Computational and statistical lower bounds for low-rank estimation under general inhomogeneous noise

Debsurya De* and Dmitriy Kunisky†

Department of Applied Mathematics & Statistics, Johns Hopkins University

October 9, 2025

Abstract

Recent work has generalized several results concerning the well-understood spiked Wigner matrix model of a low-rank signal matrix corrupted by additive i.i.d. Gaussian noise to the *inhomogeneous* case, where the noise has a variance profile. In particular, for the special case where the variance profile has a block structure, a series of results identified an effective spectral algorithm for detecting and estimating the signal, identified the threshold signal strength required for that algorithm to succeed, and proved information-theoretic lower bounds that, for some special signal distributions, match the above threshold. We complement these results by studying the computational optimality of this spectral algorithm. Namely, we show that, for a much broader range of signal distributions, whenever the spectral algorithm cannot detect a low-rank signal, then neither can any low-degree polynomial algorithm. This gives the first evidence for a computational hardness conjecture of Guionnet, Ko, Krzakala, and Zdeborová (2023). With similar techniques, we also prove sharp information-theoretic lower bounds for a class of signal distributions not treated by prior work. Unlike all of the above results on inhomogeneous models, our results do not assume that the variance profile has a block structure, and suggest that the same spectral algorithm might remain optimal for quite general profiles. We include a numerical study of this claim for an example of a smoothly-varying rather than piecewise-constant profile. Our proofs involve analyzing the *graph sums* of a matrix, which also appear in free and traffic probability, but we require new bounds on these quantities that are tighter than existing ones for non-negative matrices, which may be of independent interest.

*Email: dde4@jhu.edu

†Email: kunisky@jhu.edu.

Contents

1 Introduction

1.1 Spiked Matrix Models

Spiked matrix models are one of the fundamental examples of high-dimensional statistics, demonstrating in a relatively simple setting the important phenomena of sharp phase transitions in the effectiveness of natural estimators and statistical-to-computational gaps. The simplest version of spiked matrix model is the *spiked Wigner matrix model*, where, for $x \sim \mathcal{P}_N$ for some probability measure $\mathcal{P}_N$ over $\mathbb{R}^N$ normalized such that $\|x\| \approx \sqrt{N}$, we observe

$$Y = \frac{\beta}{\sqrt{N}} xx^\top + W,$$

where $\beta > 0$ is a scalar parameter and $W_{ij} = W_{ji} \sim \mathcal{N}(0,1)$ independently for all $1 \le i \le j \le N$.[1] We call x the *signal* of such a model. We observe this signal only through Y, and consider one of two computational tasks:

1. **Hypothesis testing** between $Y \sim \mathbb{P}_N$ the law of Y above with $\beta = c > 0$ for some given constant c (which we call the *planted* or *alternative* model) and $Y \sim \mathbb{Q}_N$ the law of Y with $\beta = 0$ (which we call the *null* model).
2. **Estimating** x from the observation Y.

We will mostly focus on hypothesis testing here, and introduce the following notions of success for that task. The first is a sensible general definition.

Definition 1.1 (Strong detection)**.** *We say that a sequence of functions $t_N : \mathbb{R}^{N \times N}_{\text{sym}} \to \{0,1\}$, which in this context we call* tests*, achieves* strong detection *in the above setting if*

$$\mathbb{P}_N[t_N(Y) = 1] + \mathbb{Q}_N[t_N(Y) = 0] = o(1).$$

The following more quantitative version is the one we will work with in our main results.

Definition 1.2 (Strong separation)**.** *We say that a sequence of functions $f_N : \mathbb{R}^{N \times N}_{\text{sym}} \to \mathbb{R}$ achieves* strong separation *in the above setting if*

$$\mathop{\mathbb{E}}_{Y \sim \mathbb{P}_N} f_N(Y) - \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}_N} f_N(Y) = \omega\left(\sqrt{\mathop{\text{Var}}_{Y \sim \mathbb{Q}_N} f_N(Y)} + \sqrt{\mathop{\text{Var}}_{Y \sim \mathbb{P}_N} f_N(Y)} \right).$$

If $f_N(Y)$ achieve strong separation, then it is easy to check that setting $t_N(Y) := \mathbf{1}\{f_N(Y) > \tau\}$ for a suitable threshold τ achieves strong detection, in essentially the same runtime as it takes to compute f_N. Conversely, if $t_N(Y)$ achieve strong detection then they themselves also achieve strong separation. Thus, the two notions are equivalent over arbitrary f_N and t_N; however, our main results and some of the prior work we discuss below concern strong separation by special classes of f_N, which does not in general have a direct correspondence with strong detection by a special class of t_N.

Since for large β one expects the rank-one deformation above to create an outlier eigenvalue in Y, it is natural to try to perform these tasks using the top eigenpair of Y. We write $\lambda_1(Y) \ge \cdots \ge \lambda_N(Y)$ for the sorted eigenvalues of a matrix, and $v_i(Y)$ for the unit eigenvector associated to eigenvalue $\lambda_i(Y)$. The following important result of random matrix theory, a variant

[1] Different prior works make different choices about how to treat the diagonal of W; this choice is almost always inconsequential in such models for the kinds of questions we will discuss.

of the celebrated *Baik–Ben Arous–Péché (BBP) transition*, characterizes how well $\lambda_1(Y)$ and $v_1(Y)$ perform at testing and estimation, respectively. We introduce a different normalization of Y,

$$\widehat{Y} := \frac{1}{\sqrt{N}} Y,$$

which is useful as standard random matrix theory results imply that $\|\widehat{Y}\| = O(1)$ with high probability [FK81]. Below, $\xrightarrow{\mathbb{P}}$ denotes convergence in probability of random variables to constants.

Theorem 1.3 ([FP07, CDMF09])**.** *In the above setting, suppose that* $\|x\|/\sqrt{N} \xrightarrow{\mathbb{P}} 1$*. Then, the following hold.*

- *If* $0 \le \beta \le 1$*, then*

$$\lambda_1(\widehat{Y}) \xrightarrow{\mathbb{P}} 2,$$
$$\left| \left\langle v_1(\widehat{Y}), \frac{x}{\sqrt{N}} \right\rangle \right| \xrightarrow{\mathbb{P}} 0.$$

- *If* $\beta > 1$*, then*

$$\lambda_1(\widehat{Y}) \xrightarrow{\mathbb{P}} \beta + \frac{1}{\beta} > 2,$$
$$\left| \left\langle v_1(\widehat{Y}), \frac{x}{\sqrt{N}} \right\rangle \right| \xrightarrow{\mathbb{P}} \sqrt{1 - \frac{1}{\beta^2}} > 0.$$

Thus the theorem implies that $\lambda_1(\widehat{Y})$ can be used to hypothesis test with high probability of success (indeed, it can be shown with a bit more argument that $\lambda_1(\widehat{Y})$ achieves strong separation in this setting) and $v_1(\widehat{Y})$ can be used to estimate x non-trivially if and only if

$$\beta > \beta_* := 1.$$

It is natural to ask whether this performance is optimal or not: maybe a better statistical procedure can achieve the above results for smaller β. The answer depends on $\mathcal{P}_N$, the "prior" distribution on x. This distribution (though not x itself) is assumed to be known to the statistical algorithm under consideration. So, for instance, if $\mathcal{P}_N = \delta_{x_\star}$ is concentrated on a single value $x_\star$, then of course estimation is trivially simple since an algorithm can just output $x_\star$, and one can show that hypothesis testing is also easy by thresholding $x_\star^\top Y x_\star$. However, the following result shows that, for "sufficiently random" $\mathcal{P}_N$, in fact no procedure can improve on the simple spectral algorithm above.

Theorem 1.4 ([PWBM18])**.** *Suppose that either* $\mathcal{P}_N = \mathrm{Unif}(\{x \in \mathbb{R}^N : \|x\| = \sqrt{N}\})$ *or* $\mathcal{P}_N = \mathrm{Unif}(\{\pm 1\}^N)$*. If* $\beta < \beta_* = 1$*, then there is no sequence of tests* t_N *achieving strong detection and no sequence of functions* f_N *achieving strong separation in the above model.*

This, too, can break down for certain $\mathcal{P}_N$. For instance, a long line of work has demonstrated that, if $\mathcal{P}_N$ is supported on vectors with pN non-zero entries for sufficiently small p (the setting of *sparse PCA*), then brute force searches requiring time $\exp(\Omega(N))$ over sparse vectors can succeed at hypothesis testing and estimation for certain values of $\beta < 1$, while conjecturing that this cannot be done by efficient algorithms [LKZ15a, LKZ15b, KXZ16, BDM+16, BMV+18, EAKJ20, LM19]. Corroborating these conjectures, the following result says that, for an even wider range of priors $\mathcal{P}_N$, the threshold $\beta = 1$ is the best for a large class of algorithms, at least for hypothesis testing (see also [MW23] for related but more intricate results on estimation). Specifically, these results and the new ones we prove below consider the class of *low-degree polynomial* separating statistics.

Theorem 1.5 ([KWB19]). *Suppose that $\mathcal{P}_N = \pi^{\otimes N}$ for some π a bounded probability measure on $\mathbb{R}$ with mean 0 and variance 1. If $\beta < \beta_* = 1$, then there is no sequence of polynomials f_N with $\deg(f_N) = o(N/\log N)$ that achieve strong separation in the above setting.*

Since a polynomial of degree D in the roughly N^2 variables of Y requires time $N^{\Theta(D)}$ to evaluate naively, this suggests that, taking polynomials as a proxy for all functions computable under a given runtime budget, even when it is possible as in sparse PCA, it should take nearly exponential time in N to compute a strongly separating statistic when $\beta < \beta_*$. We will subscribe in this paper to the belief that this kind of heuristic reasoning about low-degree polynomial separating statistics is accurate; as detailed in the surveys [KWB19, Wei25], it has given correct predictions of computational cost and hardness for numerous similar problems in prior work.

1.2 Inhomogeneous Spiked Matrix Models

The spiked Wigner matrix model is unsatisfying in various ways. One might object that one assumes that the noise is Gaussian, i.i.d., and applied additively to the signal, all of which might be inaccurate assumptions for modeling various statistical scenarios. Recent work has begun to generalize this model to allow for various other noise distributions; in addition to the references we discuss in detail below, see [BGN11, CDMFF11, CDM16, BHMS22, BCMS23, ZM24, BCXM25] for some other generalizations.

One of these directions has been to consider a noise matrix W whose entries are still independent and Gaussian, but where the variances are *inhomogeneous*, being allowed to vary from entry to entry. We summarize this in the following model, which will be the subject of this paper. We also relax the precise structural assumption on the rank-one signal from the above discussion, instead allowing for now the signal to be an arbitrary matrix (later we will consider it being both rank-one and more generally low-rank).

Definition 1.6 (Inhomogeneous spiked Wigner matrix model). *For each $N \geq 1$, let $\mathcal{X}_N$ be a probability measure on $\mathbb{R}^{N\times N}_{\mathrm{sym}}$ and let $\Delta = \Delta^{(N)} \in (\mathbb{R}_{>0} \cup \{+\infty\})^{N\times N}_{\mathrm{sym}}$. Define*

$$\mathcal{J} = \mathcal{J}^{(N)} := \{(i,j) : 1 \leq i \leq j \leq N, \Delta_{ij} \neq +\infty\}.$$

Associated to these parameters, define two probability measures $\mathbb{Q}_N$ and $\mathbb{P}_N$ on $\mathbb{R}^{\mathcal{J}}$:

1. *To draw $Y \sim \mathbb{Q}_N$, draw $Y_{ij} = Y_{ji} \sim \mathcal{N}(0, \Delta_{ij})$ independently for each $(i,j) \in \mathcal{J}$.*
2. *To draw $Y \sim \mathbb{P}_N$, first draw $X \sim \mathcal{X}_N$, and then draw $Y_{ij} = Y_{ji} \sim \mathcal{N}(X_{ij}, \Delta_{ij})$ independently for each $(i,j) \in \mathcal{J}$.*

We call Δ the variance profile *of the model. We call the model* finite-variance *if $\Delta_{ij} \neq +\infty$ for all $i,j \in [N]$, in which case it may be viewed as a pair of probability measures on $\mathbb{R}^{N\times N}_{\mathrm{sym}}$*

We allow for some entries to be observed with "infinite variance" of noise, by which in practice we just mean that those entries are not observed at all and we only observe a subset $\mathcal{J}$ of entries of a symmetric matrix. Some other work has interpreted such models as having "censorship" of some entries (e.g., [ABBS14, SLKZ15]), but we will see that, formally, it is consistent and convenient for us to view these entries as indeed having noise of variance equal to $+\infty$.

Note that if we take $\Delta_{ij} = 1$ for all $i,j \in [N]$ and $X \sim \mathcal{X}_N$ drawn as $X = \beta xx^\top$ for $x \sim \mathcal{P}_N$, then we recover the homogeneous model discussed above in Section 1.1. $\mathbb{Q}_N$ will serve as a null model in this general setting; in the rank-one case above, it is what we would obtain if we took $\beta = 0$. Such models seem to have first appeared in the concurrent series of works [BR20, BR22] and

[ACCM21b, ACCM21a, ACCM22], the former motivated by statistics and the latter by statistical physics. Further important results on this model that we will discuss below have been obtained in the recent works [PKK24, MKK24, BCSvH24, GKKZ25].

Previous work on such statistical models appears to have nearly exclusively focused on the case of finite variances having the following special *block structure* of the variance profile (the only exception we are aware of is [BM21], which as we will see studies a spectral algorithm that is likely suboptimal). In contrast, all of our results allow for infinite variances, and some of our results also do not require a block-structured variance profile. General variance profiles are actually quite frequently considered in random matrix theory outside of statistical applications; see Section 1.5 for references to such results.

Definition 1.7 (Block-structured matrices)**.** *Suppose that* $A^{(N)} \in (\mathbb{R}_{>0} \cup \{+\infty\})^{N\times N}_{\mathrm{sym}}$ *for each* $N \geq 1$*. Suppose that there exist* $n \geq 1$*,* $\rho_1, \dots, \rho_n \in (0,1)$ *with* $\rho_1 + \cdots + \rho_n = 1$*, and a matrix* $\bar{A} \in (\mathbb{R}_{>0} \cup \{+\infty\})^{n\times n}_{\mathrm{sym}}$*, such that the following hold: for each* N*, there is a function* $q : [N] \to [n]$ *so that* $A^{(N)}_{ij} = \bar{A}_{q(i)q(j)}$ *for all* $i, j \in [N]$*, and for each* $i \in [n]$*,* $|q^{-1}(i)|/N \to \rho_i$ *as* $N \to \infty$*. Then, we say that the sequence* $A^{(N)}$ *is* block-structured *with parameters* $(\bar{A}, \rho_1, \dots, \rho_n)$*.*

When $\Delta^{(N)}$ *is block-structured for an inhomogeneous spiked Wigner matrix model, we also call the model itself block-structured. We always use a bar to denote the matrix parameter associated to a block-structured sequence. For instance, we will talk about matrices* $\Delta^{(N)}$ *and* $\Phi^{(N)}$ *(the latter to be defined below) being block-structured, and* $\bar{\Delta}$ *and* $\bar{\Phi}$ *will be the associated parameters.*

It seems that the reason for the focus on block-structured models in prior work is that, as the definition makes clear, they have a well-defined entrywise limiting behavior. Our results will show, however, that such a strong notion of limit for the variance profile sequence need not hold to draw useful conclusions about the tractability of the associated spiked models.

In either the general inhomogeneous spiked Wigner matrix model or its block-structured special case, it is natural to ask all of the questions we have mentioned above in Section 1.1: what strength of signal is required for testing or estimation to be tractable, either by arbitrary functions or computationally efficient ones? Perhaps one expects that, if the signal $X \sim \mathcal{X}_N$ is low-rank and positive semidefinite (like $X = \frac{\beta}{\sqrt{N}} xx^\top$ from our earlier discussion), then it should be a good idea to simply use $\lambda_1(\widehat{Y})$ and $v_1(\widehat{Y})$ again. Or, perhaps one should "whiten" the noise and divide each entry Y_{ij} by $\sqrt{\Delta_{ij}}$ before computing the spectral decomposition. Although either approach succeeds for sufficiently large β, it turns out that neither is optimal in the inhomogeneous case (see discussion in [GKKZ25, MKK24] as well as our Section 5.5).

Motivated by the analysis of approximate message passing, [PKK24] proposed another spectral algorithm, as follows. We introduce the entrywise reciprocal of Δ, which will be useful throughout our discussion:

$$\Phi = \Phi^{(N)} := \Delta^{\odot -1}, \quad \text{i.e.,} \quad \Phi_{ij} = \left\{ \begin{array}{ll} 1/\Delta_{ij} & \text{if } \Delta_{ij} \neq +\infty, \\ 0 & \text{if } \Delta_{ij} = +\infty \end{array} \right\}.$$

Here and throughout, $\odot$ denotes the Hadamard entrywise product or power of matrices. Note that we work here and always under the convention that

$$\frac{1}{+\infty} = 0,$$

and with this convention $\Phi^{(N)} \in (\mathbb{R}_{\geq 0})^{N\times N}_{\mathrm{sym}}$, never having infinite entries even if $\Delta^{(N)}$ does.

In our normalization, where $\|x\|/\sqrt{N} \xrightarrow{\mathbb{P}} 1$ for $x \sim \mathcal{P}_N$ and $X \sim \mathcal{X}_N$ is $X = \frac{\beta}{\sqrt{N}} xx^\top$ for such x and some $\beta > 0$, the algorithm proposed by [PKK24] uses the matrix

$$\mathcal{H}(Y) := \frac{\beta}{\sqrt{N}} \Phi \odot Y - \beta^2 \operatorname{diag}\left(\frac{1}{N} \Phi 1_N\right), \tag{1}$$

where 1_N denotes the all-ones vector. The algorithm is quite unusual on its face, as the first term, in our above terminology, seems to "over-whiten" the observation Y. We remark that, unlike both alternatives discussed above, to implement this algorithm requires knowledge of the parameter β (this point is perhaps less clear in the notation used by the previous works [PKK24, MKK24, BCSvH24] studying this algorithm, who do not introduce β as a separate parameter, though [PKK24] make the same point in their Remark 3.1). The analysis of a spectral algorithm using this matrix, as conjectured by [PKK24] and proved concurrently in the two references below, is as follows.

Theorem 1.8 ([MKK24, BCSvH24])**.** *Consider a block-structured finite-variance inhomogeneous spiked Wigner matrix model and suppose that the block structure of Δ has parameters $(\bar{\Delta}, \rho_1, \dots, \rho_n)$. Suppose that $x \sim \mathcal{P}_N$ has i.i.d. entries of mean zero and variance 1, and that $X \sim \mathcal{X}_N$ is $X = \frac{\beta}{\sqrt{N}} xx^\top$ for x as above and some $\beta > 0$. Write $\bar{\Phi} := \bar{\Delta}^{\odot -1}$, and define*

$$\beta_* = \beta_*(\bar{\Delta}, \rho_1, \dots, \rho_n) := \sqrt{\frac{1}{\lambda_1(\operatorname{diag}(\rho)^{1/2} \bar{\Phi} \operatorname{diag}(\rho)^{1/2})}}. \tag{2}$$

Then, the following hold:

- *If $\beta \le \beta_*$, then*

$$\lambda_1(\mathcal{H}(Y)) - \lambda_2(\mathcal{H}(Y)) \xrightarrow{\mathbb{P}} 0,$$
$$\left| \left\langle v_1(\mathcal{H}(Y)), \frac{x}{\sqrt{N}} \right\rangle \right| \xrightarrow{\mathbb{P}} 0.$$

- *If $\beta > \beta_*$, then there exist $\delta = \delta(\beta) > 0$ and $\varepsilon = \varepsilon(\beta) > 0$ such that*

$$\lambda_1(\mathcal{H}(Y)) - \lambda_2(\mathcal{H}(Y)) \xrightarrow{\mathbb{P}} \delta > 0,$$
$$\left| \left\langle v_1(\mathcal{H}(Y)), \frac{x}{\sqrt{N}} \right\rangle \right| \xrightarrow{\mathbb{P}} \varepsilon > 0.$$

We note that, while the result above is restricted to finite variances, the quantity β_* defined in (2) is still well-defined even if some variances are infinite, which we will use below. The nature of and reason for this threshold β_* should still be mysterious; we will clarify how this value arises when we discuss our main results and analysis below.

For now, two basic observations will suffice. First, note that the values in $\bar{\Phi}$ are the reciprocals of the noise variances, so they get smaller as the amount of noise increases, and thus β_* gets larger as the amount of noise increases, meaning that the problem becomes harder (at least for this spectral algorithm), as we expect. Second, note that the result is very similar to Theorem 1.3. The reason for the slightly different form of the result on $\lambda_1(\mathcal{H}(Y))$ compared to Theorem 1.3 on $\lambda_1(\widehat{Y})$ is that, unlike for $\widehat{Y}$, the shape of the "bulk" eigenvalue distribution of $\mathcal{H}(Y)$ depends on β by construction. So, our condition on the top eigenvalue captures not whether $\lambda_1(\mathcal{H}(Y))$ is unusually large compared to the case $\beta = 0$, but whether it is an outlier from this bulk distribution. This can still be used to

achieve strong detection when $\beta > \beta_*$ by setting an appropriate threshold for $\lambda_1(\mathcal{H}(Y))$ depending on β. Actually, Theorem 1.8 generalizes Theorem 1.3, since for Δ the all-ones matrix we recover the threshold $\beta_* = 1$, and while we have not stated them, the full results of [MKK24] give explicit though complicated descriptions of δ and ε appearing above, as well as the high-probability limits of $\lambda_i(\mathcal{H}(Y))$ for $i \in \{1, 2\}$, making the result equally precise to Theorem 1.3. Thus, just as in that case, Theorem 1.8 should be read as saying that spectral algorithms using $\mathcal{H}(Y)$ succeed if and only if $\beta > \beta_*$.

