uniformly at random. It follows that the distribution of $\phi_2(M)$ conditioned on the sets $S$ and $T$ is given by

$$\mathcal{L}(\phi_2(M) | S, T) \sim \int \mathcal{L} \left(\frac{\mu}{2\ell \sqrt{2}} \cdot rc^\top + N(0, 1)^{\otimes n \times n}\right) \, d\pi(r, c)$$
where \( \pi \) is the prior in Lemma 7.1.1. As shown in Lemma 7.1.1, it follows that each pair \((r, c)\) in the support of \( \pi \) satisfies that \( \|r\|_2^2 = 2^k \|1_s\|_2^2 = 2^k k \) and \( \|c\|_2^2 = 2^k \|1_T\|_2^2 = 2^k k \) and that \( \|r\|_0, \|c\|_0 \leq 2^k \). Now consider the prior \( \pi_{S,T} \) which is \( \pi \) conditioned on the event that the following inequalities hold

\[
\|r\|_0, \|c\|_0 \geq 2^k \left( 1 - \max \left( \frac{2C \cdot \log(2^k)}{k}, \frac{2^k k}{n} \right) \right)
\]

As shown in Lemma 7.1.1, this event occurs with probability at least \( 1 - 8/k \). Since \( 2^k k < \frac{n}{\log k} \), it follows that if \( u = \frac{1}{\sqrt{2^k}} \cdot r \) and \( v = \frac{1}{\sqrt{2^k}} \cdot c \) then \( \pi_{S,T} \) induces a prior over pairs \((u, v)\) in \( V_{n,2^k} \). This follows from the fact that \( r, c \in \mathbb{Z}^n \) implies that \( u \) and \( v \) have nonzero entries with magnitudes at least \( 1/\sqrt{2^k} \). Applying Lemma 3.0.4 yields that since \( 2^{-\mu} \cdot r c^\top = \mu \cdot u v^\top \),

\[
d_{TV} \left( \mathcal{L} (\phi_2 (M) \mid S, T), \int \mathcal{L} \left( \frac{\mu k}{\sqrt{2}} \cdot u v^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi_{S,T}(u, v) \right) \leq \frac{8}{k}
\]

Let \( \pi(u, v) = E_{S,T}[\pi_{S,T}(u, v)] \) be the prior formed by marginalizing over \( S \) and \( T \). Note that \( \pi \) is also supported on pairs of unit vectors in \( V_{n,2^k} \). By the triangle inequality, it follows that

\[
d_{TV} \left( \mathcal{L} (\phi_2 (M)), \int \mathcal{L} \left( \frac{\mu k}{\sqrt{2}} \cdot u v^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi(u, v) \right)
\]

\[
\leq E_{S,T} \left[ d_{TV} \left( \mathcal{L} (\phi_2 (M) \mid S, T), \int \mathcal{L} \left( \frac{\mu k}{\sqrt{2}} \cdot u v^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi_{S,T}(u, v) \right) \right] \leq \frac{8}{k}
\]

By the triangle inequality and data processing inequality, we now have that

\[
d_{TV} \left( \mathcal{L}_{H_1} (\phi(G)), \int \mathcal{L} \left( \frac{\mu k}{\sqrt{2}} \cdot u v^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi(u, v) \right)
\]

\[
\leq d_{TV} \left( \mathcal{L}_{H_1} (\phi_2 \circ \phi_1 (G)), \mathcal{L}(\phi_2 (M)) \right) + d_{TV} \left( \mathcal{L}(\phi_2 (M)), \int \mathcal{L} \left( \frac{\mu k}{\sqrt{2}} \cdot u v^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi(u, v) \right)
\]

\[
= O \left( \frac{1}{\sqrt{\log n}} \right) + \frac{8}{k} = O \left( \frac{1}{\sqrt{\log n}} + k^{-1} \right)
\]

since \( M \) is a sample from the mixture \( \int \mathcal{L} \left( \frac{\mu}{\sqrt{2}} \cdot 1_s 1_T^\top + N(0, 1)^{\otimes n \times n} \right) \, d\pi'(S, T) \). This completes the proof of the lemma.

This lemma provides a polynomial time map from an instance of \( \text{PCD}(n, k, 1/2) \) to \( N(0, 1)^{\otimes n \times n} \) under \( H_0 \) and to a distribution in the composite hypothesis \( H_1 \) of \( \text{ROSD}(n, 2^k k, 2^{-1/2} \mu) \) under \( H_1 \). Now we deduce the hard regime of \( \text{ROSD} \) given the planted clique conjecture as in the next theorem. Here, we consider the asymptotic regime \( \frac{\mu k}{\xi} = \tilde{O}(n^{-\alpha}) \) to be consistent with Figure 2.1. The purpose of this parameterization is to focus on the factor \( \frac{\mu}{k} \) required to normalize entries in the planted submatrix to have magnitude approximately 1. This enables a valid comparison between the hardness of \( \text{ROSD} \) and \( \text{BCD} \).
Theorem 7.1.1. Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\beta < \frac{1}{2} + \alpha$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $\frac{4}{K} = \Theta(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,
   $$\lim_{n \to \infty} \frac{\log(K_n\mu_n^{-1})}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. For any sequence of randomized polynomial-time tests $\phi_n : G_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems $\text{ROSD}(N_n, K_n, \mu_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.

Therefore the computational boundary for $\text{ROSD}(n, k, \mu)$ in the parameter regime $\frac{4}{K} = \Theta(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

Proof. If $\beta < 2\alpha$ then $\text{ROSD}$ is information-theoretically impossible. Thus we may assume that $\beta \geq 2\alpha$. Let $\gamma = \beta - \alpha$ and note that $\gamma \in (0, 1/2)$. Now set

$$\ell_n = \lceil \alpha \log_2 n \rceil, \quad k_n = \lceil n^\gamma \rceil, \quad N_n = 2n, \quad K_n = 2^\ell n k_n, \quad \mu_n = \frac{\mu k_n}{\sqrt{2}}$$

where $\mu = \frac{\log 2}{2\sqrt{6} \log n + 2 \log 2}$. By Lemma 7.1.2, there is a randomized polynomial time algorithm mapping $\text{PCD}(2n, k_n, 1/2)$ to the detection problem $\text{ROSD}(N_n, K_n, \mu_n)$ under $H_0$ and to a prior over $H_1$ with total variation converging to zero as $n \to \infty$. This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n \to \infty} \frac{\log(K_n\mu_n^{-1})}{\log N_n} = \lim_{n \to \infty} \frac{[\alpha \log_2 n] \cdot \log 2 - \log(\mu/\sqrt{2})}{\log n + \log 2} = \alpha$$

$$\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{[\alpha \log_2 n] \cdot \log 2 + \log k_n}{\log n + \log 2} = \alpha + (\beta - \alpha) = \beta$$

which completes the proof. \hfill \Box

7.2 Sparse Spiked Wigner Matrix

We now show a computational lower bound for sparse spiked Wigner matrix detection given the planted clique conjecture. As observed in Chapter 2, it suffices to reduce from planted clique to SROS$D$. This is because if $M$ is an instance of SROS$D(n, k, \mu)$, then $\frac{1}{\sqrt{2}}(M + M^\top)$ is an instance of SSW$D(n, k, \mu/\sqrt{2})$. This transformation implies that any computational lower bound that applies to SROS$D(n, k, \mu)$ also applies to SSW$D(n, k, \mu/\sqrt{2})$. The rest of this section is devoted to giving a reduction to SROS$D(n, k, \mu)$. 
7.2. SPARSE SPIKED WIGNER MATRIX

Our reduction uses the symmetry preserving property of reflection cloning and yields the same computational barrier as for ROSD. However, there are two subtle differences between this reduction and that in Lemma 7.1.2. In order to show hardness for SROSD, it is important to preserve the symmetry of the planted sparse structure. This requires planting the hidden entries along the diagonal of the adjacency matrix of the input graph $G$, which we do by an averaging trick. However, this induces an arbitrarily small polynomial loss in the size of the spike, unlike in the reduction to ROSD. Although this does not affect our main theorem statement for SROSD, which only considers poly(n) size factors, it yields a weaker lower bound than that in Theorem 7.1.1 when examined up to sub-polynomial factors. This reduction to SROSD will also serve as the main sub-routine in our reduction to SSBMD.

Lemma 7.2.1. Suppose that $n$ is even and $2^t k < \frac{n}{\log k}$ where $n$ is sufficiently large. Let

$$\mu \leq \frac{\log 2}{2\sqrt{6}\log n + 2\log 2}$$

There is a randomized polynomial time computable map $\phi = \text{SROS-REDUCTION}$ with $\phi : G_n \rightarrow \mathbb{R}^{n \times n}$ such that if $G$ is an instance of $PC(n, k, 1/2)$ then under $H_0$, it holds that

$$d_{TV} \left( \mathcal{L}_{H_0}(\phi(G)), N(0,1)^{\otimes n \times n} \right) = O(n^{-1})$$

and, under $H_1$, there is a prior $\pi$ on unit vectors in $\mathcal{V}_{n, 2^t k}$ such that

$$d_{TV} \left( \mathcal{L}_{H_1}(\phi(G)), \int \mathcal{L} \left( \frac{\mu k (k-1)}{2\sqrt{(n-1)}} \cdot vu^\top + N(0,1)^{\otimes n \times n} \right) d\pi(v) \right) = O(n^{-1})$$

Proof. Let $\phi = \text{SROS-REDUCTION}$ be as in Figure 7.3. Applying the total variation bounds in Lemma 5.1.4 and entry-wise coupling as in Lemma 6.2.3 yields that

$$d_{TV} (\mathcal{L}_{H_0}(W), M_n(N(0,1))) = O(n^{-1}) \quad \text{and} \quad d_{TV} (\mathcal{L}_{H_1}(W), M_n(k, N(\mu,1), N(0,1))) = O(n^{-1})$$

Let $W' \in \mathbb{R}^{n \times n}$ be such that $W' \sim M_n(N(0,1))$ under $H_0$ and $W' \sim M_n(k, N(\mu,1), N(0,1))$ under $H_1$. Let $M'$ denote the matrix computed in Step 3 using the matrix $W'$ in place of $W$. We will first argue that under $H_0$, the variables $W'_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2}$ and $W'_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2}$ for all $i \neq j$ are independent. First note that the $\sigma$-algebras $\sigma \{ W'_{ij}, A_{ij}, B_{ij} \}$ for all $i < j$ are independent. Therefore it suffices to verify that the four variables $W'_{ij} + A_{ij} \pm B_{ij} \cdot \sqrt{2}$ and $W'_{ji} + A_{ji} \pm B_{ji} \cdot \sqrt{2}$ are independent. Observe that these four variables are jointly Gaussian and satisfy that

$$\frac{1}{2} \begin{bmatrix} W'_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2} \\ W'_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2} \\ W'_{ji} + A_{ji} + B_{ji} \cdot \sqrt{2} \\ W'_{ji} + A_{ji} - B_{ji} \cdot \sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} W'_{ij} \\ A_{ij} \\ B_{ij} \\ B_{ji} \end{bmatrix}$$
Algorithm SROS-REDUCTION

**Inputs:** Planted clique instance $G \in \mathcal{G}_n$ with clique size $k$ where $n$ is even, number of iterations $\ell$, mean parameter $0 < \mu \leq \frac{\log 2}{2\sqrt{\log n + \log 2}}$

1. Let $\text{rk}_G = \text{rk} (1 \to N(\mu, 1), 1/2 \to N(0, 1), N)$ where $N = [6 \log_2 n]$ and form the symmetric matrix $W \in \mathbb{R}^{n \times n}$ with $W_{ii} = 0$ and for all $i < j$,

$$W_{ij} = \text{rk}_G (1_{\{i,j\} \in E(G)})$$

2. Sample an antisymmetric matrix $A \in \mathbb{R}^{n \times n}$ with i.i.d. $N(0, 1)$ entries below its main diagonal and sample two matrices $B, C \in \mathbb{R}^{n \times n}$ with i.i.d. $N(0, 1)$ off-diagonal and zero diagonal entries.

3. Form the matrix $M \in \mathbb{R}^{n \times n}$ with

$$M_{ij} = \frac{k - 1}{2\sqrt{n} - 1} \left( W_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2} \right) + C_{ij} \cdot \sqrt{1 - \frac{(k - 1)^2}{n - 1}}$$

for all $i \neq j$ and

$$M_{ii} = \frac{1}{2\sqrt{n} - 1} \sum_{j=1}^{n} \left( W_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2} \right)$$

4. Update $M$ to be the output of REFLECTION-CLONING applied with $\ell$ iterations to $M$

5. Output $M^{\sigma, \sigma}$ where $\sigma$ is a permutation of $[n]$ chosen uniformly at random.

**Figure 7.3:** Reduction to sparse spiked Wigner matrix in Lemma 7.2.1.

since $W'$ is symmetric and $A$ is antisymmetric. Observe that $W'_{ij}, A_{ij}, B_{ij}, B_{ji}$ are independent Gaussians and since the coefficient matrix above is orthogonal, it follows that the vector on the left hand side above is distributed as $N(0, 1)^{\otimes 4}$ and thus has independent entries. Since the $C$ has i.i.d. $N(0, 1)$ entries off of its diagonal and $W', A, B$ and $C$ all have zero diagonals, it follows that $\text{Var}(M'_{ij}) = 1$ for all $i, j \in [n]$. Since the entries of $M'$ are independent and each entry is Gaussian with variance 1, it follows that $M' \sim N(0, 1)^{\otimes n \times n}$.

Now suppose that $H_1$ holds and let $S \subseteq [n]$ be indices of the rows and columns containing the planted $N(\mu, 1)$ entries of $W'$. It now follows that $W'_{ij} = \mu + W''_{ij}$ if $i \neq j$ and $i, j \in S$ and
7.2. SPARSE SPIKED WIGNER MATRIX

\[ W'_{ij} = W''_{ij} \] where \( W'' \sim M_n(N(0, 1)) \). Now note that if \( i \neq j \), we have that conditioned on \( S \),

\[
M'_{ij} = \frac{(k - 1)^\mu}{2\sqrt{n - 1}} \cdot 1_{\{i,j \in S\}} + \frac{k - 1}{2\sqrt{n - 1}} \left( W''_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2} \right) + C_{ij} \cdot \sqrt{1 - \frac{(k - 1)^2}{n - 1}} \]

\[
\sim \frac{(k - 1)^\mu}{2\sqrt{n - 1}} \cdot 1_{\{i,j \in S\}} + N(0, 1)
\]

by applying the previous argument to \( W'' \). Furthermore \( M'_{ii} \) has diagonal entries

\[
M'_{ii} = \frac{(k - 1)^\mu}{2\sqrt{n - 1}} \cdot 1_{\{i \in S\}} + \frac{1}{2\sqrt{n - 1}} \sum_{j=1}^{n} \left( W''_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2} \right)
\]

\[
\sim \frac{(k - 1)^\mu}{2\sqrt{n - 1}} \cdot 1_{\{i \in S\}} + N(0, 1)
\]

conditioned on \( S \). Therefore \( M'|S \sim \frac{(k-1)^\mu}{2\sqrt{n-1}} \cdot 1_{S} 1_{S}^T + N(0, 1)^{\otimes n \times n} \). Now by the data processing inequality, we now have that under \( H_0 \),

\[
d_{TV} \left( \mathcal{L}_{H_0}(M), N(0, 1)^{\otimes n \times n} \right) \leq d_{TV} \left( \mathcal{L}_{H_0}(W), M_n(N(0, 1)) \right) = O(n^{-1})
\]

By the data processing and triangle inequalities, we have that

\[
d_{TV} \left( \mathcal{L}_{H_1}(M), \int \mathcal{L} \left( \frac{(k-1)^\mu}{2\sqrt{(n-1)}} \cdot 1_{S} 1_{S}^T + N(0, 1)^{\otimes n \times n} \right) d\pi'(S) \right)
\]

\[
\leq \mathbb{E}_S d_{TV} \left( \mathcal{L}_{H_1}(M), \frac{(k-1)^\mu}{2\sqrt{(n-1)}} \cdot 1_{S} 1_{S}^T + N(0, 1)^{\otimes n \times n} \right) = O(n^{-1})
\]

where \( \pi' \) is the uniform distribution on \( k \)-subsets of \( [n] \). Now applying the same argument as in the proof of Lemma 7.1.2 and the fact that \textsc{Reflection-Cloning} preserves the fact that the rank-1 mean submatrix is symmetric as shown in Lemma 7.1.1, proves the lemma. \( \square \)

We now use this lemma to deduce the computational barriers for SROS\(_D\) and SSW\(_D\). Although the barrier matches Theorem 7.1.1, the parameter settings needed to achieve it are slightly different due to the polynomial factor loss in the reduction in Lemma 7.2.1.

**Theorem 7.2.1.** Let \( \alpha > 0 \) and \( \beta \in (0, 1) \) be such that \( \beta < \frac{1}{2} + \alpha \). There is a sequence \( \{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}} \) of parameters such that:

1. The parameters are in the regime \( \frac{\mu}{K} = \tilde{O}(N^{-\alpha}) \) and \( K = \tilde{O}(N^\beta) \) or equivalently,

\[
\lim_{n \to \infty} \frac{\log(K_n\mu_n^{-1})}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
\]
2. For any sequence of randomized polynomial-time tests \( \phi_n : \mathcal{G}_{n_n} \rightarrow \{0,1\} \), the asymptotic Type I+II error of \( \phi_n \) on the problems SROS\(_D\)\((N_n, K_n, \mu_n)\) and SSWD\(_D\)\((N_n, K_n, \mu_n/\sqrt{2})\) is at least 1 assuming the PC conjecture holds with density \( p = 1/2 \).

Therefore the computational boundaries for SROS\(_D\)\((n, k, \mu)\) and SSWD\(_D\)\((n, k, \mu)\) in the parameter regime \( \frac{K}{k} = \Theta(n^{-\alpha}) \) and \( k = \Theta(n^\beta) \) is \( \beta^* = \frac{1}{2} + \alpha \) and \( \alpha^* = 0 \) when \( \beta < \frac{1}{2} \).

**Proof.** When \( 2\alpha > \beta \), it holds that SROS\(_D\) is information-theoretically impossible. Therefore we may assume that \( 2\alpha \leq \beta \) and in particular that \( \alpha < \frac{1}{2} \). Now suppose that \( \beta < \frac{1}{2} + \alpha \). First consider the case when \( \beta \geq \frac{1}{2} \) and let \( \epsilon = \frac{1}{2}(\alpha + \frac{1}{2} - \beta) \in (0, \frac{1}{2}) \). Now set

\[
\ell_n = \left[ \left( \beta - \frac{1}{2} + \epsilon \right) \log_2 n \right], \quad k_n = \left[ n^{\frac{1}{2} - \epsilon} \right], \quad N_n = 2n, \quad K_n = 2^\ell n k_n, \quad \mu_n = \frac{\mu k_n (k_n - 1)}{2\sqrt{n - 1}}
\]

where \( \mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2} \log 2} \). By Lemma 7.2.1, there is a randomized polynomial time algorithm mapping PCD\((2n, k_n, 1/2)\) to the detection problem SROS\(_D\)\((N_n, K_n, \mu_n)\) under \( H_0 \) and to a prior over \( H_1 \) with total variation converging to zero as \( n \to \infty \). This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that

\[
\lim_{n \to \infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \lim_{n \to \infty} \frac{\left[ (\beta - \frac{1}{2} + \epsilon) \log_2 n \right] \cdot \log 2 - \log(k_n - 1) - \log(\mu/2) + \frac{1}{2} \log(n - 1)}{\log n + \log 2} = \left( \beta - \frac{1}{2} + \epsilon \right) - \left( \frac{1}{2} - \epsilon \right) + \frac{1}{2} = \alpha
\]

\[
\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{\left[ (\beta - \frac{1}{2} + \epsilon) \log_2 n \right] \cdot \log 2 + \log k_n}{\log n + \log 2} = \left( \beta - \frac{1}{2} + \epsilon \right) + \frac{1}{2} - \epsilon = \beta
\]

Now consider the case when \( \beta < \frac{1}{2} \) and \( \alpha > 0 \). In this case, set \( \ell_n = 0, k_n = \left[ n^{\frac{1}{2} - \epsilon} \right] \) and \( N_n = 2 \left[ n^{\beta - \epsilon} \right] \) and

\[
N_n = 2 \left[ n^{\beta - \epsilon} \right] \quad \text{and} \quad \mu_n = \frac{\mu' k_n (k_n - 1)}{2\sqrt{n - 1}}
\]

where \( \epsilon = \min \left( \frac{\alpha}{2(\alpha + \beta)}, \frac{1}{2} - \beta \right) \) and

\[
\mu' = \frac{\log 2}{2\sqrt{6\log n + 2\log 2} \log 2} \cdot n^{-\frac{\alpha}{\beta} \left( \frac{1}{2} - \epsilon \right)}
\]

Note that \( \epsilon \leq \frac{\alpha}{2(\alpha + \beta)} \) implies that \( \epsilon - \frac{\alpha}{\beta} \left( \frac{1}{2} - \epsilon \right) \leq 0 \). By Lemma 7.2.1, there is a randomized polynomial time algorithm mapping PCD\((2n, k_n, 1/2)\) to the detection problem SROS\(_D\)\((2n, K_n, \mu_n)\) under \( H_0 \) and to a prior over \( H_1 \) with total variation converging to zero as \( n \to \infty \). Now consider the map that pads the resulting instance with i.i.d. \( N(0,1) \) random variables until it is \( N_n \times N_n \). Note that since \( \epsilon \leq \frac{1}{2} - \beta \), we have that \( N_n \geq 2n \). By the data processing and triangle inequalities, it
Algorithm SSBM-REDUCTION

Inputs: Planted clique instance $G \in \mathcal{G}_n$ with clique size $k$ where $n$ is even, iterations $\ell$, mean parameter $\mu \leq \frac{\log 2}{2\sqrt{6}\log n + 2\log 2}$

1. Compute the output $M$ of SROS-REDUCTION applied to $G$ with $\ell$ iterations and mean parameter $\mu$

2. Sample $n$ i.i.d. Rademacher random variables $x_1, x_2, \ldots, x_n$ and update each entry of $M$ to be $M_{ij} \leftarrow x_i x_j M_{ij}$

3. Output the graph $H$ where $\{i, j\} \in E(H)$ if and only if $M_{ij} > 0$ for each $i < j$

Figure 7.4: Reduction to the subgraph stochastic block model in Lemma 7.3.1.

follows that this map takes $PC_D(2n, k_n, 1/2)$ to $SSW_D(N_n, K_n, \mu_n)$ with total variation converging to zero as $n \to \infty$. This implies that property 2 above holds and we now verify property 1. Note

$$
\lim_{n \to \infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \lim_{n \to \infty} \frac{\log 2 + \frac{1}{3} \log(n - 1) - \log(k_n - 1) - \log \mu'}{\log 2 + \frac{1}{3} \left(\frac{1}{2} - \epsilon\right) \log n}
$$

$$
= \frac{\frac{1}{2} - \left(\frac{1}{2} - \epsilon\right) - \left(\epsilon - \frac{\alpha}{3} \left(\frac{1}{2} - \epsilon\right)\right)}{\frac{1}{3} \left(\frac{1}{2} - \epsilon\right)} = \alpha
$$

$$
\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{(\frac{1}{2} - \epsilon) \log n}{(\frac{1}{2} - \epsilon) \log n} = \beta
$$

which completes the proof of the theorem. \qed

7.3 Subgraph Stochastic Block Model

In this section, we show tight hardness for the subgraph stochastic block using the reflection cloning reduction from planted clique to SROS$_D$. This captures the sharper hardness of detection in the subgraph stochastic block model over planted dense subgraph and the lack of a sum test. Note that here total variation distance under $H_1$ refers to the total variation distance to some prior over the distributions in $H_1$.

