On the Optimality and Sub-optimality of PCA in Spiked Random Matrix Models: supplementary proofs

Amelia Perry[*†1], Alexander S. Wein[*‡1], Afonso S. Bandeira[§1,3], and Ankur Moitra[¶1,2]

[1]Department of Mathematics, Massachusetts Institute of Technology
[2]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology
[3]Department of Mathematics and Center for Data Science, Courant Institute of Mathematical Sciences, New York University

April 24, 2017

A Behavior near criticality

In Gaussian Wigner settings where we have established contiguity for all $\lambda < 1$, it is natural to ask whether the spiked and unspiked models remain contiguous for a sequence $\lambda = 1 + \delta_n$ with $\delta_n \to 0$ (here δ_n may be positive or negative). However, this is never the case; it is possible to consistently distinguish the models in this critical case. By adding additional GOE noise, we can reduce to the case $\lambda = 1 - \varepsilon$ for arbitrary fixed $\varepsilon > 0$ (perhaps taking a tail of the sequence). It is known [Johnstone and Onatski, 2015] that (regardless of the spike prior) the hypothesis testing error (sum of type I and type II errors) in this case tends to 0 as $\varepsilon \to 0$; thus the minimum hypothesis testing error in the original problem cannot be bounded away from zero.

A similar result for the positively-spiked ($\beta > 0$) Wishart model follows from Onatski et al. [2013]: if γ is fixed and $\beta = \sqrt{\gamma} + \delta_n$ with $\delta_n \to 0$ then it is possible to consistently distinguish the spiked and unspiked models. (We expect the analogous result to hold for $\beta < 0$ but to the best our knowledge this has not been proven.)

B Bounds on hypothesis testing

For both the Gaussian Wigner and Wishart models, for the spherical prior (or equivalently, limited to spectral-based tests) the optimal tradeoff curve (power envelope) between type I and type II error is known exactly in the $n \to \infty$ limit [Onatski et al., 2013, Johnstone and Onatski, 2015]. For other priors, one can apply the optimal spectral-based test from above to obtain an upper bound; however, better tests (which do not depend only on the spectrum) may be possible.

---

[*]The first two authors contributed equally.

[†]Email: ameliaperry@mit.edu. This work is supported in part by NSF CAREER Award CCF-1453261 and a grant from the MIT NEC Corporation.

[‡]Email: awein@mit.edu. This research was conducted with Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.

[§]Email: bandeira@cims.nyu.edu. A.S.B. was supported by NSF Grant DMS-1317308. Part of this work was done while A.S.B. was with the Department of Mathematics at the Massachusetts Institute of Technology.

[¶]Email: moitra@mit.edu. This work is supported in part by NSF CAREER Award CCF-1453261, a grant from the MIT NEC Corporation and a Google Faculty Research Award.

In many cases we can use Proposition 2.5 to obtain lower bounds (which do not match the upper bound above). First note that Proposition 2.5 is still valid (in the $n \to \infty$ limit) in cases when we have used the *conditional* second moment. (This is because if $\tilde{Q}_n$ is obtained from Q_n by conditioning on a $(1 - o(1))$-probability event, asymptotic hypothesis testing bounds for $\tilde{Q}_n$ against P_n imply the same bounds for Q_n against P_n.)

For the Gaussian Wigner model, Theorem 3.10 (subgaussian method for i.i.d. priors) and Theorem D.2 (conditioning method) both give the limit value (as $n \to \infty$) of the (conditional) second moment, and in fact the value is $(1 - \lambda^2)^{-1/2}$ in both of these cases. Therefore, any time we have used one of those two methods, we obtain asymptotic hypothesis testing bounds from Proposition 2.5. This applies to, for instance, the i.i.d. Gaussian, Rademacher, and sparse Rademacher priors. The same bounds also hold for the spherical prior (although the exact asymptotic power envelope is known in this case) because the comparison method of Proposition 3.13 preserves the value of the second moment.

For the Wishart model, suppose we have a prior for which we know the *Wigner* second moment has limit value $(1 - \lambda^2)^{-1/2}$ (as above). Furthermore, suppose we have a Wishart lower bound for this prior via Theorem 5.7, using the Wigner second moment to control the small deviations (i.e. condition (i) of Theorem 5.7 holds). From the proof of Theorem 5.7, the limit value of the Wishart (conditional) second moment is determined by the small deviations of Section 5.5.2; the remaining large deviations contribute $o(1)$. We see from Section 5.5.2 that the asymptotic value of the small deviations is bounded by the value of the Wigner second moment with $\lambda^2 = -\log(1 - \varepsilon^2\beta^2)/\gamma\varepsilon^2 \to \beta^2/\gamma$ as $\varepsilon \to 0$. Therefore the limsup of the Wishart (conditional) second moment is at most $(1 - \beta^2/\gamma)^{-1/2}$, which yields hypothesis testing bounds via Proposition 2.5.

C Alternative proof for spherically-spiked Wigner

Here we give an alternative proof of Corollary 3.14. The proof deals with the second moment directly rather than comparing to the i.i.d. Gaussian prior.

Corollary 3.14. *Consider the spherical prior $\mathcal{X}_{\mathrm{sph}}$. If $\lambda < 1$ then* $\mathrm{GWig}(\lambda, \mathcal{X}_{\mathrm{sph}})$ *is contiguous to* $\mathrm{GWig}(0)$.