1.3 Main Results

Our goal will be to probe whether the above algorithm—which, without some study of its derivation via approximate message passing, appears quite unnatural—is optimal. The threshold β_* in the block-structured case is also a natural one for the performance of approximate message passing algorithms, and on this basis [GKKZ25] conjectured that efficient algorithms should not be able to hypothesis test or estimate effectively when $\beta < \beta_*$.

As in Theorem 1.4 on homogeneous spiked Wigner models cited above, only in some cases will we be able to show that this algorithm is *statistically* or *information-theoretically* optimal, meaning that no procedure regardless of runtime can achieve strong detection when $\beta < \beta_*$. These results of ours will generalize Theorem 1.4 to the inhomogeneous setting, complementing and extending some information-theoretic lower bounds of [GKKZ25], which apply only to the block-structured case and then only in one very special case mentioned below match the threshold β_*.

Paralleling Theorem 1.5 for the homogeneous case, we will also give evidence by analyzing low-degree polynomials that, for a much wider range of prior distributions $\mathcal{P}_N$, this algorithm is *computationally* optimal, suggesting that no polynomial-time procedure can hypothesis test when $\beta < \beta_*$ and giving evidence for the conjecture of [GKKZ25]. Notably, going beyond that conjecture, our results will treat very general variance profiles which need not be block-structured, and will suggest that in fact a threshold generalizing the above β_* might be the correct computational threshold for a broader class of inhomogeneous spiked Wigner matrix models.

A particularly important role will turn out to be played by the largest eigenvalue of Φ. Since the entries of Φ are non-negative, this eigenvalue is strictly positive unless we are in the trivial situation where $\Phi = 0$, which only arises when $\mathcal{J} = \varnothing$ and we make no observation at all. Recall that we work in a scaling where the entries of Φ have magnitude $\Phi_{ij} = \Theta(1)$, whereby we expect the largest eigenvalue to be of size $\Theta(N)$. We have the following description of the threshold β_* in terms of this eigenvalue, which we prove in Section 3.1.

Proposition 1.9. *In a block-structured inhomogeneous spiked Wigner model,*

$$\widehat{\mu}_1 := \lim_{N\to\infty} \frac{\lambda_1(\Phi^{(N)})}{N} = \lambda_1(\mathrm{diag}(\rho)^{1/2}\,\bar{\Phi}\,\mathrm{diag}(\rho)^{1/2}).$$

This is related to β_* defined for such models in (2) by

$$\beta_* = \frac{1}{\sqrt{\widehat{\mu}_1}}.$$

We now give our main result on block-structured models. In this case, we give a computational lower bound under mild conditions on the prior distribution $\mathcal{X}$, a statistical lower bound under stronger conditions, and allow for signals of any constant rank κ. Here and below, we apply the definitions of strong detection and strong separation to the probability measures $\mathbb{P}_N$ and $\mathbb{Q}_N$ described in Definition 1.6.

Theorem 1.10 (Block-structured variance profile)**.** *Consider the inhomogeneous spiked Wigner matrix model (Definition 1.6) with block-structured (but possibly infinite) variances. Suppose that $\mathcal{P}_N$ is a probability measure over $\mathbb{R}^{N\times\kappa}$ such that $V \sim \mathcal{P}_N$ has i.i.d. rows drawn from some probability measure ν on $\mathbb{R}^\kappa$. Let $X \sim \mathcal{X}_N$ be $X = \frac{\beta}{\sqrt{N}}VV^\top$ for $V \sim \mathcal{P}_N$ and some $\beta > 0$. Then, the following hold, for β_* as defined in (2):*

1. ***Computational lower bound:*** *Suppose that ν satisfies the conditions*

   (a) $\mathbb{E}_{x\sim\nu}[x] = 0$.

   (b) $\|\mathrm{Cov}_{x\sim\nu}[x]\| = \|\mathbb{E}_{x\sim\nu} xx^\top\| \le 1$.

   (c) ν has bounded support.

   If $\beta < \beta_*$, then there is no sequence of polynomials $f_N \in \mathbb{R}[Y]$ with $\deg(f_N) = o(N/\log N)$ that achieve strong separation.

2. ***Statistical lower bound:*** *Suppose that the above conditions on ν are satisfied, as well as that*

   (d) When $x, x' \sim \nu$ are i.i.d., then $x \otimes x'$ is 1-moment-subgaussian (see Definition 3.13).

   If $\beta < \beta_*$, then there is no sequence of functions f_N that achieve strong separation, there is no sequence of tests t_N that achieve strong detection, and $\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) = O(1)$.[2]

We make a few remarks on these results. Part 1 of the result generalizes Theorem 1.5 to the inhomogeneous setting, and Part 2 does the same for Theorem 1.4. As mentioned above, the conclusion of Part 1 should be read as suggesting that nearly exponential time is required to compute a strongly separating statistic. Our setting is a little bit more general than both those results as stated above and the setting treated by [PKK24], because we allow for higher-rank signals. This setting was also used (to prove different kinds of results) by [GKKZ25], and for example Theorem 1.5 for computational lower bounds in the homogeneous case was extended to higher rank in [BBK+21]. Lastly, a simple example of ν satisfying Condition (d) of Part 2 is, in the case $\kappa = 1$, $\nu = \mathrm{Unif}(\{\pm 1\})$ (as included in Theorem 1.4 for the homogeneous case). We note that [GKKZ25] (specifically, their Lemma 2.15) also prove statistical lower bounds along the lines of Part 2 of our result, except pertaining to estimation and the minimum mean squared error achievable by any estimator. In general the threshold they obtain for β for these lower bounds to hold is smaller than β_*, except in the very special case $\kappa = 1$ and $\nu = \mathcal{N}(0, 1)$ (which, however, our result does not cover).

We next give our results for the case of general variance profiles. In this case, our first description of β_* is no longer sensible, since we do not have access to the block structure parameters $\bar{\Phi}$ and ρ_i anymore. But, we may mimic the formulation of Proposition 1.9. To that end, we introduce a mild structural assumption, which we always make from now on without further mention:

Assumption A. *In an inhomogeneous spiked Wigner matrix model with general variance profile (possibly having infinite entries), the following limit exists:*

$$\widehat{\mu}_1 := \lim_{N\to\infty} \frac{\lambda_1(\Phi^{(N)})}{N} > 0.$$

[2] As we have mentioned, the first two statements here are equivalent, and as we state in Proposition 3.24, both are implied by the third.

If this is not the case, then the lim inf and lim sup of the above sequence are different, which our results will show suggests that the easiest subsequence of spiked matrix models in a given sequence is considerably easier than the hardest subsequence. Thus, in this situation, we suggest that the hardness of the sequence taken together is unclear and one should consider one or both of these subsequences individually instead. Finally, under Assumption A, we define a generalized critical β,

$$\beta_* := \frac{1}{\sqrt{\widehat{\mu}_1}} = \lim_{N\to\infty} \sqrt{\frac{N}{\lambda_1(\Phi^{(N)})}}. \tag{3}$$

By Proposition 1.9, this indeed generalizes the threshold from the block-structured case.

To handle the setting of general variance profiles, we slightly restrict the class of signals we consider: instead of low-rank signal matrices formed as the Gram matrices of i.i.d. vectors, we consider rank-one signals formed as $xx^\top$ for x having i.i.d. entries. Also, we introduce the mild but technically important assumption that the distribution of those entries of x is *symmetric*. It seems likely that both of these assumptions can be relaxed, but also we expect this to require a considerably different technical approach. We include some discussion of this following our proof in Section 5.

Theorem 1.11 (General variance profile)**.** *Consider the inhomogeneous spiked Wigner matrix model with arbitrary variance profile (possibly both non-block-structured and having some infinite entries). Suppose that $\mathcal{P}_N$ is a probability measure over $\mathbb{R}^N$ such that $x \sim \mathcal{P}_N$ has i.i.d. entries drawn from some probability measure ν on $\mathbb{R}$. Let $X \sim \mathcal{X}_N$ be $X = \frac{\beta}{\sqrt{N}} xx^\top$ for $x \sim \mathcal{P}_N$ and some $\beta > 0$. Then, the following hold:*

1. ***Computational lower bound:*** *Suppose that ν satisfies the conditions:*

   (a) $\mathbb{E}_{x\sim\nu}[x] = 0$.

   (b) $\mathbb{E}_{x\sim\nu}[x^2] \le 1$.

   (c) ν is symmetric (i.e., $x \sim \nu$ has the same law as $-x$).

   (d) If $x, x' \sim \nu$ are independent, then xx' is σ^2-subgaussian for some $\sigma^2 > 0$.

   Suppose also that the variance profiles $\Delta^{(N)}$ satisfy the conditions, for some $D(N) \in \mathbb{N}$,

   (e) There exists a constant $B > 0$ such that, for all $N \ge 1$ and $1 \le i, j \le N$, $\Delta_{ij}^{(N)} \ge \frac{1}{B}$.

   (f) $D(N)N^2/\lambda_1(\Phi^{(N)})^3 \to 0$ as $N \to \infty$.

   If $\beta < \beta_*$, then there is no sequence of polynomials $f_N \in \mathbb{R}[Y]$ with $\deg(f_N) \le D(N)$ that achieve strong separation.

2. ***Statistical lower bound:*** *Suppose that Conditions (a), (b), (c), and (e) above are satisfied, as well as:*

   (d′) When $x, x' \sim \nu$ are i.i.d., then xx' is 1-moment-subgaussian (see Definition 3.13).

   If $\beta < \beta_*$, then there exists no sequence of functions f_N that achieve strong separation, there exists no sequence of functions t_N that achieve strong detection, and $\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) = O(1)$.

We have stated the assumptions to make the results more general, but let us point out two simplifications in typical situations. First, if the measure ν is bounded (as in Theorem 1.10), then the subgaussianity in Condition (d) is automatically satisfied. Condition (d′), on the other hand, may be viewed as a variation on a measure being 1-subgaussian, defined in terms of bounds on

each moment rather than the moment generating function (see Proposition 3.14 and surrounding discussion). Second, if we have a uniform upper bound on the variance profile, say that for some $B > 0$ not depending on N we have

$$\frac{1}{B} \leq \Delta_{ij}^{(N)} \leq B \text{ for all } N \geq 1 \text{ and } i, j \in [N], \tag{4}$$

then the same bound holds for the entries of $\Phi^{(N)}$, whereby $\lambda_1(\Phi^{(N)}) \geq \frac{1}{B}N$, and we may take any degree sequence $D(N) = o(N)$ in Condition (f). But, we emphasize that our result is considerably more general and still gives non-trivial lower bounds for much more uneven variance profiles. For instance, even if we only have $\Delta_{ij}^{(N)} \leq N^{1/3-\delta}$ for all $N \geq 1$ and $i, j \in [N]$ for some $\delta > 0$, then by the same token we have $\lambda_1(\Phi^{(N)}) \geq N^{2/3+\delta}$ and we obtain a lower bound against strong separation by degree $D(N) = o(N^{3\delta})$ polynomials. Following our previous heuristic discussion, this suggests that time close to $\exp(\Omega(N^{3\delta}))$ is required to compute a strongly separating statistic, and in particular that this cannot be done in polynomial time.

1.4 Additional Results

1.4.1 Growth of Low-Degree χ^2-Divergence

To prove our computational lower bounds, we will use that degree $D = D(N)$ polynomials cannot achieve strong separation provided that a quantity we call the *low-degree χ^2-divergence* and denote $\chi^2_{\leq D}(\mathbb{P}_N \mid \mathbb{Q}_N) = O(1)$ as $N \to \infty$ (see Definition 3.23 and Proposition 3.24). It is tempting to view the opposite case, where $\chi^2_{\leq D}(\mathbb{P}_N, \mathbb{Q}_N) \to \infty$, as evidence that low-degree polynomials *do* achieve strong separation. Unfortunately, this is not the case: the low-degree χ^2-divergence is only a "one-sided" quantity, and its divergence only ensures that

$$\mathop{\mathbb{E}}_{Y \sim \mathbb{P}_N} f_N(Y) - \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}_N} f_N(Y) = \omega\left(\sqrt{\mathop{\mathrm{Var}}_{Y \sim \mathbb{Q}_N} f_N(Y)}\right).$$

In particular, it is possible that

$$\mathop{\mathbb{E}}_{Y \sim \mathbb{P}_N} f_N(Y) - \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}_N} f_N(Y) = O\left(\sqrt{\mathop{\mathrm{Var}}_{Y \sim \mathbb{P}_N} f_N(Y)}\right),$$

in which case strong separation does not hold because f_N fluctuates too much under the planted model. It is not hard to construct pathological situations where this happens, but in fact this behavior has also been observed in some reasonable models in recent work [BAH+22, COGHK+22, DMW23, DDL23].

Still, the low-degree χ^2-divergence being unbounded gives some partial evidence that low-degree polynomials or otherwise tractable functions may achieve strong separation and detection, and at the very least shows that our analysis of the divergence is tight, so we give converse results in variants of the settings of Theorems 1.10 and 1.11 stating that this happens when $\beta > \beta_*$. Of course, it would be more convincing to show that a specific low-degree polynomial actually achieves strong separation, but we expect such polynomials to be related to the spectral algorithm using $\mathcal{H}(Y)$, whose behavior at least in the case of general variance profiles (the setting of Theorem 1.13 below) is not yet well-understood, as we briefly discuss in Section 5.5.

Theorem 1.12. *Consider a block-structured inhomogeneous spiked Wigner matrix model. Suppose that $\mathcal{P}_N$ is a probability measure over $\mathbb{R}^{N \times \kappa}$ such that $V \sim \mathcal{P}_N$ has i.i.d. rows drawn from some probability measure ν on $\mathbb{R}^\kappa$. Let $X \sim \mathcal{X}_N$ be $X = \frac{\beta}{\sqrt{N}} VV^\top$ for $V \sim \mathcal{P}_N$ and some $\beta > 0$. Suppose that ν satisfies the conditions:*

(a) $\mathbb{E}_{x\sim\nu}[x] = 0$.

(b) $\|\mathrm{Cov}_{x\sim\nu}[x]\| = \|\,\mathbb{E}_{x\sim\nu}\, xx^\top\| = 1$.

(c) ν *has bounded support.*

If $\beta > \beta_*$ *and* $D = D(N) \to \infty$ *as* $N \to \infty$*, then* $\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) \to \infty$ *as* $N \to \infty$.

Theorem 1.13. *Consider an inhomogeneous spiked Wigner matrix model with arbitrary variance profile. Suppose that* $\mathcal{P}_N$ *is a probability measure over* $\mathbb{R}^N$ *such that* $x \sim \mathcal{P}_N$ *has i.i.d. entries drawn from some probability measure* ν *on* $\mathbb{R}$*. Let* $X \sim \mathcal{X}_N$ *be* $X = \frac{\beta}{\sqrt{N}} xx^\top$ *for* $x \sim \mathcal{P}_N$ *and some* $\beta > 0$*. Suppose that* ν *satisfies the conditions:*

(a) $\mathbb{E}_{x\sim\nu}[x] = 0$.

(b) $\mathbb{E}_{x\sim\nu}[x^2] = 1$.

(c) ν *has bounded support.*

Suppose also that the reciprocal variance profiles $\Phi^{(N)}$ *satisfy the conditions:*

(d) $\lim_{N\to\infty} \max_{i\in[N]} \|v_i(\Phi^{(N)})\|_\infty = 0$*, where* $v_i(\cdot)$ *are the unit eigenvectors of a symmetric matrix.*

(e) Either $\Phi^{(N)} \succeq 0$ *for all* $N \ge 1$*, or* $\mathrm{rank}(\Phi^{(N)}) \le K$ *for all* $N \ge 1$ *for some* K *not depending on* N *and the limits* $\widehat{\mu}_i := \lim_{N\to\infty} \lambda_i(\Phi^{(N)})/N$ *exist for each* $i = 1, \dots, K$.

If $\beta > \beta_*$ *and* $D = D(N) \to \infty$ *as* $N \to \infty$*, then* $\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) \to \infty$ *as* $N \to \infty$.

1.4.2 Statistical-to-Computational Gaps

Our results imply that the important phenomenon of *statistical-to-computational gaps* in some models of low-rank estimation persists in the case of inhomogeneous noise. We do not go into the details, but sketch the idea of this implication here. Consider as an illustrative example the *sparse Rademacher* prior,

$$\nu = \frac{p}{2}\delta_{-1/\sqrt{p}} + (1-p)\delta_0 + \frac{p}{2}\delta_{1/\sqrt{p}}$$

for some constant $p \in (0, 1)$ not depending on N. For any constant p this ν is bounded and has mean 0 and variance 1, and thus all of our computational lower bounds apply to signals $x \sim \mathcal{P}_N = \nu^{\otimes N}$. For a simple concrete setting, consider this choice with a variance profile that is uniformly bounded above and below, as in (4). Then (and under Assumption A), Theorem 1.11 implies that polynomials of degree $o(N)$ cannot detect such a sparse signal whenever $\beta < \beta_*$, where we emphasize that β_* depends only on the sequence of variance profiles, not on p. More heuristically, this suggests that to detect such a signal requires nearly exponential time in N.

In the homogeneous case, [BMV+18] showed that, in fact, there is a $p_0 \in (0, 1)$ such that, whenever $p < p_0$, then a brute force exponential-time computation of the likelihood ratio *does* successfully distinguish $\mathbb{Q}_N$ from $\mathbb{P}_N$ for $\beta \in (\beta'_*, \beta_*)$ for some $\beta'_* < \beta_*$. Thus, in these cases, there is a *statistical-to-computational gap*: for certain parameter regimes, it is possible to distinguish the null and planted distributions, but, based on our analysis of low-degree polynomials, we believe it is only possible to do so very inefficiently (see the discussion of this line of work on sparse PCA above in Section 1.1 for more references).

It is tedious but straightforward to reproduce the analysis of [BMV+18] for the likelihood ratio under the above class of variance profiles. This calculation, which we do not include here for

the sake of brevity, shows that the statistical-to-computational gap in this model of sparse PCA remains for any variance profile as above (specifically, for B in the constant in (4), there exists a $p_0 = p_0(B)$ such that, for any variance profile satisfying Assumption A and the condition in (4), the same result described above holds for all $\beta \in (\beta'_*, \beta_*)$ for some $\beta'_* = \beta'_*(p, B) < \beta_*$).

1.4.3 Channel Universality

Lastly, we remark that our results may be combined with a mild generalization of the universality results in [Kun25]. That work considered more general models of the observation Y, where, in the context of Definition 1.6,

$$Y_{ij} \sim \gamma_{X_{ij}}$$

for some "channel" of probability measures $(\gamma_x)_{x \in \mathbb{R}}$ through which one observes the signal X. While that work took this family to not depend on the indices (i, j), it is straightforward to extend those results to families

$$Y_{ij} \sim \gamma^{(i,j)}_{X_{ij}}$$

provided the $\gamma^{(i,j)}_x$ satisfy some uniform regularity over different (i, j). See Remark 5 of [Kun25] for more discussion.

Together with our calculations, such a generalization of the main bounds of [Kun25] gives lower bounds against low-degree polynomials for additional inhomogeneous observation models beyond ones with additive Gaussian noise, requiring only that the Y_{ij} are observed independently (and some technical regularity conditions on the entrywise channels $\gamma^{(i,j)}_x$). Per the calculations in [Kun25], the role of $\Phi = \Phi^{(N)}$, the reciprocal variance profile in our setting, would then be played by the matrix formed by the *Fisher information* of each family $(\gamma^{(i,j)}_x)_{x \in \mathbb{R}}$ at $x = 0$.

For instance, this approach applies to inhomogeneous non-Gaussian additive noise or to the degree-corrected stochastic block model discussed in [GKKZ25]. We remark also that, for that model and others, [GKKZ25] study analogous universality principles for information-theoretic quantities like the mutual information between the signal X and the observation Y, while those we propose here are for computational ones, namely the low-degree χ^2-divergences.

1.5 Related Work

Spiked matrix models The literature on spiked matrix models in general has grown very deep since their introduction to statistics by Johnstone [Joh01]. In particular, we have not mentioned many important related directions such as spiked models of rectangular or covariance matrices, for which analogs of our setting would also be sensible to analyze. A survey of many of these models from a statistical point of view is given by [JP18], and useful references for their mathematical analysis are [CDM16, Cap17]. There is a long and ongoing line of work on the algorithmic tractability of such models in the homogeneous case; some relevant results, also cited above, include [DAM15, KXZ16, BDM+16, LKZ17, BDM+18, PWBM18, LM19, EAKJ20].

Inhomogeneous spiked matrix models A rectangular version of the model we describe in Definition 1.6 was studied by [BR20], while the block-structured case of our model was studied by [BR22]. The same block-structured model is proposed in [ACCM22], following prior work on similar questions formulated in the language of statistical physics in [ACCM21a, ACCM21b]. The spectral algorithm using $\mathcal{H}(Y)$ was derived by [PKK24], who also conjectured a version of Theorem 1.8. Information-theoretic lower bounds and a derivation of the conjectural algorithmic threshold β_* were carried out by [GKKZ25], again only pertaining to block-structured models (and

also requiring some further structural assumptions). Theorem 1.8 on the BBP transition for $\mathcal{H}(Y)$ was proved concurrently by [BCSvH24, MKK24]; a similar result for the special case of $n = 2$ blocks was obtained previously in [LL24]. To the best of our knowledge, there does not yet exist a version of Theorem 1.8 for a general class of non-block-structured variance profiles. The closest known seem to be the results of [BM21], which concern Y rather than $\mathcal{H}(Y)$ and do not give closed forms for the outlier eigenvalue or the correlation between the signal vector x and $v_1(Y)$.