Lemma 7.3.1. Suppose that $n$ is even and $\delta \in (0, 1/2)$ is such that $(2^\ell k)^{1+2\delta} = O(n)$ and $k =$
\[ \Omega(n^2). \] Let \( \mu > 0 \) be such that
\[ \mu \leq \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}. \]

There is a polynomial time map \( \phi = \text{SSBM-REDUCTION with } \phi : G_n \to G_n \) such that for some mixture \( \mathcal{L}_{\text{SSBM}} \) of distributions in \( G_B(n, 2^k, 1/2, \rho) \), it holds that
\[ d_{TV}(\phi(G(n, 1/2)), G(n, 1/2)) = O(n^{-1}) \quad \text{and} \quad d_{TV}(\phi(G(n, k, 1/2)), \mathcal{L}_{\text{SSBM}}) = O(k^{-1}) \]

where \( \rho \) is given by
\[ \rho = \Phi \left( \frac{\mu(k-1)}{2^{\ell+1}\sqrt{n-1}} \right) - \frac{1}{2} \]

Proof. Given a vector \( v \in \mathbb{Z}^n \) and \( \beta > 0 \), let \( M(\beta, v) \in \mathbb{R}^{n \times n} \) be distributed as \( \beta \cdot vv^\top + N(0, 1)^{\otimes n \times n} \).

Let \( S \subseteq [n] \) and \( P \in [0, 1]^{S \times [S]} \) be a symmetric matrix with zero diagonal entries and rows and columns indexed by vertices in \( S \). Let \( G(n, S, P, q) \) be the distribution on graphs \( G \) generated as follows:

1. if \( i \neq j \) and \( i, j \in S \) then the edge \( \{i, j\} \in E(G) \) independently with probability \( P_{ij} \); and
2. all other edges of \( G \) are included independently with probability \( q \).

Now let \( \tau : \mathbb{R}^{n \times n} \to G_n \) be the map sending a matrix \( M \) to the graph \( G \) such that
\[ E(G) = \{ \{i, j\} : M_{ij} > 0 \text{ and } i < j \} \]

In other words, \( G \) is the graph formed by thresholding the entries of \( M \) below its main diagonal at zero. Suppose that \( v \in \mathbb{Z}^n \) is \( k \)-sparse and \( S = \text{supp}(v) \). Since the entries of \( M(\beta, v) \) are independent, it follows that
\[ \tau(M(\beta, v)) \sim G(n, S, P, q) \quad \text{where} \quad P_{ij} = \Phi(\beta \cdot v_i v_j) \text{ for each } i, j \in S \]

Now let \( S = A \cup B \) where \( A = A(v) = \{ i \in S : v_i > 0 \} \) and \( B = B(v) = \{ i \in S : v_i < 0 \} \). Note that since \( v \in \mathbb{Z}^n \), it follows that if \( i, j \in A \) or \( i, j \in B \) then \( v_i v_j \geq 1 \) and thus \( P_{ij} = \Phi(\beta \cdot v_i v_j) \geq \Phi(\beta) \).

Furthermore if \( (i, j) \in A \times B \) or \( (i, j) \in B \times A \) then \( v_i v_j \leq -1 \) and \( P_{ij} = \Phi(\beta \cdot v_i v_j) \leq 1 - \Phi(\beta) \).

Now note that if \( \sigma \) is a permutation of \( [n] \) chosen uniformly at random then
\[ \mathcal{L}(\tau(M(\beta, v))^\sigma) = \mathcal{L}(\tau(M(\beta, v)^{\sigma, \sigma})) \in G_B(n, k, 1/2, \Phi(\beta) - 1/2) \]

if it also holds that \( \frac{k}{2} - k^{1-\delta} \leq |A|, |B| \leq \frac{k}{2} + k^{1-\delta} \).

Let \( \phi_1 = \text{SROS-REDUCTION} \) and let \( W \sim N(0, 1)^{\otimes n \times n} \). As shown in Lemma 7.2.1,
\[ d_{TV}(\phi_1(G(n, 1/2)), \mathcal{L}(W)) = O(n^{-1}) \]
Now observe that since \( N(0,1) \) is symmetric and since the entries of \( W \) are independent, the distribution of \( W \) is invariant to flipping the signs of any subset of the entries of \( W \). Therefore \( xx^T \circ W \sim N(0,1) \) where \( \circ \) denotes the entry-wise or Schur product on \( \mathbb{R}^{n \times n} \) and \( x \in \{-1,1\}^n \) is chosen uniformly at random. Therefore the data processing inequality implies that

\[
d_{TV} \left( \mathcal{L} \left( xx^T \circ \phi_1(G(n,1/2)) \right), \mathcal{L}(W) \right) = d_{TV} \left( \mathcal{L} \left( xx^T \circ \phi_1(G(n,1/2)) \right), \mathcal{L} \left( xx^T \circ W \right) \right) \\
\leq d_{TV} \left( \phi_1(G(n,1/2)), \mathcal{L}(W) \right) = O(n^{-1})
\]

Now note that \( \phi(G(n,1/2)) \) is distributed as \( \tau \left( xx^T \circ \phi_1(G(n,1/2)) \right) \) and that \( \tau(N(0,1)^{\otimes n \times n}) \) is distributed as \( G(n,1/2) \). It follows by the data processing inequality that

\[
d_{TV} \left( \phi_1(G(n,1/2)), G(n,1/2) \right) = d_{TV} \left( \tau \left( xx^T \circ \phi_1(G(n,1/2)) \right), \tau(W) \right) = O(n^{-1})
\]

Repeating the analysis in Lemmas 7.1.2 and 7.2.1 without converting the prior in Lemma 7.1.1 to be over unit vectors yields that there is a prior \( \pi \) on vectors in \( u \in \mathbb{Z}^n \) such that

\[
d_{TV} \left( \pi_1(G(n,k,1/2)), \int \mathcal{L} \left( \frac{\mu(k-1)}{2^{t+1} \sqrt{n-1}}, uu^T + N(0,1)^{\otimes n \times n} \right) d\pi(u) \right) = O \left( n^{-1} \right)
\]

and such that with probability at least \( 1 - 8k^{-1} \) it holds that

\[
2^t k \geq \|u\|_0 \geq 2^t k \left( 1 - \max \left( \frac{2C \ell \cdot \log(2^t k)}{k}, \frac{2^t k}{n} \right) \right) = 2^t k - O \left( 2^t \log n^2 + (2^t k)^{1-2\delta} \right)
\]

where \( C > 0 \) is the constant in Lemma 7.1.1. Now let \( \pi' \) be the prior \( \pi \) conditioned on the inequality above. It follows by Lemma 3.0.4 that \( d_{TV}(\pi, \pi') \leq 8k^{-1} \). Now let \( \beta = \frac{\mu(k-1)}{2^{t+1} \sqrt{n-1}} \) and let the matrix \( M' \) be distributed as

\[
M' \sim \int \mathcal{L} \left( \beta \cdot uu^T + N(0,1)^{\otimes n \times n} \right) d\pi(u)
\]

As above, let \( x \in \{-1,1\}^n \) be chosen uniformly at random. The same argument above shows that

\[
\mathcal{L} \left( xx^T \circ M' \big| x \circ u = v \right) = \mathcal{L} \left( \beta \cdot vv^T + N(0,1)^{\otimes n \times n} \right) = \mathcal{L}(M(\beta, v))
\]

As shown above, this implies that

\[
\mathcal{L} \left( \tau \left( xx^T \circ M' \right) \big| x \circ u = v \right) \in G_B \left( n, 2^t k, 1/2, \Phi(\beta) - 1/2 \right)
\]

as long as \( 2^{t-1} k - (2^t k)^{1-\delta} \leq |A(v)|, |B(v)| \leq 2^{t-1} k + (2^t k)^{1-\delta} \). Let \( \pi''(x,u) \) be the product distribution of \( \mathcal{L}(x) \) and \( \pi' \) conditioned on the event that these inequalities hold for \( v = x \circ u \). Now note that conditioning on \( u \) yields that \( |A(x \circ u)| + |B(x \circ u)| = \|u\|_0 \) and \( |A(x \circ u)| \) is distributed as \( \text{Bin}(\|u\|_0, 1/2) \). By Hoeffding's inequality, we have that conditioned on \( u \),

\[
P \left[ \left| A(x \circ u) - \frac{1}{2} \|u\|_0 \right| > \sqrt{\|u\|_0 \log k} \right] \leq \frac{2}{k^2}
\]
Note that if this inequality holds for \(|A(x \circ u)|\), then it also holds for \(|B(x \circ u)|\) since these sum to \(|u|_0\). Therefore with probability at least \(1 - 2k^{-2}\), it holds that
\[
|A(x \circ u)| - 2^{k-1} k \leq \frac{1}{2} |u|_0 + O\left(2^k (\log n)^2 + (2^k)^{1-\delta}\right)
\]
for sufficiently large \(k\) and the same inequalities hold for \(|B(x \circ u)|\). This verifies that the desired inequalities on \(|A(x \circ u)|\) and \(|B(x \circ u)|\) hold with probability at least \(1 - 2k^{-2}\) over \(L(x) \otimes \pi(u)\). Applying Lemma 3.0.4 therefore yields that \(d_{TV}(L(x) \times \pi'(u), \pi''(x, u)) \leq 2k^{-2}\). The data processing and triangle inequalities then imply that
\[
d_{TV}(L(x) \otimes \pi(u), \pi''(x, u)) \leq d_{TV}(L(x) \otimes \pi(u), L(x) \otimes \pi'(u)) + d_{TV}(L(x) \otimes \pi'(u), \pi''(x, u))
\]
Now observe that
\[
d_{TV}(\phi(G(n, k, 1/2), \int L(\tau xx^T \circ M') \| x, u) d\pi''(x, u))
\]
\[
\leq d_{TV}(\int L(\tau xx^T \circ M') \| x, u) d\pi''(x, u)
\]
\[
\leq d_{TV}(\int L(\tau xx^T \circ \phi_1(G(n, k, 1/2))) \| x, u) d\pi''(x, u)
\]
\[
= O(n^{-1}) + O(k^{-1}) = O(k^{-1})
\]
By the definition of \(\pi''\), the distribution
\[
L_{SSBM} = \int L(\tau xx^T \circ M') \| x, u) d\pi''(x, u)
\]
is a mixture of distributions in \(G_B(n, 2^k, 1/2, \Phi(\beta) - 1/2)\), completing the proof of the lemma. 

Applying this reduction and setting parameters similarly to Theorem 7.2.1 yields the following computational lower bound for the subgraph stochastic block model.

**Theorem 7.3.1.** Let \(\alpha \in [0, 2)\) and \(\beta \in (\delta, 1 - 3\delta)\) be such that \(\beta < \frac{1}{2} + \alpha\). There is a sequence \(\{(N_n, K_n, q_n, \rho_n)\}_{n \in \mathbb{N}}\) of parameters such that:

1. The parameters are in the regime \(q = \Theta(1)\), \(\rho = \tilde{\Theta}(N^{-\alpha})\) and \(K = \tilde{\Theta}(N^\beta)\) or equivalently,
\[
\lim_{n \to \infty} \frac{\log \rho_n^{-1}}{\log N_n} = \alpha, \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta \quad \text{and} \quad \lim_{n \to \infty} q_n = q
\]
2. For any sequence of randomized polynomial-time tests $\phi_n : G_{N_n} \rightarrow \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems $SSBM_D(N_n, K_n, q_n, \rho_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.

Therefore the computational boundary for $SSBM_D(n, k, q, \rho)$ in the parameter regime $q = \Theta(1)$, $\rho = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$.

Proof. When $2\alpha > \beta$, it holds that $SSBM_D$ is information-theoretically impossible. Therefore we may assume that $2\alpha \leq \beta$ and in particular that $\alpha < \frac{1}{2}$. Now suppose that $\beta < \frac{1}{2} + \alpha$. We consider the cases $\beta \geq \frac{1}{2}$ and $\beta < \frac{1}{2}$ and $q \leq \frac{1}{2}$ and $q > \frac{1}{2}$, separately. First consider the case when $\beta \geq \frac{1}{2}$ and $q \leq \frac{1}{2}$. Let $\epsilon = \frac{1}{2} (\alpha + \frac{1}{2} - \beta) \in (0, \frac{1}{2})$ and set parameters similarly to Theorem 7.2.1 with

$$
\ell_n = \left( \beta - \frac{1}{2} + \epsilon \right) \log_2 n,
$$

$$
k_n = \left\lceil n^{\frac{1}{2} - \epsilon} \right\rceil,
$$

$$
N_n = 2n,
$$

$$
K_n = 2^{\ell_n} k_n,
$$

$$
q_n = q
$$

where $\mu = \frac{\log 2}{2\sqrt{\log n + 2\log 2}}$. By Lemma 7.3.1, there is a randomized polynomial time algorithm mapping $PC_D(2n, k_n, 1/2)$ to the detection problem $SSBM_D(N_n, K_n, 1/2, (2q)^{-1}\rho_n)$ under $H_0$ and to a prior over $H_1$ with total variation converging to zero as $n \rightarrow \infty$. Now consider the algorithm that post-processes the resulting graph by keeping each edge independently with probability $2q$. This maps any instance of $SSBM_D(N_n, K_n, 1/2, (2q)^{-1}\rho_n)$ exactly to an instance of $SSBM_D(N_n, K_n, q, \rho_n)$. Therefore the data processing inequality implies that these two steps together yield a reduction that combined with Lemma 3.0.1 implies property 2 holds. Now observe that as $n \rightarrow \infty$,

$$
\rho_n = 2q \cdot \Phi \left( \frac{\mu(k_n - 1)}{2^{\ell_n+1}\sqrt{n - 1}} \right) - q
$$

where

$$
\mu \sim 2q \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\mu(k_n - 1)}{2^{\ell_n+1}\sqrt{n - 1}}
$$

The same limit computations as in Theorem 7.2.1 show that property 1 above holds. If $q > 1/2$, then instead set

$$
\rho_n = 2(1 - q) \cdot \Phi \left( \frac{\mu(k_n - 1)}{2^{\ell_n+1}\sqrt{n - 1}} \right) - (1 - q)
$$

and post-process the graph resulting from the reduction in Lemma 7.3.1 by adding each absent edge with probability $2q - 1$. By a similar argument, the resulting reduction shows properties 1 and 2.

Now consider the case when $\beta < \frac{1}{2}$, $\alpha > 0$ and $q \leq \frac{1}{2}$. Set $\ell_n = 0, k_n = \left\lceil n^{\frac{1}{2} - \epsilon} \right\rceil, K_n = k_n$ and

$$
N_n = 2 \left\lceil n^{\frac{1}{2} \left( \frac{1}{2} - \epsilon \right)} \right\rceil
$$

and

$$
\rho_n = 2q \cdot \Phi \left( \frac{\mu'(k - 1)}{2\sqrt{n - 1}} \right) - q
$$
where $\epsilon = \min\left(\frac{\alpha}{2(\alpha+\beta)}, \frac{1}{2} - \beta\right)$ and $\mu' = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}} \cdot n^{e-\frac{3}{2}}$ as in the proof of Theorem 7.2.1. Note that since $\epsilon \leq \frac{1}{2} - \beta$, it follows that $N_n \geq 2n$. By Lemma 7.3.1, there is a randomized polynomial time algorithm mapping $PCD(2n, k_n, 1/2)$ to the detection problem $SSBM_D(2n, K_n, 1/2, (2q)^{-1}\rho_n)$ under $H_0$ and to a prior over $H_1$ with total variation converging to zero as $n \to \infty$. Now consider the map that post-processes the graph by:

1. keeping each edge independently with probability $2q$;
2. adding $N_n - 2n$ vertices to the resulting graph and includes each edge incident to these new vertices independently with probability $q$; and
3. randomly permuting the vertices of the resulting $N_n$-vertex graph.

Note that this maps any instance of $SSBM_D(2n, K_n, 1/2, (2q)^{-1}\rho_n)$ exactly to an instance of $SSBM_D(N_n, K_n, q, \rho_n)$. Thus the data processing inequality implies that this post-processing together with the reduction of Lemma 7.3.1 yields a reduction showing property 2. The same limit computations as in Theorem 7.2.1 show that property 1 above holds. The same adaptation as in the case $\beta \geq \frac{1}{2}$ also handles $q < \frac{1}{2}$. This completes the proof of the theorem. \qed
Chapter 8

Random Rotations and Sparse PCA

In this chapter, we deduce tight lower bounds for detection in sparse PCA when $k \gg \sqrt{n}$. We also show a suboptimal lower bound when $k \ll \sqrt{n}$ matching the results of [BR13a] and [GMZ17]. In both cases, we reduce from biclustering and rank-1 submatrix to biased and unbiased variants of sparse PCA. The reductions in this chapter are relatively simple, with most of the work behind the lower bounds we show residing in the reductions from PC to biclustering and rank-1 submatrix. This illustrates the usefulness of using natural problems as intermediates in constructing average-case reductions. The lower bounds we prove make use of the following theorem of Diaconis and Freedman showing that the first $m$ coordinates of a unit vector in $n$ dimensions where $m < n$ are close to independent Gaussians in total variation [DF87].

**Theorem 8.0.2** (Diaconis and Freedman [DF87]). Suppose that $(v_1,v_2,\ldots,v_n)$ is uniformly distributed according to the Haar measure on $S^{n-1}$. Then for each $1 \leq m \leq n-4$,

$$d_{TV}(\mathcal{L}(v_1,v_2,\ldots,v_m), N(0,n^{-1})^\otimes m) \leq \frac{2(m+3)}{n-m-3}$$

The next lemma is the crucial ingredient in the reductions of this chapter. Note that the procedure **RANDOM-ROTATION** in Figure 8.2 requires sampling the Haar measure on $O_{rn}$. This can be achieved efficiently by iteratively sampling a Gaussian, projecting it onto the orthogonal complement of the vectors chosen so far, normalizing it and adding it to the current set. Repeating this for $n$ iterations yields a potential implementation for the sampling part of Step 2 in Figure 8.2. In the next lemma, $N(0, I_m + \theta vv^\top^\otimes n$ is taken to denote an $m \times n$ matrix with its $n$ columns sampled i.i.d. from $N(0, I_m + \theta vv^\top)$.

**Lemma 8.0.2** (Random Rotations). Let $\tau : \mathbb{N} \to \mathbb{N}$ be an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. Consider the map $\phi : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ that sends $M$ to **RANDOM-ROTATION**(M, $\tau$). It
CHAPTER 8. RANDOM ROTATIONS AND SPARSE PCA

Algorithm RANDOM-ROTATION

Inputs: Matrix $M \in \mathbb{R}^{m \times n}$, parameter $r \in \mathbb{N}$

1. Construct the $m \times \tau n$ matrix $M'$ such that the leftmost $m \times n$ submatrix of $M'$ is $M$ and the remaining entries of $M'$ are sampled i.i.d. from $N(0, 1)$

2. Let $R$ be the leftmost $\tau n \times n$ submatrix of a random orthogonal matrix sampled from the normalized Haar measure on the orthogonal group $O_{\tau n}$

3. Output $M'R$

Figure 8.1: Random rotation procedure in Lemma 8.0.2.

follows that $\phi(N(0, 1)^{\otimes m \times n}) \sim N(0, 1)^{\otimes m \times n}$ and for any unit vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^n$ we have that

$$d_{TV}(\phi(\lambda \cdot uw^T + N(0, 1)^{\otimes m \times n}), N\left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^T\right)^{\otimes n}) \leq \frac{2(n + 3)}{\tau n - n - 3}$$

Proof. Let $R' \in O_{\tau n}$ be the original $\tau n \times \tau n$ sampled in Step 2 of RANDOM-ROTATION and let $R$ be its upper $\tau n \times n$ submatrix. Let $M$ and $M'$ be the matrices input to Random-Rotation and computed in Step 1, respectively, as shown in Figure 8.2. If $M \sim N(0, 1)^{\otimes m \times n}$, then it follows that $M' \sim N(0, 1)^{\otimes m \times \tau n}$. Since the rows of $M'$ are independent and distributed according to the isotropic distribution $N(0, 1)^{\otimes \tau n}$, multiplication on the right by any orthogonal matrix leaves the distribution of $M'$ invariant. Therefore $M'R' \sim N(0, 1)^{\otimes m \times \tau n}$ and since $M'R$ consists of the first $n$ columns of $M'R'$, it follows that $\phi(M) = M'R \sim N(0, 1)^{\otimes m \times n}$.