Proof. By symmetry, we reduce the second moment to

$$\mathop{\mathbb{E}}_{x,x'} \exp\left(\frac{n\lambda^2}{2}\langle x, x'\rangle^2\right) = \mathop{\mathbb{E}}_{x} \exp\left(\frac{n\lambda^2}{2}\langle x, e_1\rangle^2\right) = \mathop{\mathbb{E}}_{x_1} \exp\left(\frac{n\lambda^2}{2}x_1^2\right),$$

where e_1 denotes the first standard basis vector. Note that the first coordinate x_1 of a point uniformly drawn from the unit sphere in $\mathbb{R}^n$ is distributed proportionally to $(1 - x_1^2)^{(n-3)/2}$, so that its square y is distributed proportionally to $(1 - y)^{(n-3)/2}y^{-1/2}$. Hence y is distributed as $\mathrm{Beta}(\frac{1}{2}, \frac{n-1}{2})$. The second moment is thus the moment generating function of $\mathrm{Beta}(\frac{1}{2}, \frac{n-1}{2})$ evaluated at $n\lambda^2/2$, and as such, we have

$$\mathop{\mathbb{E}}_{P_n}\left(\frac{\mathrm{d}Q_n}{\mathrm{d}P_n}\right)^2 = {}_1F_1\left(\frac{1}{2}; \frac{n}{2}; \frac{\lambda^2 n}{2}\right), \tag{6}$$

where ${}_1F_1$ denotes the confluent hypergeometric function.

Suppose $\lambda < 1$. Equation 13.8.4 from [DLMF] grants us that, as $n \to \infty$,

$$\begin{aligned}{}_1F_1\left(\frac{1}{2}; \frac{n}{2}; \frac{\lambda^2 n}{2}\right) = (1 + o(1))\left(\frac{n}{2}\right)^{1/4} e^{\zeta^2 n/8}\Bigg(&\lambda^2\sqrt{\frac{\zeta}{1-\lambda^2}}U(0, \zeta\sqrt{n/2}) \\ &+ \left(-\lambda^2\sqrt{\frac{\zeta}{1-\lambda^2}} + \sqrt{\frac{\zeta}{1-\lambda^2}}\right)\frac{U(-1, \zeta\sqrt{n/2})}{\zeta\sqrt{n/2}}\Bigg),\end{aligned}$$

where $\zeta = \sqrt{2(\lambda^2 - 1 - 2\log\lambda)}$ and U is the parabolic cylinder function,

$$= (1 + o(1)) \left(\frac{n}{2}\right)^{1/4} e^{\zeta^2 n/8} \left(\lambda^2 \sqrt{\frac{\zeta}{1-\lambda^2}} e^{-\zeta^2 n/8} (\zeta\sqrt{n/2})^{-1/2} + \left(-\lambda^2\sqrt{\frac{\zeta}{1-\lambda^2}} + \sqrt{\frac{\zeta}{1-\lambda^2}}\right) \frac{e^{-\zeta^2 n/8}(\zeta\sqrt{n/2})^{1/2}}{\zeta\sqrt{n/2}}\right),$$

by Equation 12.9.1 from [DLMF],

$$= (1 + o(1))(1 - \lambda^2)^{-1/2},$$

which is bounded as $n \to \infty$, for all $\lambda < 1$. The result follows from Lemma 2.3. □

D Conditioning method for Gaussian Wigner model

In this section we give the full details of the conditioning method for the Gaussian Wigner model. We assume that the prior is $\mathcal{X} = \mathrm{iid}(\pi/\sqrt{n})$ where π is a finitely-supported distribution on $\mathbb{R}$ with mean zero and variance one.

The argument that we will use is based on Banks et al. [2016], in particular their Proposition 5. Suppose ω_n is a set of 'good' x values so that $x \in \omega_n$ with probability $1 - o(1)$. Let $Q_n = \mathrm{GWig}_n(\lambda, \mathcal{X})$ and let $P_n = \mathrm{GWig}_n(0)$. Let $\tilde{\mathcal{X}}_n$ be the conditional distribution of $\mathcal{X}_n$ given ω_n. Let $\tilde{Q}_n = \mathrm{GWig}_n(\lambda, \tilde{\mathcal{X}})$. Our goal is to show $\tilde{Q}_n \lhd P_n$, from which it follows that $Q_n \lhd P_n$ (see Lemma 2.4). If we let Ω_n be the event that x and x' are both in ω_n, our second moment becomes

$$\mathop{\mathbb{E}}_{P_n} \left(\frac{\mathrm{d}\tilde{Q}_n}{\mathrm{d}P_n}\right)^2 = \mathop{\mathbb{E}}_{\tilde{x},\tilde{x}' \sim \tilde{\mathcal{X}}} \left[\exp\left(\frac{n\lambda^2}{2}\langle \tilde{x}, \tilde{x}'\rangle^2\right)\right] = (1 + o(1)) \mathop{\mathbb{E}}_{x,x' \sim \mathcal{X}} \left[\mathbb{1}_{\Omega_n} \exp\left(\frac{n\lambda^2}{2}\langle x, x'\rangle^2\right)\right].$$