Inhomogeneity in random matrix theory Inhomogeneous variance profiles have been studied more thoroughly in random matrix theory without the statistical setting we work in here. A more general family of variance profile that is common in those works is to have $\Delta_{ij} = f(i/N, j/N)$ for some symmetric $f : [0,1]^2 \to \mathbb{R}_{\geq 0}$, usually subject to various further regularity assumptions. Block-structured models can then be expressed as step functions, but "smoother" variance profiles can also be expressed by continuous f. We give a numerical study of BBP transitions for an example of such a model in Section 5.5. Prior work has established characterizations of the limiting empirical spectral distribution for matrices with independent entries under certain such variance profiles as well as various local properties of the eigenvalues; see, e.g., [Shl96, AEK17, AEK19, Zhu20]. Completely general variance profiles have also appeared in various results in random matrix theory. The most common instance of these is the case of *generalized Wigner matrices*, which have the boundedness condition (4) and $\sum_{j=1}^{N} \Delta_{ij} = N$ for all $i \in [N]$. Such matrices are known to share many structural properties, including the limiting semicircle law for the empirical spectral distribution, with the case $\Delta_{ij} \equiv 1$ of i.i.d. matrices (see, e.g., [AEK19] or Corollary 2.11 of [BBvH21] for a remarkable further generalization). Outside of this special class, most results seek only to control only some partial information about the spectrum like its moments or the spectral norm [BVH16, vH17, LvHY18, BBvH21].

Low-degree polynomial algorithms Studying low-degree polynomials as algorithms for hypothesis testing problems originates in certain technical considerations in the study of sum-of-squares semidefinite programming relaxations [BHK+19]. Such algorithms were then first studied as a function class of independent interest by works including [HKP+17, HS17, Hop18]. In recent years, their analysis has become a popular approach to giving evidence of computational hardness of various problems. We do not attempt to survey the full literature here, but the surveys [KWB19, Wei25] give a useful overview.

2 Notation

Basic notions We denote by $\mathbb{R}$ the set of real numbers, and its non-negative and strictly positive subsets by $\mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{>0} := \{x \in \mathbb{R} : x > 0\}$. For a positive integer n, we write $[n] = \{1, 2, \ldots, n\}$. For even $2n$, we will sometimes identify $[2n]$ with the paired index set $\{1, 1', 2, 2', \ldots, n, n'\}$, where the correspondence $i \leftrightarrow i'$ indicates paired elements of $[2n]$ (see Section 5.1). The notation 1 may denote either the scalar 1 or the all-ones vector of appropriate dimension, depending on the context. For an event E, we denote by $\mathbf{1}\{E\}$ the indicator function of E, i.e.,

$$\mathbf{1}\{E\} := \begin{cases} 1 & \text{if } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Linear algebra For any set $S \subseteq \mathbb{R} \cup \{+\infty\}$, we write $S^{m \times n}$ for the collection of all $m \times n$ matrices with entries in S. When $m = n$, the subset of symmetric matrices in $S^{n \times n}$ is denoted $S^{n \times n}_{\text{sym}}$. For

a symmetric matrix $A \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$, we write $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ for the sorted (real) eigenvalues of A. The operation $\mathrm{vec}(A)$ denotes the vectorization of A, obtained by stacking its columns into a single vector in $\mathbb{R}^{n^2}$. The trace of a square matrix A is denoted by $\mathrm{Tr}(A) = \sum_{i=1}^n A_{ii}$. For a matrix $A \in \mathbb{R}^{m\times n}$, we denote by $\|A\|$ its spectral norm, and by $\|A\|_{\ell_\infty}$ its entrywise ℓ_∞ norm, defined as

$$\|A\| := \sup_{\|x\|=1} \|Ax\|, \quad \|A\|_{\ell_\infty} := \sup_{i,j} |A_{i,j}|\,.$$

For a vector $x \in \mathbb{R}^n$, we write $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$ for its Euclidean norm, and $\|x\|_{\ell_\infty} = \max_{1\leq i\leq n} |x_i|$ for its ℓ_∞-norm. The standard Euclidean inner product between $x, y \in \mathbb{R}^n$ is denoted by $\langle x, y\rangle = x^\top y$. The standard Euclidean or Frobenius inner product between matrices $A, B \in \mathbb{R}^{m\times n}$ is denoted by $\langle A, B\rangle = \mathrm{Tr}(A^\top B)$. The operation $\mathrm{diag}(\cdot)$ is defined as follows:

- when applied to a vector $v \in \mathbb{R}^n$, $\mathrm{diag}(v)$ denotes the diagonal matrix with diagonal entries $(v_1, \ldots, v_n)$;
- when applied to a matrix $A \in \mathbb{R}^{n\times n}$, $\mathrm{diag}(A)$ denotes the diagonal matrix that retains the diagonal entries of A while setting all off-diagonal entries to zero.

For element-wise operations on matrices of the same dimension, we write $A \odot B$ for their Hadamard product, defined by $(A \odot B)_{ij} = A_{ij}B_{ij}$, and $A^{\odot -1}$ for the Hadamard inverse, whenever all entries of A are nonzero, defined by $(A^{\odot -1})_{ij} = A_{ij}^{-1}$. $A \otimes B$ denotes the Kronecker product between matrices A and B, defined by

$$A \otimes B = [A_{ij}B]_{i,j},$$

yielding a block matrix of size $(mp) \times (nq)$ when $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$.

Asymptotics We use the standard asymptotic notations $O(\cdot), o(\cdot), \Omega(\cdot), \omega(\cdot)$, and $\Theta(\cdot)$ in their usual sense, always referring to the limit $N \to \infty$.

Probability We write $\mathcal{N}(\mu, \Sigma)$ for the multivariate Gaussian distribution with mean μ and covariance Σ, $\mathrm{Pois}(\lambda)$ for the Poisson distribution with rate λ, $\chi^2(k)$ for the chi-squared distribution with k degrees of freedom, and $\Gamma(\alpha, \beta)$ for the Gamma distribution with shape parameter α and scale parameter β. Throughout, the notation $X_N \xrightarrow{\mathbb{P}} X$ denotes convergence in probability, and $X_N \xrightarrow{(\mathrm{d})} X$ denotes convergence in distribution. We write $\chi^2(\mathbb{P} \mid \mathbb{Q})$ for the chi-squared divergence between two probability measures $\mathbb{P}$ and $\mathbb{Q}$, defined as

$$\chi^2(\mathbb{P} \mid \mathbb{Q}) = \int \left(\frac{d\mathbb{P}}{d\mathbb{Q}} - 1\right)^2 d\mathbb{Q},$$

whenever $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$.

3 Preliminaries

3.1 Linear Algebra

We will apply the following result to the reciprocal variance profile matrix $\Phi^{(N)}$ throughout without mention.

Theorem 3.1 (Non-negative Perron-Frobenius)**.** *Let $A \in \mathbb{R}^{N\times N}_{\mathrm{sym}}$ have non-negative entries $A_{ij} \geq 0$. Then, $\|A\| = \lambda_1(A) \geq 0$, i.e., for all $i = 2, \ldots, N$, we have $|\lambda_i(A)| \leq \lambda_1(A)$.*

The following general results about block-structured matrix sequences will be useful. Suppose that $A^{(N)} \in \mathbb{R}^{N\times N}_{\mathrm{sym}}$ is block-structured with parameters $(\bar{A}, \rho_1, \dots, \rho_n)$. For a given $N \geq 1$, recall that this means there is a $q : [N] \to [n]$ such that $A^{(N)}_{ij} = \bar{A}^{(N)}_{q(i)q(j)}$.

Proposition 3.2. *Let $S_i := q^{-1}(i)$ for $i = 1, \dots, n$, let $s_i := \sqrt{|S_i|}$, and $s = s^{(N)}$ be the vector of the s_i. Consider the matrix*

$$B = B^{(N)} = (ss^\top) \circ \bar{A} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}.$$

Then, the following hold:

1. *The non-zero eigenvalues of $A^{(N)}$ are the same as those of B.*
2. *Suppose $v = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$ is an unit eigenvector of B with eigenvalue ν. Then, $w \in \mathbb{R}^N$ with entries*

$$w_i = \frac{v_{q(i)}}{s_{q(i)}}$$

   is a unit eigenvector of $A^{(N)}$ with the same eigenvalue ν.

3. *Let ρ be the vector of the ρ_i, and $\sqrt{\rho}$ be its entrywise square root. Let*

$$\widetilde{B} := (\sqrt{\rho}\sqrt{\rho}^\top) \circ \bar{A} = \mathrm{diag}(\rho)^{1/2} \bar{A}\, \mathrm{diag}(\rho)^{1/2}.$$

   Then,

$$\lim_{N\to\infty} \frac{\lambda_1(A^{(N)})}{N} = \lambda_1(\widetilde{B}).$$

Proof. Let $R \in \{0,1\}^{N\times n}$ denote the block indicator matrix, defined by

$$R_{j,i} = \mathbf{1}\{q(j) = i\}, \quad \text{for } j \in [N] \text{ and } i \in [n].$$

By construction we have $R^\top R = G := \mathrm{diag}(|S_1|, \dots, |S_n|)$, and we may express the block structure of $A^{(N)}$ by writing $A^{(N)} = R\bar{A}R^\top$. It follows from the construction that

$$B := (ss^\top) \circ \bar{A} = G^{1/2}\bar{A}G^{1/2}.$$

Part 1: Equality of non-zero spectra. Recall that for any pair of matrices X, Y with appropriate dimensions, the non-zero eigenvalues of XY and YX coincide. Applying this observation, we infer that $A^{(N)} = R\bar{A}R^\top$ has the same non-zero eigenvalues as $\bar{A}R^\top R = \bar{A}G = \bar{A}G^{1/2}G^{1/2}$, which in turn has the same non-zero eigenvalues as $G^{1/2}\bar{A}G^{1/2} = B$. Hence, $A^{(N)}$ and B share the same non-zero spectrum.

Part 2: Relation between eigenvectors. Let w be a unit eigenvector of $A^{(N)}$ associated with a non-zero eigenvalue μ, so that $A^{(N)}w = \mu w$. Since $A^{(N)} = R\bar{A}R^\top$, we obtain $R(\bar{A}R^\top w) = \mu w$, which shows that w must lie in the column space of R. That is, there exists $x \in \mathbb{R}^n$ with $w = Rx$. Substituting, $R^\top R\bar{A}R^\top Rx = \mu R^\top Rx$, which simplifies to $G\bar{A}Gx = \mu Gx$. Equivalently,

$$G^{1/2}\bar{A}G^{1/2}(G^{1/2}x) = \mu(G^{1/2}x),$$

which is $Bv = \mu v$, where $v := G^{1/2}x$. Thus v is an unit eigenvector of B with eigenvalue μ, and conversely, $w = Rx = RG^{-1/2}v$. In coordinates this reads

$$w_i = \frac{v_{q(i)}}{c_{q(i)}} \quad \text{for all } i \in [N],$$

as claimed.

Part 3: Asymptotic spectral limit. Recall that the block structure assumption implies that $|S_i|/N \to \rho_i$ as $N \to \infty$ for each block $i \in [n]$. Then, entrywise,

$$\frac{1}{N} B^{(N)} \to \widetilde{B},$$

where $\widetilde{B}$ is the deterministic matrix with entries $\widetilde{B}_{ij} = \sqrt{\rho_i}\sqrt{\rho_j}\bar{A}_{ij}$. Since these matrices have a constant dimension n and converge entrywise, we equivalently have

$$\lim_{N\to\infty} \left\| \frac{1}{N} B^{(N)} - \widetilde{B} \right\| = 0,$$

which implies

$$\lim_{N\to\infty} \frac{\lambda_1(B^{(N)})}{N} = \lambda_1(\widetilde{B}).$$

By Part 1, the same conclusion holds for $A^{(N)}$, namely

$$\lim_{N\to\infty} \frac{\lambda_1(A^{(N)})}{N} = \lambda_1(\widetilde{B}),$$

completing the proof. □

3.2 Combinatorics

Definition 3.3 (Partitions)**.** *A* partition *of a finite set $\mathcal{I}$ is a set of sets $\pi = \{S_1, \ldots, S_m\}$ where each $S_i \subseteq \mathcal{I}$, each S_i is non-empty, the S_i are pairwise disjoint, and the disjoint union of the S_i equals $\mathcal{I}$. We call the S_i the* blocks *of π, and $\mathcal{I}$ the* base set*. We write $b_\pi : \mathcal{I} \to \pi$ for the function that maps $x \in \mathcal{I}$ to the (unique) block $S_i \in \pi$ such that $x \in S_i$.*

We write $\mathrm{Part}(\mathcal{I})$ *for the set of all partitions of $\mathcal{I}$,* $\mathrm{Part}_m(\mathcal{I})$ *for the set of all partitions into m blocks, and* $\mathrm{EvenPart}(\mathcal{I})$ *and* $\mathrm{EvenPart}_m(\mathcal{I})$ *for the same notions but only including* even partitions*, ones where $|S_i|$ is even for all $i \in [m]$. We write* $\mathrm{Match}(\mathcal{I})$ *for the partitions all of whose blocks have size exactly 2 (which can be viewed as perfect matchings of $\mathcal{I}$).*

We remark, since it will often come up in our calculations, that when $|\mathcal{I}| = 2d$ then, by definition, $\mathrm{Match}(\mathcal{I}) = \mathrm{EvenPart}_d(\mathcal{I})$.

Definition 3.4 (Partial ordering of partitions)**.** *Two partitions ρ, π on the same base set have the relation $\rho \preceq \pi$ if there exists a function $q : \pi \to \rho$ such that $S \subset q(S)$ for all $S \in \pi$. In this case, we say that π is* finer than *or* a refinement of *ρ, and conversely that ρ is* coarser than *or* a coarsening of *π. The relation $\preceq$ defines a partial ordering on the sets* $\mathrm{Part}(\mathcal{I})$ *and* $\mathrm{EvenPart}(\mathcal{I})$*.*

Definition 3.5 (Stirling numbers)**.** *The* Stirling numbers of the second kind *are*

$$S_2(d, m) := |\mathrm{Part}_m([d])|.$$

Definition 3.6 (Touchard polynomials)**.** *The* Touchard polynomials *are the polynomials*

$$T_d(x) := \sum_{m=1}^{d} S_2(d, m) x^m.$$

As we will see below, the Touchard polynomials also have a probabilistic interpretation that will be useful to us.

3.3 Probability

To prove our lower bounds, we will use that certain random variables are asymptotically Gaussian by the central limit theorem. The following multidimensional sufficient condition will be useful in some cases.

Theorem 3.7 (Cramér-Wold central limit theorem, Corollary 4.5 of [Kal97])**.** *Let $X^{(k)} \in \mathbb{R}^n$ be a sequence of random vectors and let X be another random vector. Then, the $X^{(k)}$ converge in distribution to X as $k \to \infty$ if and only if $\langle c, X^{(k)} \rangle = \sum_{i=1}^n c_i X_i^{(k)}$ converges in distribution to $\langle c, X \rangle = \sum_{i=1}^n c_i X_i$ for all $c \in \mathbb{R}^n$.*

In particular, suppose that $\langle c, X^{(k)} \rangle$ converges in distribution to $\mathcal{N}(0, \|c\|^2)$ for all $c \in \mathbb{R}^n$. Then, $X^{(k)}$ converges in distribution to $\mathcal{N}(0, I_n)$.

To use these results, we will also want to give lower bounds on the moments of Gaussian quadratic forms. The following result follows easily from Corollary 1 of [BU06].

Proposition 3.8. *Let $A \in \mathbb{R}^{n \times n}_{\mathrm{sym}}$ satisfy that $\mathrm{Tr}(A^k) \geq 0$ for all $k \geq 1$. For instance, we can either have $A \succeq 0$ or A having only non-negative entries. Let $y \in \mathbb{R}^n$ be a random vector with $y_i \sim \mathcal{N}(0,1)$ i.i.d. Then, we have for all $d \geq 1$ that*

$$\mathbb{E}[(y^\top A y)^d] \geq 2^{d-1}(d-1)! \,\mathrm{Tr}(A^d).$$

Our calculations will involve various sums over partitions which will lead to evaluations of the Touchard polynomials, defined above. We use the following probabilistic interpretation to derive sufficiently tight bounds on them.

Proposition 3.9. *For any $\lambda > 0$ and $d \geq 1$,*

$$\mathop{\mathbb{E}}_{X \sim \mathrm{Pois}(\lambda)} X^d = T_d(\lambda).$$

We will use the following bound, which in fact, as discussed in the reference, holds for all "sub-Poissonian" random variables (a notion similar to subgaussian and subexponential).

Proposition 3.10 ([Ahl22])**.** *For all $d \geq 1$ and $\lambda > 0$, we have*

$$\mathop{\mathbb{E}}_{X \sim \mathrm{Pois}(\lambda)} \left(\frac{X}{\lambda}\right)^d \leq \left(\frac{\frac{d}{\lambda}}{\log(1 + \frac{d}{\lambda})}\right)^d.$$

Corollary 3.11. *For all $d \geq 1$ and $\lambda > 0$, we have*

$$T_d(\lambda) = \mathop{\mathbb{E}}_{X \sim \mathrm{Pois}(\lambda)} X^d \leq \lambda^d \exp\left(\frac{d^2}{\lambda}\right).$$

Proof. From Proposition 3.10, to reach this claim it suffices to show the scalar inequality $x/\log(1+x) \leq \exp(x)$ for all $x > 0$ (which we apply with $x = d/\lambda$). This result follows since the function $f(x) := \log(1+x) - x\exp(-x)$ has $f(0) = 0$ and

$$f'(x) = \frac{1}{1+x} + x\exp(-x) - \exp(-x) = \frac{\exp(x) - 1 + x^2}{(1+x)\exp(x)} \geq 0$$

for all $x \geq 0$. □

Lastly, we will need some properties of subgaussian random variables, as well as a few variants thereof. The following is the standard definition of subgaussianity.

Definition 3.12 (Subgaussianity)**.** *We say that a centered random variable X is σ^2-*subgaussian *if $\mathbb{E}\exp(tX) \le \exp(\frac{\sigma^2}{2}t^2)$ for all $t \in \mathbb{R}$. We say that a centered random vector x is σ^2-subgaussian if $\langle x, v\rangle$ is σ^2-subgaussian for all $\|v\| = 1$.*

For some of our applications, the following version will also be useful.

Definition 3.13 (Moment subgaussianity)**.** *We say that a symmetric random variable X is σ^2-*moment-subgaussian *if, for all $k \ge 1$,*

$$\mathbb{E}X^{2k} \le \sigma^{2k}(2k-1)!! = \mathop{\mathbb{E}}_{Z\sim\mathcal{N}(0,\sigma^2)} Z^{2k}.$$

We say that a symmetric random vector x is σ^2-moment-subgaussian if $\langle x, v\rangle$ is σ^2-moment-subgaussian for all $\|v\| = 1$.

These definitions are equivalent up to constants in σ^2; however, for some purposes we will need to be careful about those constants.

Proposition 3.14. *The following hold for a symmetric random variable X:*

1. *If X is σ^2-subgaussian, then X is $4\sigma^2$-moment-subgaussian.*
2. *If X is σ^2-moment-subgaussian, then X is σ^2-subgaussian.*

Proof. For the second implication, we may consider the moment generating function directly: for $G \sim \mathcal{N}(0, \sigma^2)$,

$$\begin{aligned}
\mathbb{E}\exp(tX) &= \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}X^k \\
&= \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!}\mathbb{E}X^{2k} && \text{(symmetry of } X\text{)} \\
&\le \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!}\mathbb{E}G^{2k} \\
&= \mathbb{E}\exp(tG) \\
&= \exp\left(\frac{\sigma^2}{2}t^2\right).
\end{aligned}$$

For the first implication, by Lemma 1.4 of [RH17], we have

$$\begin{aligned}
\mathbb{E}X^{2k} &\le (2\sigma^2)^k \cdot 2 \cdot k! \\
&\le (4\sigma^2)^k \cdot k! \\
&\le (4\sigma^2)^k \cdot (2k-1)!!,
\end{aligned}$$

giving the result. □

Proposition 3.15. *If X is σ^2-subgaussian and has $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$, then there exists a $\tau > 0$ such that, for all $k \ge 1$,*

$$\mathbb{E}X^{2k} \le \tau^{2k-2}(2k-1)!!.$$

Proof. By Proposition 3.14, X is $4\sigma^2$-moment-subgaussian. Take $\tau^2 = \max\{1, (4\sigma^2)^3\}$, so that $\mathbb{E}X^{2k} \le \tau^{2k/3}(2k-1)!!$. Since $2k/3 \le 2k-2$ for all $k \ge 2$, the result follows. □

We will also use the following result about moment-subgaussian vectors.

Proposition 3.16. *If a random vector $x \in \mathbb{R}^k$ is 1-moment-subgaussian then, for any $\delta \in (0,1)$ and $\gamma \in (0, \frac{1-\delta}{2})$,*

$$\begin{aligned}\mathbb{E}\left[\exp\left(\gamma\|x\|^2\right)\right] &\le C(\delta,k) \mathop{\mathbb{E}}_{Z\sim\mathcal{N}(0,1)}\left[\exp\left(\frac{\gamma}{1-\delta}Z^2\right)\right] \\ &= C(\delta,k)\frac{1}{\sqrt{1-\frac{2\gamma}{1-\delta}}},\end{aligned}$$

where $C(\delta,k)$ is a constant depending only on δ and k. In particular, we have

$$\mathbb{E}\left[\exp\left(\gamma\|x\|^2\right)\right] < \infty$$

for all $\gamma < \frac{1}{2}$.

Proof. Let $\mathcal{N} \subset \mathbb{R}^k$ be a δ-net of the unit sphere of $\mathbb{R}^k$. Standard bounds give

$$\|x\| \le \frac{1}{1-\delta}\max_{u\in\mathcal{N}}\langle u, \boldsymbol{x}\rangle.$$

Therefore, we may bound the moments by

$$\begin{aligned}\mathbb{E}\left[\|x\|^{2k}\right] &\le \left(\frac{1}{1-\delta}\right)^{2k}\mathbb{E}\left[\max_{u\in\mathcal{N}}\langle u,x\rangle^{2k}\right] \\ &\le \left(\frac{1}{1-\delta}\right)^{2k}\sum_{u\in\mathcal{N}}\mathbb{E}\left[\langle u,x\rangle^{2k}\right] \\ &\le \left(\frac{1}{1-\delta}\right)^{2k}\sum_{u\in\mathcal{N}}\mathbb{E}[Z^{2k}] \\ &= C(\delta,k)\left(\frac{1}{1-\delta}\right)^{2k}\mathbb{E}[Z^{2k}]\end{aligned}$$

where $C(\delta,k) = |\mathcal{N}|$ is a constant depending only on δ, k. The result then follows by expanding $\exp(\gamma\|x\|^2)$ in a Taylor series. □

Now we will state some definitions and propositions proposed in [BBK+21] concerning a "local" version of subgaussianity, formalizing ideas appearing previously in [PWBM18].