Now suppose that $M$ is distributed as $\lambda \cdot uw^T + N(0, 1)^{\otimes m \times n}$ for some unit vectors $u, v$. Let $v'$ be the unit vector in $\mathbb{R}^{\tau n}$ formed by appending $\tau n - n$ zeros to the end of $v$. It follows that $M'$ is distributed as $\lambda \cdot uv'^T + N(0, 1)^{\otimes m \times \tau n}$. Let $M' = \lambda \cdot uv'^T + W$ where $W \sim N(0, 1)^{\otimes m \times \tau n}$. Now let $W' = WS^{-1}$ where $S \in O_{\tau n}$ is sampled according to the Haar measure and independently of $R'$. Now note that

$$M'R' = \lambda \cdot u \cdot (R'^T v')^T + W'SR'$$

Now note that conditioned on $R'$, the product $SR'$ is distributed according to the Haar measure on $O_{\tau n}$. This implies that $SR'$ is independent of $R'$. Therefore $W'SR'$ is independent of $R'$ and distributed according to $N(0, 1)^{\otimes m \times \tau n}$. This also implies that $R'^T v'$ is independent of $W'SR'$ and distributed uniformly over $S^{\tau n-1}$. Let $r \in \mathbb{R}^n$ denote the vector consisting of the first $n$ coordinates
of $R^T v'$ and let $W''$ denote the $m \times n$ matrix consisting of the first $n$ columns of $W'SR'$. It follows that $\phi(M) = M'R = \lambda \cdot ur^T + W''$ where $r$ and $W''$ are independent. Now let $g \in \mathbb{R}^n$ be a Gaussian vector with entries i.i.d. sampled from $N(0,n^{-1})$. Also let $Z = \lambda \cdot ug^T + W''$ and note that by Diaconis-Freedman's theorem and coupling the noise terms $W''$, the data processing inequality implies

$$d_{TV}(\mathcal{L}\left(\lambda \cdot ug^T + W''\right), \mathcal{L}\left(\lambda \cdot ur^T + W''\right)) \leq d_{TV}(\mathcal{L}(r), \mathcal{L}(g)) \leq \frac{2(n+3)}{\tau n - n - 3}$$

Now note that since the entries of $g$ are independent, the matrix $\lambda \cdot ug^T + W''$ has independent columns. Its ith row has jointly Gaussian entries with covariance matrix

$$E\left[(\lambda \cdot ug_i + W_i)(\lambda \cdot ug_i + W_i)^T\right] = E\left[\lambda^2 \cdot uu^T g_i^2 + \lambda \cdot g_i \cdot uW_i^T + \lambda \cdot g_i \cdot W_i u^T + W_i W_i^T\right]$$

$$= \frac{\lambda^2}{\tau n} \cdot uu^T + I_m$$

Therefore $\lambda \cdot ug^T + W'' \sim N\left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^T\right)^{\otimes n}$. Combining these results yields that

$$d_{TV}\left(\phi(M), N\left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^T\right)^{\otimes n}\right) \leq \frac{2(n+3)}{\tau n - n - 3}$$

which completes the proof of the lemma. □

Applying reflection cloning to an instance of rank-1 submatrix and then randomly rotating to obtain an instance of sparse PCA yields the following reduction. This establishes tight lower bounds in the regime $k \gg \sqrt{n}$. This reduction is stated in the next lemma, which maps to an instance of sparse PCA with $\theta = \frac{\mu k^2}{2 \tau n}$ where $\tau$ is an arbitrarily slow-growing function of $n$.

**Lemma 8.0.3.** Suppose that $n$ and $\ell$ are such that $\ell = O(\log n)$ and are sufficiently large, $\mu \leq \frac{1}{2\sqrt{6 \log n + 2 \log 2}}$ and $\tau : \mathbb{N} \to \mathbb{N}$ is an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. The map $\phi =$ SPFA-HIGH-SPARSITY is a randomized polynomial time computable map $\phi : G_n \to \mathbb{R}^{n \times n}$ such that if $G$ is an instance of PC($n, k, 1/2$) then under $H_0$, it holds that $\phi(G) \sim N(0,1)^{\otimes n \times n}$ and, under $H_1$, there is a prior $\pi$ such that

$$d_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2 \tau n} \cdot uu^T\right)^{\otimes n} d\pi(u)\right) \leq \frac{2(n+3)}{\tau n - n - 3} + O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

where $\pi$ is supported on pairs of unit vectors in $V_{n, 2^\ell k}$.

**Proof.** Let $M'$ be the matrix output in Step 2. Under $H_0$, Lemma 7.1.2 implies that $M \sim N(0,1)^{\otimes n \times n}$ and Lemma 8.0.2 implies that we also have $M' \sim N(0,1)^{\otimes n \times n}$. It suffices to consider $H_1$. By Lemma 7.1.2,

$$d_{TV}\left(\mathcal{L}_{H_1}(M), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uu^T + N(0,1)^{\otimes n \times n}\right) d\pi'(u, v)\right) = O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$
**Algorithm SPCA-High-Sparsity**

*Inputs:* Graph $G \in \mathcal{G}_n$, number of iterations $\ell$, function $\tau : \mathbb{N} \to \mathbb{N}$ with $\tau(n) \to \infty$

1. Compute the output $M$ of ROS-REDUCTION applied to $G$ with $\ell$ iterations
2. Output the matrix returned by RANDOM-ROTATION($M, \tau$)

**Algorithm SPCA-Low-Sparsity**

*Inputs:* Graph $G \in \mathcal{G}_n$, number of iterations $\ell$, function $\tau : \mathbb{N} \to \mathbb{N}$ with $\tau(n) \to \infty$

1. Compute the output $M$ of BC-REDUCTION applied to $G$ with $\ell$ iterations
2. Output the matrix returned by RANDOM-ROTATION($M, \tau$)

**Algorithm SPCA-Recovery**

*Inputs:* Graph $G \in \mathcal{G}_n$, density bias $\rho$

1. Let $M$ be the output of BC-RECOVERY applied to $G$ with density bias $\rho$
2. Output the matrix RANDOM-ROTATION($M, \tau$)

**Figure 8.2:** Reductions to SPCA$_D$ when $k \gtrsim \sqrt{n}$ and $k \lesssim \sqrt{n}$ in Lemmas 8.0.3 and 8.0.4 and reduction to SPCA$_R$ in Theorem 8.0.5.

where $\pi'(u, v)$ is a prior supported on pairs of unit vectors in $\mathbb{V}_{n,2^\ell k}$. Let $W \sim \frac{\mu k}{\sqrt{2}} uv^\top + N(0, 1)^{\otimes n \times n}$ where $(u, v)$ is distributed according to $\pi'$ and let $\varphi$ denote the map in Step 2 taking $A$ to RANDOM-ROTATION($A, \tau$). Conditioning on $(u, v)$ yields by Lemma 8.0.2 that

$$d_{TV}\left(L(\varphi(W)|u, v), N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n}\right) \leq \frac{2(n + 3)}{\tau n - n - 3}$$

Now consider the measure $\pi(u) = \mathbb{E}_v \pi'(u, v)$ given by marginalizing over $v$ in $\pi'$. The triangle
inequality implies that
\[
\text{d}_{TV}\left(\mathcal{L}(\phi(W)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right) \otimes^n d\pi(u)\right)
\leq \mathbf{E}_{u,v} \left[\text{d}_{TV}\left(\mathcal{L}(\phi(W)|u,v), N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right) \otimes^n\right)\right] \leq \frac{2(n+3)}{\tau n - n - 3}
\]
By the data processing inequality and triangle inequality, we now have that
\[
\text{d}_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right) \otimes^n d\pi(u)\right)
\leq \text{d}_{TV}(\mathcal{L}_{H_1}(M), \mathcal{L}(W)) + \text{d}_{TV}\left(\mathcal{L}(\phi(W)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right) \otimes^n d\pi(u)\right)
\leq \frac{2(n+3)}{\tau n - n - 3} + O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)
\]
since \(\mathcal{L}_{H_1}(\phi(G)) \sim \mathcal{L}_{H_1}(\phi(M))\). This completes the proof of the lemma.

The next lemma maps from planted clique to an instance of biclustering and then to sparse PCA. This reduction shows hardness for the simple vs. simple hypothesis testing formulations of sparse PCA. In particular, the output in Lemma 8.0.2 is close in total variation to the simple vs. simple model UBSPCA. After multiplying the rows of the matrix output in Lemma 8.0.2 by ±1, each with probability 1/2, this also yields a reduction to USPCA. The next lemma can be proven with the same argument as in 8.0.3 using the total variation bounds in Lemma 6.2.2.

**Lemma 8.0.4.** Suppose that \(n\) and \(\ell\) are such that \(\ell = O(\log n)\) and are sufficiently large, \(\mu \leq \frac{\log^2}{2\sqrt{\log n + 2\log^2}}\) and \(\tau : \mathbb{N} \rightarrow \mathbb{N}\) is an arbitrary function with \(\tau(n) \rightarrow \infty\) as \(n \rightarrow \infty\). Then \(\phi = \text{SPCA-LOW-SPARSITY}\) is a randomized polynomial time computable map \(\phi : G_n \rightarrow \mathbb{R}^{2^n \times 2^n}\) such that if \(G\) is an instance of PC\((n, k, 1/2)\) then under \(H_0\), it holds that \(\phi(G) \sim N(0,1)^{\otimes 2^n \times 2^n}\) and
\[
\text{d}_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n+1} \cdot 1_S 1_S^\top\right) \otimes^n d\pi(S)\right) \leq \frac{2(2^n n + 3)}{\tau \cdot 2^n n - 2^n n - 3} + O\left(\frac{1}{\sqrt{\log n}}\right)
\]
where \(\pi\) is the uniform distribution over all ordered pairs of size \(2^{\ell} k\) elements subsets of \([2^n]\).

We now apply these reductions to deduce planted clique hardness for sparse PCA and its variants. Note that when \(k \ll \sqrt{n}\), the lower bounds for sparse PCA are not tight. For biased sparse PCA, the lower bounds are only tight at the single point when \(\theta = \hat{\Theta}(1)\). The next theorem deduces hardness for SPCA\(_D\) when \(k \gtrsim \sqrt{n}\) with Lemma 8.0.3.
Theorem 8.0.3. Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\alpha > \max(1 - 2\beta, 0)$. There is a sequence $(N_n, K_n, d_n, \theta_n)_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $d = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

$$
\lim_{n \to \infty} \log \frac{\theta_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \log \frac{K_n}{\log N_n} = \beta
$$

2. For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic $Type I+II$ error of $\phi_n$ on the problems $\text{SPCAD}(N_n, K_n, d_n, \theta_n)$ is at least 1 assuming the PC conjecture holds for $p = 1/2$.

Therefore the computational boundary for $\text{SPCAD}(n, k, d, \theta)$ in the parameter regime $d = \Theta(n)$, $\theta = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\alpha^* = \max(1 - 2\beta, 0)$.

Proof. Note that if $\alpha \geq 1$, then sparse PCA is information theoretically impossible. Thus we may assume that $\alpha \in (0, 1)$ and $\beta > \frac{1-\alpha}{2}$. Let $\gamma = \frac{1-\alpha}{2} \in (0, 1/2)$. Now set $N_n = d_n = n$

$$
\ell_n = \left[ \left( \beta - \frac{1-\alpha}{2} \right) \log_2 n \right], \quad k_n = \lceil n^{\gamma} \rceil, \quad K_n = 2^{\ell_n} k_n, \quad \theta_n = \frac{\mu^2 k_n^2}{2\tau n}
$$

where $\mu = \frac{\log 2}{2\sqrt{6} \log n + 2 \log 2}$ and $\tau$ is a sub-polynomially growing function of $n$. By Lemma 8.0.3, there is a randomized polynomial time algorithm mapping $\text{PCD}(n, k_n, 1/2)$ to the detection problem $\text{SPCAD}(N_n, K_n, d_n, \theta_n)$, exactly under $H_0$ and to a prior over $H_1$, with total variation converging to zero as $n \to \infty$. This map with Lemma 3.0.1 now implies that property 2 above holds. We now verify property 1. Note that

$$
\lim_{n \to \infty} \log \frac{K_n}{\log N_n} = \lim_{n \to \infty} \left[ \left( \beta - \frac{1-\alpha}{2} \right) \log_2 n \right] \cdot \log 2 + \left( \frac{1-\alpha}{2} \right) \log n = \beta
$$

$$
\lim_{n \to \infty} \log \frac{\theta_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{1 - 2\gamma - 2\log \mu + \log(2\tau)}{\log n} = \alpha
$$

which completes the proof. \qed

Similarly, varying the parameters $\ell$ and $k$ in Lemma 8.0.4 yields the following hardness for the simple vs. simple hypothesis testing formulations of biased and ordinary sparse PCA.

Theorem 8.0.4. Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\frac{1-\alpha}{2} < \beta < \frac{1+\alpha}{2}$. There is a sequence $(N_n, K_n, d_n, \theta_n)_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $d = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

$$
\lim_{n \to \infty} \log \frac{\theta_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \log \frac{K_n}{\log N_n} = \beta
$$
2. For any sequence of randomized polynomial-time tests $\phi_n : G_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of $\phi_n$ on the problems USPCAD($N_n, K_n, d_n, \theta_n$) and UBSPCA$_D$($N_n, K_n, d_n, \theta_n$) is at least 1 assuming the PC conjecture holds for $p = 1/2$.

Therefore the computational boundary for USPCAD($n, k, d, \theta$) and UBSPCA$_D$($N_n, K_n, d_n, \theta_n$) in the parameter regime $d = \Theta(n)$, $\theta = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\alpha^* = |1 - 2\beta|$.

To conclude this section, we observe that the reduction SPCA-RECOVERY shows that recovery in UBSPCA$_R$ is hard if $\theta \ll 1$ given the PDS recovery conjecture. The proof of the following theorem follows the same structure as Lemma 8.0.3 and Theorem 6.2.3. Note that BC-RECOVERY approximately maps from PDS$_R$($n, k, 1/2 + \rho, 1/2$) to BC$_R$($n, k, \mu$) where $\mu = \frac{\log(1+2\rho)}{2\sqrt{6\log n} + 2\log 2} = \tilde{\Theta}(\rho)$. Furthermore, this map preserves the support of the planted dense subgraph in the row support of the planted matrix in BC$_R$. Then RANDOM-ROTATIONS approximately maps from this BC$_R$ instance to a UBSPCA$_R$($n, k, n, \theta$) instance with $\theta = \frac{k^2\mu^2}{\mu} = \tilde{\Theta} \left( \frac{k^2\mu^2}{\mu} \right)$. This map ensures that the planted vector $u$ is supported on the same indices as the original PDS$_R$ instance. Furthermore, the PDS conjecture is that the original PDS$_R$ instance is hard if $\rho^2 \ll \frac{n}{k^2}$ which corresponds to the barrier $\theta \ll 1$ under this reduction. Therefore we obtain the following theorem.

**Theorem 8.0.5.** Let $\alpha \in \mathbb{R}$ and $\beta \in (0, 1)$. There is a sequence $\{(N_n, K_n, D_n, \theta_n)\}_{n \in \mathbb{N}}$ of parameters such that:

1. The parameters are in the regime $D = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

   \[
   \lim_{n \to \infty} \frac{\log \theta_n^{-1}}{\log N_n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta
   \]

2. If $\alpha > 0$ and $\beta > \frac{1}{2}$, then the following holds. Let $\epsilon > 0$ be fixed and let $X_n$ be an instance of UBSPCA$_R$($N_n, K_n, D_n, \theta_n$). There is no sequence of randomized polynomial-time computable functions $\phi_n : \mathbb{R}^{D_n \times N_n} \to \binom{[N_n]}{k}$ such that for all sufficiently large $n$ the probability that $\phi_n(X_n)$ is exactly the pair of latent row and column supports of $X_n$ is at least $\epsilon$, assuming the PDS recovery conjecture.

Therefore, given the PDS recovery conjecture, the computational boundary for UBSPCA$_R$($n, k, d, \theta$) in the parameter regime $\theta = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\alpha^* = 0$ when $\beta > \frac{1}{2}$. 
CHAPTER 8. RANDOM ROTATIONS AND SPARSE PCA
Chapter 9

Algorithms and Information-Theoretic Thresholds

In this chapter, we give the algorithms and information-theoretic lower bounds necessary to prove Theorem 2.3.1. Specifically, for each problem, we give an information-theoretic lower bound, an inefficient algorithm that achieves the information-theoretic lower bound and a polynomial-time algorithm. As the computational lower bounds and reductions previously presented are the main novel contribution of the paper, the details in this chapter are succinctly presented only as needed for Theorem 2.3.1.

Many of the problems we consider have pre-existing algorithms and information-theoretic lower bounds. In these cases, we cite the relevant literature and state the results needed for Theorem 2.3.1. Note that we only require algorithms and lower bounds optimal up to sub-polynomial factors for Theorem 2.3.1. For some problems, we only give an information-theoretic lower bound for detection and show that this implies the recovery lower bound in the next section.

9.1 Biclustering, Planted Dense Subgraph and Independent Set

Information-Theoretic Lower Bounds The information-theoretic lower bound for $BC_D$ was shown in [BI13]. More precisely, they showed the following theorem re-written in our notation. Note that they showed a lower bound for a composite hypothesis testing version of $BC_D$, but took a uniform prior over the support of the hidden submatrix, matching our formulation.

**Theorem 9.1.1** (Theorem 2.2 in [BI13]). Suppose that $k, \mu$ are such that as $n \to \infty$, it holds that
CHAPTER 9. ALGORITHMS AND INFORMATION-THEORETIC THRESHOLDS

$k/n \to 0$ and one of the following holds

$$\frac{\mu k^2}{n} \to 0 \quad \text{and} \quad \lim sup_{n \to \infty} \frac{\mu}{2\sqrt{k^{-1}\log(n/k)}} < 1$$

Then if $M_n$ denotes an instance of $\text{BC}_D(n,k,\mu)$,

$$d_{TV}(\mathcal{L}_{H_0}(M_n),\mathcal{L}_{H_1}(M_n)) \to 0 \quad \text{as} \quad n \to \infty$$

This corresponds exactly to the information-theoretic barrier of $\mu \ll \frac{1}{\sqrt{k}}$ and $\mu \ll \frac{n}{k^2}$. We remark that the information-theoretic lower bounds for biclustering can also be deduced from the information-theoretic lower bounds for planted dense subgraph with $q = 1/2$ using the reduction from Lemma 6.2.3 and the data-processing inequality. Tight information-theoretic lower bounds for $\text{PDS}_D$ and $\text{PIS}_D$ can be deduced from a mild adaptation of [HWX15]. The argument from the proof of Proposition 3 in [HWX15] yields the following lemma.

**Lemma 9.1.1** (Proposition 3 in [HWX15]). If $G$ is an instance of $\text{PDS}_D(n,k,p,q)$, then

$$d_{TV}(\mathcal{L}_{H_0}(G),\mathcal{L}_{H_1}(G)) \leq \frac{1}{2} \sqrt{\mathbb{E}\left[\exp\left(\frac{(p-q)^2}{q(1-q)} \cdot H^2\right) - 1\right]}$$

where $H \sim \text{Hypergeometric}(n,k,k)$.

This lemma can be derived by combining the $\chi^2$ computation in the proof of Proposition 3 with Cauchy-Schwarz. In [HWX15], Proposition 3 is specifically for the case when $p = cq$ where $c > 1$ and also for a mixture over $\text{PDS}_D(n,K,p,q)$ where $K \sim \text{Bin}(n,k/n)$. However, the first step in the proof of Proposition 3 is to condition on $K$ and prove this bound for each fixed $K$. When combined with Lemma 14 from [HWX15], we obtain the desired information-theoretic lower bounds.

**Lemma 9.1.2** (Lemma 14 in [HWX15]). There is an increasing function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{x \to 0^+} \tau(x) = 1$ and

$$\mathbb{E}[\exp(\lambda H^2)] \leq \tau(b)$$

where $H \sim \text{Hypergeometric}(n,k,k)$, $\lambda = b \cdot \max\left\{\frac{1}{k} \log\left(\frac{en}{k}\right), \frac{n^2}{k^4}\right\}$ and $0 < b < (16e)^{-1}$.

Combining these two lemmas and setting $b$ as

$$b = \frac{(p-q)^2}{q(1-q)} \cdot \left(\max\left\{\frac{1}{k} \log\left(\frac{en}{k}\right), \frac{n^2}{k^4}\right\}\right)^{-1}$$

yields the following theorem on the information-theoretic lower bound for the general regime of $\text{PDS}_D$. 