Let $\Sigma \subseteq \mathbb{R}$ (a finite set) be the support of π, and let $s = |\Sigma|$. We will index Σ by $[s] = \{1, 2, \ldots, s\}$ and identify π with the vector of probabilities $\pi \in \mathbb{R}^s$. For $a, b \in \Sigma$, let N_{ab} denote the number of indices i for which $x_i = a/\sqrt{n}$ and $x'_i = b/\sqrt{n}$ (recall x_i is drawn from $\pi/\sqrt{n}$). Note that N follows a multinomial distribution with n trials, s^2 outcomes, and with probabilities given by $\bar{\alpha} = \pi\pi^\top \in \mathbb{R}^{s \times s}$. We have

$$\frac{n\lambda^2}{2}\langle x, x'\rangle^2 = \frac{\lambda^2}{2n}\left(\sum_{a,b\in\Sigma} abN_{ab}\right)^2 = \frac{\lambda^2}{2n}\sum_{a,b,a',b'} aba'b'N_{ab}N_{a'b'} = \frac{1}{n}N^\top AN$$

where A is the $s^2 \times s^2$ matrix $A_{ab,a'b'} = \frac{\lambda^2}{2}aba'b'$, and the quadratic form $N^\top AN$ is computed by treating N as a vector of length s^2.

We are now in a position to apply Proposition 5 from Banks et al. [2016]. Define $Y = (N - n\bar{\alpha})/\sqrt{n}$. Let Ω_n be the event defined in Appendix A of Banks et al. [2016], which enforces that the empirical distributions of $\sqrt{n}x$ and $\sqrt{n}x'$ are close to π; namely,

$$\max_j \left|\sum_i N_{ij} - n\pi_j\right| \le \eta_n \quad \text{and} \quad \max_i \left|\sum_j N_{ij} - n\pi_i\right| \le \eta_n$$

where (for concreteness) $\eta_n = \sqrt{n}\log n$.

Note that $\bar{\alpha}$ (treated as a vector of length s^2) is in the kernel of A because π is mean-zero: the inner product between $\bar{\alpha}$ and the (a, b) row of A is

$$\sum_{a',b'} A_{ab,a'b'}\bar{\alpha}_{a'b'} = \frac{\lambda^2}{2}\sum_{a',b'} aba'b'\pi_{a'}\pi_{b'} = \frac{\lambda^2}{2}ab\left(\sum_{a'} a'\pi_{a'}\right)\left(\sum_{b'} b'\pi_{b'}\right) = 0.$$

Therefore we have $\frac{1}{n}N^\top AN = Y^\top AY$ and so we can write our second moment as $(1+o(1))\mathbb{E}[\mathbb{1}_{\Omega_n}\exp(Y^\top AY)]$.

Let $\Delta_{s^2}(\pi)$ denote the set of nonnegative vectors $\alpha \in \mathbb{R}^{s^2}$ with row- and column-sums prescribed by π, i.e. treating α as an $s \times s$ matrix, we have (for all i) that row i and column i of α each sum to π_i. Let $D(u,v)$ denote the KL divergence between two vectors: $D(u,v) = \sum_i u_i \log(u_i/v_i)$. For convenience, we restate Proposition 5 in Banks et al. [2016].

Proposition D.1 (Banks et al. [2016], Proposition 5)**.** *Let $\pi \in \mathbb{R}^s$ be any vector of probabilities. Let A be any $s^2 \times s^2$ matrix. Define N, Y, $\bar{\alpha}$, and Ω_n as above (depending on π). Let*

$$m = \sup_{\alpha \in \Delta_{s^2}(\pi)} \frac{(\alpha - \bar{\alpha})^\top A(\alpha - \bar{\alpha})}{D(\alpha, \bar{\alpha})}.$$

If $m < 1$ then $\lim_{n\to\infty} \mathbb{E}[\mathbb{1}_{\Omega_n}\exp(Y^\top AY)] = \mathbb{E}[\exp(Z^\top AZ)] < \infty$, where $Z \sim \mathcal{N}(0, \mathrm{diag}(\bar{\alpha}) - \bar{\alpha}\bar{\alpha}^\top)$. If $m > 1$ then $\lim_{n\to\infty} \mathbb{E}[\mathbb{1}_{\Omega_n}\exp(Y^\top AY)] = \infty$.

We apply Proposition D.1 to our specific choice of π and A:

Theorem D.2 (conditioning method)**.** *Let $\mathcal{X} = \mathrm{iid}(\pi)$ where π has mean zero, unit variance, and finite support $\Sigma \subseteq \mathbb{R}$ with $|\Sigma| = s$. Let $Q_n = \mathrm{GWig}_n(\lambda, \mathcal{X})$, $\tilde{Q}_n = \mathrm{GWig}_n(\lambda, \tilde{\mathcal{X}})$, and $P_n = \mathrm{GWig}_n(0)$. Define the $s \times s$ matrix $\beta_{ab} = ab$ for $a, b \in \Sigma$. Let*

$$\overline{\lambda}_{\mathcal{X}} = \left[\sup_{\alpha \in \Delta_{s^2}(\pi)} \frac{\langle \alpha, \beta\rangle^2}{2D(\alpha, \bar{\alpha})}\right]^{-1/2}.$$

If $\lambda < \overline{\lambda}_{\mathcal{X}}$ then $\lim_{n\to\infty} \mathbb{E}_{P_n}(\mathrm{d}\tilde{Q}_n/\mathrm{d}P_n)^2 = (1-\lambda^2)^{-1/2} < \infty$ and so $Q_n \lhd P_n$. Conversely, if $\lambda > \overline{\lambda}_{\mathcal{X}}$ then $\lim_{n\to\infty} \mathbb{E}_{P_n}\left(\mathrm{d}\tilde{Q}_n/\mathrm{d}P_n\right)^2 = \infty$.