Definition 3.17. *A random vector $x \in \mathbb{R}^k$ is ε-locally c-subgaussian if, for any vector $v \in \mathbb{R}^k$ with $\|v\| \le \varepsilon$,*

$$\mathbb{E}\exp\left(\langle v,x\rangle\right) \le \exp\left(\frac{c}{2}\|v\|^2\right).$$

Proposition 3.18. *Suppose x is ε-locally c-subgaussian. For a (non-random) scalar $\alpha \neq 0$, αx is $\varepsilon/|\alpha|$-locally $c\alpha^2$-subgaussian. Also, the sum of n independent copies of x is ε-locally cn-subgaussian.*

Proposition 3.19. *If $x \in \mathbb{R}^k$ is ε-locally c-subgaussian, then for any $\delta > 0$ and any $0 \le t \le \varepsilon c/(1-\delta)$,*

$$\mathbb{P}\left[\|x\| \ge t\right] \le C(\delta, k) \exp\left(-\frac{1}{2c}(1-\delta)^2 t^2\right)$$

where $C(\delta, k)$ is a constant depending only on δ and k.

Proposition 3.20. *Let $\delta > 0$. If a random vector $x \in \mathbb{R}^k$ is ε-locally c-subgaussian with $c < (1-\delta)^2/2$ then*

$$\mathbb{E}\left[\mathbf{1}\{\|x\| \le \varepsilon c/(1-\delta)\} \cdot \exp(\|x\|^2)\right] \le 1 + \frac{C(\delta, k)}{(1-\delta)^2/2c - 1}$$

where $C(\delta, k)$ is a constant depending only on δ and k.

3.4 Calculus

The following is a standard series evaluation, a special case of Newton's generalized binomial theorem.

Proposition 3.21 (Hypergeometric sum)**.** *For $|a| < 1$ and $c > 0$,*

$$\sum_{d=0}^{\infty} \frac{a^d}{d!} \frac{\Gamma(d+c)}{\Gamma(c)} = (1-a)^{-c}.$$

We will also need to bound the output of the above evaluation, for which we apply the following useful bound.

Proposition 3.22. *For $a \in (0,1)$ and $c > 0$,*

$$\exp(ca) \le (1-a)^{-c} \le \exp\left(\frac{1 - \frac{a}{2}}{1-a} \cdot ca\right).$$

This result may be viewed as saying that the simple lower bound above is actually tight up to constant factors in the exponent, provided that a is not too close to 1.

Proof. The lower bound follows immediately from the bound $1 - a \le \exp(-a)$. For the upper bound, we take the Taylor expansion of the logarithm,

$$\begin{aligned}
(1-a)^{-c} &= \exp\left(-c \log(1-a)\right) \\
&= \exp\left(c \sum_{p=1}^{\infty} \frac{a^p}{p}\right) \\
&\le \exp\left(c\left(a + a^2 \sum_{p=0}^{\infty} \frac{a^p}{2}\right)\right) \\
&= \exp\left(c\left(a + \frac{1}{2} \cdot \frac{a^2}{1-a}\right)\right) \\
&= \exp\left(ca\left(1 + \frac{1}{2} \cdot \frac{a}{1-a}\right)\right) \\
&= \exp\left(ca \cdot \frac{1}{2} \cdot \frac{2-a}{1-a}\right),
\end{aligned}$$

giving the result. □

3.5 Low-Degree Polynomial Algorithms

We introduce some tools, by now standard in the literature, for showing that low-degree polynomials cannot strongly separate sequences of probability measures. See the surveys [KWB19, Wei25] for many more details. The technique we will use revolves around the following quantities.

Definition 3.23 (Low-degree χ^2-divergence)**.** *Consider a pair of probability measures $\mathbb{P}, \mathbb{Q}$ over $\mathbb{R}^{N\times N}_{\mathrm{sym}}$ (say in the setting of Definition 1.6). The* degree-D χ^2-divergence *is given by*

$$\begin{aligned}\chi^2_{\le D}(\mathbb{P} \mid \mathbb{Q}) &:= \sup_{f \in \mathbb{R}[Y]_{\le D}\setminus\{0\}} \frac{(\mathbb{E}_{\mathbb{P}}[f(Y)])^2}{\mathbb{E}_{\mathbb{Q}}[f(Y)^2]} - 1 \\ &= \sup_{\substack{f \in \mathbb{R}[Y]_{\le D}\setminus\{0\} \\ \mathbb{E}_{\mathbb{Q}} f(Y) = 0}} \frac{(\mathbb{E}_{\mathbb{P}}[f(Y)])^2}{\mathbb{E}_{\mathbb{Q}}[f(Y)^2]}.\end{aligned}$$

If we remove the restriction to low-degree polynomials, then this expression gives the classical χ^2-divergence $\chi^2(\mathbb{P} \mid \mathbb{Q})$ (provided that $\frac{d\mathbb{P}}{d\mathbb{Q}} \in L^2(\mathbb{Q})$). The kind of analysis we carry out is more commonly formulated in terms of the low-degree "advantage" $1 + \chi^2_{\le D}(\mathbb{P} \mid \mathbb{Q})$, but we follow the above notation (also appearing, for instance, in [COGHK+22]) to emphasize the analogy with the χ^2 divergence, which appears in our statistical lower bounds.

The low-degree and classical χ^2-divergences govern strong separation by low-degree polynomials and by arbitrary functions in the following way.

Proposition 3.24 (Proposition 6.2 of [BAH+22], Lemma 1.13 of [KWB19])**.** *For a sequence of pairs of probability measures $\mathbb{P}_N, \mathbb{Q}_N$ over $\mathbb{R}^{N\times N}_{\mathrm{sym}}$, the following hold:*

1. *If $\chi^2_{\le D(N)}(\mathbb{P}_N \mid \mathbb{Q}_N) = O(1)$ as $N \to \infty$, then there is no sequence of polynomials $f_N \in \mathbb{R}[Y]$ with $\deg(f_N) \le D(N)$ achieving strong separation.*
2. *If $\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) = O(1)$ as $N \to \infty$, then there is no sequence of (measurable) functions $f_N : \mathbb{R}^{N\times N}_{\mathrm{sym}} \to \mathbb{R}$ achieving strong separation, and no sequence of tests $t_N : \mathbb{R}^{N\times N}_{\mathrm{sym}} \to \{0, 1\}$ achieving strong detection.*

Importantly, neither converse statement is true in general, as discussed earlier in Section 1.4.1.

3.6 Tools for Additive Gaussian Models

It turns out that convenient and elegant closed forms for the low-degree and classical χ^2-divergences are known for the special case of models with additive Gaussian noise. Let us first state a general result for a general such model.

Definition 3.25 (Additive Gaussian Noise Model)**.** *Let $X \sim \mathcal{X} = \mathcal{X}_M$ for some probability measure $\mathcal{X}$ over $\mathbb{R}^M$ and $Z \sim \mathcal{N}(0, \Sigma)$ for some $\Sigma \in \mathbb{R}^{M\times M}_{\mathrm{sym}}$ with $\Sigma \succeq 0$. The associated additive Gaussian model consists of the probability measures $\mathbb{P} = \mathbb{P}_M$ and $\mathbb{Q} = \mathbb{Q}_M$ given by:*

- *To sample $Y \sim \mathbb{P}$, observe $Y = X + Z$.*
- *To sample $Y \sim \mathbb{Q}$, observe $Y = Z$.*

Proposition 3.26 (Theorem 2.6 of [KWB19])**.** *Suppose in the above setting that $\Sigma = I_M$. Then,*

$$\begin{aligned}\chi^2(\mathbb{P} \mid \mathbb{Q}) &= \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}} \exp\left(\langle X^1, X^2\rangle\right) - 1, \\ \chi^2_{\le D}(\mathbb{P} \mid \mathbb{Q}) &= \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}} \exp^{\le D}\left(\langle X^1, X^2\rangle\right) - 1,\end{aligned}$$

where we define

$$\exp^{\le D}(x) := \sum_{d=0}^{D} \frac{x^d}{d!}$$

and X^1, X^2 are drawn independently from $\mathcal{X}$.

Since in the more general model with arbitrary covariance Σ we may just multiply through by $\Sigma^{-1/2}$ to obtain a model with $\Sigma = I_M$, it is easy to derive the following similar identity.

Corollary 3.27. *For a general $\Sigma \succ 0$, in the above setting,*

$$\chi^2(\mathbb{P} \mid \mathbb{Q}) = \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}} \exp\left(X^{1\top} \Sigma^{-1} X^2\right) - 1,$$
$$\chi^2_{\le D}(\mathbb{P} \mid \mathbb{Q}) = \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}} \exp^{\le D}\left(X^{1\top} \Sigma^{-1} X^2\right) - 1.$$

We now apply these results to our matrix models. Since we work over symmetric matrices, it suffices to consider the entries on and above the diagonal, of which there are $M = \frac{N(N+1)}{2}$. But, to work in the inhomogeneous setting, we must indeed use the extra flexibility of Corollary 3.27. Applying this, we obtain the following, which will be the starting point of our coming calculations.

Lemma 3.28. *Let $\mathbb{P}_N$ and $\mathbb{Q}_N$ be the planted and null distributions for an inhomogeneous spiked Wigner matrix model with signal distribution $\mathcal{X}_N$ and variance profile $\Delta = \Delta^{(N)}$. Recall that we set $\Phi := \Delta^{\odot -1}$. Then, we have*

$$\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) = \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}_N} \exp\left(R(X^1, X^2)\right) - 1,$$
$$\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) = \mathop{\mathbb{E}}_{X^1, X^2 \sim \mathcal{X}_N} \exp^{\le D}(R(X^1, X^2)) - 1,$$

where $R(X^1, X^2) \in \mathbb{R}$ is the "overlap" random variable, which may be written equivalently as

$$\begin{aligned} R(X^1, X^2) &= \sum_{1 \le i \le j \le N} \Phi_{ij} X^1_{ij} X^2_{ij} \\ &= \frac{1}{2} \sum_{i,j=1}^{N} (\Phi + \operatorname{diag}(\Phi))_{ij} X^1_{ij} X^2_{ij}. \end{aligned}$$

Further, suppose we are in the case of any of our Theorems in Section 1.3, where there is a further probability measure $\mathcal{P}_N$ on $\mathbb{R}^{N \times \kappa}$ such that $X \sim \mathcal{X}_N$ can be drawn as $X = \frac{\beta}{\sqrt{N}} V V^\top$ for $V \sim \mathcal{P}_N$. Let $V^1, V^2 \sim \mathcal{P}_N$ be i.i.d. draws, and write $x^a_1, \dots, x^a_N \in \mathbb{R}^\kappa$ for the rows and $v^a_1, \dots, v^a_\kappa \in \mathbb{R}^N$ for the columns of V^a for $a \in \{1, 2\}$. In this case, the overlap may be rewritten as

$$\begin{aligned} R(V^1, V^2) &:= R\left(\frac{\beta}{\sqrt{N}} V^1 V^{1\top}, \frac{\beta}{\sqrt{N}} V^2 V^{2\top}\right) \\ &= \frac{\beta^2}{2N} \sum_{i,j=1}^{N} (\Phi + \operatorname{diag}(\Phi))_{ij} \langle x^1_i, x^1_j \rangle \langle x^2_i, x^2_j \rangle \\ &= \frac{\beta^2}{2N} \sum_{a,b=1}^{\kappa} (v^1_a \odot v^2_b)^\top (\Phi + \operatorname{diag}(\Phi))(v^1_a \odot v^2_b). \end{aligned}$$

We note that, conveniently, this way of expressing the full and low-degree χ^2-divergences captures the possibility of censored observations by setting certain entries of Φ to zero, letting us not need to explicitly keep track of the set $\mathcal{J}$ of entries with finite variance from now on.

The following basic facts about the truncated exponential polynomials $\exp^{\leq D}(x)$ will be useful (while simple and natural-looking, the second turns out to be surprisingly non-trivial to prove).

Proposition 3.29. *For all $D \geq 1$, $\exp^{\leq D}(x)$ is monotonically increasing over $x \geq 0$.*

Proposition 3.30 (Proposition 5.3 of [Kun25])**.** *For all $D \geq 2$ even and $x \in \mathbb{R}$, we have*

$$0 < \exp^{\leq D}(x) \leq \exp^{\leq D}(|x|).$$

Further, for all $x, y \in \mathbb{R}$, we have

$$\exp^{\leq D}(x)\exp(-100|y|) \leq \exp^{\leq D}(x+y) \leq \exp^{\leq D}(x)\exp(100|y|)$$

While this result requires D to be even, for all of our purposes it will be possible to assume this without loss of generality by just increasing or decreasing D by one as needed and using that the low-degree χ^2-divergence is monotone in D. So, we do not explicitly verify this assumption when we apply the Proposition later.

Finally, the following is a useful general tool for controlling expectations of the form that we will be dealing with, which abstracts a style of analysis of such expressions first (to the best of our knowledge) developed by [BMV+18, PWBM18].

Lemma 3.31 (Lemma 2.10 of [Kun24])**.** *Let $R_N \geq 0$ be a sequence of real bounded random variables and $D(N) \in \mathbb{N}$. Suppose there exists a sequence $A(N) \in \mathbb{R}_{\geq 0}$ such that the following conditions hold:*

1. *For all sufficiently large N,*

$$A(N) \geq D(N)\left(2 \vee \log\left(\frac{\|R_N\|_\infty}{D(N)}\right)\right).$$

   (Recall that $\|R_N\|_\infty$ is the smallest $C > 0$ such that $R_N \leq C$ almost surely.)

2. *For some bounded $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\int_0^\infty f(t)\,dt < \infty$, for sufficiently large N, for all $t \in [0, A(N)]$,*

$$\mathbb{P}[R_N \geq t] \leq f(t)\exp(-t).$$

Then, we have

$$\limsup_{N\to\infty} \mathbb{E}\exp^{\leq D(N)}(R_N) < \infty.$$

4 Block-Structured Variance Profiles: Proof of Theorem 1.10

Recall that Theorem 1.10 concerns the inhomogeneous spiked Wigner matrix model (Definition 1.6) where the variance profiles $\Delta = \Delta^{(N)}$ are a block-structured sequence of matrices. Consequently, the Hadamard inverses $\Phi = \Phi^{(N)}$ are also block-structured. The Theorem includes upper bounds on $\chi^2_{\leq D}(\mathbb{P}_N \mid \mathbb{Q}_N)$ (implying computational lower bounds of the form that polynomials of degree at most D cannot achieve strong separation) and upper bounds on $\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N)$ (implying statistical lower bounds of the form that arbitrary functions cannot achieve strong separation). Also, in Theorem 1.12, we give complementary lower bounds on $\chi^2_{\leq D}(\mathbb{P}_N \mid \mathbb{Q}_N)$, implying that the analysis in the first part is tight. We consider these three results separately in the following three sections.

4.1 Preliminaries on Block-Structured Models

Before continuing, let us introduce some further notation to deal with the block structure of the model. Recall that the block structure parameters are $(\bar{\Delta}, \rho_1, \dots, \rho_n)$ such that, for each N, there is a function $q : [N] \to [n]$ such that $\Delta_{ij}^{(N)} = \bar{\Delta}_{q(i)q(j)}$ for all $i, j \in [N]$. Let us write $S_a = S_{a,N} := q^{-1}(a)$ for each $a \in [n]$, so that the S_a form a partition of $[N]$. Then, the block structure assumption asks that $|S_a|/N \to \rho_a \in (0,1)$ for each $a \in [n]$.

We have previously defined $\bar{\Phi} := \bar{\Delta}^{\odot -1}$. Let us also define

$$\widetilde{\Gamma} := \operatorname{diag}(\rho)^{1/2} \, \bar{\Phi} \operatorname{diag}(\rho)^{1/2},$$

so that our threshold parameter is

$$\widehat{\mu}_1 = \lambda_1(\widetilde{\Gamma}).$$

We also define a matrix $\Gamma = \Gamma^{(N)} \in \mathbb{R}^{n \times n}$, given by

$$\Gamma := \operatorname{diag}(|S_1|, \dots, |S_n|)^{1/2} \, \bar{\Phi} \operatorname{diag}(|S_1|, \dots, |S_n|)^{1/2}.$$

From the block structure assumption, since $|S_a|/N \to \rho_a$, one may verify that

$$\begin{aligned} \lim_{N \to \infty} \frac{1}{N} \Gamma^{(N)} &= \widetilde{\Gamma}, \\ \lim_{N \to \infty} \frac{1}{N} \lambda_1(\Gamma^{(N)}) &= \lambda_1(\widetilde{\Gamma}) \\ &= \widehat{\mu}_1, \end{aligned}$$

where the first convergence occurs entrywise (noting that the dimension n of $\Gamma^{(N)}$ does not depend on N) and the second convergence follows from the first.

Finally, let us define the matrix $P \in \mathbb{R}^{n \times N}$ such that, for all $v \in \mathbb{R}^N$,

$$(Pv)_a = \frac{1}{\sqrt{|S_a|}} \sum_{i \in S_a} v_i.$$

This matrix averages vectors over blocks, with a normalization *a la* the central limit theorem (whose role will become clear in the proof). This satisfies the identity

$$v^\top \Phi v = (Pv)^\top \Gamma (Pv)$$

for all $v \in \mathbb{R}^N$.

4.2 Computational Lower Bound

Proof. Let us introduce a bit of notation for constants that will appear throughout. Since the entries of Φ all appear in the finite matrix $\bar{\Phi}$, they are uniformly bounded:

$$\|\Phi\|_{\ell^\infty} \le \|\bar{\Phi}\|_{\ell^\infty} =: B.$$

Also, since the distribution ν of the rows of matrices drawn from $\mathcal{P}_N$ is of bounded support, when $x \sim \nu$ we may assume that almost surely

$$\|x\| \le C.$$

Recall that our goal is to show that

$$\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) = \mathbb{E}\exp^{\le D}(R) - 1 \overset{(?)}{=} O(1).$$

Let us begin by proving some preliminary bounds on the overlap. We may decompose

$$R(V^1, V^2) = R^{(0)}(V^1, V^2) + R^{(1)}(V^1, V^2),$$

where

$$\begin{aligned} R^{(0)}(V^1, V^2) &:= \frac{\beta^2}{2N} \sum_{a,b=1}^{\kappa} (v_a^1 \odot v_b^2)^\top \Phi (v_a^1 \odot v_b^2) \\ &= \frac{\beta^2}{2N} \sum_{a,b=1}^{\kappa} (P(v_a^1 \odot v_b^2))^\top \Gamma (P(v_a^1 \odot v_b^2)) \end{aligned}$$

and

$$R^{(1)}(V^1, V^2) := \frac{\beta^2}{2N} \sum_{a,b=1}^{\kappa} (v_a^1 \odot v_b^2)^\top \mathrm{diag}(\Phi)(v_a^1 \odot v_b^2).$$

We drop the arguments (V^1, V^2) from $R, R^{(0)}$, and $R^{(1)}$ below for the sake of clarity. It will turn out that $R^{(0)}$ is the "main term" of R, while $R^{(1)}$ makes a negligible contribution.