Theorem 9.1.2. Suppose $p, q, k$ are such that as $n \to \infty$, it holds that $\frac{(p-q)^2}{q(1-q)} \ll \frac{1}{k}$ and $\frac{(p-q)^2}{q(1-q)} \ll \frac{n^2}{k^4}$. Then if $G_n$ is an instance of $\text{PDS}_D(n, k, p, q)$, it follows that

$$d_{TV} (\mathcal{L}_{H_0}(G_n), \mathcal{L}_{H_1}(G_n)) \to 0 \quad \text{as} \quad n \to \infty$$

Note that when $p = cq$ for some constant $c > 1$ or when $p = 0$, this barrier is $q \ll \frac{1}{k}$ and $q \ll \frac{n^2}{k^4}$. This recovers the information-theoretic lower bounds for $\text{PISD}$ and $\text{PDS}_D$ when $p = cq$. The information-theoretic lower bounds for the weak and strong recovery variants $\text{BCR}$ and $\text{PDSR}$ are derived in [HWX16b]. The following theorems of [HWX16b] characterize these lower bounds.

Theorem 9.1.3 (Corollary 2 in [HWX16b]). Suppose $k$ and $\mu$ are such that as $n \to \infty$,

$$k\mu^2 \to \infty \quad \text{and} \quad \liminf_{n \to \infty} \frac{(k-1)\mu^2}{\log \frac{n}{k}} > 4$$

then weak recovery in $\text{BCR}(n, k, \mu)$ is possible. If weak recovery is possible, then (9.1) holds as a non-strict inequality.

Theorem 9.1.4 (Corollary 4 in [HWX16b]). Suppose $k$ and $\mu$ are such that as $n \to \infty$, condition (9.1) holds and

$$\liminf_{n \to \infty} \frac{k\mu^2}{(\sqrt{2\log n} + \sqrt{2\log k})^2} > 1$$

then exact recovery in $\text{BCR}(n, k, \mu)$ is possible. If exact recovery is possible, then (9.1) and (9.2) hold as non-strict inequalities.

Theorem 9.1.5 (Corollary 1 in [HWX16b]). Suppose $p$ and $q$ are such that the ratios $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ are bounded as $n \to \infty$. If $k$ satisfies

$$k \cdot d_{KL}(p, q) \to \infty \quad \text{and} \quad \liminf_{n \to \infty} \frac{k \cdot d_{KL}(p, q)}{\log \frac{n}{k}} > 2$$

then there is an algorithm achieving weak recovery for $\text{PDSR}(n, k, p, q)$ is possible. If weak recovery is possible, then (9.3) holds as a non-strict inequality.

Theorem 9.1.6 (Corollary 3 in [HWX16b]). Suppose $p$ and $q$ are such that the ratios $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ are bounded as $n \to \infty$. If $k$ satisfies

$$\tau = \frac{\log\frac{1-q}{1-p} + \frac{1}{k} \log \frac{n}{k}}{\log \frac{p(1-q)}{q(1-p)}}$$

If (9.3) holds and

$$\liminf_{n \to \infty} \frac{k \cdot d_{KL}(\tau, q)}{\log n} > 1$$

then exact recovery in $\text{PDSR}(n, k, p, q)$ is possible. If exact recovery is possible, then (9.3) and (9.4) hold as non-strict inequalities.
These theorems show that both weak and strong recovery for \( BC_R(n, k, \mu) \) are information-theoretically impossible when \( \mu \lesssim \frac{1}{\sqrt{k}} \) by the first condition in (9.1). Now note that if \( p - q = O(q) \) and \( q \to 0 \) as \( n \to \infty \), then

\[
d_{KL}(p, q) = p \log \left( \frac{p}{q} \right) + (1 - p) \cdot \log \left( \frac{1 - p}{1 - q} \right)
\]

\[
= p \cdot \left( \frac{p - q}{q} \right) - O \left( p \cdot \left( \frac{p - q}{1 - q} \right)^2 \right) - (1 - p) \cdot \left( \frac{p - q}{1 - q} \right) - O \left( (1 - p) \cdot \left( \frac{p - q}{1 - q} \right)^2 \right)
\]

\[
= \frac{(p - q)^2}{q(1 - q)} + O \left( p \cdot \left( \frac{p - q}{q} \right)^2 + (p - q)^2 \right) = O \left( \frac{(p - q)^2}{q(1 - q)} \right)
\]

as \( n \to \infty \). Therefore it follows that if \( p - q = O(q) \), \( q \to 0 \) as \( n \to \infty \) and \( \log \frac{p}{q} \) and \( \log \frac{1 - p}{1 - q} \) are bounded as \( n \to \infty \) then both weak and strong recovery in \( PDS_R(n, k, p, q) \) are information-theoretically impossible if \( \frac{(p - q)^2}{q(1 - q)} \lesssim \frac{1}{k} \) by the first condition in (9.3). Note that these theorems of \([HWX16b]\) do not imply the necessary information-theoretic lower bound for \( PIS_R \) since \( p = 0 \) violates the condition that \( \log \frac{p}{q} \) is bounded. However, the genie argument in the necessary part of Theorem 1 in \([HWX16b]\) can be mildly adapted to obtain the following theorem.

**Theorem 9.1.7.** If \( k \geq 2 \), \( q \ll \frac{1}{k} \) and \( n - k = \Omega(n) \) as \( n \to \infty \), then weak recovery in \( PIS_R(n, k, q) \) is impossible.

**Proof.** Let \( G \sim G_1(n, k, q) \) where \( S \subseteq [n] \) denotes the indices of the planted independent set in \( G \) and satisfies \( |S| = k \). Fix a vertex \( i \in [n] \) and consider a random \( J \in [n] \) where \( J \) is chosen uniformly at random from \( S \) if \( i \not\in S \) and \( J \) is chosen uniformly at random from \( S^c \) if \( i \in S \). Now consider the binary hypothesis testing problem with observations \( (G, J, S\backslash\{i, J\}) \) and the task of distinguishing between \( H_0 : i \not\in S \) and \( H_1 : i \in S \).

Since \( S \) is chosen uniformly at random, it follows that the identity of the vertex \( J \) is uniformly at random chosen from \( [n] \backslash \{i\} \) and is independent of the event \( \{i \in S\} \). It also holds that conditioned on \( J \) and the outcome of the event \( \{i \in S\} \), the set \( S\backslash\{i, J\} \) is uniformly distributed on \((k - 1)\)-subsets of \( [n] \backslash \{i, J\} \). Therefore for any \( J \in [n] \backslash \{i\} \) and \((k - 1)\) subset \( S\backslash\{i, J\} \) of \( [n] \backslash \{i, J\} \), it holds that

\[
\frac{P[G, J, S\backslash\{i, J\} | i \in S]}{P[G, J, S\backslash\{i, J\} | i \not\in S]} = \frac{P[G, S\backslash\{i, J\} | i \in S]}{P[G, S\backslash\{i, J\} | i \not\in S]} = \frac{P[G | i \in S, J, S\backslash\{i, J\}]}{P[G | i \not\in S, J, S\backslash\{i, J\}]} = \prod_{k \in S\backslash\{i, J\}} \frac{(1 - A_{ik}) \cdot q^{A_{jk}} (1 - q)^{1 - A_{jk}}}{(1 - A_{ik}) \cdot q^{A_{jk}} (1 - q)^{1 - A_{ik}}}
\]

where \( A = A(G) \) is the adjacency matrix of \( G \). From this factorization, it follows that the vector \( v \) of values \( A_{ik} \) and \( A_{jk} \) for all \( k \in S \backslash \{i, J\} \) is therefore a sufficient statistic for this binary hypothesis.
testing problem. Furthermore, if \( i \in S \) then \( v \) has its first \( k - 1 \) coordinates equal to zero and its last \( k - 1 \) coordinates distributed as \( \text{Bern}(q)^{\otimes (k-1)} \). If \( i \not\in S \), then \( v \) has its first \( k - 1 \) coordinates distributed as \( \text{Bern}(q)^{\otimes (k-1)} \) and its last \( k - 1 \) coordinates equal to zero. Thus the given hypothesis testing problem is equivalent to testing between these two distributions. Note that

\[
\mathbb{P}_{H_0}[v = 0] = \mathbb{P}_{H_1}[v = 0] = (1 - q)^{k-1} \geq 1 - (k - 1)q \to 1 \quad \text{as } n \to \infty
\]

by Bernoulli's inequality if \( k \geq 2 \). Taking any coupling of \( \mathcal{L}_{H_0}(v) \) and \( \mathcal{L}_{H_1}(v) \) such that the events \( \{v = 0\} \) under \( H_0 \) and \( H_1 \) coincide yields that

\[
d_{TV}(\mathcal{L}_{H_0}(v), \mathcal{L}_{H_1}(v)) \leq 1 - (1 - q)^{k-1} \to 0 \quad \text{as } n \to \infty
\]

Now let \( p_{e_1} \) and \( p_{e_2} \) be the optimal Type I and Type II error probabilities. Note that the prior on the hypotheses \( H_0 : i \in S \) and \( H_1 : i \not\in S \) is \( \mathbb{P}[i \in S] = k/n. \) Let \( \mathcal{E} \) be the optimal average probability of testing error under this prior. Also note that \( p_{e_1} + p_{e_2} = 1 - d_{TV}(\mathcal{L}_{H_0}(v), \mathcal{L}_{H_1}(v)) \to 1 \) as \( n \to \infty. \)

Now assume for contradiction that there is some algorithm \( A : G_n \to \binom{[n]}{k} \) that achieves weak recovery with \( \mathbb{E}[|S \cap A(G)|] = k - o(k) \) as \( n \to \infty. \) It follows that

\[
k - \mathbb{E}[|S \cap A(G)|] = \sum_{i=1}^{n} \mathbb{P}[\mathbf{1}_{\{i \in A(G)\}} \neq \mathbf{1}_{\{i \in S\}}] \geq \sum_{i=1}^{n} \min_{\phi_i(G)} \mathbb{P}[\phi_i(G) \neq 1_{\{i \in S\}}]
\geq \sum_{i=1}^{n} \min_{\phi_i(G, J, S \setminus \{i, J\})} \mathbb{P}[\phi_i(G, J, S \setminus \{i, J\}) \neq 1_{\{i \in S\}}] = n\mathcal{E}
\]

The first minimum is over all functions \( \phi_i : G_n \to \{0, 1\} \) that only observe the graph \( G \), while the second minimum is over all functions \( \phi_i \) that also observe \( J \) and \( S \setminus \{i, J\}. \) From these inequalities, it must follow that \( \mathcal{E} = o(k/n) \) which implies the test achieving \( \mathcal{E} \) must have Type I and Type II errors that are \( o(1) \) as \( n \to \infty \) since \( 1 - k/n = \Omega(1) \). This implies that \( p_{e_1} + p_{e_2} = o(1), \) which is a contradiction. \( \square \)

### Information-Theoretically Optimal Algorithms

A corresponding algorithm achieving the information-theoretic lower bound for \( \text{BCD} \) was also shown in [BI13]. Their algorithm outputs the hypothesis \( H_1 \) if either the maximum sum over all \( k \times k \) submatrices of the input exceeds a threshold or if the total sum of the input exceeds another threshold. The guarantees of this algorithm are summarized in the following theorem.

**Theorem 9.1.8 (Theorem 2.1 in [BI13]).** Suppose that \( k, \mu \) are such that as \( n \to \infty, \) it holds that \( k/n \to 0 \) and one of the following holds

\[
\frac{\mu k^2}{n} \to \infty \quad \text{or} \quad \limsup_{n \to \infty} \frac{\mu}{2\sqrt{k^{-1}\log(n/k)}} > 1
\]
Then there is an algorithm solving $\text{BCD}(n, k, \mu)$ with Type I+II error tending to zero as $n \to \infty$.

A very similar algorithm is optimal for $\text{PDS}_D$. Generalizing the concentration bounds in the proof of Proposition 4 in [HWX15] to any $p, q$ with $p - q = O(q)$ and $q \to 0$ yields the following theorem.

**Theorem 9.1.9.** Suppose that $p$, $q$ and $k$ are such that $|p - q| = O(q)$, $q \to 0$ and

$$
\frac{(p - q)^2}{q(1 - q)} = \omega \left( \frac{n^2}{k^4} \right) \quad \text{or} \quad \frac{(p - q)^2}{q(1 - q)} = \omega \left( \frac{\log(n/k)}{k} \right)
$$

as $n \to \infty$. Then there is an algorithm solving $\text{PDS}_D(n, k, p, q)$ with Type I+II error tending to zero as $n \to \infty$.

**Proof.** First suppose that $p > q$. Let $G$ be an instance of $\text{PDS}_D(n, k, p, q)$. Note that under $H_0$, the edge count $|E(G)| \sim \text{Bin}(\binom{n}{2}, q)$ and under $H_1$, $|E(G)|$ is the independent sum of $\text{Bin}(\binom{n}{2} - \binom{k}{2}, q)$ and $\text{Bin}(\binom{k}{2}, p)$. By Bernstein’s inequality, we have that

$$
P_{H_0} \left[ |E(G)| > \binom{n}{2} q + \binom{k}{2} \cdot \frac{p - q}{2} \right] \leq \exp \left( -\frac{\binom{k}{2}^2 (p - q)^2 / 4}{2 \binom{n}{2} q + \binom{k}{2} \cdot (p - q) / 3} \right)
$$

By the multiplicative Chernoff bound, it follows that

$$
P_{H_1} \left[ |E(G)| \leq \binom{n}{2} q + \binom{k}{2} \cdot \frac{p - q}{2} \right] \leq \exp \left( -\frac{\binom{k}{2}^2 (p - q)^2 / 4}{2 \binom{n}{2} q + 2 \binom{k}{2} \cdot (p - q) / 3} \right)
$$

By Bernstein’s inequality, it follows that

$$
P_{H_0} \left[ X \geq \binom{k}{2} \cdot \frac{p + q}{2} \right] \leq \sum_{R \in \binom{[n]}{k}} P_{H_0} \left[ |E(G[R])| \geq \binom{k}{2} \cdot \frac{p + q}{2} \right]
$$

$$
\leq (\frac{en}{k})^k \exp \left( -\frac{\binom{k}{2}^2 (p - q)^2 / 4}{2 \binom{k}{2} q + \binom{k}{2} \cdot (p - q) / 3} \right)
$$

where the second inequality uses the fact that for any fixed $R$, $|E(G[R])| \sim \text{Bin}(\binom{k}{2}, q)$. Under $H_1$, it holds that if $S$ is the vertex set of the latent planted dense subgraph then $|E(G[S])| \sim \text{Bin}(\binom{k}{2}, p)$. 


By the multiplicative Chernoff bound, it follows that
\[
P_{H_1} \left[ X < \binom{k}{2} \cdot \frac{p + q}{2} \right] \leq P_{H_1} \left[ |E(G[S])| < \binom{k}{2} \cdot \frac{p + q}{2} \right]
\leq \exp \left( -\frac{\binom{k}{2}^2 (p - q)^2 / 4}{2 \binom{k}{2} p} \right)
= \exp \left( -\Omega \left( k^2 \cdot \frac{(p - q)^2}{q(1 - q)} \right) \right)
\]

Therefore the test that outputs \( H_1 \) if \(|E(G)| > \binom{n}{2} q + \binom{k}{2} \cdot \frac{p - q}{2} \) or \( X \geq \binom{k}{2} \cdot \frac{p + q}{2} \) and \( H_0 \) otherwise has Type I+II error tending to zero as \( n \to \infty \) if one of the two given conditions holds. In the case when \( p < q \), this test with inequalities reversed can be shown to have Type I+II error tending to zero as \( n \to \infty \) by analogous concentration bounds.

This theorem gives the necessary algorithm matching the information-theoretic lower bound for PDSD in the general regime \( p - q = O(q) \), including \( p = cq \) for some constant \( c > 1 \). The algorithm needed for PISD can be obtained by setting \( p = 0 \) in this theorem. An algorithm matching the information-theoretic lower bound for BCR follows from Theorem 9.1.4, which asserts that exact recovery is possible as long as
\[
\mu > (1 + \epsilon) \cdot \frac{\sqrt{2 \log n} + \sqrt{2 \log k}}{\sqrt{k}}
\]
for some fixed \( \epsilon > 0 \). Specializing Corollary 2.4 in [CX16] to the case of \( r = 1 \) clusters yields an analogous algorithm for PDSR.

**Theorem 9.1.10** (Corollary 2.4 in [CX16]). Suppose that \( p, q \) and \( k \) are such that \( p > q \) and
\[
\frac{(p - q)^2}{q(1 - q)} \geq \frac{C \log n}{k}, \quad q \geq \frac{C \log k}{k} \quad \text{and} \quad kq \log \frac{p}{q} \geq C \log n
\]
for some sufficiently large constant \( C > 0 \). Then the maximum likelihood estimator for the planted dense subgraph in \( PDSD(n, k, p, q) \) solves strong recovery with error probability tending to zero.

This implies that if \( p > q \), \( p - q = O(q) \), \( q \to 0 \) and \( \frac{(p - q)^2}{q(1 - q)} \geq \frac{1}{k} \) as \( n \to \infty \), then exact recovery is possible. Specializing the result to \( p = 1 \) and applying this algorithm to the complement graph of a PISR instance yields that there is an algorithm for PISR if \( q \geq \frac{1}{k} \). We remark that the necessary algorithm for PDSR can also be deduced from Theorem 9.1.6. However, the constraints that \( \log \frac{p}{q} \) and \( \log \frac{1 - p}{1 - q} \) must be bounded does not yield the desired algorithm for PISR.
**Polynomial-Time Algorithms**  The polynomial time algorithm matching our planted clique lower bound for $BC_D$ is another simple algorithm thresholding the maximum and sum of the input matrix. Given an instance $M$ of $BC_D(n, k, \mu)$, let $\max(M) = \max_{i,j \in [n]} M_{ij}$ and $\sum(M) = \sum_{i,j=1}^{n} M_{ij}$. Specializing Lemma 1 of [MW15b] to our setup yields the following lemma.

**Lemma 9.1.3** (Lemma 1 in [MW15b]). If $M$ is an instance of $BC_D(n, k, \mu)$ then

$$\mathbb{P}_{H_0}[\text{sum}(M) > \mu k^2 \frac{2}{2}] + \mathbb{P}_{H_1}[\text{sum}(M) \leq \mu k^2] \leq \exp \left( -\frac{\mu^2 k^4}{8n^2} \right)$$

If $c > 0$ is any absolute constant and $\tau = \sqrt{(4 + c) \log n}$, then

$$\mathbb{P}_{H_0}[\max(M) > \tau] + \mathbb{P}_{H_1}[\max(M) \leq \tau] \leq n^{-c/2} + \exp \left( -\frac{1}{2} |\mu - \tau| \right)$$

It follows that the algorithm that outputs $H_1$ if $\max(M) > 5 \log n$ or $\sum(M) > \mu k^2 \frac{2}{2}$ solves $BC_D$ with Type I+II error tending to zero as $n \to \infty$ if either $\mu \geq \sqrt{6 \log n}$ or $\mu = \omega \left( \frac{n}{k^2} \right)$. By Theorem 9.1.9, if $\frac{p - q}{q(1 - q)} = \omega \left( \frac{n}{k^2} \right)$ it follows that thresholding the number of edges of an instance $G$ of $PDS_D(n, k, p, q)$ has Type I+II error tending to zero as $n \to \infty$. Setting $p = 0$ recovers the computational barrier of $PIS_D$. Polynomial-time algorithms for the recovery variants of these problems were given in [CX16]. The following are three theorems of [CX16] written in our notation.

**Theorem 9.1.11** (Theorem 2.5 in [CX16]). A polynomial-time convex relaxation of the MLE solves exact recovery in $PDS_R(n, k, p, q)$ with error probability at most $n^{-10}$ if $p > q$ and

$$k^2(p - q) \geq C [p(1 - q)k \log n + q(1 - q)n]$$

where $C > 0$ is a fixed constant.

**Theorem 9.1.12** (Theorem 3.3 in [CX16]). A polynomial-time convex relaxation of the MLE solves exact recovery in $BC_R(n, k, \mu)$ with error probability at most $n^{-10}$ if

$$\mu^2 \geq C \left[ \frac{\log n}{k} + \frac{n}{k^2} \right]$$

where $C > 0$ is a fixed constant.

**Theorem 9.1.13** (Theorem 3.3 in [CX16]). A polynomial-time element-wise thresholding algorithm solves exact recovery in $BC_R(n, k, \mu)$ with error probability at most $n^{-3}$ if $\mu^2 \geq C \log n$ where $C > 0$ is a fixed constant.
Since $k \log n = \tilde{O}(n)$ and $p = O(q)$ if $p - q = O(q)$, the first theorem above implies that exact recovery is possible in polynomial time for the general regime of PDSR if $\frac{(p-q)^2}{(1-q)} \gg \frac{n}{k^2}$. Taking the complement graph of the input and setting $p = 1$ and $q = 1 - \tilde{O}(n^{-\alpha})$ in the first theorem yields that PISR$(n, k, 1 - q)$ can be solved in polynomial time if $1 - q \gg \frac{n}{k^2}$. The second and third theorems above imply that exact recovery for BC$_R$ is possible in polynomial time if $\mu \gg \frac{1}{\sqrt{k}}$ or $\mu \gg 1$. These polynomial-time algorithms for detection and recovery match the computational lower bounds shown in previous sections.

9.2 Rank-1 Submatrix, Sparse Spiked Wigner and Subgraph SBM

Information-Theoretic Lower Bounds  Applying a similar $\chi^2$ computation as in information-theoretic lower bounds for sparse PCA and planted dense subgraph, we can reduce showing an information-theoretic lower bound for SSWD to bounding an MGF. In the case of SSWD, this MGF turns out to be that of the square of a symmetric random walk on $\mathbb{Z}$ terminated after a hypergeometric number of steps. An asymptotically tight upper bound on this MGF was obtained in [CMW15] through the following lemma.