Note that this is a tight characterization of when the conditional second moment is bounded, but not necessarily of when contiguity holds.

The intuition behind this matrix optimization problem is the following. The matrix α represents the 'type' of a pair of spikes (x, x') in the sense that for any $a, b \in \Sigma$, α_{ab} is the fraction of entries i for which $x_i = a$ and $x'_i = b$. A pair (x, x') of type α contributes the value $\exp(\frac{n\lambda^2}{2}\langle \alpha, \beta\rangle^2)$ to the second moment $\mathbb{E}_{x,x'}\exp\left(\frac{n\lambda^2}{2}\langle x, x'\rangle^2\right)$. The probability (when $x, x' \sim \mathrm{iid}(\pi/\sqrt{n})$) that a particular type α occurs is asymptotically $\exp(-nD(\alpha, \bar{\alpha}))$. Due to the exponential scaling, the second moment is dominated by the worst α value: the second moment is unbounded if there is some α such that $\frac{\lambda^2}{2}\langle \alpha, \beta\rangle^2 > D(\alpha, \bar{\alpha})$. (This idea is often referred to as Laplace's method or the saddle point method.) Rearranging this yields the optimization problem in the theorem. The fact that we are conditioning on 'good' values of x (that have close-to-typical proportions of entries) allows us to add the constraint $\alpha \in \Delta_{s^2}(\pi)$. If we were not conditioning, we would have the same optimization problem over $\alpha \in \Delta_{s^2}$ (the simplex of dimension s^2), which in some cases gives a worse threshold.

Unfortunately we do not have a good general technique to understand the value of the matrix optimization problem. However, in certain special cases we do. Namely, in Appendix E we show, for the sparse Rademacher prior, how to use symmetry to reduce the problem to only two variables so that it can be easily solved numerically. In other applications, closed form solutions to related optimization problems have been found [Achlioptas and Naor, 2004, Banks et al., 2016].

Above we have computed the limit value of the second moment in the case $\lambda < \overline{\lambda}_{\mathcal{X}}$ as follows. Defining Z as in Proposition D.1 we have $\langle Z, \beta\rangle \sim \mathcal{N}(0, \sigma^2)$ where

$$\sigma^2 = \beta^\top(\mathrm{diag}(\bar{\alpha}) - \bar{\alpha}\bar{\alpha}^\top)\beta = \sum_{ab} \beta_{ab}^2 \bar{\alpha}_{ab} + \left(\sum_{ab} \beta_{ab}\bar{\alpha}_{ab}\right)^2$$

$$= \left(\sum_a a^2 \pi_a\right)\left(\sum_b b^2 \pi_b\right) + \left(\sum_a a\pi_a \sum_b b\pi_b\right)^2 = 1,$$

since π is mean-zero and unit-variance, and so

$$\mathbb{E}[\exp(Z^\top AZ)] = \mathbb{E}\left[\exp\left(\frac{\lambda^2}{2}\langle Z, \beta\rangle^2\right)\right] = \mathbb{E}\left[\exp\left(\frac{\lambda^2}{2}\chi_1^2\right)\right] = (1-\lambda^2)^{-1/2}.$$

E Sparse Rademacher prior

In this section we give details for our results on the spiked Gaussian Wigner model with the i.i.d. sparse Rademacher prior: $\mathrm{iid}(\pi/\sqrt{n})$ where $\pi = \sqrt{1/\rho}\,\mathcal{R}(\rho)$ where $\mathcal{R}(\rho)$ is the sparse Rademacher distribution with sparsity $\rho \in (0,1]$:

$$\mathcal{R}(\rho) = \begin{cases} 0 & \text{w.p.} & 1-\rho \\ +1 & \text{w.p.} & \rho/2 \\ -1 & \text{w.p.} & \rho/2 \end{cases}.$$

First we apply the subgaussian method (Theorem 3.10). The subgaussian constant σ^2 for π needs to satisfy

$$\exp\left(\frac{1}{2}\sigma^2 t^2\right) \ge \mathbb{E}\exp(t\pi) = 1 - \rho + \rho\cosh(t/\sqrt{\rho}) \tag{7}$$

for all $t \in \mathbb{R}$ so the best (smallest) choice for σ^2 is

$$(\sigma^*)^2 \triangleq \sup_{t\in\mathbb{R}} \frac{2}{t^2}\log\left[1-\rho+\rho\cosh(t/\sqrt{\rho})\right].$$

Recall that Theorem 3.10 (subgaussian method) gives contiguity for all $\lambda < 1/\sigma^*$. We now show that for sufficiently large ρ, we have $\sigma^* = 1$, implying that PCA is tight:

Proposition E.1. *When $\rho \ge 1/3$, we have $\sigma^* = 1$, yielding contiguity for all $\lambda < 1$. On the other hand, if $\rho < 1/3$, then $\sigma^* > 1$.*

Proof. We equivalently consider the following reformulation of (7):

$$\frac{1}{2}\sigma^2 t^2 \overset{?}{\ge} \log\left(1-\rho+\rho\cosh(t/\sqrt{\rho})\right) \triangleq k_\rho(t). \tag{8}$$

Both sides of the inequality are even functions of t, agreeing in value at $t = 0$. When $\sigma^2 < 1$, the inequality fails, by comparing their second-order behavior about $t = 0$. When $\sigma^2 = 1$ but $\rho < 1/3$, the inequality fails, as the two sides have matching behavior up to third order, but $k_\rho^{(4)}(0) = 3 - 1/\rho < 0$.