Using Propositions 3.29 and 3.30, we may bound

$$\begin{aligned} \chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) &\le \mathbb{E}\exp^{\le D}(R) \\ &\le \mathbb{E}\exp^{\le D}(|R|) \\ &\le \mathbb{E}\exp^{\le D}(|R^{(0)}| + |R^{(1)}|). \end{aligned}$$

Let us show first that we may effectively get rid of the $|R^{(1)}|$ term here. We have, by our boundedness assumption on $x \sim \nu$, that we have almost surely

$$\begin{aligned} |R^{(1)}| &\le \frac{\beta^2}{2N} \|\Phi\|_{\ell^\infty} \sum_{a,b=1}^{\kappa} \|v_a^1 \odot v_b^2\|^2 \\ &= \frac{\beta^2}{2N} \|\Phi\|_{\ell^\infty} \sum_{a,b=1}^{\kappa} \sum_{i=1}^{N} (v_a^1)_i^2 (v_b^2)_i^2 \\ &= \frac{\beta^2}{2N} \|\Phi\|_{\ell^\infty} \sum_{i=1}^{N} \|x_i^1\|^2 \|x_i^2\|^2 \\ &\le \frac{\beta^2 B C^4}{2}, \end{aligned}$$

which is just a constant. Using Proposition 3.30, we may then bound

$$\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) \le \exp(50\beta^2 B C^4) \cdot \mathbb{E}\exp^{\le D}(|R^{(0)}|),$$

and thus we have reduced our task to just showing that

$$\mathbb{E}\exp^{\le D}(|R^{(0)}|) \overset{(?)}{=} O(1).$$

We now work on bounding $R^{(0)}$. Since Γ has non-negative entries whereby, by the Perron-Frobenius theorem (Theorem 3.1) we have $\lambda_1(\Gamma) = \|\Gamma\|$, we also have

$$\begin{aligned}
|R^{(0)}| &\leq \frac{\beta^2 \lambda_1(\Gamma)}{2N} \sum_{a,b=1}^{\kappa} \|P(v_a^1 \odot v_b^2)\|^2 \\
&= \frac{\beta^2 \lambda_1(\Gamma)}{2N} \sum_{a,b=1}^{\kappa} \sum_{h=1}^{n} \left(\frac{1}{\sqrt{|S_h|}} \sum_{i \in S_h} (v_a^1)_i (v_b^2)_i \right)^2 \\
&= \frac{\beta^2 \lambda_1(\Gamma)}{2N} \sum_{h=1}^{n} \frac{1}{|S_h|} \sum_{a,b=1}^{\kappa} \left(\sum_{i \in S_h} (v_a^1)_i (v_b^2)_i \right)^2,
\end{aligned}$$

where we note that the factor of $\lambda_1(\Gamma)/N$ on the outside will, in the limit $N \to \infty$, lead to the factor of $\widehat{\mu}_1$ that we expect. Now, let us define the matrix $T^{(h)} \in \mathbb{R}^{\kappa \times \kappa}$ for each $h \in [n]$ whose entries are the "partial inner products" appearing here:

$$T_{ab}^{(h)} := \sum_{i \in S_h} (v_a^1)_i (v_b^2)_i.$$

Then, we may rewrite the above bound, also separating the leading factors per the above observation, as

$$|R^{(0)}| \leq \beta^2 \cdot \frac{\lambda_1(\Gamma)}{N} \cdot \frac{1}{2} \sum_{h=1}^{n} \frac{\|T^{(h)}\|_F^2}{|S_h|}. \tag{5}$$

We would like to apply Lemma 3.31 to $|R^{(0)}|$. Recall that the Lemma asks for two conditions: a weak but almost sure bound on this random variable, and a bound on the probability of certain large deviations. For the former, note that we can write

$$T^{(h)} = U^{(h,1)\top} U^{(h,2)},$$

where $U^{(h,\ell)} \in \mathbb{R}^{S_h \times \kappa}$ has entries

$$U_{ia}^{(h,\ell)} = (v_a^\ell)_i.$$

Equivalently, the rows of $U^{(h,\ell)}$ are those x_i^ℓ for $i \in S_h$. Thus, $\|U^{(h,\ell)}\|_F^2 \leq |S_h| \cdot C^2$, since each of these rows has norm at most C by our assumption on the boundedness of ν. Now, we have

$$\|T^{(h)}\|_F^2 \leq \|U^{(h,1)}\|_F^2 \|U^{(h,2)}\|_F^2 \leq C^4 |S_h|^2,$$

and therefore, almost surely,

$$\begin{aligned}
|R^{(0)}| &\leq \beta^2 \cdot \frac{\lambda_1(\Gamma)}{N} \cdot \frac{1}{2} \sum_{h=1}^{n} \frac{\|T^{(h)}\|_F^2}{|S_h|} \\
&\leq \beta^2 C^4 \cdot \frac{\lambda_1(\Gamma)}{N} \cdot \frac{1}{2} \sum_{h=1}^{n} |S_h| \\
&= \frac{1}{2} \beta^2 C^4 \cdot \lambda_1(\Gamma)
\end{aligned}$$

and finally, a simple bound gives that, since $|S_i| \leq N$, we have $\lambda_1(\Gamma) \leq \lambda_1(\bar{\Phi}) N = \widehat{\mu}_1 N$, so

$$\leq \frac{1}{2} \beta^2 C^4 \widehat{\mu}_1 N$$

The details of the constants here will be inconsequential, so let us just write

$$=: C'N.$$

Since we plan to take $D(N) \leq N/\log(N)$ (indeed, even smaller, but at least satisfying this inequality for sufficiently large N), the condition of Lemma 3.31 requires that we take the value $A(N)$ to be at least $D(N)\log\log N \leq N \cdot \frac{\log\log N}{\log N}$, for sufficiently large N. In particular, it is permissible to take $A(N) := \delta N$ for a small constant $\delta > 0$ to be chosen later.

Recall that, to apply the Lemma, we then need to bound the tail probability $\mathbb{P}[|R^{(0)}| \geq t]$ for all $t \in [0, \delta N]$. For the Lemma to apply, we need a tail bound on this probability that is "slightly better" than $\exp(-t)$. We first take some initial steps towards showing this. Starting from (5), which we repeat below, we have

$$|R^{(0)}| \leq \beta^2 \cdot \frac{\lambda_1(\Gamma)}{N} \cdot \frac{1}{2}\sum_{h=1}^{n} \frac{\|T^{(h)}\|_F^2}{|S_h|}$$

Now, as we noted earlier, as $N \to \infty$ we have $\lambda_1(\Gamma)/N \to \widehat{\mu}_1$, and we have also assumed that $\beta^2 < 1/\widehat{\mu}_1$. Thus, there exists an $\varepsilon > 0$ such that, for all sufficiently large N, we have

$$\leq (1-\varepsilon) \cdot \frac{1}{2}\sum_{h=1}^{n} \frac{\|T^{(h)}\|_F^2}{|S_h|}. \tag{6}$$

Since $T^{(h)}$ involves only the rows of V^1 and V^2 indexed by S_h, the different S_h are disjoint (forming a partition of $[N]$), and the rows of V^1 and V^2 are all i.i.d. by assumption, we see that the $T^{(h)}$ are independent random matrices.

As a heuristic aside, by the central limit theorem we expect that each $\|T^{(h)}\|_F^2/|S_h|$ has roughly the distribution $\chi^2(\kappa^2)$. Thus, the entire sum above should have roughly the distribution $\chi^2(\kappa^2 n)$. If this were indeed exactly the case, then the tail bound we want would follow by (a simple variant of) Bernstein's inequality, the tails of this distribution looking, for large enough deviations, like those of $\chi^2(1)$, which precisely decay as $\exp(-t)$.

Now let us actually implement this reasoning. By our previous observation, we may rewrite

$$T^{(h)} = \sum_{i \in S_h} x_i^1 {x_i^2}^\top,$$

or, upon vectorizing this matrix,

$$\mathrm{vec}(T^{(h)}) = \sum_{i \in S_h} x_i^1 \otimes x_i^2.$$

The summands here are i.i.d. random vectors. Let us write $\nu^{(2)}$ for their law, which is the law of $x \otimes x'$ with $x, x' \sim \nu$ drawn independently. Consider the moment generating function of such a random vector, $M : \mathbb{R}^{\kappa^2} \to \mathbb{R}$ given by

$$M(\xi) := \mathop{\mathbb{E}}_{x,x' \sim \nu} \exp(\langle \xi, x \otimes x' \rangle).$$

Since ν has bounded support, this expectation is always finite. By the generating function expansion

of the moment generating function, the gradient and Hessian of M at zero are

$$\begin{aligned}\nabla M(0) &= \mathop{\mathbb{E}}_{x,x'\sim\nu} x \otimes x' \\ &= 0, \\ \nabla^2 M(0) &= \mathop{\mathbb{E}}_{x,x'\sim\nu} (x\otimes x')(x\otimes x')^\top \\ &= \mathbb{E}(xx^\top \otimes x'x'^\top) \\ &= (\mathop{\mathrm{Cov}}_{x\sim\nu}[x])^{\otimes 2} \\ &\preceq I_{\kappa^2},\end{aligned}$$

the final observation following by our assumption on the covariance of ν. Taking a Taylor expansion, it then follows that, for all $\eta > 0$, there exists $\gamma > 0$ such that

$$M(\xi) \le \exp\left(\frac{1}{2}(1+\eta)\|\xi\|^2\right) \text{ for all } \|\xi\| \le \gamma.$$

Thus, for any choice of these parameters, each summand appearing in $\mathrm{vec}(T^{(h)})$ is γ-locally $(1+\eta)$-subgaussian. Accordingly, since these summands are independent, $\mathrm{vec}(T^{(h)})$ itself is γ-locally $(1+\eta)|S_h|$-subgaussian, and $\mathrm{vec}(T^{(h)})/\sqrt{|S_h|}$ is $\gamma\sqrt{|S_h|}$-locally $(1+\eta)$-subgaussian. Finally, define a vector $w \in \kappa^2 n$ to be the concatenation of the $\mathrm{vec}(T^{(h)})/\sqrt{|S_h|}$ over $h = 1,\dots,n$. Then, w is $(\gamma \min_{h\in[n]}\sqrt{|S_h|})$-locally $(1+\eta)$-subgaussian, and we have

$$\|w\|^2 = \sum_{h=1}^n \frac{\|T^{(h)}\|_F^2}{|S_h|},$$

precisely the expression from our bound on $|R^{(0)}|$. Since $|S_h|/N \to \rho_h > 0$ for all $h \in [n]$, for sufficiently large N we will have $|S_h| \ge \frac{1}{2}\rho_{\min} N$ where $\rho_{\min} := \min_{h\in[n]}\rho_h$. Thus, we find that, for N sufficiently large, w is $(\gamma\sqrt{\frac{1}{2}\rho_{\min}N})$-locally $(1+\eta)$-subgaussian.

Now, we have

$$|R^{(0)}| \le (1-\varepsilon)\cdot\frac{1}{2}\|w\|^2.$$

By Proposition 3.19, taking $\delta = \eta$, we have that, for a constant $C'' = C''(\eta)$, for all $0 \le t \le \gamma\frac{1+\eta}{1-\eta}\sqrt{\frac{1}{2}\rho_{\min}N}$

$$\mathbb{P}[\|w\| \ge t] \le C''\exp\left(-\frac{1-\eta}{2(1+\eta)}t^2\right)$$

Thus we find a tail bound on $R^{(0)}$:

$$\begin{aligned}\mathbb{P}[|R^{(0)}| \ge t] &\le \mathbb{P}\left[\|w\|^2 \ge \frac{2}{1-\varepsilon}t\right] \\ &\le C''\exp\left(-\frac{1-\eta}{(1+\eta)(1-\varepsilon)}t\right),\end{aligned} \tag{7}$$

which holds whenever $\sqrt{\frac{2}{1-\varepsilon}t} \le \gamma\sqrt{\frac{1}{2}\rho_{\min}N}$, or equivalently whenever $t \le \frac{1}{4}\gamma^2(1-\varepsilon)\rho_{\min}N$.

Let us review the sequence of choices of parameters. We determine $\varepsilon \in (0,1)$ according to the relation between β^2 and $\widehat{\mu}_1$, both of which are given in the setting of the Theorem. We then choose η depending on ε such that $\frac{1-\eta}{(1+\eta)(1-\varepsilon)} = 1+\zeta > 1$ for some $\zeta > 0$. The above reasoning then gives

γ depending on η the parameter of local subgaussianity above. Finally, the tail bound (7) then holds for all $0 \le t \le \delta N$, for $\delta := \frac{1}{4}\gamma^2(1-\varepsilon)\rho_{\min}$.

In summary, with these choices, we may take $A(N) := \delta N$ such that, for all $t \in [0, A(N)]$ we have

$$\mathbb{P}[|R^{(0)}| \ge t] \le C'' \exp\left(-(1+\zeta)t\right).$$

Thus, the conditions of Lemma 3.31, and finally we find that

$$\limsup_{N\to\infty} \mathbb{E}\exp^{\le D}(|R^{(0)}|) < \infty,$$

concluding the proof along with our previous reduction from an expectation over R to the above one over $R^{(0)}$. □

4.3 Statistical Lower Bound

Proof. We follow the same outline as before; in this case, we want to show that

$$\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) = \mathbb{E}\exp(R) - 1 \overset{(?)}{=} O(1).$$

Bounding $|R| \le |R^{(0)}| + |R^{(1)}|$ as before and using that $|R^{(1)}| = O(1)$ almost surely under our assumptions, it suffices to show that $\mathbb{E}\exp(|R^{(0)}|) = O(1)$. We reuse the bound from (6) from the previous proof, which gives that, for some $\varepsilon > 0$ depending on β,

$$\mathbb{E}\exp(|R^{(0)}|) \le \mathbb{E}\exp\left(\frac{1-\varepsilon}{2}\sum_{h=1}^{n}\frac{\|T^{(h)}\|_F^2}{|S_h|}\right)$$

and, using the independence of the matrices $T^{(h)}$, we may factorize this as

$$= \prod_{h=1}^{n} \mathbb{E}\exp\left(\frac{1-\varepsilon}{2}\cdot\frac{\|T^{(h)}\|_F^2}{|S_h|}\right)$$

Recall that for this result we have the extra assumption that, when $x, x' \sim \nu$ are independent, then $x \otimes x'$ is a 1-subgaussian random vector. Further, as we used above, the vectorization $\mathrm{vec}(T^{(h)})$ is a sum of $|S_h|$ independent vectors having this same distribution, so in particular $\mathrm{vec}(T^{(h)})/\sqrt{|S_h|}$ is again a 1-subgaussian random vector. By Proposition 3.16, all of the expectations in this product are finite, and thus since there is a fixed number n of terms in the product, we have

$$= O(1),$$

completing the proof. □

4.4 Growth of Low-Degree χ^2-Divergence: Proof of Theorem 1.12

We adapt some of the calculations from Section 4.2. Recall that the low-degree χ^2-divergence is

$$\begin{aligned} 1 + \chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) &= \mathbb{E}\exp^{\le D}(R) \\ &= \mathbb{E}\exp^{\le D}(R^{(0)} + R^{(1)}), \end{aligned}$$

where $|R^{(1)}|$ is bounded by an absolute constant. Thus, by Proposition 3.30, it suffices to show that

$$\mathbb{E}\exp^{\le D}(R^{(0)}) = \omega(1)$$

as $N \to \infty$.

Recall also that, setting $u_{ab} := P(v_a^1 \odot v_b^2)$, we have

$$R^{(0)} = \frac{\beta^2}{2N} \sum_{a,b=1}^{\kappa} u_{ab}^\top \Gamma u_{ab} = \left\langle \Gamma, \sum_{a,b=1}^{\kappa} u_{ab} u_{ab}^\top \right\rangle.$$

Let us define

$$M := \sum_{a,b=1}^{\kappa} u_{ab} u_{ab}^\top \in \mathbb{R}^{n \times n}_{\mathrm{sym}}.$$

The entries of this matrix are

$$\begin{aligned}
M_{rs} &= \sum_{a,b=1}^{\kappa} (u_{ab})_r (u_{ab})_s \\
&= \sum_{a,b=1}^{\kappa} \left(\frac{1}{\sqrt{|S_r|}} \sum_{i \in S_r} (v_a^1)_i (v_b^2)_i \right) \left(\frac{1}{\sqrt{|S_s|}} \sum_{i \in S_s} (v_a^1)_i (v_b^2)_i \right) \\
&= \sum_{a,b=1}^{\kappa} \left(\frac{1}{\sqrt{|S_r|}} \sum_{i \in S_r} (x_i^1 \otimes x_i^2)_a \right) \left(\frac{1}{\sqrt{|S_s|}} \sum_{i \in S_s} (x_i^1 \otimes x_i^2)_b \right) \\
&= \left\langle \frac{1}{\sqrt{|S_r|}} \sum_{i \in S_r} x_i^1 \otimes x_i^2, \frac{1}{\sqrt{|S_s|}} \sum_{i \in S_s} x_i^1 \otimes x_i^2 \right\rangle.
\end{aligned}$$

Recall that here the x_i^1 and x_i^2 for $i \in [N]$ are all i.i.d. draws from the distribution ν on $\mathbb{R}^\kappa$, and the S_r are a partition of $[N]$ each of whose blocks has diverging size. Let us write $\Sigma := \mathrm{Cov}_{x \sim \nu}[x]$. Since we also assume that ν is centered, we have $\mathrm{Cov}_{x^1, x^2 \sim \nu}[x^1 \otimes x^2] = \Sigma^{\otimes 2}$. By the central limit theorem, we then have the convergence in distribution

$$\left(\frac{1}{\sqrt{|S_1|}} \sum_{i \in S_1} x_i^1 \otimes x_i^2, \ldots, \frac{1}{\sqrt{|S_n|}} \sum_{i \in S_n} x_i^1 \otimes x_i^2 \right) \xrightarrow{\text{(d)}} (g_1, \ldots, g_n)$$

where $g_i \sim \mathcal{N}(0, \Sigma \otimes \Sigma)$ are i.i.d. Gaussian random vectors in $\mathbb{R}^{\kappa^2}$. Writing $G \in \mathbb{R}^{\kappa^2 \times n}$ for the matrix with these g_i as its columns, we then also have

$$M \xrightarrow{\text{(d)}} G^\top G,$$

and therefore, using also that $\frac{1}{N}\Gamma \to \widetilde{\Gamma}$ entrywise,

$$R^{(0)} = \frac{\beta^2}{2N} \langle \Gamma, M \rangle \xrightarrow{\text{(d)}} \frac{\beta^2}{2} \langle \widetilde{\Gamma}, G^\top G \rangle = \frac{\beta^2}{2} \mathrm{Tr}(G \widetilde{\Gamma} G^\top).$$

Since the underlying distribution ν is bounded, a routine approximation argument shows that the moments of $R^{(0)}$ converge to the corresponding moments of the right-hand side as well: for any fixed $d \geq 0$,

$$\mathbb{E} R^{(0)^d} \to \left(\frac{\beta^2}{2} \right)^d \mathbb{E} \left(\mathrm{Tr}(G \widetilde{\Gamma} G^\top) \right)^d$$

Next, we "whiten" the distribution of G: introducing H of the same shape with i.i.d. entries distributed as $\mathcal{N}(0,1)$, G has the same law as $(\Sigma^{\otimes 2})^{1/2}H$. In terms of H, we may rewrite

$$= \left(\frac{\beta^2}{2}\right)^d \mathbb{E}\left(\mathrm{Tr}((\Sigma^{\otimes 2})^{1/2}H\widetilde{\Gamma}H^\top(\Sigma^{\otimes 2})^{1/2})\right)^d$$
$$= \left(\frac{\beta^2}{2}\right)^d \mathbb{E}\left(\mathrm{Tr}(\Sigma^{\otimes 2}H\widetilde{\Gamma}H^\top)\right)^d$$

Further, by the rotational invariance of H, we may diagonalize the two deterministic matrices appearing here without affecting the value. Letting $\lambda(X)$ denote the vector of eigenvalues of a symmetric X, we have

$$= \left(\frac{\beta^2}{2}\right)^d \mathbb{E}\left(\mathrm{Tr}(\mathrm{diag}(\lambda(\Sigma^{\otimes 2}))H\mathrm{diag}(\lambda(\widetilde{\Gamma}))H^\top)\right)^d$$
$$= \left(\frac{\beta^2}{2}\right)^d \mathbb{E}\left(\sum_{a,b=1}^{\kappa}\sum_{r=1}^{n}\lambda_a(\Sigma)\lambda_b(\Sigma)\lambda_r(\widetilde{\Gamma})H^2_{(a,b),r}\right)^d,$$

where we identify $[\kappa^2]$ with $[\kappa]\times[\kappa]$ in indexing the rows of H. Finally, Proposition 3.8 gives a lower bound on such moments of quadratic forms,

$$\geq \left(\frac{\beta^2}{2}\right)^d \cdot 2^{d-1}(d-1)!\sum_{a,b=1}^{\kappa}\sum_{r=1}^{n}\lambda_a(\Sigma)^d\lambda_b(\Sigma)^d\lambda_r(\widetilde{\Gamma})^d$$

Here, we have $\|\Sigma\| = 1$ and $\Sigma \succeq 0$, so $\lambda_1(\Sigma) = 1$, while $\lambda_1(\widetilde{\Gamma}) = \widehat{\mu}_1$ and so, for d even,

$$\geq \frac{(d-1)!}{2}(\beta^2\widehat{\mu}_1)^d$$

Now, we have

$$\mathbb{E}\exp^{\leq D}(R^{(0)}) = \sum_{d=0}^{D}\frac{1}{d!}\mathbb{E}R^{(0)^d}$$

Here, since we can write $R^{(0)} = \langle U^1, U^2\rangle$ for two suitable i.i.d. random vectors U^i, every term is non-negative. So, we may bound below by only the even terms, and fixing some $2d_0$ not depending on N, for sufficiently large N we have $D = D(N) \geq 2d_0$ as $D(N)\to\infty$ by assumption. So, we may further bound below by only the even d terms between $d = 2$ and $d = 2d_0$, obtaining after reindexing

$$\geq \sum_{d=1}^{d_0}\frac{1}{(2d)!}\mathbb{E}R^{(0)^{2d}}$$
$$\geq \sum_{d=1}^{d_0}\frac{1}{(2d)!}\cdot\frac{(2d-1)!}{2}(\beta^2\widehat{\mu}_1)^{2d}$$
$$= \frac{1}{4}\sum_{d=1}^{d_0}\frac{1}{d}(\beta^2\widehat{\mu}_1)^{2d}.$$

Since by assumption $\beta^2\widehat{\mu}_1 > 1$, we obtain a diverging lower bound taking $d_0\to\infty$.

5 General Variance Profiles: Proof of Theorem 1.11

In this section, we prove our main result on the case of general variance profiles Δ, not necessarily having block structure. Our approach in this case will be quite different. In the block-structured case, we have seen that the low-degree χ^2-divergence calculations become essentially finite-dimensional after appropriately reorganizing to take advantage of the averaging occurring over blocks. Here, recall that we assume that the signal has rank 1, and thus our overlap random variable R takes the form

$$R = \frac{\beta^2}{2N}(x^1 \odot x^2)^\top \Phi'(x^1 \odot x^2),$$

where $\Phi' = \Phi + \mathrm{diag}(\Phi)$ is our slight modification of the entrywise reciprocal of the variance profile matrix (as we will see, just like in the block-structured case, the second term here is inconsequential).

Recall that our initial calculations gave the formula

$$\chi^2_{\le D}(\mathbb{P} \mid \mathbb{Q}) = \mathbb{E}\left[\exp^{\le D}(R)\right] - 1$$

In the coming proof, we will expand this into a power series,

$$= \sum_{d=1}^{D} \frac{1}{d!}\mathbb{E}R^d,$$

and will further expand $\mathbb{E}R^d$ and evaluate or bound the expectations over x^1 and x^2 that appear. In each such term, we will then be left with a complicated polynomial of Φ. So, before continuing to the proof itself, we develop some tools for bounding the polynomials that will appear.

5.1 Tools for Graph Sums of a Matrix

5.1.1 Algebraic Properties

The specific polynomials we will need to control are as follows.