Lemma 9.2.1 (Lemma 1 in [CMW15]). Suppose that $d \in \mathbb{N}$ and $k \in [p]$. Let $B_1, B_2, \ldots, B_k$ be independent Rademacher random variables. Let the symmetric random walk on $\mathbb{Z}$ stopped at the $m$th step be

$$G_m = \sum_{i=1}^{m} B_i$$

If $H \sim \text{Hypergeometric}(d, k, k)$ then there is an increasing function $g : (0, 1/36) \rightarrow (1, \infty)$ such that $\lim_{x \rightarrow 0^+} g(x) = 1$ and for any $a \in (0, 1/36)$, it holds that

$$\mathbb{E} \left[ \exp \left( G_H^2 \cdot \frac{a}{k} \log \frac{ed}{k} \right) \right] \leq g(a)$$

With this bound, we obtain the following information-theoretic lower bound for SSWD, which matches Theorem 2.3.1.

Theorem 9.2.1. Suppose that $M$ is an instance of SSWD$(n, k, \mu)$ where under $H_1$, the planted vector $v$ is chosen uniformly at random from all $k$-sparse unit vectors in $\mathbb{R}^n$ with nonzero coordinates equal to $\pm \frac{1}{\sqrt{k}}$. Suppose it holds that $\mu \leq \sqrt{\beta_0 k \log \frac{nk}{k}}$ for some $0 < \beta_0 < (16e)^{-1}$. Then there is a function $w : (0, 1) \rightarrow (0, 1)$ satisfying that $\lim_{\beta_0 \rightarrow 0^+} w(\beta_0) = 0$ and

$$d_{TV}(\mathcal{L}_{H_0}(M), \mathcal{L}_{H_1}(M)) \leq w(\beta_0)$$
Proof. Let \( P_0 \) denote \( \mathcal{L}_{H_0}(M) = N(0,1)^{\otimes n \times n} \) and \( P_u \) denote \( \mathcal{L}(\mu \cdot uu^T + N(0,1)^{\otimes n \times n}) \) where \( u \) is in the set \( S \) of \( k \)-sparse unit vectors \( u \) with nonzero entries equal to \( \pm 1/\sqrt{k} \). Now let \( P_1 \) denote \( \mathcal{L}_{H_1}(M) \) which can also be written as

\[
P_1 = \frac{1}{|S|} \sum_{u \in S} P_u
\]

Given two matrices \( A, B \in \mathbb{R}^{n \times n} \), let \( \langle A, B \rangle = \sum_{i,j=1}^{n} A_{ij}B_{ij} \) denote their inner product. Now note that for any \( X \in \mathbb{R}^{n \times n} \),

\[
\frac{dP_u}{dP_0}(X) = \exp \left( -\frac{1}{2} \sum_{i,j=1}^{n} (X_{ij} - \mu \cdot u_i u_j)^2 + \frac{1}{2} \sum_{i,j=1}^{n} X_{ij}^2 \right) = \exp \left( \mu \cdot \langle X, uu^T \rangle - \frac{\mu^2}{2} \|u\|_2^4 \right) = \exp \left( \mu \cdot \langle X, uu^T \rangle - \frac{\mu^2}{2} \right)
\]

since \( \|u\|_2 = 1 \). Now observe that

\[
\chi^2(P_1, P_0) = E_{X \sim P_0} \left[ \left( \frac{dP_1}{dP_0}(X) - 1 \right)^2 \right] = -1 + \frac{1}{|S|^2} \sum_{u,v \in S} E_{X \sim P_0} \left[ \exp \left( \mu \cdot \langle X, uu^T + vv^T \rangle - \mu^2 \right) \right]
\]

\[
= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp \left( \frac{\mu^2}{2} \|uu^T + vv^T\|_F^2 - \mu^2 \right)
\]

\[
= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp \left( \frac{\mu^2}{2} \langle uu^T, uu^T \rangle + \mu^2 \langle vv^T, vv^T \rangle + \frac{\mu^2}{2} \langle vv^T, vv^T \rangle - \mu^2 \right)
\]

\[
= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp \left( \mu^2 \langle u, v \rangle^2 \right) = -1 + E_{u,v \sim \text{Unif}[S]} \left[ \exp \left( \mu^2 \langle u, v \rangle^2 \right) \right]
\]

where the third inequality follows since \( E[\exp (\langle t, X \rangle)] = \exp \left( \frac{1}{2} \|t\|_2^2 \right) \) and the last inequality follows since \( \langle uu^T, uu^T \rangle = \|u\|_2^4 = \|v\|_2^4 = 1 \) and \( \langle uu^T, vv^T \rangle = \langle u, v \rangle^2 \). Let \( G_m \) denote a symmetric random walk on \( \mathbb{Z} \) stopped at the \( m \)th step and \( H \sim \text{Hypergeometric}(n,k,k) \) as in Lemma 9.2.1. Now note that if \( u, v \sim \text{Unif}[S] \) are independent, then \( \langle u, v \rangle \) is distributed as \( G_{H/k} \). Now let \( a = \mu^2 \left( k \log \frac{en}{k} \right)^{-1} \leq \beta_0 \) and note that Lemma 9.2.1 along with Cauchy-Schwarz implies that

\[
d_{TV}(P_0, P_1) \leq \frac{1}{2} \sqrt{\chi^2(P_1, P_0)} \leq \frac{1}{2} \sqrt{g(\beta_0) - 1}
\]

where \( g \) is the function from Lemma 9.2.1. Setting \( w(\beta_0) = \frac{1}{2} \sqrt{g(\beta_0) - 1} \) proves the theorem. \( \square \)

Note that any instance of SSWD is also an instance of ROSD and thus the information theoretic lower bound in Theorem 9.2.1 also holds for ROSD. Now consider the function \( \tau : \mathbb{R}^{n \times n} \rightarrow \mathcal{G}_n \) that
such that if \( \tau(M) = G \) then \( \{i, j\} \in E(G) \) if and only if \( M_{ij} > 0 \) for all \( i < j \). In other words, \( \tau \) thresholds the above-diagonal entries of \( M \) as in Step 3 of SSBM-REDUCTION from Lemma 7.3.1. Note that \( \tau \) maps \( \mathcal{N}(0, 1)^{n \times n} \) to \( G(n, 1/2) \) and takes \( \mathcal{L}_{H_1}(M) \) from Theorem 9.2.1 to a distribution in \( \mathcal{L}_{SSBM} \in G_B(n, k, 1/2, \rho') \) where

\[
\rho = \Phi\left(\frac{\mu}{k}\right) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \cdot \frac{\mu}{k}
\]

As in the proof of Theorem 7.3.1 there is a method \( e : G_n \to G_n \) that either adds or removes edges with a fixed probability mapping \( G(n, 1/2) \) to \( G(n, q) \) and \( \mathcal{L}_{SSBM} \) to \( e(\mathcal{L}_{SSBM}) \in G_B(n, k, 1/2, \rho') \) where \( \rho' = \Theta(\rho) \) as long as \( q = \Theta(1) \). By the data processing inequality, we now have that

\[
d_{TV}(G(n, 1/2), \mathcal{L}_{SSBM}) \leq d_{TV}(\mathcal{L}_{H_0}(M), \mathcal{L}_{H_1}(M)) \to 0 \quad \text{as } n \to \infty
\]

if \( \mu \ll \sqrt{k} \) which corresponds to \( \rho \ll 1/\sqrt{k} \), establishing the information theoretic lower bound for SSBM in the regime \( q = \Theta(1) \) matching Theorem 2.1.

**Information-Theoretically Optimal Algorithms** Unlike existing maximum likelihood estimators for recovery such as those for BC_R in [CX16] and [CLR15], the definition of \( \psi_{n,k} \) requires that algorithms solving ROS_R are adaptive to the sparsity level \( k \). We introduce a modified exhaustive search algorithm ROS-SEARCH that searches over all possible sparsity levels and checks whether each resulting output is reasonable using an independent copy \( B \) of the data matrix.

We first establish the notation that will be used in this section. Given some \( v \in \mathbb{R}^n \), let \( \text{supp}_+(v) \) denote the set of \( i \) with \( v_i > 0 \) and \( \text{supp}_-(v) \) denote the set of \( i \) with \( v_i < 0 \). If \( A, B \in \mathbb{R}^{n \times n} \), let \( \langle A, B \rangle = \text{Tr}(A^T B) \). Let \( S_t \) be the set of \( v \in \mathbb{R}^n \) with exactly \( t \) nonzero entries each in \( \{-1, 1\} \). In order to show that ROS-SEARCH succeeds at solving ROS_R asymptotically down to its information-theoretic limit, we begin by showing the following lemma.

**Lemma 9.2.2.** Let \( R \) and \( C \) be subsets of \( [n] \) such that \( |R| = k_1 \) and \( |C| = k_2 \) where \( k_1, k_2 \in [c_1 k, k] \) for some constant \( c_1 \in (0, 1) \). Let \( \rho > 0 \) and \( M \in \mathbb{R}^{n \times n} \) be a random matrix and with independent sub-Gaussian entries with sub-Gaussian norm at most 1 such that:

- \( \mathbb{E}[M_{ij}] \geq \rho \) if \( (i, j) \in R \times C \); and
- \( \mathbb{E}[M_{ij}] = 0 \) if \( (i, j) \notin R \times C \).

There is an absolute constant \( c_2 > 0 \) such that if \( k\rho^2 \geq c_2 \log n \), then

\[
\arg\max_{(u,v) \in S_{k_1} \times S_{k_2}} \left\{ u^T M v \right\}
\]

is either \((1_R, 1_C)\) and \((-1_R, -1_C)\) with probability at least \( 1 - n^{-1} \) for sufficiently large \( n \).
Algorithm ROS-SEARCH

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$, sparsity upper bound $k$, threshold $\rho > 0$, constant $c_1 \in (0, 1)$

1. Sample $G \sim N(0, 1)^{\otimes n \times n}$ and form $A = \frac{1}{\sqrt{2}} (M + G)$ and $B = \frac{1}{\sqrt{2}} (M - G)$

2. For each pair $k_1, k_2 \in [c_1 k, k]$ do:
   a. Let $S_t$ be the set of $v \in \mathbb{R}^n$ with exactly $t$ nonzero entries each in $\{-1, 1\}$ and compute
      \[
      (u, v) = \arg \max_{(u, v) \in S_{k_1} \times S_{k_2}} \left\{ u^T A v \right\}
      \]
   b. Mark the pair $(u, v)$ if it satisfies that
      - The set of $i$ with $\sum_{j=1}^n u_i v_j B_{ij} \geq \frac{1}{2} k_2 \rho$ is exactly $\text{supp}(u)$
      - The set of $j$ with $\sum_{i=1}^n u_i v_j B_{ij} \geq \frac{1}{2} k_1 \rho$ is exactly $\text{supp}(v)$

3. Output $\text{supp}(u)$, $\text{supp}(v)$ where $(u, v)$ is the marked pair maximizing $|\text{supp}(u)| + |\text{supp}(v)|$

Figure 9.1: Exhaustive search algorithm for sparse rank-1 submatrix recovery in Theorem 9.2.2.

Proof. For each pair $(u, v) \in S_{k_1} \times S_{k_2}$, let $A_1(u, v)$ be the set of pairs $(i, j) \in R \times C$ with $u_i v_j = -1$, let $A_2(u, v)$ be the set of pairs $(i, j) \in \text{supp}(u) \times \text{supp}(v)$ that are not in $R \times C$ and let $A_3(u, v)$ be the set of $(i, j) \in R \times C$ that are not in $\text{supp}(u) \times \text{supp}(v)$. Now observe that

\[
\mathbf{1}^T R \text{vec} C - u^T M v = \langle M, \mathbf{1}^T C - w^T \rangle = \sum_{(i,j) \in A_1(u,v)} 2 M_{ij} - \sum_{(i,j) \in A_2(u,v)} u_i v_j M_{ij} + \sum_{(i,j) \in A_3(u,v)} M_{ij}
\geq \rho \left( 2 \left| A_1(u,v) \right| + \left| A_3(u,v) \right| \right) + \sum_{(i,j) \in A_1(u,v)} 2 \left( M_{ij} - \mathbb{E}[M_{ij}] \right)
- \sum_{(i,j) \in A_2(u,v)} u_i v_j M_{ij} + \sum_{(i,j) \in A_3(u,v)} (M_{ij} - \mathbb{E}[M_{ij}])
\]

Since $R \times C$ and $\text{supp}(u) \times \text{supp}(v)$ both have size $k_1 k_2$, it follows that $|A_2(u,v)| = |A_3(u,v)|$. Note that the random variables in the sum above are independent, zero mean and sub-Gaussian with norm at most 1. By Hoeffding's inequality for sub-Gaussian random variables as in Proposition
5.10 in [Ver10], it follows that
\[
\mathbb{P}\left[(M, 1_R 1_C^T - uv^T) \leq 0\right] \leq e \cdot \exp\left(-\frac{c \rho^2 (2|A_1(u,v)| + |A_2(u,v)|)^2}{4|A_1(u,v)| + |A_2(u,v)| + |A_3(u,v)|}\right)
\]
\[
= e \cdot \exp\left(-\frac{1}{2c} \rho^2 (2|A_1(u,v)| + |A_2(u,v)|)^2\right)
\]
\[
\leq e \cdot n^{-c_1^{-2} \frac{16}{k}(2|A_1(u,v)| + |A_2(u,v)|)}
\]
for some absolute constant \(c > 0\) as long as \(c \rho^2 \geq 16k c_1^{-2} \log n\). Let \(S(a_1, a_2, b_1, b_2)\) be the set of all pairs \((u, v)\) such that \(a_1 = |\text{supp}(u) \setminus R|, a_2 = |\text{supp}(v) \cap R|, b_1 = |\text{supp}(v) \setminus C|\) and \(b_2 = |\text{supp}(v) \cap C|\). Suppose that \(a_2 \leq \frac{1}{2}(k_1 - a_1)\). Note that for any \((u, v) \in S(a_1, a_2, b_1, b_2)\), we have

\[
|A_1(u,v)| = a_2(k_2 - b_1 - b_2) + b_2(k_1 - a_1 - a_2) \quad \text{and} \quad |A_2(u,v)| = a_1 k_2 + b_1 k_1 - a_1 b_1
\]

Note that \(k_1, k_2 \geq c_1 k, a_1 + a_2 \leq k_1\) and \(b_1 + b_2 \leq k_2\). Therefore we have that \(\frac{1}{k} |A_2(u,v)| \geq a_1 \cdot \frac{k_2}{k} \geq c_1 a_1\) and \(\frac{1}{k} |A_2(u,v)| \geq c_1 b_1\). This implies that \(\frac{1}{k} |A_2(u,v)| \geq \frac{1}{2} c_1 (a_1 + b_1)\). Now note that if \(b_2 \geq c_1 a_2\), then it holds that

\[
\frac{1}{k} |A_1(u,v)| \geq \frac{b_2}{k} (k_1 - a_1 - a_2) \geq \frac{1}{2k} b_2 (k_1 - a_1) \geq \frac{c_1}{2} \cdot b_2 - \frac{a_1}{2} \geq \frac{c_1^2}{4} (a_2 + b_2) - \frac{a_1}{2}
\]

Otherwise if \(b_2 < c_1 a_2\) then it follows that \(b_2 < c_1 a_2 \leq c_1 \cdot \frac{1}{2} (k_1 - a_1) \leq \frac{k_2}{2}\) since \(k_2 \geq c_1 k \geq c_1 k_1\).

Now we have that

\[
\frac{1}{k} |A_1(u,v)| \geq \frac{a_2}{k} (k_2 - b_1 - b_2) \geq \frac{a_2}{k} (k_2 - b_2) - b_1 \geq \frac{c_1}{2} a_2 - b_1 \geq \frac{c_1^2}{4} (a_2 + b_2) - b_1
\]

Therefore in either case it follows that

\[
\frac{1}{k} |A_1(u,v)| \geq \frac{c_1^2}{4} (a_2 + b_2) - a_1 - b_1
\]

Combining these inequalities and the fact that \(c_1 \in (0, 1)\) yields that

\[
\frac{2}{k} |A_1(u,v)| + \frac{1}{k} |A_2(u,v)| \geq \frac{c_1}{4k} |A_1(u,v)| + \frac{1}{k} |A_2(u,v)|
\]

\[
\geq \frac{c_1^2}{4} (a_2 + b_2) + \frac{c_1}{4} (a_1 + b_1) \geq \frac{c_1^2}{4} (a_1 + a_2 + b_1 + b_2)
\]

as long as \(a_2 \leq \frac{1}{2}(k_1 - a_1)\). Furthermore, we have that

\[
|S(a_1, a_2, b_1, b_2)| = \begin{pmatrix} k_1 \end{pmatrix}_{a_1} \begin{pmatrix} n - k_1 \end{pmatrix}_{a_1} \begin{pmatrix} k_2 \end{pmatrix}_{a_2} \begin{pmatrix} n - k_2 \end{pmatrix}_{b_1} \begin{pmatrix} k_2 - b_1 \end{pmatrix}_{b_2} \leq n^{2a_1 + 2b_1 + a_2 + b_2}
\]

since \(\binom{n}{k} \leq n^k\) and \(k_1, k_2 \leq n\). Let \(T\) be the set of \((a_1, a_2, b_1, b_2) \neq (0, 0, 0, 0)\) such that \(a_1, a_2, b_1, b_2 \geq 0, a_1 + a_2 \leq k_1, b_1 + b_2 \leq k_2\) and \(a_2 \leq \frac{1}{2}(k_1 - a_1)\). Now note for all \((u, v) \in S^2\), it holds that at
least one of the pairs $(u, v)$ or $(-u, -v)$ satisfies that $a_2 \leq \frac{1}{2}(k_1 - a_1)$. Since $(u, v)$ and $(-u, -v)$ yield the same value of $u^T M v$, we can restrict to $T$ in the following union bound. Now note that

$$ \mathbb{P} \left[ \text{there is } (u, v) \in S^2 \text{ with } (u, v) \neq \pm (1_R, 1_C) \text{ and } \langle M, 1_R 1_C^T - w^T \rangle \leq 0 \right] $$

\[
\leq \sum_{(a_1, a_2, b_1, b_2) \in T} \left( \sum_{(u, v) \in S(a_1, a_2, b_1, b_2)} e \cdot n^{-c_1 - \frac{1}{2} \log \frac{2 |A_1(u, v)| + |A_2(u, v)|}{k}} \right) 
\leq \sum_{(a_1, a_2, b_1, b_2) \in T} |S(a_1, a_2, b_1, b_2)| \cdot n^{-4(a_1 + a_2 + b_1 + b_2)} 
\leq \sum_{(a_1, a_2, b_1, b_2) \in T} n^{-2a_1 - 3a_2 - 2b_1 - 3b_2} 
\leq -1 + \sum_{a_1, a_2, b_1, b_2 = 0}^{\infty} n^{-2a_1 - 3a_2 - 2b_1 - 3b_2} 
= -1 + \left( \sum_{i=0}^{\infty} n^{-2i} \right)^2 \left( \sum_{j=0}^{\infty} n^{-3j} \right)^2 
= -1 + (1 - n^{-2})^{-2} (1 - n^{-3})^{-2} = O(n^{-2})
\]

which as at most $n^{-1}$ for sufficiently large $n$, completing the proof of the lemma.

We now use this lemma to prove the following theorem, which shows that ROS-SEARCH solves ROSR and SSWR as long as $\mu \geq \sqrt{k} \log n$, asymptotically matching their information theoretic limits.

**Theorem 9.2.2.** Suppose that $M \sim \mu \cdot r c^T + N(0, 1)^{n \times n}$ where $r, c \in \mathcal{Y}_{n,k}$. There is an an absolute constant $c > 0$ such that if $\mu \geq c \sqrt{k} \log n$, then ROS-SEARCH applied with $c_1 = 1/2$ and $\rho = \mu / k$ outputs $\text{supp}(r)$ and $\text{supp}(c)$ with probability at least $1 - 4n^{-1}$ for sufficiently large $n$.