It remains to show that the inequality (8) does hold for $\rho > 1/3$ and $\sigma^2 = 1$. As the left and right sides agree to first order at $t = 0$, and are both even functions, it suffices to show that for all $t \ge 0$,

$$1 \overset{?}{\ge} k_\rho''(t) = \frac{\rho + (1-\rho)\cosh(t/\sqrt{\rho})}{(1-\rho+\rho\cosh(t/\sqrt{\rho}))^2}.$$

Completing the square for cosh, we have the equivalent inequality:

$$0 \overset{?}{\le} 1 - 3\rho + \rho^2 + \Bigl(\underbrace{\rho\cosh(t/\sqrt{\rho}) + \frac{(2\rho-1)(1-\rho)}{2\rho}}_{(*)}\Bigr)^2 - \frac{(2\rho-1)^2(1-\rho)^2}{4\rho^2}.$$

Note that cosh is bounded below by 1; thus for $\rho > 1/3$, the underbraced term $(*)$ is nonnegative, and hence minimized in absolute value when $t = 0$. It then suffices to show the above inequality in the case $t = 0$, so that $\cosh(t/\sqrt{\rho}) = 1$; but here the inequality is an equality, by simple algebra. ☐

Using the conditioning method of Section 3.6, we will now improve the range of ρ for which PCA is optimal, although our argument here relies on numerical optimization.

Example E.2. Let $\mathcal{X}$ be the sparse Rademacher prior $\mathrm{iid}(\sqrt{1/\rho}\,\mathcal{R}(\rho))$. There exists a critical value $\rho^* \approx 0.184$ (numerically computed) such that if $\rho \ge \rho^*$ and $\lambda < 1$ then $\mathrm{GWig}(\lambda, \mathcal{X})$ is contiguous to $\mathrm{GWig}(0, \mathcal{X})$. When $\rho < \rho^*$ we are only able to show contiguity when $\lambda < \lambda^*_\rho$ for some $\lambda^*_\rho < 1$.

Details. Consider the optimization problem of Theorem 3.16 (conditioning method). We will first use symmetry to argue that the optimal α must take a simple form. Abbreviate the support of π as $\{0, +, -\}$. For a given α matrix, define its complement by swapping $+$ and $-$, e.g. swap α_{0+} with α_{0-} and swap α_{-+} with α_{+-}. Note that if we average α with its complement, the numerator $\langle \alpha, \beta \rangle^2$ remains unchanged, the denominator $D(\alpha, \overline{\alpha})$ can only decrease, and the row- and column-sum constraints remain satisfied; this means the new solution is at least as good as the original α. Therefore we only need to consider α values satisfying $\alpha_{++} = \alpha_{--}$ and $\alpha_{+-} = \alpha_{-+}$. Note that the remaining entries of α are uniquely determined by the row- and column-sum constraints, and so we have reduced the problem to only two variables. It is now easy to solve the optimization problem numerically, say by grid search. $\square$

F Proof of non-Gaussian Wigner lower bound

In this section we prove Theorem 4.4, and verify its hypotheses for spherical and i.i.d. priors.

Theorem 4.4. *Under Assumption 4.3,* $\mathrm{Wig}(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$ *is contiguous to* $\mathrm{Wig}(0, \mathcal{P}, \mathcal{P}_d)$ *for all* $\lambda < \lambda^*_{\mathcal{X}}/\sqrt{F_{\mathcal{P}}}$.

Proof. We begin by conditioning the prior $\mathcal{X}$ on the high-probability events that $\|x\|_q^q \le \alpha_q n^{\frac{1}{q}-\frac{1}{2}}$ for $q = 2, 4, 6, 8$, and on the event that no entry of x exceeds $5\sqrt{\log n/n}$, which is true with high probability by the subgaussian hypothesis; let $\tilde{\mathcal{X}}$ be this conditioned prior. Hence if $\mathrm{Wig}(\lambda, \mathcal{P}, \tilde{\mathcal{X}})$ is contiguous to $\mathrm{Wig}(0, \mathcal{P})$ then so is $\mathrm{Wig}(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$. Let $Q_n = \mathrm{Wig}_n(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$, $\tilde{Q}_n = \mathrm{Wig}_n(\lambda, \mathcal{P}, \mathcal{P}_d, \tilde{\mathcal{X}})$, and $P_n = \mathrm{Wig}_n(0, \mathcal{P}, \mathcal{P}_d)$.

For convenience, let p_{ij} denote p if $i \ne j$ and p_d if $i = j$, the density of the noise on the ij entry. Likewise let τ_{ij} denote τ or τ_d as appropriate. We proceed from the second moment:

$$\begin{aligned}
\mathop{\mathbb{E}}_{P_n}\left(\frac{\mathrm{d}\tilde{Q}_n}{\mathrm{d}P_n}\right)^2 &= \mathop{\mathbb{E}}_{Y\sim P_n}\left[\mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\prod_{i\le j}\frac{p_{ij}(\sqrt{n}Y_{ij}-\lambda\sqrt{n}x_ix_j)}{p_{ij}(\sqrt{n}Y_{ij})}\frac{p_{ij}(\sqrt{n}Y_{ij}-\lambda\sqrt{n}x'_ix'_j)}{p_{ij}(\sqrt{n}Y_{ij})}\right]\\
&= \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\left[\prod_{i\le j}\mathop{\mathbb{E}}_{\sqrt{n}Y_{ij}\sim\mathcal{P}}\frac{p_{ij}(\sqrt{n}Y_{ij}-\lambda\sqrt{n}x_ix_j)}{p_{ij}(\sqrt{n}Y_{ij})}\frac{p_{ij}(\sqrt{n}Y_{ij}-\lambda\sqrt{n}x'_ix'_j)}{p_{ij}(\sqrt{n}Y_{ij})}\right]\\
&= \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\left[\exp\left(\sum_{i\le j}\tau_{ij}(\lambda\sqrt{n}x_ix_j,\lambda\sqrt{n}x'_ix'_j)\right)\right].
\end{aligned}$$