Definition 5.1 (Scalar graph functions of a matrix)**.** *Let $G = (V, E)$ be a multigraph (that is, allowing self-loops and parallel edges) with labelled vertex set $V = [M]$, and let $\Phi \in \mathbb{R}^{N\times N}_{\mathrm{sym}}$.*

$$f_G(\Phi) := \sum_{i:[M]\hookrightarrow[N]} \prod_{\{x,y\}\in E(G)} \Phi_{i(x)i(y)}$$
$$\tilde{f}_G(\Phi) := \sum_{i:[M]\to[N]} \prod_{\{x,y\}\in E(G)} \Phi_{i(x)i(y)}$$

The difference between the two expressions is that, in the first one, the "labelling function" i is constrained to be injective, while in the second it is not.

The $f_G(\Phi)$ are scalar-valued versions of the *graph matrices* that have found applications in the analysis of sum-of-squares optimization and, more recently, other algorithmic settings [AMP16, BHK+19, HKP+17, HK22, PVTW22, HKPX23]. The $\tilde{f}_G(\Phi)$ have appeared in the theories of free probability and traffic probability as well [BS10, MS12, ACD+21]. Both can be used as bases for the polynomials of (the entries of) Φ that are invariant under the two-sided action of the symmetric group; for further discussion on this perspective see [KMW24].

For many of the results we will need, it will turn out to be more convenient to form graphs by "gluing" certain vertices of a perfect matching. We define notation to work with this notion as follows. The below will rely on our definitions concerning partitions from Section 3.2.

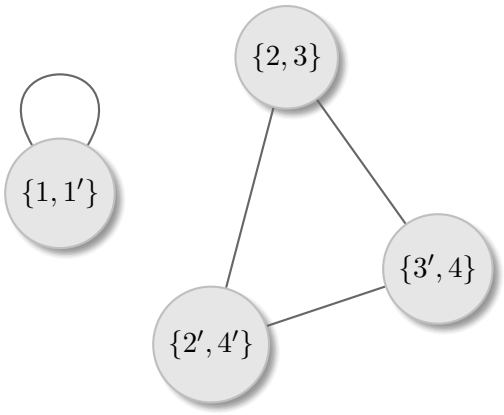

Figure 1: The graph $G(\pi)$ represented by the partition $\pi = \{\{1, 1'\}, \{2, 3\}, \{3', 4\}, \{4', 2'\}\}$ of the set $\mathcal{I} = \{1, 1', 2, 2', 3, 3', 4, 4'\}$.

Definition 5.2 (Partition representing a graph)**.** *Let us identify* $[2d]$ *with the more convenient indexing set* $\mathcal{I} = \{1, 1', 2, 2', \ldots, d, d'\}$*. Let* π *be a partition of* $\mathcal{I}$*, and let* $b : \mathcal{I} \to \pi$ *be the function so that* $b(x)$ *equals the block of* π *that* x *belongs to. We write* $G(\pi) = (V, E)$ *for the multigraph with* $|\pi|$ *vertices, identified with* π *itself, whose multiset of edges is* $E = \{\{b_\pi(x), b_\pi(x')\} : x \in [d]\}$*.*

We use the same notation to refer to sums associated to graphs associated to a partition:

Definition 5.3. *For* π *a partition of* $\mathcal{I}$ *as above, we write*

$$f_\pi(\Phi) := f_{G(\pi)}(\Phi) = \sum_{i:\pi \hookrightarrow [N]} \prod_{x=1}^{d} \Phi_{i(b(x))i(b(x'))},$$

$$\tilde{f}_\pi(\Phi) := \tilde{f}_{G(\pi)}(\Phi) = \sum_{i:\pi \to [N]} \prod_{x=1}^{d} \Phi_{i(b(x))i(b(x'))}.$$

Example 5.4. *To illustrate, for the case* $d = 4$ *and* $\pi = \{(1, 1'), (2, 3), (3', 4), (2', 4')\}$*, the corresponding graph* $G(\pi)$ *is represented by Figure 1. The associated polynomials of a matrix are:*

$$f_\pi(\Phi) = f_{G(\pi)}(\Phi) = \sum_{i,j,k,l \text{ distinct}} \Phi_{ii}\Phi_{jl}\Phi_{jk}\Phi_{kl},$$

$$\tilde{f}_\pi(\Phi) = \tilde{f}_{G(\pi)}(\Phi) = \sum_{i,j,k,l} \Phi_{ii}\Phi_{jl}\Phi_{jk}\Phi_{kl} = \mathrm{Tr}(\Phi)\mathrm{Tr}(\Phi^3).$$

The following describes the simple but important relationship between the two polynomial families f_G and $\tilde{f}_G$.

Proposition 5.5. *Let* π *be a partition of* $\mathcal{I}$ *as above. Then, for any* $\Phi \in \mathbb{R}^{N \times N}_{\mathrm{sym}}$*,*

$$\tilde{f}_\pi(\Phi) = \sum_{\rho \preceq \pi} f_\pi(\Phi).$$

Proof. Given $\tilde{i} : \pi \to [N]$, there are a unique $\rho \preceq \pi$ and $i : \rho \hookrightarrow [N]$ such that, for $q : \pi \to \rho$ as above, $\tilde{i}(S) = i(q(S))$ for each $S \in \pi$. In words, this ρ is formed by merging the blocks of π on

which $\tilde{i}$ is constant, and i is the same function of $\tilde{i}$ but on these coarser blocks. Now, working from the definition,

$$\begin{aligned}\tilde{f}_\pi(\Phi) &= \sum_{\tilde{i}:\pi\to[N]} \prod_{x=1}^{d} \Phi_{\tilde{i}(b_\pi(x))\tilde{i}(b_\pi(x'))} \\ &= \sum_{\rho\preceq\pi} \sum_{i:\rho\hookrightarrow[N]} \prod_{x=1}^{d} \Phi_{i(b_\rho(x))i(b_\rho(x'))} \\ &= \sum_{\rho\preceq\pi} f_\rho(\Phi),\end{aligned}$$

completing the proof. □

One may also invert the above relation using Möbius inversion in the partially ordered set of partitions, but we will only need the simpler above direction.

5.1.2 Matchings and 2-Regular Graphs

The case of perfect matchings $\pi \in \mathrm{Match}(\mathcal{I})$ in the above setting will play a crucial role throughout. There are two coincident reasons for the importance of the perfect matchings: on the one hand, they are the even partitions with the greatest number of parts, and thus "entropically dominate" certain calculations we will encounter; on the other hand, the only cases of $f_\pi(\Phi)$ or $\tilde{f}_\pi(\Phi)$ that have a natural interpretation in terms of the spectrum of Φ are the $\tilde{f}_\pi(\Phi)$ where $\pi \in \mathrm{Match}(\mathcal{I})$, as appears in the above example.

Indeed, for π a matching we have that $G(\pi)$ is a 2-regular multigraph on d vertices, or a disjoint union of several cycles (this is a multigraph because some of the cycles may be self-loops or pairs of parallel edges forming 2-cycles).

Definition 5.6. *For G a 2-regular graph, we write $\mathrm{cyc}(G)$ for the set of cycles of G. Formally, this may be viewed as $\mathrm{cyc}(G) \in \mathrm{Part}(V(G))$, a partition of the vertices into those belonging to each cycle of G.*

It is well-known that one may sample uniformly from the set of 2-regular multigraphs on d vertices using the *configuration model*: to each vertex we assign two "half-edges" or "stubs" (for a total of $2d$ half-edges), and then form a graph by linking half-edges along a uniformly random perfect matching. Analyzing this probability measure, the following result has been derived in prior work.

Proposition 5.7 ([TBKK23])**.** *Let $\mathcal{U}$ be the set of all 2-regular multigraphs on the labelled vertex set $[d]$ (which also must have d edges). Let $\mathcal{G}$ be the uniform distribution over $\mathcal{U}$. Then, for any $a > 0$,*

$$\begin{aligned}|\mathcal{U}| &= (2d-1)!!, \\ \mathop{\mathbb{E}}_{G\sim\mathcal{G}}\left[a^{|\mathrm{cyc}(G)|}\right] &= \frac{2^d}{(2d-1)!!}\frac{\Gamma\left(d+\frac{a}{2}\right)}{\Gamma\left(\frac{a}{2}\right)}.\end{aligned}$$

Equivalently,

$$\sum_{G\in\mathcal{U}} a^{|\mathrm{cyc}(G)|} = 2^d \cdot \frac{\Gamma\left(d+\frac{a}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} = \prod_{j=1}^{d}(a+2j-2).$$

One may also take a "dual" viewpoint to the configuration model: suppose that we have a fixed perfect matching on a set of vertices of size $2d$ (such as $\mathcal{I}$ discussed above). Form a uniformly random perfect matching of those vertices, where the d pairs of this perfect matching are also uniformly labelled by $[d]$. Identifying each pair of vertices joined in this perfect matching and giving the resulting vertex the label of that pair yields a random graph in $\mathcal{U}$. By symmetry considerations we see that this is again a uniformly random graph in $\mathcal{U}$, i.e., that this has the law $\mathcal{G}$. But, what we have described above is precisely the process of drawing $\pi \sim \mathrm{Unif}(\mathrm{Match}(\mathcal{I}))$ and forming $G(\pi)$, except that we endow the pairs of π with a random labelling. Since this labelling has no effect on $|\mathrm{cyc}(G(\pi))|$, we find the following corollary that will be useful in our calculations.

Corollary 5.8. *For any $a > 0$,*

$$\sum_{\pi \in \mathrm{Match}(\mathcal{I})} a^{|\mathrm{cyc}(G(\pi))|} = 2^d \cdot \frac{\Gamma\left(d + \frac{a}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} = \prod_{j=1}^{d} (a + 2j - 2).$$

5.1.3 Sharpened Bounds on Positive Graph Sums

We will be interested in bounding the $f_G(\Phi)$ and $\tilde{f}_G(\Phi)$ expressions for Φ the inverse variance profile matrix discussed previously. Recall that this matrix has non-negative entries, and in the block-structured case is low-rank. Our assumptions also roughly impose that all entries of Φ are of order $\Theta(1)$ (we actually only require a uniform upper bound, but we think of all entries being of the same constant order as the "typical case," which for example is true in the case of block-structured models).

While these polynomials have appeared in the literature before and some bounds have been proved on them, usually these bounds are more effective when neither of the above assumptions holds. The following is the one of the main and most useful bounds on these expressions stemming from the free probability literature mentioned above; for the sake of exposition we restrict our attention to a special class of graphs.

Proposition 5.9 ([BS10, MS12])**.** *Suppose that G is a 2-edge-connected graph. Then,*

$$\tilde{f}_G(\Phi) \leq N \cdot \|\Phi\|^{|E(G)|}.$$

The following example shows that, in the simplest choice of Φ that will arise in our setting, this bound is already quite loose.

Example 5.10 (All-ones matrix)**.** *If $\Phi = 11^\top$, then we have $\|\Phi\| = N$, so the right-hand side of the bound above is $N^{|E(G)|+1}$. On the other hand, the actual value is easily computed to be $\tilde{f}_G(\Phi) = N^{|V(G)|}$. Since for a connected graph G we have $|V(G)| = |E(G)| + 1$ if and only if G is a tree, which is never 2-edge-connected, we see that on 2-edge-connected graphs the above bound is always suboptimal on this Φ.*

Our main goal here will be to develop bounds that behave, for the Φ with the rough structure outlined above, more like $\tilde{f}_G(\Phi) \lesssim N^{|V(G)|}$.

We build up some tools for increasingly more general bounds on the $\tilde{f}_G(\Phi)$, starting from simple G and using those bounds to treat more general G.

Proposition 5.11 (Paths)**.** *Let G be a path on $k \geq 1$ vertices (where the path on $k = 1$ vertex is a single isolated vertex). Then,*

$$\tilde{f}_G(\Phi) \leq N \|\Phi\|^{k-1} = \left(\frac{N}{\|\Phi\|}\right) \|\Phi\|^k.$$

Proposition 5.12 (Cycles)**.** *Let G be a cycle on $k \geq 1$ vertices (where the cycle on $k = 1$ vertex is a single self-loop). Then, both of the following hold:*

$$\begin{aligned}\tilde{f}_G(\Phi) &\leq \operatorname{rank}(\Phi)\|\Phi\|^k, \\ \tilde{f}_G(\Phi) &\leq N\|\Phi\|_{\ell^\infty}\|\Phi\|^{k-1} \\ &= \left(\frac{N\|\Phi\|_{\ell^\infty}}{\|\Phi\|}\right)\|\Phi\|^k.\end{aligned}$$

To compare the above bounds, recall that we are interested in Φ that behave like the all-ones matrix. In particular, we expect to have $\|\Phi\|_{\ell^\infty} = O(1)$, while $\|\Phi\| = \Theta(N)$. Thus when $\operatorname{rank}(\Phi) = O(1)$ then the above two bounds are of the same order, while when $\operatorname{rank}(\Phi) = \omega(1)$ the second bound can be smaller than the first. For this reason, and to avoid our results depending on the more fragile property of the rank, we will rely on the second bound above.

In general, one may of course also bound $\tilde{f}_G(\Phi) \leq N^k\|\Phi\|_{\ell^\infty}^k$ in the above setting, but we will see that to obtain sharp results it is important for us that it is $\|\Phi\|^k$ that appears in our bound, not $(N \cdot \|\Phi\|_{\ell^\infty})^k$, which may be viewed as replacing $\|\Phi\|$ by a coarse entrywise bound on this norm. The situation in our application will be delicate: it will be acceptable for this coarse bound expression to appear linearly, as it does above, but not with the possibly much larger exponent k.

Finally, we show how to control much more general graph sums by reducing them to the above case.

Proposition 5.13. *Suppose that G and H are graphs on the same vertex set, and H is formed by deleting t edges from G. Then,*

$$\tilde{f}_G(\Phi) \leq \|\Phi\|_{\ell^\infty}^t \tilde{f}_H(\Phi).$$

Proposition 5.14. *Let G be a multigraph with $|V(G)| = m$ and $|E(G)| = d$ where every vertex has degree at least 2. Then, it is possible to remove at most $2d - 2m \geq 0$ edges from G to leave a graph consisting of a disjoint union of paths (including isolated vertices) and cycles (including isolated self-loops), and having at most $2d - 2m + |\operatorname{conn}(G)|$ such connected components.*

Proof. For each vertex v of G, choose an arbitrary $\deg(v) - 2 \geq 0$ edges adjacent to that vertex to remove, allowing edges to be repeated in these choices. The total number of edges removed is at most $\sum_{v \in V(G)}(\deg(v) - 2) = 2d - 2m$. In the resulting graph, every vertex has degree at most 2, and thus all connected components are either paths or cycles. Further, removing a single edge from a graph can only increase the number of connected components by at most 1. Since initially G has $|\operatorname{conn}(G)|$ connected components, the result on the number of connected components follows as well. □

Corollary 5.15. *Let G be a multigraph with $|V(G)| = m$ and $|E(G)| = d$ where every vertex has degree at least 2. Then,*

$$\tilde{f}_G(\Phi) \leq \left(\frac{N(1 + \|\Phi\|_{\ell^\infty})^2}{\|\Phi\|}\right)^{2d-2m+|\operatorname{conn}(G)|}\|\Phi\|^m.$$

Proof. Let H be the graph formed from G by deleting at most $2d - 2m$ edges to leave a disjoint union of paths and cycles. For any connected component C of H, by the above results we have

$$\tilde{f}_C(\Phi) \leq \left(\frac{N(1 + \|\Phi\|_{\ell^\infty})}{\|\Phi\|}\right)\|\Phi\|^{|V(C)|}.$$

Combining the previous results and using the multiplicativity of $\tilde{f}_G$ over disjoint unions of graphs, we have

$$\begin{aligned}
\tilde{f}_G(\Phi) &\leq (1+\|\Phi\|_{\ell^\infty})^{2d-2m} \tilde{f}_H(\Phi) \\
&\leq (1+\|\Phi\|_{\ell^\infty})^{2d-2m} \prod_{C \in \mathrm{conn}(H)} \tilde{f}_C(\Phi) \\
&\leq (1+\|\Phi\|_{\ell^\infty})^{2d-2m} \prod_{C \in \mathrm{conn}(H)} \left(\frac{N(1+\|\Phi\|_{\ell^\infty})}{\|\Phi\|} \right) \|\Phi\|^{|V(C)|} \\
&= (1+\|\Phi\|_{\ell^\infty})^{2d-2m} \left(\frac{N(1+\|\Phi\|_{\ell^\infty})}{\|\Phi\|} \right)^{|\mathrm{conn}(H)|} \|\Phi\|^m \\
&\leq (1+\|\Phi\|_{\ell^\infty})^{2d-2m} \left(\frac{N(1+\|\Phi\|_{\ell^\infty})}{\|\Phi\|} \right)^{2d-2m+|\mathrm{conn}(G)|} \|\Phi\|^m \\
&\leq \left(\frac{N(1+\|\Phi\|_{\ell^\infty})^2}{\|\Phi\|} \right)^{2d-2m+|\mathrm{conn}(G)|} \|\Phi\|^m,
\end{aligned}$$

as claimed. □

Specifically, the following is the way in which we will apply the above ideas. Recall that we have mentioned that our arguments will involve showing that some sums over general partitions are dominated by the contribution of the matchings. To implement this proof idea, it will be useful to *compare* the $\tilde{f}_{G(\rho)}$ for general partitions ρ to the same terms for matchings. The following precisely gives a way to do this.

Lemma 5.16. *Suppose that $\rho \preceq \pi$ is such that $\pi \in \mathrm{Match}(\mathcal{I})$. Then,*

$$\tilde{f}_{G(\rho)}(\Phi) \leq \left(\frac{N(1+\|\Phi\|_{\ell^\infty})^2}{\|\Phi\|} \right)^{2d-2m+|\mathrm{cyc}(G(\pi))|} \|\Phi\|^m$$

Proof. The result is immediate from the previous one once we note that, since $G(\rho)$ is formed by identifying certain vertices of $G(\pi)$, we have $|\mathrm{conn}(G(\rho))| \leq |\mathrm{conn}(G(\pi))| = |\mathrm{cyc}(G(\pi))|$. □

5.2 Computational Lower Bound

Let us recall the assumptions of the theorem: we have

$$\begin{aligned}
D = D(N) &\ll \frac{\lambda_1(\Phi)^3}{N^2}, \\
\|\Phi\|_{\ell^\infty} &\leq B,
\end{aligned}$$

where B is a constant independent of N. Recall from our preliminary calculations in Section 3.6 that we will in effect need to replace Φ by $\Phi' := \Phi + \mathrm{diag}(\Phi)$. However, as the following relations express, both the entrywise ℓ^∞ and operator norms are not changed much by this modification:

$$\begin{aligned}
\lambda_1(\Phi) \leq \lambda_1(\Phi') &\leq \lambda_1(\Phi) + B, \\
B \leq \|\Phi'\|_{\ell^\infty} &\leq 2B.
\end{aligned}$$

We take as a starting point our formula for the low-degree χ^2-divergence in a general inhomogeneous spiked Wigner matrix model from Lemma 3.28. Let us denote

$$u := x^1 \odot x^2,$$

where we recall that x^1 and x^2 have i.i.d. entries drawn from ν. Therefore, u has i.i.d. entries drawn from $\nu^{(2)}$, the law of xx' for $x, x' \sim \nu$. Then, the Lemma says that we have

$$\chi^2_{\leq D}(\mathbb{P}_N \mid \mathbb{Q}_N) = \sum_{d=1}^{D} \frac{1}{d!} \mathbb{E} R^d = \sum_{d=1}^{D} \frac{1}{d!} \left(\frac{\beta^2}{2N}\right)^d \mathbb{E}(u^\top \Phi' u)^d.$$

Expanding the remaining expectations, we have

$$\mathbb{E}(u^\top \Phi' u)^d = \sum_{i_1,\dots,i_d,j_1,\dots,j_d \in [N]} \left(\prod_{k=1}^{d} \Phi'_{i_k j_k}\right) \cdot \mathbb{E}\left[\prod_{k=1}^{d} u_{i_k} u_{j_k}\right]$$

Rephrasing, let us introduce the indexing set

$$\mathcal{I} := \{1, 1', 2, 2', \dots, d, d'\}.$$

Then, we may write

$$\mathbb{E}(u^\top \Phi' u)^d = \sum_{i : \mathcal{I} \to [N]} \left(\prod_{k=1}^{d} \Phi'_{i(k) i(k')}\right) \cdot \mathbb{E}\left[\prod_{k=1}^{d} u_{i(k)} u_{i(k')}\right]$$

and, since $\nu^{(2)}$ is a symmetric distribution, any term where any element of $[N]$ occurs an odd number of times in the image of i is zero and we may restrict

$$= \sum_{\substack{i : \mathcal{I} \to [N] \\ |i^{-1}(j)| \text{ is even for all } j \in [N]}} \left(\prod_{k=1}^{d} \Phi'_{i(k) i(k')}\right) \cdot \mathbb{E}\left[\prod_{k=1}^{d} u_{i(k)} u_{i(k')}\right]$$

Equivalently, we may associate to each i a partition ρ of $\mathcal{I}$ into even parts, where the elements of $\mathcal{I}$ that i maps to the same element of $[N]$ belong to the same block of ρ. In these terms, we can rewrite

$$= \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \sum_{i : \rho \hookrightarrow [N]} \left(\prod_{k=1}^{d} \Phi'_{i(k) i(k')}\right) \cdot \prod_{A \in \pi} \mathbb{E}\left[u_1^{|A|}\right]$$

where we have used that the entries of u are i.i.d. Recall that u_1 has the law of xx' for $x, x' \sim \nu$, and so by assumption u_1 is σ^2-subgaussian. In particular, by Proposition 3.15, there exists $\tau > 0$ such that its even moments are bounded by $\mathbb{E} u_1^{2k} \leq \tau^{2k-2}(2k-1)!!$. We may then expand and bound, noting that all terms appearing are non-negative,

$$\begin{aligned}
&= \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \sum_{i : \rho \hookrightarrow [N]} \left(\prod_{k=1}^{d} \Phi'_{i(k) i(k')}\right) \cdot \prod_{A \in \rho} \tau^{2|A|-2}(|A|-1)!! \\
&= \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \tau^{2d-2|\rho|} \prod_{A \in \rho} (|A|-1)!! \sum_{i : \rho \hookrightarrow [N]} \prod_{k=1}^{d} \Phi'_{i(k) i(k')} \\
&= \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \tau^{2d-2|\rho|} \cdot \prod_{A \in \rho} (|A|-1)!! \cdot f_\rho(\Phi') \qquad (8)
\end{aligned}$$