**Proof.** Suppose that $(u, v) \in S_{k_1} \times S_{k_2}$ where $k_1, k_2 \in [c_1 k, k]$ are random vectors that are independent of $B$ and either $\text{supp}(u) \nsubseteq \text{supp}(r)$ or $\text{supp}(v) \nsubseteq \text{supp}(c)$. Note that the definition of $\mathcal{Y}_{n,k}$ is such that any fixed $c_1 \in (0, 1)$ suffices for sufficiently large $k$. We first observe that if $\rho = \mu / k$ and $\mu \geq c \sqrt{k} \log n$ then $(u, v)$ is not marked in Step 2b of ROS-SEARCH with probability at least $1 - n^{-3}$. If $\text{supp}(u) \nsubseteq \text{supp}(r)$, then let $i \in \text{supp}(u) \setminus \subseteq \text{supp}(r)$. It follows that $\sum_{j=1}^{n} u_i v_j B_{ij} \sim N(0, k_2)$ since $\|v\|_0 = k_2$ and by Gaussian tail bounds that

$$ \mathbb{P} \left[ \sum_{j=1}^{n} u_i v_j B_{ij} \geq \frac{1}{2} k_2 \rho \right] \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\rho \sqrt{k_2}} \cdot \exp \left( -\frac{\rho^2 k_2}{8} \right) \leq n^{-3} $$

if $\rho^2 \geq \sqrt{3c_1 k \log n}$. This implies that if $\text{supp}(u) \nsubseteq \text{supp}(r)$, then $(u, v)$ is marked in Step 2b of ROS-SEARCH with probability at most $n^{-3}$. A symmetric argument shows that the same is true
if \( \text{supp}(v) \not\subseteq \text{supp}(c) \). Now for each pair \( k_1, k_2 \in \{c_1, k \} \), let \( u_{k_1} \) and \( v_{k_2} \) be such that

\[
(u_{k_1}, v_{k_2}) = \arg\max_{(u, v) \in S_{k_1} \times S_{k_2}} \left\{ u^\top A v \right\}
\]

if the pair is marked and let \( (u_{k_1}, v_{k_2}) = (0, 0) \) otherwise. By Lemma ?? in the next section, \( A \) and \( B \) from Step 1 of ROS-SEARCH are i.i.d. and distributed as \( \frac{1}{\sqrt{2}} \cdot r c^\top + N(0, 1)^{\otimes n \times n} \). In particular, the pairs \((u_{k_1}, v_{k_2})\) are in the \( \sigma \)-algebra generated by \( A \) and hence are independent of \( B \). By a union bound, we have

\[
P\left[ \text{supp}(u_{k_1}) \subseteq \text{supp}(r) \text{ and } \text{supp}(v_{k_2}) \subseteq \text{supp}(c) \text{ for all } k_1, k_2 \in \{c_1, k \} \right] 
\geq 1 - \sum_{k_1, k_2 \in \{c_1, k \}} P\left[ \text{supp}(u_{k_1}) \not\subseteq \text{supp}(r) \text{ or } \text{supp}(v_{k_2}) \not\subseteq \text{supp}(c) \right] 
\geq 1 - k^2 \cdot n^{-3} \geq 1 - n^{-1}
\]

Now let \( k_1' = ||r||_0 \) and \( k_2' = ||c||_0 \) and let \( r' \) be the vector such that \( r'_i = 0 \) if \( v_i = 0 \), \( r'_i = 1 \) if \( v_i > 0 \) and \( r'_i = -1 \) if \( v_i < 0 \). Define the map \( \tau_r : \mathbb{R}^n \to \mathbb{R}^n \) such that \( \tau_r \) maps \( v \) to the vector with \( i \)-th entry \( r'_i v_i \). Define \( c' \) and \( c_c \) analogously. Now let \( M' \) be the matrix with \((i, j)\)th entry \( r'_i c'_j A_{ij} \). Since the entries of \( M \) are independent Gaussians, it follows that \( M' \sim \mu \cdot \tau_r(r) \tau_c(c)^\top + N(0, 1)^{\otimes n \times n} \). Now observe that since \( S_{k_1'} \) and \( S_{k_2'} \) are preserved by \( \tau_r \) and \( \tau_c \), respectively, we have that \((u_{k_1}', v_{k_2}') = (\tau_r(u), \tau_c(v))\) where

\[
(u, v) = \arg\max_{(u, v) \in S_{k_1'} \times S_{k_2'}} \left\{ u^\top M' v \right\}
\]

Now note that the mean matrix \( \mu \cdot \tau_r(r) \tau_c(c)^\top \) of \( M' \) has all nonnegative entries. Furthermore, the entries in its support are at least \( \mu \cdot \min_{(i, j) \in \text{supp}(r) \times \text{supp}(c)} |r_i c_j| \geq \frac{k_2 \rho}{k} = \rho \) by the definition of \( \mathcal{V}_{n, k} \). Applying Lemma 9.2.2 now yields that with probability at least \( 1 - n^{-1} \), it follows that \((u, v) = (1_{\text{supp}(r)}, 1_{\text{supp}(c)})\) or \((u, v) = (-1_{\text{supp}(r)}, -1_{\text{supp}(c)})\). This implies that \( u_{k_1}' = r' \) and \( v_{k_2}' = c' \) and thus are supported on all of \text{supp}(r) and \text{supp}(c), respectively.

We now will show that \( r' \) and \( c' \) are marked by the test in Step 2b with high probability. If \( i \in \text{supp}(r) \), then \( \sum_{j=1}^n r'_i c'_j B_{ij} \sim N\left( \sum_{j=1}^n \mu \cdot |r_i| \cdot |c_j|, k_2 \right) \). Since \( \sum_{j=1}^n \mu \cdot |r_i| \cdot |c_j| \geq \mu \cdot \frac{k_2}{k} = k_2 \rho \) by the definition of \( \mathcal{V}_{n, k} \). Therefore it follows by the same Gaussian tail bound as above that

\[
P\left[ \sum_{j=1}^n r'_i c'_j B_{ij} < \frac{1}{2} k_2 \rho \right] \leq P\left[ N(0, k_2) < -\frac{1}{2} k_2 \rho \right] \leq n^{-3}
\]

Furthermore, if \( i \not\in \text{supp}(r) \) it follows that \( \sum_{j=1}^n r'_i c'_j B_{ij} \sim N(0, k_2) \) and thus

\[
P\left[ \sum_{j=1}^n r'_i c'_j B_{ij} \geq \frac{1}{2} k_2 \rho \right] = P\left[ N(0, k_2) \geq \frac{1}{2} k_2 \rho \right] \leq n^{-3}
\]
Now a union bound yields that
\[
P \left[ \text{supp}(r) = \left\{ i \in [n] : \sum_{j=1}^{n} r_i' c_j B_{ij} \geq \frac{1}{2} k_2 \rho \right\} \right] \geq 1 - \sum_{i \in \text{supp}(r)} P \left[ \sum_{j=1}^{n} r_i' c_j B_{ij} < \frac{1}{2} k_2 \rho \right] - \sum_{i \notin \text{supp}(r)} P \left[ \sum_{j=1}^{n} r_i' c_j B_{ij} \geq \frac{1}{2} k_2 \rho \right] \geq 1 - k_1 \cdot n^{-3} - (n - k_1)n^{-3} = 1 - n^{-2}
\]

An identical argument yields that
\[
P \left[ \text{supp}(c) = \left\{ j \in [n] : \sum_{i=1}^{n} r_i' c_j B_{ij} \geq \frac{1}{2} k_2 \rho \right\} \right] \geq 1 - n^{-2}
\]

A union bound now yields that \((r', c')\) is marked by the test in Step 2b with probability at least \(1 - 2n^{-2}\). A further union bound now yields that with probability at least \(1 - 2n^{-1} - 2n^{-2}\), the following events all hold:

- \(\text{supp}(u_{k_1}) \subseteq \text{supp}(r)\) and \(\text{supp}(v_{k_2}) \subseteq \text{supp}(c)\) for all \(k_1, k_2 \in [c_1 k, k]\);
- \((u_{k_1}', v_{k_2}') = (r', c')\); and
- \((r', c')\) is marked when input to the test in Step 2b.

These three events imply that that the vector \((r', c')\) is marked in ROS-SEARCH and hence the maximum of \(|\text{supp}(u_{k_1})| + |\text{supp}(v_{k_2})|\) over marked pairs \((u_{k_1}, v_{k_2})\) is \(k_1' + k_2\). Furthermore, first event implies that any marked pair \((u_{k_1}, v_{k_2})\) with \(|\text{supp}(u_{k_1})| + |\text{supp}(v_{k_2})| = k_1' + k_2\) must satisfy that \(\text{supp}(u_{k_1}) = \text{supp}(r)\) and \(\text{supp}(v_{k_2}) = \text{supp}(v)\). Thus the algorithm correctly recovers the supports of \(r\) and \(c\) with probability at least \(1 - 2n^{-1} - 2n^{-2}\), proving the theorem. \(\square\)

The last theorem of this section gives a simple test solving the detection problems \(\text{SSBM}_D,\) \(\text{ROS}_D\) and \(\text{SSW}_D\) asymptotically down to their information-theoretic limits. More precisely, this test solves \(\text{SSBM}_D\) if \(\rho \gtrsim \frac{1}{\sqrt{k}}\) and by setting \(\rho = \mu/k\), solves \(\text{ROS}_D\) and \(\text{SSW}_D\) as long as \(\mu \gtrsim \sqrt{k}\).

**Theorem 9.2.3.** Suppose that \(c_1 \in (0, 1)\) is a fixed constant and let \(R_+\) and \(R_-\) be disjoint subsets of \([n]\) with \(c_1 k \leq |R_+| + |R_-| \leq k\). Let \(C_+\) and \(C_-\) be defined similarly. Let \(M \in \mathbb{R}^{n \times n}\) be a random matrix with independent sub-Gaussian entries with sub-Gaussian norm at most 1 such that:

- \(\mathbb{E}[M_{ij}] \geq \rho\) if \((i, j) \in R_+ \times C_+\) or \((i, j) \in R_- \times C_-\);
- \(\mathbb{E}[M_{ij}] \leq -\rho\) if \((i, j) \in R_+ \times C_-\) or \((i, j) \in R_- \times C_+\); and
9.2. RANK-1 SUBMATRIX, SPARSE SPIKED WIGNER AND SUBGRAPH SBM

- \( \mathbb{E}[M_{ij}] = 0 \) if \((i, j) \notin (R_+ \cup R_-) \times (C_+ \cup C_-) \).

There is a constant \( c_2 > 0 \) such that if \( kp \geq c_2 \sqrt{\log n} \), then \( \max_{(u,v) \in S^2_k} u^T M v \geq \frac{1}{2} c_1^2 k^2 \rho \) with probability at least \( 1 - e^{-n} \). If \( \mathbb{E}[M_{ij}] = 0 \) for all \((i, j) \in [n]^2 \), then there is some constant \( c_3 > 0 \) such that if \( kp^2 \geq c_2 \log n \) then \( \max_{(u,v) \in S^2_k} u^T M v < \frac{1}{2} c_1^2 k^2 \rho \) with probability at least \( 1 - e^{-n} \).

**Proof.** Let \( u \in S_k \) satisfy that \( u_i = 1 \) for each \( i \in R_+ \) and \( u_i = -1 \) for each \( i \in R_- \). Similarly let \( v \in S_k \) satisfy that \( v_i = 1 \) for each \( i \in C_+ \) and \( v_i = -1 \) for each \( i \in C_- \). It follows that

\[
\sum_{i,j=1}^{n} u_i v_j \cdot \mathbb{E}[M_{ij}] \geq (|R_+| + |R_-|)(|C_+| + |C_-|) \rho \geq c_2^2 k^2 \rho
\]

Now note that if \( c \cdot c_1^2 k^2 \rho^2 \geq 4 \log n \), then

\[
\mathbb{P} \left[ u^T M v < \frac{1}{2} c_1^2 k^2 \rho \right] = \mathbb{P} \left[ \sum_{i,j=1}^{n} u_i v_j (M_{ij} - \mathbb{E}[M_{ij}]) < \frac{1}{2} c_1^2 k^2 \rho - \sum_{i,j=1}^{n} u_i v_j \cdot \mathbb{E}[M_{ij}] \right]
\]
\[
\leq \mathbb{P} \left[ \sum_{i,j=1}^{n} u_i v_j (M_{ij} - \mathbb{E}[M_{ij}]) < -\frac{1}{2} c_1^2 k^2 \rho \right]
\]
\[
\leq e \cdot \exp \left( -\frac{c \cdot (c_1^2 k^2 \rho)^2}{4k^2} \right) \leq e^{-n}
\]

for some constant \( c > 0 \) by Hoeffding’s inequality for sub-Gaussian random variables as in Proposition 5.10 in [Ver10]. This proves the first claim of the theorem.

Now suppose that \( \mathbb{E}[M_{ij}] = 0 \) for all \((i, j) \in [n]^2 \). For a fixed pair \((u, v) \in S^2_k \), we have by the same application of Hoeffding’s inequality that

\[
\mathbb{P} \left[ u^T M v \geq \frac{1}{2} c_1^2 k^2 \rho \right] \leq e \cdot \exp \left( -\frac{c \cdot c_1^4 k^2 \rho^2}{4} \right)
\]

Now note that \( |S_k| = 2^k \binom{n}{k} \leq (2n)^k \). Thus a union bound yields that

\[
\mathbb{P} \left[ \max_{(u,v) \in S^2_k} u^T M v \geq \frac{1}{2} c_1^2 k^2 \rho \right] \leq |S_k| \cdot e \cdot \exp \left( -\frac{c \cdot c_1^4 k^2 \rho^2}{4} \right) \leq e \cdot \exp \left( k \log(2n) - \frac{c \cdot c_1^4 k^2 \rho^2}{4} \right) \leq e^{-n}
\]

if \( c \cdot c_1^4 k^2 \rho^2 \geq 4k \log(2n) + 4 \log n \). This completes the proof of the theorem. \( \square \)

**Polynomial-Time Algorithms** The polynomial-time algorithms achieving the tight boundary for the problems in this section are very different in the regimes \( k \lesssim \sqrt{n} \) and \( k \gtrsim \sqrt{n} \). When \( k \lesssim \sqrt{n} \), the simple linear-time algorithm thresholding the absolute values of the entries of the data matrix matches the planted clique lower bounds in Theorem 2.3.1 up to polylogarithmic factors.
This is captured in the following simple theorem, which shows that if $\mu \gtrsim k$ then this algorithm solves $\text{ROS}_R$ and $\text{SSW}_R$. Setting $u = v = 0$ in the Theorem yields that the test outputting $H_1$ if $\max_{i,j} |M_{ij}| > \sqrt{6 \log n}$ solves the detection variants $\text{ROS}_D$ and $\text{SSW}_D$ if $\mu \gtrsim k$.

**Theorem 9.2.4.** Let $M \sim \mu u v^T + N(0,1)^{\otimes n \times n}$ where $u, v \in V_{n,k}$ and suppose that $\mu \geq 2k\sqrt{\log n}$, then the set of $(i, j)$ with $|M_{ij}| > \sqrt{6 \log n}$ is exactly $\text{supp}(u) \times \text{supp}(v)$ with probability at least $1 - O(n^{-1})$.

**Proof.** This theorem follows from the Gaussian tail bound $1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} \cdot t^{-1}e^{-t^2/2}$ for all $t \geq 1$ and a union bound. Now observe that if $(i, j) \not\in \text{supp}(u) \times \text{supp}(v)$, then $M_{ij} \sim N(0,1)$ and thus

$$P\left[|M_{ij}| > \sqrt{6 \log n}\right] = 2 \left(1 - \Phi(\sqrt{6 \log n})\right) \leq \frac{2}{\sqrt{2\pi}} \cdot e^{-3 \log n} = O(n^{-3})$$

If $(i, j) \not\in \text{supp}(u) \times \text{supp}(v)$, then $M_{ij} \sim N(\mu \cdot u_i v_j, 1)$ where $|\mu \cdot u_i v_j| \geq \sqrt{6 \log n}$ since $|u_i|, |v_j| \geq 1/\sqrt{k}$ by the definition of $V_{n,k}$. This implies that

$$P\left[|M_{ij}| \leq \sqrt{6 \log n}\right] \leq \left(1 - \Phi(\sqrt{6 \log n})\right) + \left(1 - \Phi(3\sqrt{6 \log n})\right) \leq \frac{2}{\sqrt{2\pi}} \cdot e^{-3 \log n} = O(n^{-3})$$

Now the probability that the set of $(i, j)$ with $|M_{ij}| > \sqrt{6 \log n}$ is not exactly $\text{supp}(u) \times \text{supp}(v)$ is, by a union bound, at most

$$\sum_{(i,j) \in \text{supp}(u) \times \text{supp}(v)} P\left[|M_{ij}| \leq \sqrt{6 \log n}\right] + \sum_{(i,j) \not\in \text{supp}(u) \times \text{supp}(v)} P\left[|M_{ij}| > \sqrt{6 \log n}\right] = O(n^{-1})$$

which completes the proof of the theorem. \hfill $\square$

In the regime $k \gtrsim \sqrt{n}$, the spectral projection algorithm in Figure 9.2 from [CLR15] achieves exact recovery in $\text{ROS}_R$ down to the planted clique lower bounds in Theorem 2.3.1. This method is described in Algorithm 1 and its guarantees established in Lemma 1 of [CLR15]. Although it is stated as a recovery algorithm for a sub-Gaussian variant of $\text{BC}_R$, as indicated in Remark 2.1 in [CLR15], the guarantees of the algorithm extend more generally to rank one perturbations of a sub-Gaussian noise matrix. Our model of $\text{ROS}_R$ does not exactly fit into the extended model in Remark 2.1, but the argument in Lemma 1 can be applied to show $\text{ROS-SPECTRAL-PROJECTION}$ solves $\text{ROS}_R$. The details of this argument are show below. For brevity, we omit parts of the proof that are identical to [CLR15].

**Theorem 9.2.5.** Suppose that $M = \mu r c^T + N(0,1)^{\otimes n \times n}$ where $r, c \in V_{n,k}$. There is a constant $C_1 > 0$ such that if $\mu \geq C_1(\sqrt{n} + \sqrt{k \log n})$ then the algorithm $\text{ROS-SPECTRAL-PROJECTION}$ correctly outputs $\text{supp}(r)$ and $\text{supp}(c)$ with probability at least $1 - 2n^{-C_2} - 2\exp(-2C_2n)$ for some constant $C_2 > 0$. 

Algorithm ROS-SPECTRAL-PROJECTION

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$

1. Let $G \sim N(0,1)^{\otimes n \times n}$ and let $A = \frac{1}{\sqrt{2}}(M + G)$ and $B = \frac{1}{\sqrt{2}}(M - G)$
2. Compute the top left and right singular vectors $U$ and $V$ of $A$
3. Sort the $n$ entries of $U^T B$ in decreasing order and separate the entries into two clusters $R$ and $[n] \setminus R$ at the largest gap between consecutive values.
4. Sort the $n$ entries of $B V$ in decreasing order and separate the entries into two clusters $C$ and $[n] \setminus C$ at the largest gap between consecutive values.
5. Output $R$ and $C$

Figure 9.2: Algorithm for sparse rank-1 submatrix recovery from [CLR15] and in Theorem 9.2.5.

Proof. Let $P_u$ denote the projection operator onto the vector $u$ and let $\|r\|_0 = k_1$ and $\|c\|_0 = k_2$.

By the definition of $V_{n,k}$, it follows that

$$k \left(1 - \frac{1}{\log k}\right) \leq k_1, k_2 \leq k$$

By the argument in Lemma 1 of [CLR15], there are constants $C_2, C_3 > 0$ such that

$$\|P_{U} B_{ij} - \mu c_j r\|_2 \leq C_3 \sqrt{\log n} + C_3 \sqrt{\frac{n}{k_1}}$$

$$\|P_{V} B_{i'j'} - \mu r_{i'} c\|_2 \leq C_3 \sqrt{\log n} + C_3 \sqrt{\frac{n}{k_2}}$$

hold for all $1 \leq i, j \leq n$ with probability at least $1 - 2n^{-C_2} - 2\exp(-2C_2 n)$. Now note that if $j \in \text{supp}(c)$ and $j' \notin \text{supp}(c)$, then it follows that

$$\|\mu c_j r - \mu c_{j'} r\|_2 = \mu \cdot |c_j| \geq \frac{\mu}{\sqrt{k}}$$

by the definition of $V_{n,k}$. Similarly, if $i \in \text{supp}(r)$ and $i' \notin \text{supp}(r)$ then $\|\mu r_i c - \mu r_{i'} c\|_2 \geq \mu / \sqrt{k}$.

Therefore if for both $i = 1, 2$

$$\frac{\mu}{\sqrt{k}} \geq 6C_3 \left(\sqrt{\log n} + \sqrt{\frac{n}{k_i}}\right)$$
then it holds that
\[
2 \max_{j,j' \in \text{supp}(c)} \| \mathcal{P}_U B_j - \mathcal{P}_U B_{j'} \|_2 \leq \max_{j \in \text{supp}(c), j' \notin \text{supp}(c)} \| \mathcal{P}_U B_j - \mathcal{P}_U B_{j'} \|_2
\]
\[
2 \max_{i,i' \in \text{supp}(c)} \| \mathcal{P}_V B_i^\top - \mathcal{P}_U B_{i'}^\top \|_2 \leq \max_{i \in \text{supp}(c), i' \notin \text{supp}(c)} \| \mathcal{P}_V B_i^\top - \mathcal{P}_U B_{i'}^\top \|_2
\]
and Steps 3 and 4 succeed in recovering supp(r) and supp(c).

A simple singular value thresholding algorithm solves ROSD if \( p \geq \frac{\sqrt{n}}{4} \) and is comparatively simpler to analyze. As previously mentioned, this algorithm also solves SSWD since any instance of SSWD is an instance of ROSD. Let \( \sigma_1(M) \) denote the largest singular value of the matrix \( M \).

**Theorem 9.2.6.** Suppose that \( M \) is an instance of ROSD \((n, k, \mu)\). There is a constant \( C_1 > 0 \) such that if \( \mu > 4\sqrt{n} + 2\sqrt{2\log n} \) then the algorithm that outputs \( H_1 \) if \( \sigma_1(M) \geq \frac{1}{2} \mu \) and \( H_0 \) otherwise has Type I+II error tending to zero as \( n \to \infty \).

**Proof.** Under \( H_0 \), it holds that \( M \sim N(0, 1)^{\otimes n \times n} \). By Corollary 5.35 in [Ver10], we have
\[
\sigma_1(M) \leq 2\sqrt{n} + \sqrt{2\log n}
\]
with probability at least \( 1 - 2n^{-1} \). Now consider the case of \( H_1 \) and suppose that \( M = \mu \cdot rc^\top + Z \) where \( r, c \in \mathcal{V}_{n,k} \) and \( Z \sim N(0, 1)^{\otimes n \times n} \). By Weyl's interlacing inequality, it follows that
\[
|\mu - \sigma_1(M)| = |\sigma_1(M \cdot rc^\top) - \sigma_1(M)| \leq \sigma_1(Z) \leq 2\sqrt{n} + \sqrt{2\log n}
\]
with probability at least \( 1 - 2n^{-1} \). If \( \mu > 4\sqrt{n} + 2\sqrt{2\log n} \) then the Type I+II error of the algorithm is at most \( 4n^{-1} \), proving the theorem. \( \square \)

To complete this section, we give a simple spectral thresholding algorithm for SSBM\(_D\) if \( \rho \geq \frac{\sqrt{n}}{k} \). Let \( \lambda_1(X) \) denote the largest eigenvalue of \( X \) where \( X \) is a square symmetric matrix.