We will expand τ and τ_d using Taylor's theorem, using the C^4 assumption:

$$\tau(a,b) = \sum_{0\le k+\ell\le 3}\frac{1}{(k+\ell)!}\frac{\partial^{k+\ell}\tau}{\partial a^k\partial b^\ell}(0,0)\,a^kb^\ell + \sum_{k+\ell=4}\frac{1}{4!}\left(\frac{\partial^4\tau}{\partial a^k\partial b^\ell}(0,0)+h_{k,\ell}(a,b)\right)a^kb^\ell$$

$$\tau_d(a,b) = \sum_{0\le k+\ell\le 1}\frac{1}{(k+\ell)!}\frac{\partial^{k+\ell}\tau}{\partial a^k\partial b^\ell}(0,0)\,a^kb^\ell + \sum_{k+\ell=2}\frac{1}{2!}\left(\frac{\partial^2\tau}{\partial a^k\partial b^\ell}(0,0)+h_{d;k,\ell}(a,b)\right)a^kb^\ell$$

for some remainder function $h_{k,\ell}(a,b)$ tending to 0 as $(a,b) \to (0,0)$. As x and x' are entrywise $O(\sqrt{\log n/n})$, these remainder terms $h_{k,\ell}(\lambda\sqrt{n}x_ix_j, \lambda\sqrt{n}x'_ix'_j)$ are $o(1)$ as $n \to \infty$. Note that $\tau(a, 0) = 0 = \tau(0, b)$, so that the non-mixed partials of τ vanish, and likewise for τ_d. We note also that $\frac{\partial^2 \tau}{\partial a \partial b}(0,0) = F_{\mathcal{P}}$, the Fisher information defined above. Thus,

$$\mathop{\mathbb{E}}_{P_n}\left(\frac{\mathrm{d}\tilde{Q}_n}{\mathrm{d}P_n}\right)^2 = \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\Bigg[\exp\Bigg(F_{\mathcal{P}}\lambda^2 n\sum_{i<j} x_i x_j x'_i x'_j + \sum_{\substack{k+\ell=3\\k,\ell>0}} \frac{\partial^3\tau}{\partial a^k \partial b^\ell}(0,0)\frac{\lambda^3 n^{3/2}}{k!\ell!}\sum_{i<j} x_i^k x_j^k (x'_i)^\ell (x'_j)^\ell + \sum_{k+\ell=4}\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0)+o(1)\right)\frac{\lambda^4 n^2}{k!\ell!}\sum_{i<j} x_i^k x_j^k (x'_i)^\ell (x'_j)^\ell + \left(\frac{\partial^2\tau_d}{\partial a \partial b}(0,0)+o(1)\right)\lambda^2 n\sum_i x_i^2 (x'_i)^2 \Bigg)\Bigg].$$

We can separate these four terms using a weighted AM–GM inequality. For all $\varepsilon > 0$:

$$\mathop{\mathbb{E}}_{P_n}\left(\frac{\mathrm{d}\tilde{Q}_n}{\mathrm{d}P_n}\right)^2 \le \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\exp\left((1-\varepsilon)^{-1}F_{\mathcal{P}}\lambda^2 n\sum_{i<j} x_i x_j x'_i x'_j\right) \tag{9}$$

$$+ \sum_{\substack{k+\ell=3\\k,\ell>0}} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\exp\left(\frac{8}{\varepsilon}\frac{\partial^3\tau}{\partial a^k \partial b^\ell}(0,0)\frac{\lambda^3 n^{3/2}}{k!\ell!}\sum_{i<j} x_i^k x_j^k (x'_i)^\ell (x'_j)^\ell\right) \tag{10}$$

$$+ \sum_{k+\ell=4} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\exp\left(\frac{8}{\varepsilon}\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0)+o(1)\right)\frac{\lambda^4 n^2}{k!\ell!}\sum_{i<j} x_i^k x_j^k (x'_i)^\ell (x'_j)^\ell\right) \tag{11}$$

$$+ \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\exp\left(\frac{8}{\varepsilon}\left(\frac{\partial^2\tau_d}{\partial a \partial b}(0,0)+o(1)\right)\lambda^2 n\sum_i x_i^2 (x'_i)^2\right) \tag{12}$$

so it suffices to control terms (9–12) individually.

By hypothesis, $\lambda < \lambda^*_{\mathcal{X}}/\sqrt{F_{\mathcal{P}}}$, implying that we can choose $\varepsilon > 0$ such that $(1-\varepsilon)^{-1}F_{\mathcal{P}}\lambda^2 < (\lambda^*_{\mathcal{X}})^2$. But $\tilde{\mathcal{X}}$ is dominated as a measure by $(1+o(1))\mathcal{X}$; it follows that $\lambda_{\mathcal{X}} \le \lambda_{\tilde{\mathcal{X}}}$, and the first term (9) is bounded.