Now, note that the product of $(|A|-1)!!$ precisely counts the number of matchings that are refinements of ρ. Therefore, we may reorganize

$$= \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \sum_{\substack{\pi \in \mathrm{Match}(\mathcal{I}) \\ \rho \preceq \pi}} \tau^{2d-2|\rho|} f_\rho(\Phi')$$
$$= \sum_{\pi \in \mathrm{Match}(\mathcal{I})} \sum_{\substack{\rho \in \mathrm{EvenPart}(\mathcal{I}) \\ \rho \preceq \pi}} \tau^{2d-2|\rho|} f_\rho(\Phi')$$

Here, our idea will be to show that the dominant contribution in each inner sum comes from π itself. It turns out that we do not need to treat π specially to see this; this fact is already contained in the estimate of Corollary 5.15. We first rewrite

$$= \sum_{\pi \in \mathrm{Match}(\mathcal{I})} \sum_{m=1}^{d} \tau^{2d-2m} \sum_{\substack{\rho \in \mathrm{EvenPart}_m(\mathcal{I}) \\ \rho \preceq \pi}} f_\rho(\Phi')$$

Where we may now bound by first bounding all f by corresponding $\tilde{f}$ and then using Corollary 5.15, setting $K := \frac{N(1+\|\Phi'\|_{\ell^\infty})^2}{\|\Phi'\|}$ and $L := \tau^2(1+\|\Phi'\|_{\ell^\infty})^4$,

$$\leq \sum_{\pi \in \mathrm{Match}(\mathcal{I})} \sum_{m=1}^{d} \tau^{2d-2m} \sum_{\substack{\rho \in \mathrm{EvenPart}_m(\mathcal{I}) \\ \rho \preceq \pi}} \left(\frac{N(1+\|\Phi'\|_{\ell^\infty})^2}{\|\Phi'\|} \right)^{2d-2m+|\mathrm{cyc}(G(\pi))|} \|\Phi'\|^m$$
$$= \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|} \sum_{m=1}^{d} L^{d-m} \sum_{\substack{\rho \in \mathrm{EvenPart}_m(\mathcal{I}) \\ \rho \preceq \pi}} N^{2d-2m} \|\Phi'\|^{3m-2d}$$

and, using that the number of ρ with m parts that are coarsenings of π (which, being a matching, has d parts) is equal to $|\mathrm{Part}_m([d])| = S_2(d,m)$, a Stirling number of the second kind, we can further rewrite upon recognizing the Touchard polynomials (Definition 3.6)

$$= \left(\frac{LN^2}{\|\Phi'\|^2} \right)^d \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|} \sum_{m=1}^{d} S_2(d,m) \left(\frac{\|\Phi'\|^3}{LN^2} \right)^m$$
$$= \left(\frac{LN^2}{\|\Phi'\|^2} \right)^d T_d\left(\frac{\|\Phi'\|^3}{LN^2} \right) \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|}$$
$$\leq \left(\frac{LN^2}{\|\Phi'\|^2} \right)^d \left(\frac{\|\Phi'\|^3}{LN^2} \right)^d \exp\left(\frac{LN^2 d^2}{\|\Phi'\|^3} \right) \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|} \qquad \text{(Corollary 3.11)}$$
$$= \|\Phi'\|^d \exp\left(\frac{LN^2 d^2}{\|\Phi'\|^3} \right) \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|}$$
$$= (2\|\Phi'\|)^d \exp\left(\frac{LN^2 d^2}{\|\Phi'\|^3} \right) \frac{\Gamma(d+\frac{K}{2})}{\Gamma(\frac{K}{2})} \qquad \text{(Corollary 5.8)}$$

Finally, we return to the original sum we wanted to bound. We have

$$\begin{aligned}\chi^2_{\le D}(\mathbb{P}_N \mid \mathbb{Q}_N) &= \sum_{d=1}^{D} \frac{1}{d!}\left(\frac{\beta^2}{2N}\right)^d \mathbb{E}(u^\top \Phi' u)^d \\ &\le \sum_{d=1}^{D} \frac{\Gamma(d+\frac{K}{2})}{d!\cdot\Gamma(\frac{K}{2})}\left(\beta^2\cdot\frac{\|\Phi'\|}{N}\cdot\exp\left(\frac{LN^2D}{\|\Phi'\|^3}\right)\right)^d\end{aligned}$$

By assumption we have, as $N\to\infty$, that $L = O(1)$ and $\frac{N^2D}{\|\Phi'\|^3} \le \frac{N^2D}{\|\Phi\|^3} \to 0$, whereby for any $\eta > 0$, for sufficiently large N, we will have $\exp(\frac{LN^2D}{\|\Phi'\|^3}) \le 1+\eta$. Also, we have $\limsup_{N\to\infty}\beta^2\cdot\frac{\|\Phi\|'}{N} \le \limsup_{N\to\infty}\beta^2\cdot(\frac{\|\Phi\|}{N}+\frac{\|\Phi\|_{\ell^\infty}}{N}) \le 1-2\varepsilon$ for some $\varepsilon > 0$. Let us choose η such that $(1-2\varepsilon)(1+\eta)\le 1-\varepsilon$. For sufficiently large N, we may bound

$$\begin{aligned}&\le \sum_{d=1}^{D} \frac{\Gamma(d+\frac{K}{2})}{d!\cdot\Gamma(\frac{K}{2})}\left(\beta^2\cdot\frac{\|\Phi'\|}{N}\cdot(1+\eta)\right)^d \\ &\le \sum_{d=0}^{\infty} \frac{\Gamma(d+\frac{K}{2})}{d!\cdot\Gamma(\frac{K}{2})}\left(\beta^2\cdot\frac{\|\Phi'\|}{N}\cdot(1+\eta)\right)^d \\ &= \left(1-\beta^2\cdot\frac{\|\Phi'\|}{N}\cdot(1+\eta)\right)^{-K/2} && \text{(Proposition 3.21)} \\ &\le \exp\left(\frac{K}{2}\cdot\beta^2\cdot\frac{\|\Phi'\|}{N}\cdot(1+\eta)\cdot\frac{1+\varepsilon}{\varepsilon}\right) && \text{(Proposition 3.22)}\end{aligned}$$

Finally, expanding the definition of K again gives a last cancellation, leaving us with

$$\begin{aligned}&= \exp\left(\frac{\beta^2}{2}\cdot(1+\|\Phi'\|_{\ell^\infty})^2\cdot(1+\eta)\cdot\frac{1+\varepsilon}{\varepsilon}\right) \\ &= O(1),\end{aligned}$$

completing the proof, since all factors above are either constant parameters of the model, constants introduced earlier in the proof, or, for $\|\Phi'\|_{\ell^\infty} \le 2\|\Phi\|_{\ell^\infty}$, assumed to be $O(1)$ as $N\to\infty$.

5.2.1 Gaussian Intuition and Symmetry Assumption

Let us give some intuition about the part of the above manipulations where we insistently seek out the Touchard polynomials. We have in mind that the quadratic form $u^\top\Phi' u \approx u^\top\Phi u$ behaves as though u were a standard Gaussian vector. In this case, by rotational invariance, we could diagonalize Φ without loss of generality, and, keeping only the contribution of the largest eigenvalue $\lambda_1(\Phi)$, we would expect to have $\mathbb{E}(u^\top\Phi' u)^d \approx \lambda_1(\Phi)^d\mathbb{E}h^d$ for $h\sim\chi^2(1)$.

When we identify that Poisson moments (Touchard polynomials) appear in our calculation, we are aiming to implement this intuition by bounding them by the corresponding moments of the χ^2 distribution. Indeed, using that $\chi^2(d) = \Gamma(d/2, 2)$ for the family of Gamma distributions, we may embed the χ^2 distributions into a continuous family, and we have observed numerically that, for all $\lambda > 0$ and $d \ge 1$,

$$\mathop{\mathbb{E}}_{X\sim\mathrm{Pois}(\lambda)} X^d \le \mathop{\mathbb{E}}_{X\sim\Gamma(\frac{\lambda}{2},2)} X^d,$$

both sides of which may be evaluated as polynomials in λ for any given d. We have not been able to establish this intriguing empirical inequality, but fortunately Corollary 3.11 gives a sufficient approximation for our purposes.

This circle of ideas is also the reason that the symmetry assumption on the distribution ν is important to our argument. Without it, we would not be able to perform the crucial regrouping of a sum over $\rho \in \mathrm{EvenPart}(\mathcal{I})$ into one over $\pi \in \mathrm{Match}(\mathcal{I})$ with $\pi \succeq \rho$, which in turn is what eventually gives rise to the Stirling numbers and Touchard polynomials. While this approach seems exceedingly algebraically delicate, and we suspect that the symmetry condition could be removed, we have not yet been able to find an alternative approach to the one taken here that is sufficiently precise.

5.3 Statistical Lower Bound

We repeat the beginning of the previous proof, but make some changes starting with (8) above. Using the extra assumption on moments of ν, we find that this equation holds without the term $\tau^{2d-2|\rho|}$, so we merely have

$$\mathbb{E}(u^\top \Phi' u)^d \leq \sum_{\rho \in \mathrm{EvenPart}(\mathcal{I})} \prod_{A \in \rho} (|A|-1)!! \cdot f_\rho(\Phi')$$

Using as we did in the previous proof that the combinatorial factor here is precisely the number of $\pi \in \mathrm{Match}(\mathcal{I})$ such that $\rho \preceq \pi$, we may bound by the dramatically simpler

$$\begin{aligned} &= \sum_{\pi \in \mathrm{Match}(\mathcal{I})} \sum_{\rho \preceq \pi} f_\rho(\Phi') \\ &= \sum_{\pi \in \mathrm{Match}(\mathcal{I})} \tilde{f}_\pi(\Phi') && \text{(Proposition 5.5)} \end{aligned}$$

and setting $K := N\|\Phi'\|_{\ell^\infty}/\lambda_1(\Phi')$, we have

$$\begin{aligned} &\leq \lambda_1(\Phi')^d \sum_{\pi \in \mathrm{Match}(\mathcal{I})} K^{|\mathrm{cyc}(G(\pi))|}. && \text{(Proposition 5.12)} \\ &= (2\lambda_1(\Phi'))^d \frac{\Gamma(d+\frac{K}{2})}{\Gamma(\frac{K}{2})}, && \text{(Corollary 5.8)} \end{aligned}$$

reaching a simplified version of our final bound on this expectation from the previous proof.

Remark. *The first bound above would hold with equality if the vector u consisted of i.i.d. entries distributed as $\mathcal{N}(0,1)$, as may be verified by applying Wick's formula. This observation is the motivation for our approach to these calculations and to the form of our assumption on the moments of ν for this proof. Note, however, that $u = x^1 \odot x^2$ for two draws of x^i from the distribution of our prior $\mathcal{P}_N$, so this intuition is* not *saying that the above calculation behaves as it would for $\mathcal{P}_N = \mathcal{N}(0, I_N)$ (to which case this proof does not apply). Rather, the idea is that a situation like $\mathcal{P}_N = \mathrm{Unif}(\{\pm 1\}^N)$, in which case the law of u is again $\mathrm{Unif}(\{\pm 1\}^N)$, should resemble the case where $u \sim \mathcal{N}(0, I_N)$.*

Repeating the same manipulations from the previous proof, we find, letting $\varepsilon > 0$ be such that

$\limsup_{N\to\infty} \beta^2 \cdot \frac{\lambda_1(\Phi')}{N} \le 1-\varepsilon$,

$$\begin{aligned}\chi^2(\mathbb{P}_N \mid \mathbb{Q}_N) &\le \left(1-\beta^2\cdot\frac{\lambda_1(\Phi')}{N}\right)^{-K/2}\\ &\le \exp\left(\frac{K}{2}\cdot\beta^2\cdot\frac{\lambda_1(\Phi')}{N}\cdot\frac{1+\varepsilon}{\varepsilon}\right)\\ &= \exp\left(\frac{\|\Phi'\|_{\ell^\infty}}{2}\cdot\beta^2\cdot\frac{1+\varepsilon}{\varepsilon}\right)\\ &= O(1),\end{aligned}$$

completing the proof.

5.4 Growth of Low-Degree χ^2-Divergence: Proof of Theorem 1.13

We follow the same general approach as in the proof of Theorem 1.12 in Section 4.4. As we argued there, it suffices to show that

$$\mathbb{E}\exp^{\le D}(R^{(0)}) = \omega(1),$$

for

$$R^{(0)} = \frac{\beta^2}{2N}u^\top \Phi u,$$

where $u = x^1 \odot x^2$ for $x^1, x^2 \sim \mathcal{P}_N$ two independent draws from the signal distribution.

The main difference is that, in Theorem 1.12, we could use the block structure of Φ to group this sum into a finite number of sums to each of which the central limit theorem would apply. Here, we cannot do that, but we take a similar approach with a generalized central limit theorem. Let $\mu_1 \ge \cdots \ge \mu_N$ be the ordered eigenvalues of $\Phi^{(N)}$, and $v_1, \dots, v_N$ be the associated unit eigenvectors. We may then rewrite

$$R^{(0)} = \frac{\beta^2}{2N}\sum_{i=1}^N \mu_i \langle v_i, u\rangle^2$$

We now claim that we have the convergence in distribution

$$(\langle v_1, u\rangle, \dots, \langle v_K, u\rangle) \xrightarrow{\text{(d)}} \mathcal{N}(0, I_K),$$

for any K fixed as $N\to\infty$. Recall that we assume either that $\Phi^{(N)} \succeq 0$, in which case we will use this with $K=1$, or that $\Phi^{(N)}$ has bounded rank, in which case we will take K to be that bound. By Theorem 3.7, it suffices to show that, for any $c \in \mathbb{R}^K$ with $\|c\| = 1$, we have

$$\left\langle \sum_{i=1}^K c_i v_i, u\right\rangle \xrightarrow{\text{(d)}} \mathcal{N}(0,1). \tag{9}$$

Recall that we have assumed that

$$\lim_{N\to\infty} \max_{i=1}^N \|v_i\|_\infty = 0.$$

So, we also have

$$\lim_{N\to\infty} \left\|\sum_{i=1}^K c_i v_i\right\|_\infty = 0.$$

Thus, since the entries of u are i.i.d., bounded, and have mean zero and variance 1, the convergence in (9) follows by the Lindeberg central limit theorem. Let us write $(g_1, \dots, g_K) \sim \mathcal{N}(0, I_K)$ for this distributional limit.

We also have in general that

$$\mathbb{E}\exp^{\leq D}(R^{(0)}) = \sum_{d=0}^{D} \frac{1}{d!}\mathbb{E}R^{(0)^d}$$

and every term here is non-negative. Further, since $D = D(N) \to \infty$ by assumption, for any given $d_0 \geq 1$ we will eventually have $D \geq d_0$ for sufficiently large N, and so we can bound

$$\geq \frac{1}{d_0!}\mathbb{E}R^{(0)^{d_0}}.$$

We now consider our two possible assumptions on $\Phi^{(N)}$ separately. If $\Phi^{(N)} \succeq 0$, then $R^{(0)} \geq 0$ almost surely, and may bound by just the term of the top eigenvalue μ_1,

$$R^{(0)} \geq \frac{\beta^2}{2N}\mu_1\langle v_1, u\rangle^2 \xrightarrow{\text{(d)}} \frac{\beta^2\widehat{\mu}_1}{2}g_1^2.$$

As in the proof of Theorem 1.12, a standard approximation argument implies that this convergence also holds for moments, and thus

$$\begin{aligned}
\mathbb{E}\exp^{\leq D}(R^{(0)}) &\geq \frac{1}{d_0!}\mathbb{E}R^{(0)^{d_0}} \\
&\geq \frac{1}{d_0!}\mathbb{E}\left(\frac{\beta^2}{2N}\mu_1\langle v_1, u\rangle^2\right)^{d_0} \\
&\to \frac{1}{d_0!}\mathbb{E}\left(\frac{\beta^2\widehat{\mu}_1}{2}g_1^2\right)^{d_0} \\
&= \frac{(2d_0-1)!!}{(2d_0)!!}(\beta^2\widehat{\mu}_1)^{d_0}.
\end{aligned}$$

Stirling-type approximations imply that the first term is polynomial in d, while by assumption $\beta^2\widehat{\mu}_1 > 1$, so taking $d_0 \to \infty$ gives a diverging lower bound.

Now, suppose we instead have the second possible condition on $\Phi^{(N)}$, that $\operatorname{rank}(\Phi^{(N)}) \leq K$ and $\lambda_i(\Phi^{(N)})/N \to \widehat{\mu}_i$ for each $i = 1, \dots, K$. Then, we have that $R^{(0)}$ itself converges in distribution:

$$R^{(0)} = \frac{\beta^2}{2N}\sum_{i=1}^{K}\mu_i\langle v_i, u\rangle^2 \xrightarrow{\text{(d)}} \frac{\beta^2}{2}\sum_{i=1}^{K}\widehat{\mu}_i g_i^2.$$

By the same argument as above, we have

$$\begin{aligned}
\liminf_{N\to\infty}\exp^{\leq D}(R^{(0)}) &\geq \frac{1}{d_0!}\mathbb{E}\left(\frac{\beta^2}{2}\sum_{i=1}^{K}\widehat{\mu}_i g_i^2\right)^{d_0} \\
&= \frac{1}{d_0!}\mathbb{E}\left(\frac{\beta^2}{2}g^\top D g\right)^{d_0}
\end{aligned}$$

for $D = \operatorname{diag}(\widehat{\mu}_1, \dots, \widehat{\mu}_K)$. Note that, for any $\ell \geq 1$, $\operatorname{Tr}(D^\ell) = \lim_{N\to\infty}\operatorname{Tr}(\Phi^{(N)}/N)^k \geq 0$ since all entries of $\Phi^{(N)}$ are non-negative. Thus, we may apply Proposition 3.8, which gives

$$\geq \frac{1}{d_0!}\left(\frac{\beta^2}{2}\right)^{d_0} 2^{d_0-1}(d_0-1)!\operatorname{Tr}(D^{d_0})$$

and, provided we take d_0 even,

$$\geq \frac{1}{2d_0}(\beta^2\widehat{\mu}_1)^{d_0},$$

and the proof concludes after taking $d_0 \to \infty$ as in the first case.

5.5 Numerical Experiment

While our results suggest that the computational threshold for general variance profiles of noise is given by a straightforward extension of the formula for the threshold for block-structured profiles, they do not say anything about whether the spectral algorithm studied by [PKK24, MKK24, BCSvH24] for block-structured profiles should still be effective for general profiles. Still, the formal similarity between the thresholds does suggest the intriguing possibility that the spectral algorithm using $\mathcal{H}(Y)$ might remain optimal for general variance profiles. Here we give some numerical evidence that, at least, that spectral algorithm is superior to two other natural choices. The theoretical discussion in this section is all at a heuristic level and is meant merely to justify the setup of our experiment and perhaps to suggest possible explanations of the results.

5.5.1 Alternative Algorithms and Predicted BBP Thresholds

The two alternative algorithms we consider (also discussed by [GKKZ25, MKK24]) are as follows. The first is just to consider the top eigenpair of $\widehat{Y} = Y/\sqrt{n}$ itself. We call this the *direct* spectral algorithm, and denote its BBP threshold (i.e., the value of β at which the transition from Theorems 1.3 and 1.8 occurs) by β_*^{Direct}. The second is to consider the different and more directly intuitive modification of Y,

$$\mathcal{G}(Y) := \frac{1}{\sqrt{N}}\Delta^{\odot -1/2} \odot Y,$$

which divides each entry of Y by the standard deviation of the noise applied to that entry. We call this the *whitening* spectral algorithm, because it makes the "noise term" of $\mathcal{G}(Y)$ i.i.d. $\mathcal{N}(0,1)$, in exchange for modifying the "signal term" in a possibly complicated way. We call the BBP threshold of this algorithm β_*^{White}. Finally, we call the *linearized AMP* or *LinAMP* spectral algorithm the one using $\mathcal{H}(Y)$, and call its BBP threshold β_*^{LinAMP}.

Let us assume for the sake of simplicity that the signal x in our model has law $x \sim \text{Unif}(\{\pm 1\}^n)$, i.e., we take the entrywise distribution $\nu = \text{Unif}(\{\pm 1\})$ in the notation of the main results. We now describe the values we expect each of these three thresholds to take. To avoid needing to carefully prescribe what sequences of variance profiles "converge" in a suitable sense, we keep the discussion entirely heuristic and make non-asymptotic predictions about these thresholds.

First, from our results in this paper, we are led to hope for the LinAMP spectral algorithm that

$$\beta_*^{\text{LinAMP}} \approx \sqrt{\frac{N}{\lambda_1(\Phi)}} = \sqrt{\frac{N}{\lambda_1(\Delta^{\odot -1})}}. \tag{10}$$

Second, for the whitening spectral algorithm, note that

$$\mathcal{G}(Y) = \frac{1}{\sqrt{N}}\beta\Delta^{\odot -1/2} \odot xx^\top + W^{(0)} = \frac{1}{\sqrt{N}}\beta D_x\Delta^{\odot -1/2}D_x + W^{(0)}$$

where the entries of $W^{(0)}$ on and above the diagonal are i.i.d. with distribution $\mathcal{N}(0,1)$. This is a spiked matrix model with homogeneous noise but where the signal part $\frac{1}{\sqrt{N}}\beta D_x\Delta^{\odot -1/2}D_x$

does not necessarily have low rank. Conveniently, by our assumption on x we have that D_x is an orthogonal matrix, so the eigenvalues of this signal part up to rescaling are those of $\Delta^{\odot -1/2}$. Provided this matrix is approximately low rank, per the analysis of, for example, [CDMF09], we expect the threshold

$$\beta_*^{\text{White}} \approx \frac{N}{\lambda_1(\Delta^{\odot -1/2})}. \tag{11}$$

Note that [GKKZ25] show that, if we take the above as definitions of β_*^{White} and β_*^{LinAMP}, then for any variance profile Δ we have $\beta_*^{\text{White}} \geq \beta_*^{\text{LinAMP}}$, with equality if and only if Δ is constant.