**Theorem 9.2.7.** Let \( G \) be an instance of SSBM\(_D\)(\( n, k, q, \rho \)) where \( q = \Theta(1) \), \( k = \Omega(\sqrt{n}) \) and \( \rho \geq \frac{6\sqrt{n}}{k} \). Let \( A \in \mathbb{R}^{n \times n} \) be the adjacency matrix of \( G \) and \( J \in \mathbb{R}^{n \times n} \) be the matrix with zeros on its diagonal and ones elsewhere. Then the algorithm that outputs \( H_1 \) if \( \lambda_1(A - qJ) \geq 2\sqrt{n} \) and \( H_0 \) otherwise has Type I+II error tending to zero as \( n \to \infty \).

**Proof.** Suppose that \( G \) is drawn from some distribution in \( H_1 \) and that the two hidden communities have index sets \( S_1, S_2 \subseteq [n] \) where \( k_1 = |S_1| \) and \( k_2 = |S_2| \). For the remainder of the analysis of \( H_1 \), consider \( A \) and \( G \) conditioned on \( S_1 \) and \( S_2 \). Now let \( v \) be the vector
\[
v_i = \begin{cases} 
\frac{1}{\sqrt{k_1 + k_2}} & \text{if } i \in S_1 \\
-\frac{1}{\sqrt{k_1 + k_2}} & \text{if } i \in S_2 \\
0 & \text{otherwise}
\end{cases}
\]
for each $i \in [n]$. Now observe that

$$v^T (A - qJ)v = \frac{2}{k_1 + k_2} \left( \sum_{(i,j) \in E(G)} (1_{(i,j)} - q) + \sum_{(i,j) \in S_1 \times S_2} (q - 1_{(i,j)} \in E(G)) \right)$$

By the definition of $H_1$ in SSBMD, the expression above is the sum of $\binom{k_1 + k_2}{2}$ independent shifted Bernoulli random variables each with expectation at least $\rho$. Therefore it follows that

$$\mathbb{E} \left[ v^T (A - qJ)v \right] \geq \frac{2}{k_1 + k_2} \cdot \binom{k_1 + k_2}{2} \cdot \rho = (k_1 + k_2 - 1)\rho \geq 3\sqrt{n}$$

since $k_1 + k_2 - 1 \geq k - 2k^{1-\delta_{SSBM}} - 1 \geq \frac{k}{2}$ for sufficiently large $k$, as defined in Section 2.2. Now note that each of the centered random variables $1_{(i,j) \in E(G)} - \mathbb{E}[1_{(i,j) \in E(G)}]$ and $\mathbb{E}[1_{(i,j) \in E(G)}] - 1_{(i,j) \in E(G)}$ are bounded in $[-1, 1]$ and therefore Bernstein's inequality implies that for all $t > 0$,

$$\mathbb{P} \left[ v^T (A - qJ)v < \mathbb{E}[v^T (A - qJ)v] - \frac{2t}{k_1 + k_2} \right] \leq \exp \left( -\frac{\frac{1}{3}t^2}{n + \frac{1}{3}t} \right)$$

Note that $v$ is a unit vector and thus $\lambda_1 (A - qJ) \geq v^T (A - qJ)v$. Setting $t = \frac{1}{2}(k_1 + k_2)\sqrt{n}$ now yields that

$$\mathbb{P} \left[ \lambda_1 (A - qJ) < 2\sqrt{n} \right] \leq \exp \left( -\frac{\frac{1}{8}(k_1 + k_2)^2n}{n + \frac{1}{6}(k_1 + k_2)\sqrt{n}} \right) = \exp (-\Omega(n))$$

since $k_1 + k_2 \geq \frac{k}{2} = \Omega(\sqrt{n})$ for sufficiently large $k$. Now suppose that $G$ is drawn from $G(n, q)$ as in $H_0$. By Theorem 1.5 in [Vu05], it follows that with probability $1 - o_n(1)$, we have that

$$\lambda_1 (A - qJ) \leq 2\sqrt{q(1 - q)n} + C(q - q^2)^{1/4}n^{1/4} \log n$$

for some constant $C > 0$. Therefore $\lambda_1 (A - qJ)$ is less than $2\sqrt{n}$ for sufficiently large values of $n$ since $q(1 - q) \leq 1/4$. Therefore the Type I+II error of this algorithm on SSBMD is $o_n(1) + \exp (-\Omega(n)) = o_n(1)$ as $n \to \infty$.

### 9.3 Sparse PCA and Biased Sparse PCA

**Information-Theoretic Lower Bounds** In [BR13b], information-theoretic lower bounds for $\text{SPCA}_D(n, k, d, \theta)$ were considered and it was shown that if

$$\theta \leq \min \left\{ \frac{1}{\sqrt{2}} \sqrt{\frac{k \log(1 + o(d/k^2))}{n}} \right\}$$
CHAPTER 9. ALGORITHMS AND INFORMATION-THEORETIC THRESHOLDS

then the optimal Type I+II error of any algorithm for SPCA \( D \) tends to 1 as \( n \to \infty \). The proof of this information-theoretic lower bound follows a similar \( \chi^2 \) and MGF argument as in the previous section. If \( d \ll k^2 \), then this bound degrades to \( \theta = o \left( \frac{d}{kn} \right) \). When \( d = \Theta(n) \), there is a gap between this information-theoretic lower bound and the best known algorithm based on thresholding the \( k \)-sparse eigenvalue of the empirical covariance matrix, which only requires \( \theta \gtrsim \sqrt{k/n} \). This information-theoretic lower bound was improved in [CMW15] to match the algorithmic upper bound. The following is their theorem in our notation.

**Theorem 9.3.1** (Proposition 2 in [CMW15]). Let \( \beta_0 \in (0, 1/36) \) be a constant. Suppose that \( X = (X_1, X_2, \ldots, X_n) \) is an instance of SPCA \( D(n, k, d, \theta) \) where under \( H_1 \), the planted vector \( v \) is chosen uniformly at random from all \( k \)-sparse unit vectors with nonzero coordinates equal to \( \pm \frac{1}{\sqrt{k}} \). If it holds that

\[
\theta \leq \min \left\{ 1, \frac{\beta_0 k}{n} \log \left( \frac{ed}{k} \right) \right\}
\]

then there is a function \( w : (0, 1/36) \to (0, 1) \) satisfying that \( \lim_{\beta_0 \to 0^+} w(\beta_0) = 0 \) and

\[
d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \leq w(\beta_0)
\]

In particular if \( \theta \ll \sqrt{k/n} \), then this inequality eventually applies for every \( \beta_0 > 0 \) and \( d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \to 0 \), establishing the desired information-theoretic lower bound for SPCA \( D \). We now show that BSPCA \( D \) satisfies a weaker lower bound. The prior on the planted vector \( v \) used in Theorem 5.1 of [BR13b] to derive the suboptimal lower bound for SPCA \( D \) shown above placed entries equal to \( 1/\sqrt{k} \) on a random \( k \)-subset of the \( d \) coordinates of \( v \). Therefore their bound also applies to BSPCA \( D \). In order strengthen this to the optimal information-theoretic lower bound for BSPCA \( D \), we need apply Lemma 9.1.2 in place of the weaker hypergeometric squared MGF bounds used in [BR13b]. To simplify the proof, we will use the following lemma from [CMW15].

**Lemma 9.3.1** (Lemma 7 in [CMW15]). Let \( v \) be a distribution on \( d \times d \) symmetric random matrices \( M \) such that \( \|M\| \leq 1 \) almost surely. If \( \mathbb{E}_v[N(0, I_d + M)^{\otimes n}] = \int N(0, I_d + M)^{\otimes n} dv(M) \), then

\[
\chi^2(\mathbb{E}_v[N(0, I_d + M)^{\otimes n}], N(0, I_d)^{\otimes n}) + 1 = \mathbb{E} \left[ \det(I_d - M_1 M_2)^{-n/2} \right]
\]

where \( M_1 \) and \( M_2 \) are independently drawn from \( v \).

Applying this lemma with the hypergeometric squared MGF bounds in Lemma 9.1.2 yields the following information-theoretic lower bound for BSPCA \( D \).
Theorem 9.3.2. Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of BSPCA$_D(n, k, d, \theta)$ where under $H_1$, the planted vector $v$ is chosen uniformly at random from all $k$-sparse unit vectors in $\mathbb{R}^d$ with nonzero coordinates equal to $\frac{1}{\sqrt{k}}$. Suppose it holds that $\theta \leq 1/\sqrt{2}$ and

$$\theta \leq \min \left\{ \sqrt{\frac{\theta_0 k}{n} \log \left( \frac{ed}{k} \right)}, \sqrt{\frac{\theta_0 d^2}{nk^2}} \right\}$$

for some $0 < \beta_0 < (16e)^{-1}$. Then there is a function $w : (0, 1) \rightarrow (0, 1)$ satisfying that $\lim_{\beta_0 \rightarrow 0^+} w(\beta_0) = 0$ and

$$d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \leq w(\beta_0)$$

Proof. Let $u_S$ denote the $d$-dimensional unit vector with entries in $S$ equal to $1/\sqrt{k}$ and all other entries equal to zero where $S$ is some $k$-subset of $d$. Let $v$ be the distribution on matrices $\theta u_S u_S^\top$ where $\theta \leq 1$ and $S$ is chosen uniformly at random from the set of $k$-subsets of $[d]$. Note that $\mathcal{L}_{H_0}(X) = N(0, I_d)^{\otimes n}$ and $\mathcal{L}_{H_1}(X) = \mathbb{E}_v[N(0, I_d + \theta u_S u_S^\top)^{\otimes n}]$. By Lemma 9.3.1, it follows that

$$\chi^2(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X)) = \mathbb{E} \left[ \det \left( I_d - \theta^2 u_S u_S^\top u_T u_T^\top \right)^{-n/2} \right] - 1$$

$$= \mathbb{E} \left[ \left( 1 - \frac{\theta^2}{k^2} \cdot |S \cap T|^2 \right)^{-n/2} \right] - 1$$

$$\leq \mathbb{E} \left[ \exp \left( \frac{n \theta^2}{k^2} \cdot |S \cap T|^2 \right) \right] - 1$$

where $S$ and $T$ are independent random $k$-subsets of $[d]$. The last inequality above follows from the fact that $(1 - t)^{-1/2} \leq e^t$ if $t \leq 1/2$, $\theta^2 \leq 1/2$ and $|S \cap T| \leq k$. Now note that $|S \cap T| \sim \text{Hypergeometric}(d, k, k)$ and let

$$b = \frac{n \theta^2}{k^2} \cdot \left( \max \left\{ \frac{1}{k} \log \left( \frac{ed}{k} \right), \frac{d^2}{k^2} \right\} \right)^{-1}$$

The given condition on $\theta$ implies that $b \leq \beta_0$. It follows by Lemma 9.1.2 that $\chi^2(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X)) \leq r(\beta_0) - 1$ and by Cauchy-Schwarz that if $w(\beta) = \frac{1}{2} \sqrt{r(\beta) - 1}$ then

$$d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \leq \frac{1}{2} \sqrt{\chi^2(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X))} \leq w(\beta_0)$$

where $w(\beta_0) \rightarrow 0$ as $\beta_0 \rightarrow 0^+$, proving the theorem. \qed

We remark that this same proof technique applied to the ensemble of $k$-sparse unit vectors $v$ chosen uniformly at random from those with nonzero coordinates equal to $\pm 1/\sqrt{k}$ with Lemma 9.2.1 proves Theorem 9.3.1. This difference in the lower bounds resulting from these two choices
of ensembles illustrates the information-theoretic difference between SPCAD and BSPCAD. We now will show information-theoretic lower bounds for the weak recovery problems SPCAWR and BSCPAWR. The argument presented here is similar to the proof of Theorem 3 in [WBS16]. For this argument, we will need a variant of the Gilbert-Varshamov lemma and generalized Fano’s lemma as in [WBS16]. Given two \( u, v \in \mathbb{R}^d \), let \( d_H(u, v) \) denote the Hamming distance between \( u \) and \( v \).

**Lemma 9.3.2** (Gilbert-Varshamov, Lemma 4.10 in [Mas07]). Suppose that \( \alpha, \beta \in (0, 1) \) and \( k \leq \alpha \beta d \). Let

\[
\rho = \frac{\alpha}{\log(\alpha \beta)} - 1 (\beta - \log \beta - 1)
\]

Then there is a subset \( S \) of \( \{v \in \{0, 1\}^d : \|v\|_0 = k\} \) of size at least \( \left( \frac{\rho}{\rho_k} \right)^{\rho k} \) such that for any two \( u, v \in S \) with \( u \neq v \), it holds that \( d_H(u, v) \geq 2(1 - \alpha)k \).

**Lemma 9.3.3** (Generalized Fano’s Lemma, Lemma 3 in [Yu97]). Let \( P_1, P_2, \ldots, P_M \) be probability distributions on a measurable space \((X, \mathcal{B})\) and assume that \( d_{KL}(P_i, P_j) \leq \beta \) for all \( i \neq j \). Any measurable function \( \phi : X \to \{1, 2, \ldots, M\} \) satisfies that

\[
\max_{1 \leq i \leq M} P_i(\phi \neq i) \geq 1 - \frac{\beta + \log 2}{\log M}
\]

With these two lemmas, we now will show that the weak recovery problems SPCAWR and BSCPAWR cannot be solved if \( \theta \lesssim \sqrt{k/n} \), matching Theorem 2.3.1. Note that in the theorem below, \( \phi(X) \) and \( \text{supp}(v) \) have size \( k \) for all \( X \in \mathbb{R}^{d \times n} \) and \( v \in S \). Therefore it holds that \( |\phi(X) \Delta \text{supp}(v)| = 2k - 2|\phi(X) \cap \text{supp}(v)| \) for all such \( X \) and \( v \).

**Theorem 9.3.3.** Fix positive integers \( n, k, d \) and real numbers \( \theta > 0 \) and a constant \( \epsilon \in (0, 1) \) such that \( k \leq \epsilon d/4 \). Let \( P_v \) denote the distribution \( N(0, I_d + \theta vv^T) \) and let \( S \) be the set of all \( k \)-sparse unit vectors with nonzero entries equal to \( 1/\sqrt{k} \). If

\[
\frac{n \theta^2}{2(1 + \theta)} + \log 2 \leq \frac{\epsilon^2}{2 \log 4 \epsilon^{-1}} \cdot k \log \left( \frac{d}{k} \right)
\]

then for any function \( \phi : \mathbb{R}^{d \times n} \to \left( \frac{[n]}{k} \right) \), it holds that

\[
\min_{v \in S} \mathbb{E}_{X \sim P_v^{\otimes n}} [|\phi(X) \cap \text{supp}(v)|] \leq \left( \frac{1}{2} + \epsilon \right) k
\]

**Proof.** Note that \( d_{KL}(N(0, \Sigma_0), N(0, \Sigma_1)) = \frac{1}{2} \left[ \text{Tr} \left( \Sigma_1^{-1} \Sigma_0 \right) - d + \ln \frac{\text{det}(\Sigma_1)}{\text{det}(\Sigma_0)} \right] \) for any positive semidef-
9.3. **Sparse PCA and Biased Sparse PCA**

Finite $\Sigma_0, \Sigma_1 \in \mathbb{R}^{d \times d}$. Therefore for any $\|u\|_2 = \|v\|_2 = 1$,

$$
\begin{align*}
    d_{KL}(P_u^{\otimes n}, P_v^{\otimes n}) &= n \cdot d_{KL}(P_u, P_v) = \frac{n}{2} \cdot \text{Tr} \left( \left( I_d + \theta uu^\top \right)^{-1} \left( I_d + \theta vv^\top \right) - I_d \right) \\
    &= \frac{n\theta}{2} \cdot \text{Tr} \left( \left( I_d + \theta uu^\top \right)^{-1} \left( vv^\top - uu^\top \right) \right) \\
    &= \frac{n\theta}{2} \cdot \text{Tr} \left( \left( I_d - \frac{\theta}{1+\theta} \cdot uu^\top \right) \left( vv^\top - uu^\top \right) \right) \\
    &= \frac{n\theta}{2} \cdot \text{Tr} \left( vv^\top - uu^\top - \frac{\theta}{1+\theta} \cdot \langle u, v \rangle \cdot uu^\top + \frac{\theta}{1+\theta} \cdot uu^\top \right) \\
    &= \frac{n\theta^2}{4(1+\theta)} \cdot \|u - v\|_2^2 \leq \frac{n\theta^2}{2(1+\theta)}
\end{align*}
$$

since $\det(I_d + \theta uu^\top) = \det(I_d + \theta vv^\top) = 1 + \theta$. Let $S_0$ be the subset from Gilbert-Varshamov's lemma applied with $\alpha = \epsilon$ and $\beta = \frac{1}{4}$. It follows that

$$
\log |S_0| \geq \frac{\epsilon}{2\log 4e^{-1}} \cdot k \log \left( \frac{d}{k} \right)
$$

For each $u \in S$, let $\hat{u}$ be an element of $S_0$ such that $|\text{supp}(u) \cap \text{supp}(\hat{u})|$ is maximal. Let $\hat{\phi}$ denote the function that maps $X \in \mathbb{R}^{d \times n}$ to $\text{supp}(\hat{u})$ where $u = 1_{\phi(X)}$. Suppose that $v \in S_0$ is such that $v \neq \hat{u}$. Observe by the triangle inequality that

$$4k - 2|\phi(X) \cap \text{supp}(v)| - 2|\phi(X) \cap \phi(X)| = d_H(\hat{u}, u) + d_H(u, v) \geq d_H(\hat{u}, v) \geq 2(1 - \epsilon)k$$

Rearranging and using the fact that $|\phi(X) \cap \text{supp}(v)| \leq |\hat{\phi}(X) \cap \phi(X)|$ yields that

$$k - |\phi(X) \cap \text{supp}(v)| \geq \frac{1}{2} (1 - \epsilon) k$$

Note that $k - |\phi(X) \cap \text{supp}(v)| \geq 0$ is true for all $v$. Therefore if $v \in S_0$, we have

$$\mathbb{E}_{X \sim P_\phi^{\otimes n}} [k - |\phi(X) \cap \text{supp}(v)|] \geq \mathbb{P}_{X \sim P_\phi^{\otimes n}} [\phi(X) \neq v] \cdot \frac{1}{2} (1 - \epsilon) k$$

Now observe by generalized Fano's Lemma,

$$\max_{v \in S} \mathbb{E}_{X \sim P_\phi^{\otimes n}} [k - |\phi(X) \cap \text{supp}(v)|] \geq \max_{v \in S_0} \mathbb{E}_{X \sim P_\phi^{\otimes n}} [k - |\phi(X) \cap \text{supp}(v)|]$$

$$\geq \frac{1}{2} (1 - \epsilon) k \cdot \max_{v \in S_0} \mathbb{P}_{X \sim P_\phi^{\otimes n}} [\phi(X) \neq v]$$

$$\geq \frac{1}{2} (1 - \epsilon) k \cdot \left[ 1 - \frac{n\theta^2}{2(1+\theta)} + \log 2 \right]$$

$$\geq \frac{1}{2} (1 - \epsilon) k \cdot (1 - \epsilon) \geq \left( \frac{1}{2} - \epsilon \right) k$$

Rearranging completes the proof of the lemma. \(\square\)
**Information-Theoretically Optimal Algorithms**  Given a positive semidefinite matrix $M \in \mathbb{R}^{d \times d}$, let the $k$-sparse maximum eigenvalue and eigenvector be

$$
\lambda_{\text{max}}^k(M) = \max_{\|u\|_2 = 1, \|u\|_0 = k} u^T M u \quad \text{and} \quad v_{\text{max}}^k(M) = \arg \max_{\|u\|_2 = 1, \|u\|_0 = k} u^T M u
$$

Note that $\lambda_{\text{max}}^k(M)$ and $v_{\text{max}}^k(M)$ can be computed by searching over all principal $k \times k$ minors for the maximum eigenvalue and corresponding eigenvector. In [BR13a] and [WBS16], the $k$-sparse maximum eigenvalue and eigenvector of the empirical covariance matrix are shown to solve sparse PCA detection and estimation in the $\ell_2$ norm under general distributions satisfying a restricted covariance concentration condition. Specializing these results to Gaussian formulations of sparse PCA yields the following theorems. Let $L(u, v) = \sqrt{1 - \langle u, v \rangle}$ for $u, v \in \mathbb{R}^d$ with $\|u\|_2 = \|v\|_2 = 1$. The statement of Theorem 2 in [BR13a] is for $k < \sqrt{d}$, but the same argument also shows the result for $k \ll d$. Note that the following theorem applies to BSPCAD since any instance of BSPCAD is also an instance of SPCAD.

**Theorem 9.3.4** (Theorem 2 in [BR13a]). Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of SPCAD($n, k, d, \theta$) and let $\hat{\Sigma}$ be the empirical covariance matrix of $X$. If $\theta, \delta \in (0, 1)$ are such that

$$
\theta > 15 \sqrt{\frac{k \log \left( \frac{3ed}{k^3} \right)}{n}}
$$

then the algorithm that outputs $H_1$ if $\lambda_{\text{max}}^k(\hat{\Sigma}) > 1 + 8\sqrt{\frac{k \log \left( \frac{3ed}{k^3} \right)}{n}}$ and $H_0$ otherwise has Type I+II error at most $\delta$.