We bound the second term (10) using the subgaussian assumption:

$$\begin{aligned}(10) &\le 2 \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}}\exp\left(\frac{2\lambda^3 n^{3/2}}{\varepsilon}\frac{\partial^3\tau}{\partial a^2 \partial b}(0,0)\langle x^2, x'\rangle^2\right)\\ &= 2\mathbb{E}_{x\sim\tilde{\mathcal{X}}}\mathbb{E}_{x'\sim\tilde{\mathcal{X}}}\exp(\langle v, x'\rangle^2)\\ &= 2\mathbb{E}_{x\sim\tilde{\mathcal{X}}}(1+o(1))\mathbb{E}_{x'\sim\mathcal{X}}\exp(\langle v, x'\rangle^2)\end{aligned}$$

where $v = \sqrt{2/\varepsilon}\lambda^{3/2}n^{3/4}\sqrt{\frac{\partial^3\tau}{\partial a^2 \partial b}(0,0)}\, x^2$. We thus have

$$\|v\|_2^2 = \frac{2\lambda^3 n^{3/2}}{\varepsilon}\frac{\partial^3\tau}{\partial a^2 \partial b}(0,0)\|x\|_4^4 = O(n^{1/2}).$$

By subgaussian hypothesis on $\mathcal{X}$, the inner expectation over x' is $O(1)$, so that the overall term (10) is bounded.

We bound the third term (11) using Cauchy–Schwarz:

$$\begin{aligned}
(11) &\leq \sum_{k+\ell=4} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0) + o(1)\right) \frac{8\lambda^4 n^2}{2\varepsilon k!\ell!} \langle x^k, (x')^\ell\rangle^2\right) \\
&\leq \sum_{k+\ell=4} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0) + o(1)\right) \frac{8\lambda^4 n^2}{2\varepsilon k!\ell!} \|x^k\|_2^2 \, \|(x')^\ell\|_2^2\right) \\
&= \sum_{k+\ell=4} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0) + o(1)\right) \frac{8\lambda^4 n^2}{2\varepsilon k!\ell!} \|x\|_{2k}^{2k} \, \|x'\|_{2\ell}^{2\ell}\right) \\
&\leq \sum_{k+\ell=4} \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0) + o(1)\right) \frac{8\lambda^4 n^2}{2\varepsilon k!\ell!} \alpha_{2k}^{2k} n^{1-k} \alpha_{2\ell}^{2\ell} n^{1-\ell}\right) \\
&= \sum_{k+\ell=4} \exp\left(\left(\frac{\partial^4\tau}{\partial a^k \partial b^\ell}(0,0) + o(1)\right) \frac{8\lambda^4}{2\varepsilon k!\ell!} \alpha_{2k}^{2k} \alpha_{2\ell}^{2\ell}\right),
\end{aligned}$$

due to the norm restrictions on prior $\tilde{\mathcal{X}}$. This evidently remains bounded as $n \to \infty$.

The fourth term proceeds similarly:

$$\begin{aligned}
(12) &\leq \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\frac{8\lambda^2 n}{\varepsilon}\left(\frac{\partial^2\tau_d}{\partial a \partial b}(0,0) + o(1)\right) \langle x^2, (x')^2\rangle\right) \\
&\leq \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\frac{8\lambda^2 n}{\varepsilon}\left(\frac{\partial^2\tau_d}{\partial a \partial b}(0,0) + o(1)\right) \|x\|_4^2 \|x'\|_4^2\rangle\right) \\
&\leq \mathop{\mathbb{E}}_{x,x'\sim\tilde{\mathcal{X}}} \exp\left(\frac{8\lambda^2 n}{\varepsilon}\left(\frac{\partial^2\tau_d}{\partial a \partial b}(0,0) + o(1)\right) \alpha_4^4 n^{-1}\right)
\end{aligned}$$

which likewise remains bounded.

With the overall second moment $\mathbb{E}_{P_n}\left(\frac{\mathrm{d}\tilde{Q}_n}{\mathrm{d}P_n}\right)^2$ bounded as $n \to \infty$, the result follows from Lemma 2.4. ☐

Proposition 4.5. *Conditions (i) and (ii) in Assumption 4.3 are satisfied for the spherical prior* $\mathcal{X}_{\mathrm{sph}}$.

Proof. Note that one can sample $x \sim \mathcal{X}_{\mathrm{sph}}$ by first sampling $y \sim \mathcal{N}(0, I_n)$ and then taking $x = y/\|y\|_2$. By Chebyshev, $\left|\|y\|_2^2 - n\right| < n^{3/4}$ with probability $1 - o(1)$. For $q \in \{4, 6, 8\}$, $\|y\|_q^q$ has expectation $n(q-1)!!$ and variance

$$n[(2q-1)!! - ((q-1)!!)^2].$$

Supposing that $\|y\|_2^2 > n - n^{3/4} > n/2$, which occurs with probability $1 - o(1)$, we have for any α_q that

$$\begin{aligned}
\Pr[\|x_q\| > \alpha_q n^{\frac{1}{q}-\frac{1}{2}}] &= \Pr[\|x\|_q^q > \alpha_q^q n^{1-\frac{q}{2}}] \\
&= \Pr[\|y\|_q^q > \alpha_q^q n^{1-\frac{q}{2}} \|y\|_2^q] \\
&\leq \Pr[\|y\|_q^q > \alpha_q^q 2^{-q/2} n] \\
&\leq \frac{n((2q-1)!! - ((q-1)!!)^2)}{n^2(2^{-q}\alpha_q^{2q} - (q-1)!!)^2},
\end{aligned}$$

by Chebyshev. This probability is $o(1)$ so long as we take $\alpha_q^{2q} > 2^q(q-1)!!$.