Finally, we propose that a reasonable proxy for the BBP threshold of the original inhomogeneous model $\widehat{Y}$ is to apply the theory for spiked matrix models with orthogonally invariant noise, in particular the results of [BGN11]. (The more involved analysis of [BM21] also applies to this question, but we take this more heuristic approach to lighten the numerical computations required.) While our matrix W is not orthogonally invariant, if for instance we instead had x drawn uniformly from the sphere of radius $\sqrt{N}$, then we could consider $Q\widehat{Y}Q^\top$ instead of $\widehat{Y}$ for Q a Haar-distributed orthogonal matrix and use that $QWQ^\top$ is then orthogonally invariant. In any case, writing $\widehat{W} := \widehat{W}/\sqrt{N}$, the results of [BGN11] describe the BBP threshold for such $\widehat{Y}$ in terms of the limiting empirical eigenvalue distribution of $\widehat{W}$ over a suitably converging sequence. Showing that certain sequences of matrices with variance profiles have these distributions converging is itself non-trivial, so let us just suppose we believe this distribution to be some compactly supported γ. If b is the right edge of the support of γ, then [BGN11] yields the prediction

$$\beta_*^{\text{Direct}} \approx \frac{1}{\lim_{z \to b^+} G(z)}, \tag{12}$$

where $G(z)$ is the *Cauchy transform* of γ, given by

$$G(z) = \mathop{\mathbb{E}}_{x \sim \gamma}\left[\frac{1}{z-x}\right].$$

For our purposes here, given a specific variance profile matrix Δ, we would like to compute these three thresholds to compare with numerical results. While for the LinAMP and whitening algorithms (10) and (11) respectively give simple such formulas, it is unclear how to compute with (12), since we do not have access to the measure γ or the number b. Given a concrete Δ and W sampled from Δ, we will simply approximate

$$b \approx \lambda_1(\widehat{W}).$$

Then, following the heuristic derivation discussed, for instance, on the first pages of [AEK19], we use that $G(z)$ may be estimated by first solving the recursion

$$-\frac{1}{m_i} = z + \frac{1}{N}\sum_{j=1}^N \Delta_{ij} m_j = z + \frac{1}{N}(\Delta m)_i$$

for a vector $m = (m_1, \dots, m_N) \in \mathbb{R}^N$, and then taking the approximation

$$G(z) \approx -\frac{1}{N}\sum_{i=1}^N m_i.$$

See the citation above for a justification of this procedure involving approximating the resolvent matrix of $\widehat{W}$. We then estimate β_*^{Direct} by performing this procedure (concretely, we obtain m by fixed point iterations of the above equations) with z equal to our estimate of b, $z := \lambda_1(\widehat{W})$.

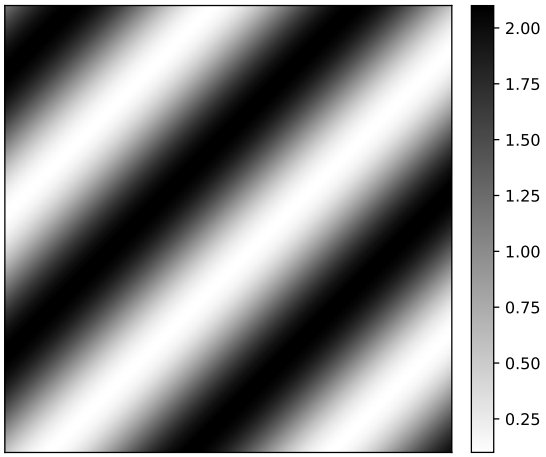

Figure 2: An illustration of the variance profile matrix Δ, plotted as a heatmap of an $N \times N$ matrix with $N = 12\,000$, defined by (13) with the choice of the function f given in (14). For such large N, this heatmap is visually indistinguishable from a plot of the continuous function f itself on the domain $[0,1]^2$.

5.5.2 Empirical Results for a Continuous Variance Profile

Now let us propose a specific non-block-structured variance profile and describe the above predicted thresholds for the three spectral algorithms. In general, a widely studied type of variance profile in random matrix theory (though, to the best of our knowledge, not yet in the context of spiked matrix models) is a profile of the form

$$\Delta_{ij} := f\left(\frac{i}{N}, \frac{j}{N}\right) \tag{13}$$

for a continuous symmetric function $f : [0,1]^2 \to \mathbb{R}_{\geq 0}$ (see the discussion in Section 1.5). If f is also bounded away from zero, then such a variance profile will satisfy the condition (4) discussed earlier, and our Theorem 1.11 will apply (provided that Assumption A holds, which also may be shown to hold for this family of choices). Such variance profiles are also the context in which the description of γ mentioned above has been shown to hold; they are convenient for the above description since $\frac{1}{N}\sum_{j=1}^{N} \Delta_{ij} m_j$ then becomes a Riemann sum if $m_j = m(j/N)$ is associated to another function.

We will consider such a variance profile for the function

$$f(x,y) = \frac{11}{10} + \sin(10(x+y)). \tag{14}$$

Note that we indeed have the uniform bounds $1/10 \leq f(x,y) \leq 21/10$. We illustrate the variance profile associated to such an f in Figure 2. Performing the above calculations for this variance profile evaluated with $N = 12\,000$ gives

$$\begin{aligned} \beta_*^{\text{LinAMP}} &\approx 0.653, \\ \beta_*^{\text{White}} &\approx 0.759, \\ \beta_*^{\text{Direct}} &\approx 1.158. \end{aligned}$$

In Figure 3, we show eigenvalue distributions of the matrices constructed by each algorithm for a sequence of values

$$\beta_1 < \beta_*^{\mathrm{LinAMP}} < \beta_2 < \beta_*^{\mathrm{White}} < \beta_3 < \beta_*^{\mathrm{Direct}} < \beta_4.$$

Specifically, we take

$$\begin{aligned}\beta_1 &= 0.65,\\ \beta_2 &= 0.75,\\ \beta_3 &= 0.85,\\ \beta_4 &= 1.25.\end{aligned}$$

We observe that, as expected, for $\beta = \beta_1$ none of the matrices has an outlier eigenvalue, for $\beta = \beta_2$ only the LinAMP algorithm's matrix has an outlier eigenvalue, for $\beta = \beta_3$ only the LinAMP and whitening algorithm's matrices have outlier eigenvalues, and for $\beta = \beta_4$ all three matrices have outlier eigenvalues.

Remark. *Note that, as is visible in the figure, the bulk distribution of eigenvalues for the direct and whitening algorithms' matrices $\widehat{Y}$ and $\mathcal{G}(Y)$ does not depend on β, while the bulk distribution for the LinAMP algorithm's matrix $\mathcal{H}(Y)$ does. This is just because the construction of $\mathcal{H}(Y)$ presumes access to β and includes β in its very definition; see* (1).

We emphasize the finding that, at least on this particular variance profile, the LinAMP algorithm is indeed superior to the other two, and its performance is consistent with the predicted threshold $\beta_*^{\mathrm{LinAMP}}$. We leave it as an intriguing open question to determine whether the actual BBP threshold for $\mathcal{H}(Y)$ coincides with our β_* parameter more generally, to understand whether a spectral algorithm using $\mathcal{H}(Y)$ is indeed optimal for many more variance profiles (and if so, under what conditions on variance profiles) or whether this matrix construction needs to be modified further in the non-block-structured case.

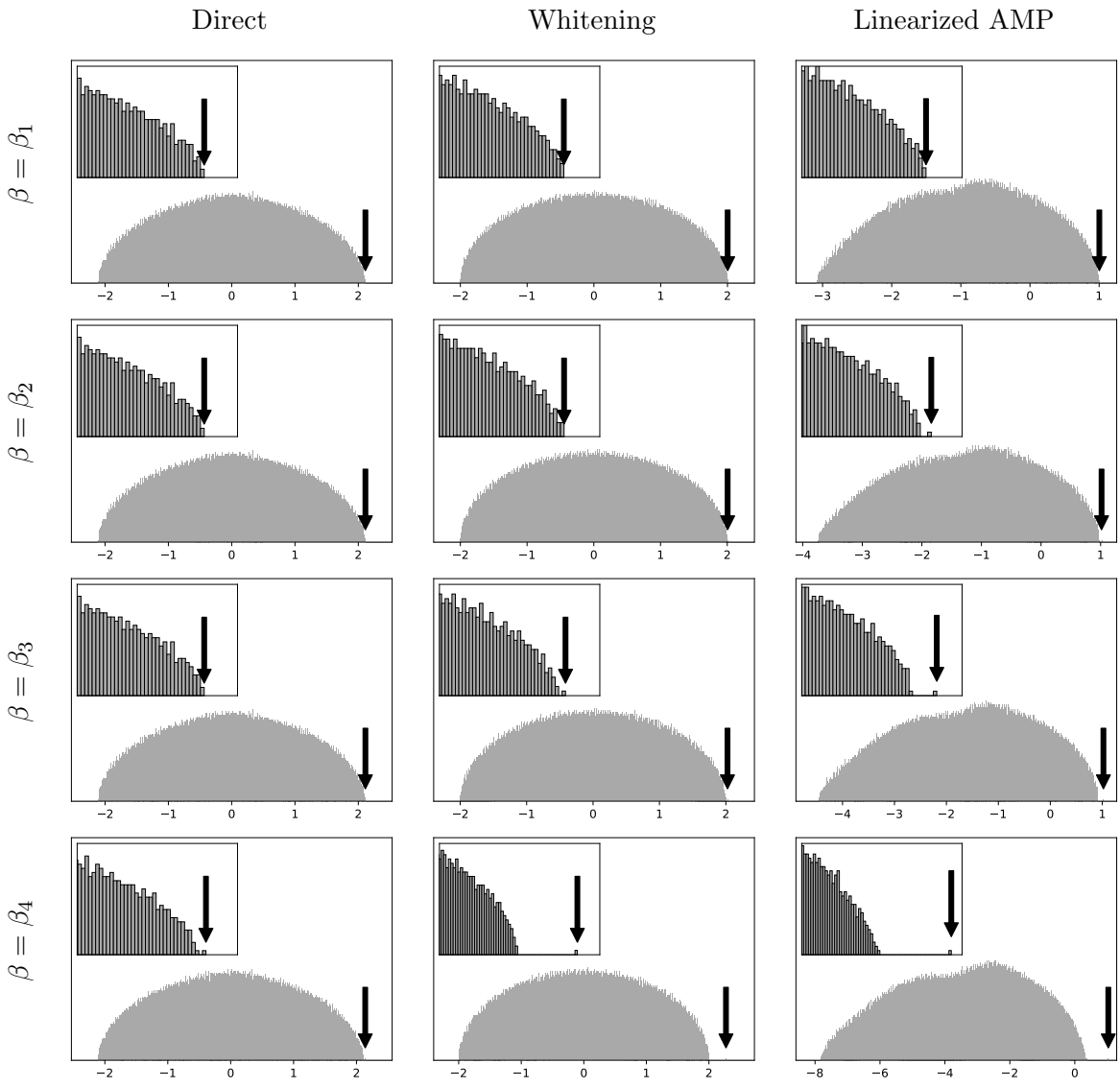

Figure 3: We compare the direct, whitening, and linearized AMP (LinAMP) spectral algorithms discussed in Section 5.5 for several signal strengths $\beta_1 < \beta_2 < \beta_3 < \beta_4$ and $N = 12\,000$, plotting the eigenvalue distribution of the matrix each constructs with the maximum eigenvalue indicated with an arrow. As predicted, the LinAMP algorithm requires the smallest threshold signal strength to observe an outlier eigenvalue, the whitening algorithm the next-smallest strength, and the direct algorithm the largest threshold strength.

References

[ABBS14] Emmanuel Abbe, Afonso S Bandeira, Annina Bracher, and Amit Singer. Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering*, 1(1):10–22, 2014.

[ACCM21a] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The multi-species mean-field spin-glass on the Nishimori line. *Journal of Statistical Physics*, 182(1):2, 2021.

[ACCM21b] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The solution of the deep Boltzmann machine on the Nishimori line. *Communications in Mathematical Physics*, 387(2):1191–1214, 2021.

[ACCM22] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. A statistical physics approach to a multi-channel Wigner spiked model. *Europhysics Letters*, 136(4):48001, 2022.

[ACD+21] Benson Au, Guillaume Cébron, Antoine Dahlqvist, Franck Gabriel, and Camille Male. Freeness over the diagonal for large random matrices. *The Annals of Probability*, 49(1):157–179, 2021.

[AEK17] Oskari H Ajanki, László Erdős, and Torben Krüger. Universality for general Wigner-type matrices. *Probability Theory and Related Fields*, 169(3):667–727, 2017.

[AEK19] Oskari Ajanki, László Erdős, and Torben Krüger. *Quadratic vector equations on complex upper half-plane*, volume 261. American Mathematical Society, 2019.

[Ahl22] Thomas D Ahle. Sharp and simple bounds for the raw moments of the binomial and Poisson distributions. *Statistics & Probability Letters*, 182:109306, 2022.

[AMP16] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: norm bounds and applications. *arXiv preprint arXiv:1604.03423*, 2016.

[BAH+22] Afonso S Bandeira, Ahmed El Alaoui, Samuel B Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *arXiv preprint arXiv:2205.09727*, 2022.

[BBK+21] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Christopher Moore, and Alex Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory*, pages 410–473. PMLR, 2021.

[BBvH21] Afonso S Bandeira, March T Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *arXiv preprint arXiv:2108.06312*, 2021.

[BCMS23] Jean Barbier, Francesco Camilli, Marco Mondelli, and Manuel Sáenz. Fundamental limits in structured principal component analysis and how to reach them. *Proceedings of the National Academy of Sciences*, 120(30), 2023.

[BCSvH24] Afonso S Bandeira, Giorgio Cipolloni, Dominik Schröder, and Ramon van Handel. Matrix concentration inequalities and free probability II. Two-sided bounds and applications. *arXiv preprint arXiv:2406.11453*, 2024.

[BCXM25] Jean Barbier, Francesco Camilli, Yizhou Xu, and Marco Mondelli. Information limits and thouless-anderson-palmer equations for spiked matrix models with structured noise. *Physical Review Research*, 7(1):013081, 2025.

[BDM+16] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *arXiv preprint arXiv:1606.04142*, 2016.

[BDM+18] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method. *arXiv preprint arXiv:1812.02537*, 2018.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[BHK+19] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[BHMS22] Jean Barbier, TianQi Hou, Marco Mondelli, and Manuel Sáenz. The price of ignorance: how much does it cost to forget noise structure in low-rank matrix estimation? *Advances in Neural Information Processing Systems*, 35:36733–36747, 2022.

[BM21] Jérémie Bigot and Camille Male. Freeness over the diagonal and outliers detection in deformed random matrices with a variance profile. *Information and Inference: A Journal of the IMA*, 10(3):863–919, 2021.

[BMV+18] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Transactions on Information Theory*, 64(7):4872–4894, 2018.

[BR20] Jean Barbier and Galen Reeves. Information-theoretic limits of a multiview low-rank symmetric spiked matrix model. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2771–2776. IEEE, 2020.

[BR22] Joshua K Behne and Galen Reeves. Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *International Conference on Artificial Intelligence and Statistics*, pages 8650–8672. PMLR, 2022.

[BS10] Zhidong Bai and Jack William Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, 2010.

[BU06] Yong Bao and Aman Ullah. Expectation of quadratic forms in normal and nonnormal variables with econometric applications. *Unpublished manuscript, University of California, Riverside*, 2006.

[BVH16] Afonso S Bandeira and Ramon Van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. 2016.

[Cap17] Mireille Capitaine. *Deformed ensembles, polynomials in random matrices and free probability theory*. PhD thesis, Université Paul Sabatier-Toulouse 3, 2017.

[CDM16] Mireille Capitaine and Catherine Donati-Martin. Spectrum of deformed random matrices and free probability. *arXiv preprint arXiv:1607.05560*, 2016.

[CDMF09] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1):1–47, 2009.

[CDMFF11] Mireille Capitaine, Catherine Donati-Martin, Delphine Féral, and Maxime Février. Free convolution with a semicircular distribution and eigenvalues of spiked deformations of wigner matrices. *Electron. J. Probab*, 16(64):1750–1792, 2011.

[COGHK+22] Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.

[DAM15] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the two-groups stochastic block model. *arXiv preprint arXiv:1507.08685*, 2015.

[DDL23] Jian Ding, Hang Du, and Zhangsong Li. Low-degree hardness of detection for correlated erd\h {o} sr\'enyi graphs. *arXiv preprint arXiv:2311.15931*, 2023.

[DMW23] Abhishek Dhawan, Cheng Mao, and Alexander S Wein. Detection of dense subhypergraphs by low-degree polynomials. *arXiv preprint arXiv:2304.08135*, 2023.

[EAKJ20] Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. Fundamental limits of detection in the spiked Wigner model. *Annals of Statistics*, 48(2):863–885, 2020.

[FK81] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[FP07] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.

[GKKZ25] Alice Guionnet, Justin Ko, Florent Krzakala, and Lenka Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *Information and Inference: A Journal of the IMA*, 14(2):iaaf010, 2025.

[HK22] Jun-Ting Hsieh and Pravesh K Kothari. Algorithmic thresholds for refuting random polynomial systems. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1154–1203. SIAM, 2022.

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *58th Annual Symposium on Foundations of Computer Science (FOCS 2017)*, pages 720–731, 2017.

[HKPX23] Jun-Ting Hsieh, Pravesh K Kothari, Aaron Potechin, and Jeff Xu. Ellipsoid fitting up to a constant. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2023.

[Hop18] Samuel B Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

[HS17] Samuel B Hopkins and David Steurer. Efficient Bayesian estimation from few samples: community detection and related problems. In *58th Annual Symposium on Foundations of Computer Science (FOCS 2017)*, pages 379–390. IEEE, 2017.

[Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, pages 295–327, 2001.

[JP18] Iain M Johnstone and Debashis Paul. Pca in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292, 2018.

[Kal97] Olav Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.

[KMW24] Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein. Tensor cumulants for statistical inference on invariant distributions. *arXiv preprint arXiv:2404.18735*, 2024.

[Kun24] Dmitriy Kunisky. Low coordinate degree algorithms II: Categorical signals and generalized stochastic block models. *arXiv preprint arXiv:2412.21155*, 2024.

[Kun25] Dmitriy Kunisky. Low coordinate degree algorithms I: Universality of computational thresholds for hypothesis testing. *The Annals of Statistics*, 53(2):774–801, 2025.

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

[KXZ16] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016.

[LKZ15a] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton 2015)*, pages 680–687. IEEE, 2015.

[LKZ15b] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse PCA. In *IEEE International Symposium on Information Theory (ISIT 2015)*, pages 1635–1639. IEEE, 2015.

[LKZ17] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[LL24] Sunmin Lee and Ji Oon Lee. Phase transition for the generalized two-community stochastic block model. *Journal of Applied Probability*, 61(2):385–400, 2024.

[LM19] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3):859–929, 2019.

[LvHY18] Rafał Latała, Ramon van Handel, and Pierre Youssef. The dimension-free structure of nonhomogeneous random matrices. *Inventiones mathematicae*, 214:1031–1080, 2018.

[MKK24] Pierre Mergny, Justin Ko, and Florent Krzakala. Spectral phase transition and optimal PCA in block-structured spiked models. *arXiv preprint arXiv:2403.03695*, 2024.

[MS12] James A Mingo and Roland Speicher. Sharp bounds for sums associated to graphs of matrices. *Journal of Functional Analysis*, 262(5):2272–2288, 2012.

[MW23] Ankur Moitra and Alexander S Wein. Precise error rates for computationally efficient testing. *arXiv preprint arXiv:2311.00289*, 2023.

[PKK24] Aleksandr Pak, Justin Ko, and Florent Krzakala. Optimal algorithms for the inhomogeneous spiked Wigner model. *Advances in Neural Information Processing Systems*, 36, 2024.

[PVTW22] Aaron Potechin, Prayaag Venkat, Paxton Turner, and Alexander S Wein. Near-optimal fitting of ellipsoids to random points. *arXiv preprint arXiv:2208.09493*, 2022.

[PWBM18] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA I: Spiked random matrix models. *Annals of Statistics*, 46(5):2416–2451, 2018.

[RH17] Phillippe Rigollet and Jan-Christian Hütter. Lecture notes on high dimensional statistics, 2017.

[Shl96] Dimitri Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *IMRN: International Mathematics Research Notices*, 1996(20), 1996.

[SLKZ15] Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Spectral detection in the censored block model. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1184–1188. IEEE, 2015.

[TBKK23] Ido Tishby, Ofer Biham, Eytan Katzav, and Reimer Kühn. Distribution of the number of cycles in directed and undirected random regular graphs of degree 2. *Physical Review E*, 107(2):024308, 2023.

[vH17] Ramon van Handel. Structured random matrices. *Convexity and concentration*, pages 107–156, 2017.

[Wei25] Alexander S Wein. Computational complexity of statistics: New insights from low-degree polynomials. *arXiv preprint arXiv:2506.10748*, 2025.

[Zhu20] Yizhe Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, 56(1):251–279, 2020.

[ZM24] Yihan Zhang and Marco Mondelli. Matrix denoising with doubly heteroscedastic noise: Fundamental limits and optimal spectral methods. *Advances in Neural Information Processing Systems*, 37:93060–93117, 2024.