**Theorem 9.3.5** (Theorem 2 in [WBS16]). Suppose that $k, d$ and $n$ are such that $2k \log d \leq n$. Let $P_v$ denote the distribution $N(0, I_d + \theta vv^T)$ and given some $X = (X_1, X_2, \ldots, X_n) \sim P_v^\otimes n$, let $\hat{\Sigma}(X)$ be the empirical covariance matrix of $X$. It follows that

$$
\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v} L \left( v_{\text{max}}^k \left( \hat{\Sigma}(X) \right), v \right) \leq 7 \sqrt{\frac{k \log d}{n \theta^2}}
$$

The latter result on estimation in the $\ell_2$ norm yields a weak recovery algorithm for SPCAWR and BSPCAwR by thresholding the entries of $v_{\text{max}}^k(\hat{\Sigma})$, as in the following theorem. If $\frac{k \log d}{n \theta^2} \to 0$ then this algorithm achieves weak recovery.

**Theorem 9.3.6.** Suppose that $k, d$ and $n$ are such that $2k \log d \leq n$. Let $S(X) \subseteq [d]$ be the set of coordinates of $v_{\text{max}}^k \left( \hat{\Sigma}(X) \right)$ with magnitude at least $\frac{1}{2\sqrt{k}}$. It follows that

$$
\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v} \left[ \frac{1}{k} |S(X)\Delta\text{supp}(v)| \right] \leq 56 \sqrt{\frac{2k \log d}{n \theta^2}}
$$
9.3. SPARSE PCA AND BIASED SPARSE PCA

Proof. Let \( u = v_{\max}^k \left( \hat{\Sigma}(X) \right) \) and note that
\[
\|u\|_2^2 = \|v\|_2^2 + \|v_{\max}^k \left( \hat{\Sigma}(X) \right)\|_2^2 - 2 \left( v, v_{\max}^k \left( \hat{\Sigma}(X) \right) \right) = 2 \cdot L \left( v_{\max}^k \left( \hat{\Sigma}(X) \right), v \right).
\]
If \( i \in \text{supp}(v) \) where \( v \in \mathcal{V}_{d,k} \), then \( |v_i| \geq \frac{1}{\sqrt{k}} \). Therefore each \( i \in S(X)\Delta \text{supp}(v) \) satisfies that
\[
|u_i| \geq \frac{1}{2\sqrt{k}},
\]
which implies that
\[
\frac{1}{k} |S(X)\Delta \text{supp}(v)| \leq 4 \cdot \sum_{i \in S(X)\Delta \text{supp}(v)} |u_i|^2 \leq 4\|u\|_2 \leq 8\sqrt{2} \cdot L \left( v_{\max}^k \left( \hat{\Sigma}(X) \right), v \right)
\]
using the fact that \( L(u, v) \leq \sqrt{2} \) if \( \|u\|_2 = \|v\|_2 = 1 \). This inequality along with the previous theorem completes the proof.

We next analyze an algorithm thresholding the sum of the entries of the empirical covariance matrix. For instances of BSPCAD, this sum test solves the detection problem when \( \theta \gtrsim \sqrt[3]{n} \) in the regime \( d = \Theta(n) \) and becomes optimal when \( k \gtrsim n^{2/3} \). Recall that \( \delta = \delta_{\text{BSPCA}} > 0 \) is the constant in the definition of \( \mathcal{B}_d,k \).

**Theorem 9.3.7.** Suppose that \( X = (X_1, X_2, \ldots, X_n) \) is an instance of BSPCAD \( (n, k, d, \theta) \) and let \( \hat{\Sigma}(X) \) be the empirical covariance matrix of \( X \). Suppose that \( 2\delta^2 k \theta \leq d \) and \( \frac{n \delta^2 \theta^2}{d^2} \rightarrow \infty \) as \( n \rightarrow \infty \). Then the test that outputs \( H_1 \) if \( 1^T \hat{\Sigma}(X)1 > d + 2 \delta^2 k \theta \) and \( H_0 \) otherwise has Type I+II error tending to zero as \( n \rightarrow \infty \).

Proof. First assume that \( H_0 \) holds and \( X \sim N(0, I_d)^{\otimes n} \). Observe that
\[
\frac{n}{d} \cdot 1^T \hat{\Sigma}(X)1 = \frac{1}{d} \sum_{i=1}^{n} (1, X_i)^2
\]
where the values \( \frac{1}{\sqrt{d}} (1, X_i) \) are independent and distributed as \( N(0, 1) \). Therefore \( \frac{n}{d} \cdot 1^T \hat{\Sigma}(X)1 \) is distributed as a \( \chi^2 \) distribution with \( n \) degrees of freedom. Since \( \frac{1}{d}(1, X_i)^2 - 1 \) is zero-mean and sub-exponential with norm 1, Bernstein's inequality implies that for all \( t \geq 0 \)
\[
\mathbb{P} \left[ \frac{n}{d} \cdot 1^T \hat{\Sigma}(X)1 \geq n + t \right] \leq 2 \exp \left( -c \cdot \min \left( \frac{t^2}{n}, t \right) \right)
\]
for some constant \( c > 0 \). Substituting \( t = \frac{2n \delta^2 k \theta}{d} \leq n \) yields that
\[
\mathbb{P} \left[ 1^T \hat{\Sigma}(X)1 \geq d + 2 \delta^2 k \theta \right] \leq 2 \exp \left( -c \cdot \min \left( \frac{4n \delta^4 k^2 \theta^2}{d^2}, \frac{2n \delta^2 k \theta}{d} \right) \right) = 2 \exp \left( -\frac{4cn \delta^4 k^2 \theta^2}{d^2} \right)
\]
which tends to zero as \( n \to \infty \). Now assume that \( H_1 \) holds and \( X \sim N(0, I_d + \theta vv^T)^{\oplus n} \) for some \( v \in \mathcal{B}_d,k \). Note that each \( X_i \) can be written as \( X_i = \sqrt{\theta} \cdot g_i v + Z_i \) where \( g_1, g_2, \ldots, g_n \sim \text{i.i.d.} N(0, 1) \) and \( Z_1, Z_2, \ldots, Z_n \sim \text{i.i.d.} N(0, I_d) \). If \( s(v) = \sum_{j=1}^{d} v_j \) is the sum of the entries of \( v \), then

\[
(1, X_i) = \sqrt{\theta} \cdot g_i s(v) + \sum_{j=1}^{d} Z_{ij} \sim N(0, \theta s(v)^2)
\]

Furthermore, these inner products are independent for \( i = 1, 2, \ldots, n \). Therefore \( \frac{n}{d+\theta s(v)^2} \cdot 1 \Sigma(X) 1 \) is also distributed as a \( \chi^2 \) distribution with \( n \) degrees of freedom. Since \( v \in \mathcal{B}_d,k \), it either follows that \( |\text{supp}_+(v)| \geq (\frac{1}{2} + \delta) k \) or \( |\text{supp}_-(v)| \geq (\frac{1}{2} + \delta) k \). If \( |\text{supp}_+(v)| \geq (\frac{1}{2} + \delta) k \), then by Cauchy-Schwarz we have that

\[
s(v) = \sum_{i \in \text{supp}_+(v)} v_i - \sum_{i \in \text{supp}_-(v)} |v_i| \geq (\frac{1}{2} + \delta) \sqrt{k} - \left( \sum_{i \in \text{supp}_-(v)} |v_i|^2 \right)^{1/2} \cdot \sqrt{|\text{supp}_-(v)|}
\]

\[
= \left( \frac{1}{2} + \delta \right) \sqrt{k} - \left( 1 - \sum_{i \in \text{supp}_+(v)} |v_i|^2 \right)^{1/2} \cdot \sqrt{|\text{supp}_-(v)|}
\]

\[
\geq \left( \frac{1}{2} + \delta \right) \sqrt{k} - \left( 1 - \frac{|\text{supp}_+(v)|}{k} \right)^{1/2} \cdot \sqrt{\left( \frac{1}{2} - \delta \right) k} \geq 2\delta \sqrt{k}
\]

Bernstein’s inequality with \( t = \frac{2n\sigma^2 k\theta}{d+\theta s(v)^2} \leq n \) now implies that

\[
P \left[ 1 \Sigma(X) 1 \leq d + 2\sigma^2 k\theta \right] \leq 2 \exp \left( -\frac{4cn\delta^4 k^2 \theta^2}{(d+\theta s(v)^2)^2} \right)
\]

which tends to zero as \( n \to \infty \) since \( \theta s(v)^2 \leq \theta k \leq \frac{d}{2\sigma^2} \) by Cauchy-Schwarz. This completes the proof of the theorem.

\[ \square \]

**Polynomial-Time Algorithms**  As shown in [BR13a], SPCA\(_D\) and BSPCAD\(_D\) can be solved with a semidefinite program in the regime \( k \lesssim \sqrt{n} \). Their algorithm computes a semidefinite relaxation of the maximum \( k \)-sparse eigenvalue, first forming the empirical covariance matrix \( \hat{\Sigma}(X) \) and solving the convex program

\[
\text{SDP}(X) = \max_Z \text{Tr} \left( \hat{\Sigma}(X) Z \right)
\]

\[
s.t. \quad \text{Tr}(Z) = 1, |Z|_1 \leq k, Z \succeq 0
\]

Thresholding SDP\((X)\) yields a detection algorithm for SPCA\(_D\) with guarantees captured in the following theorem. In [BR13a], a more general model is considered with sub-Gaussian noise. We specialize the more general theorem in [BR13a] to our setup.
Algorithm SPCA-SDP

Inputs: $X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^{d \times n}$

1. Compute the empirical covariance matrix $\hat{\Sigma} = \hat{\Sigma}(X)$ and set
   $$\epsilon = \frac{\log d}{4n} \quad \text{and} \quad \lambda = 4\sqrt{\frac{\log d}{n}}$$

2. Let $f : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}$ be
   $$f(M) = \text{Tr} \left( \hat{\Sigma} M \right) - \lambda \| M \|_1$$
   and compute an $\epsilon$-maximizer $\hat{M}^\epsilon$ of $f(M)$ subject to the constraints that $M$ is symmetric, $M \geq 0$ and $\text{Tr}(M) = 1$

3. Output $v_{\text{SDP}} = \arg \max_{\|u\|_2 = 1} u^\top \hat{M}^\epsilon u$

Figure 9.3: SDP algorithm for weak $\ell_2$ estimation in sparse PCA from [WBS16] and in Theorem 9.3.9.

Theorem 9.3.8 (Theorem 5 in [BR13a]). Suppose that $\delta \in (0, 1)$. Let $X = (X_1, X_2, \ldots, X_n)$ be an instance of $\text{SPCA}_D(n, k, d, \theta)$ and suppose that $\theta \in [0, 1]$ satisfies that
   $$\theta \geq 23 \sqrt{\frac{k^2 \log(d^2/\delta)}{n}}$$
Then the algorithm that outputs $H_1$ if $\text{SDP}(X) \geq 16 \sqrt{\frac{k^2 \log(d^2/\delta)}{n}} + \frac{1}{\sqrt{n}}$ and $H_0$ otherwise has Type I+II error at most $\delta$.

In [WBS16], a semidefinite programming approach is shown to solve the sparse PCA estimation task under the $\ell_2$ norm. As in the proof of Theorem 9.3.6, this yields a weak recovery algorithm for $\text{SPCA}_{WR}$ and $\text{BSPCA}_{WR}$, achieving the tight barrier when $k \lesssim \sqrt{n}$. The SDP algorithm in [WBS16] is shown in Figure 9.3 and its guarantees specialized to the case of Gaussian data are in the following theorem.

Theorem 9.3.9 (Theorem 5 in [WBS16]). Suppose that $k, d$ and $n$ are such that $4 \log d \leq n \leq k^2 d^2 \theta^{-2} \log d$ and $0 < \theta \leq k$. Let $P_\theta$ denote the distribution $N(0, I_d + \theta vv^\top)$ and given some $X = (X_1, X_2, \ldots, X_n) \sim P_\theta^\otimes n$, let $v_{\text{SDP}}(X)$ be the output of SPCA-SDP applied to $X$. Then
   $$\sup_{v \in V_{d,k}} \mathbb{E}_{X \sim P_\theta} L(v_{\text{SDP}}(X), v) \leq \min \left( 16\sqrt{2} + 2, \sqrt{\frac{k^2 \log d}{n \theta^2}}, 1 \right)$$
Using an identical argument to Theorem 9.3.6, we obtain the following theorem.

**Theorem 9.3.10.** Suppose that \( k, d \) and \( n \) are such that \( 4 \log d \leq n \leq k^2 d^2 \theta^{-2} \log d \) and \( 0 < \theta \leq k \). Let \( S(X) \subseteq [d] \) be the set of coordinates of \( v_{\text{SDP}}(X) \) with magnitude at least \( \frac{1}{2\sqrt{k}} \). It follows that

\[
\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim \mathcal{P}_n} \left[ \frac{1}{k} |S(X)\Delta \text{supp}(v)| \right] \leq 8\sqrt{2} \cdot \min \left( (16\sqrt{2} + 2) \sqrt{\frac{k^2 \log d}{n\theta^2}}, 1 \right)
\]

To establish the polynomial-time upper bounds for SPCA and BSPCA in Theorem 2.3.1, it now suffices to consider \( k \gtrsim \sqrt{n} \). First consider the detection problems SPCAD and BSPCAD. The following theorem establishes that a spectral algorithm directly applied to the empirical covariance matrix solves SPCAD when \( \theta \gtrsim 1 \).

**Theorem 9.3.11.** Suppose that \( X = (X_1, X_2, \ldots, X_n) \) is an instance of SPCAD\( (n, k, d, \theta) \) and let \( \hat{\Sigma}(X) \) be the empirical covariance matrix of \( X \). Suppose that \( d \leq cn \) for some constant \( c > 0 \), \( d \to \infty \) and \( n(1 + \theta)^{-2} \to \infty \) as \( n \to \infty \) and it holds that \( \theta > 4\sqrt{c} \). Then the test that outputs \( H_1 \) if \( \lambda_1(\hat{\Sigma}(X)) > 1 + 2\sqrt{c} \) and \( H_0 \) otherwise has Type I+II error tending to zero as \( n \to \infty \).

**Proof.** First observe that \( \hat{\Sigma}(X) = \frac{1}{n} XX^\top \) and thus \( \lambda_1(\hat{\Sigma}(X)) = \frac{1}{n} \sigma_1(X)^2 \). Under \( H_0 \), it follows that \( X \sim N(0, 1)^{d \times n} \). By Corollary 5.35 in [Ver10], it follows that

\[
\mathbb{P} \left[ \lambda_1(\hat{\Sigma}(X)) > 1 + 2\sqrt{c} \right] \leq \mathbb{P} \left[ \sigma_1(X) > \sqrt{n} + 2\sqrt{d} \right] \leq 2e^{-d/2}
\]

Under \( H_1 \), suppose that \( X \sim N(0, I_d + \theta vv^\top)^{\otimes n} \) where \( v \in \mathcal{V}_{d,k} \). As in the proof of Theorem 9.3.7, write \( X_i = \sqrt{\theta} \cdot g_i v + Z_i \) where \( g_1, g_2, \ldots, g_n \sim_{\text{i.i.d.}} N(0, 1) \) and \( Z_1, Z_2, \ldots, Z_n \sim_{\text{i.i.d.}} N(0, I_d) \). Now observe that

\[
v^\top \hat{\Sigma}(X)v = \frac{1}{n} \sum_{i=1}^{n} \langle v, X_i \rangle^2 = \frac{1}{n} \sum_{i=1}^{n} \left( \sqrt{\theta} \cdot g_i + \langle v, Z_i \rangle \right)^2
\]

Note that since \( \|v\|_2 = 1 \), it holds that \( \sqrt{\theta} \cdot g_i + \langle v, Z_i \rangle \sim N(0, 1 + \theta) \) and are independent for \( i = 1, 2, \ldots, n \). Now note that \( (1 + \theta)^{-1} \langle v, X_i \rangle^2 - 1 \) is zero-mean and sub-exponential with norm 1. Therefore Bernstein’s inequality implies that

\[
\mathbb{P} \left[ \lambda_1(\hat{\Sigma}(X)) \leq 1 + 2\sqrt{c} \right] \leq \mathbb{P} \left[ v^\top \hat{\Sigma}(X)v \leq 1 + 2\sqrt{c} \right] \leq \mathbb{P} \left[ \sum_{i=1}^{n} [(1 + \theta)^{-1} \langle v, X_i \rangle^2 - 1] \leq \frac{2n\sqrt{c}}{1 + \theta} \right] \leq 2 \exp \left( -c_1 \cdot \frac{4cn}{(1 + \theta)^2} \right) \to 0 \text{ as } n \to \infty
\]

for some constant \( c_1 > 0 \) and since \( \frac{2n\sqrt{c}}{1 + \theta} < n \). This completes the proof of the theorem. \( \square \)
The algorithm summing the entries of the empirical covariance matrix in Theorem 9.3.7 runs in polynomial time and shows that BSPCA$_D$ can be solved in polynomial time as long as $\theta \gtrsim \frac{\sqrt{n}}{k}$.

Note that this algorithm gives an upper bound matching Theorem 2.3.1 and can detect smaller signal levels $\theta$ in BSPCA$_D$ than the spectral algorithm.

For the recovery problem when $k \gtrsim \sqrt{n}$, the spectral algorithm considered in Theorem 1.1 of [KNV15] achieves the upper bound in Theorem 2.3.1 for the exact recovery problems SPCA$_R$ and BSPCA$_R$. As given in [KNV15], this spectral algorithm is not adaptive to the support size of the planted vector and assumes that the planted sparse vector has nonzero entries of the form $\pm 1/\sqrt{k}$.

We mildly adapt this algorithm to only require that the planted vector is in $V_{d,k}$. The proof of its correctness follows a similar argument as Theorem 1.1 in [KNV15]. We omit details that are identical for brevity.

**Theorem 9.3.12.** Suppose that $k, d$ and $n$ are such that $k, d \to \infty$, $\frac{d}{n} \to c$ and $\frac{k \log d}{n} \to 0$ as $n \to \infty$ for some constant $c > 0$. Let $X = (X_1, X_2, \ldots, X_n) \sim P_{\omega}^{\otimes n}$ and let $\hat{v}$ be the leading eigenvector of $\hat{\Sigma}(X)$. Let $S \subseteq [d]$ be the set of $i$ such that $|\hat{v}_i|^4 \geq \frac{\log d}{kd}$. If $\theta > \sqrt{c}$ is fixed then $S = \text{supp}(v)$ with probability tending to one as $n \to \infty$.

**Proof.** Let $\hat{v} = g \cdot v + \sqrt{1 - g^2} \cdot u$ where $u$ is the projection of $\hat{v}$ onto the space orthogonal to $v$. Also assume that $g \in [0, 1]$, negating $v$ if necessary. By Theorem 4 in [Pau07], it follows that

$$g \to \sqrt{\frac{\theta^2 - c}{\theta^2 + \theta c}} \quad \text{with probability } 1 - o(1) \text{ as } n \to \infty$$

By Theorem 6 in [Pau07], $u$ is distributed uniformly on the $(d - 1)$-dimensional unit sphere of vectors in $\mathbb{R}^d$ orthogonal to $v$. By Lemma 4.1 in [KNV15], it holds that $|u_i| \leq h \sqrt{\frac{\log d}{d}}$ for all $i \in [d]$ with probability tending to one. Condition on this event. Since $\frac{k \log d}{d} \to 0$, it follows that $\frac{\log d}{d} = o \left( \sqrt{\frac{\log d}{kd}} \right)$. Therefore with probability tending to one, each $i \notin \text{supp}(v)$ satisfies that $|\hat{v}_i|^4 < \frac{\log d}{kd}$ for sufficiently large $n, k$ and $d$. Now note that each $i \in \text{supp}(v)$ satisfies that

$$|v_i| \geq \frac{g}{\sqrt{k}} - \sqrt{1 - g^2} \cdot |u_i| \geq \frac{g}{\sqrt{k}} - h \sqrt{\frac{\log d}{d}} \geq 4 \sqrt{\frac{\log d}{kd}}$$

for sufficiently large $n, k$ and $d$ with probability tending to one. This is because $g = \Omega(1)$ with probability $1 - o(1)$ and $\frac{1}{\sqrt{k}} = \omega \left( \sqrt{\frac{\log d}{kd}} \right)$. Therefore it follows that $S = \text{supp}(v)$ with probability tending to one as $n \to \infty$. \hfill \Box

Note that the theorem statement assumes that $d/n \to c$ where $c > 1$. The algorithm can more generally accommodate inputs with $d = \Theta(n)$ by padding the input $X$ with i.i.d. $N(0, 1)$ entries so that $d/n \to c$ where $c > 1$. 

Chapter 10

Conclusions and Future Directions

We have introduced a number of new techniques for average-case reductions and applied them to demonstrate that a web of average-case reductions among problems with statistical-computational gaps is possible even for showing strong computational lower bounds. This suggests that using natural problems as intermediates as is commonplace in worst-case complexity can be useful in average-case complexity. This work also has left number of problems open, including the following.

1. Collisions between support elements in REFLECTION-CLONING causes our formulations of SSW\textsubscript{D}, ROS\textsubscript{D}, SSBM\textsubscript{D} and SPCA\textsubscript{D} to be as composite hypothesis testing problems rather than the canonical simple hypothesis testing formulations. Is there an alternative reduction from planted clique that can strengthen these lower bounds to hold for the simple hypothesis testing analogues?

2. All previous planted clique lower bounds for SPCA\textsubscript{D} are not tight over the parameter regime \(k \ll \sqrt{n}\). Is there a reduction from planted clique yielding tight computational lower bounds for SPCA\textsubscript{D} in this highly sparse regime?

3. Is there a polynomial time algorithm for the recovery variant of SSBM matching the computational barrier for the detection variant?

4. Can the PDS recovery conjecture be shown to follow from the planted clique conjecture?
Bibliography