The spherical prior is appropriately subgaussian: the inner product $\langle x, v\rangle$ is distributed as $2z - 1$ with $z \sim \mathrm{Beta}(n/2, n/2)$, which is known to be $O(1/n)$-subgaussian (see e.g. Elder [2016]). ☐

Proposition 4.6. *Consider an i.i.d. prior* $\mathcal{X} = \mathrm{iid}(\pi/\sqrt{n})$ *where* π *is zero-mean, unit-variance, and subgaussian with some constant* σ^2. *Then conditions (i) and (ii) in Assumption 4.3 are satisfied.*

Proof. We have $x_i = \frac{1}{\sqrt{n}}\pi_i$ where π_i are independent copies of π. For $q \in \{2,4,6,8\}$,

$$\begin{aligned}\Pr[\|x\|_q > \alpha_q n^{\frac{1}{q}-\frac{1}{2}}] = \Pr[\|x\|_q^q > \alpha_q^q n^{1-\frac{q}{2}}] &= \Pr\left[\sum_i x_i^q > \alpha_q^q n^{1-\frac{q}{2}}\right] \\ = \Pr\left[\sum_i \pi_i^q > \alpha_q^q n\right] &= \Pr\left[\sum_i \pi_i^q - n\mathbb{E}[\pi^q] > (\alpha_q^q - \mathbb{E}[\pi^q])n\right].\end{aligned}$$

Choose α_q so that $C \equiv \alpha_q^q - \mathbb{E}[\pi^q] > 0$, and apply Chebyshev's inequality:

$$\leq \frac{\mathrm{Var}[\sum_i \pi_i^q]}{C^2 n^2} = \frac{n\mathrm{Var}[\pi^q]}{C^2 n^2} = \mathcal{O}(1/n).$$

Here we needed $\mathbb{E}[\pi^{2q}] < \infty$ (which follows from subgaussianity) so that $\mathrm{Var}[\pi^q] < \infty$. □

G Non-Gaussian Wigner with discrete noise

In this section we show that in the non-Gaussian Wigner model, if the noise distribution has a point mass then the detection problem becomes easy for any $\lambda > 0$.

Theorem G.1. *Let $\mathcal{P}$ be a (mean-zero, unit-variance) distribution on $\mathbb{R}$ with a point mass: $\Pr_{w\sim\mathcal{P}}[w = c] = m$ for some c and some $m > 0$. Let $\mathcal{P}_d$ be any distribution on $\mathbb{R}$. Let $\mathcal{X}$ be a spike prior such that for some $\delta > 0$ and $\alpha > 0$, with probability $1 - o(1)$, $x \sim \mathcal{X}_n$ satisfies both (i) $\|x\|_0 \geq \delta n$ and (ii) $|x_i| \leq n^{-1/4-\alpha}$ $\forall i$. Then for any $\lambda > 0$, there exists a test that consistently distinguishes $\mathrm{Wig}(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$ from $\mathrm{Wig}(0, \mathcal{P}, \mathcal{P}_d)$.*

Here, $\|x\|_0$ denotes the ℓ_0 norm, i.e. the number of nonzero entries.

Proof. Let the test statistic $T(Y)$ be the fraction of entries of Y that are exactly equal to $c/\sqrt{n}$. Under the unspiked model $Y \sim \mathrm{Wig}(0, \mathcal{P}, \mathcal{P}_d)$, we have $T(Y) \to m$ in probability. Let $\varepsilon > 0$. Under the spiked model $Y \sim \mathrm{GWig}(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$ we have with probability $1 - o(1)$ that at least $(\delta^2 - \varepsilon)n^2$ entries of $xx^\top$ lie in the set $[-n^{-1/2-2\alpha}, n^{-1/2-2\alpha}] \smallsetminus \{0\}$. With probability $1 - o(1)$, at most εn^2 of the corresponding entries of Y take the value (exactly) $c/\sqrt{n}$ because by continuity of measure,

$$\lim_{d\to 0^+} \Pr_{w\sim\mathcal{P}}[w \in [c-d, c+d] \smallsetminus \{c\}] = 0.$$

Therefore, taking ε sufficiently small, we have $T(Y) \leq m - \varepsilon$ with probability $1 - o(1)$ and thus T consistently distinguishes the spiked and unspiked models. □

H Proof of pre-transformed PCA

In this section we prove our upper bound for the non-Gaussian Wigner model via pre-transformed PCA. We make the following assumptions on the spike prior $\mathcal{X}$ and the entrywise noise distribution $\mathcal{P}$.

Assumption 4.7. *Of the prior $\mathcal{X}$ we require (as usual) $\|x\| \to 1$ in probability, and we also assume that with probability $1 - o(1)$, all entries of x are small: $|x_i| \leq n^{-1/2+\alpha}$ for some fixed $\alpha < 1/8$. Of the noise $\mathcal{P}$, we assume the following:*

(i) $\mathcal{P}$ has a non-vanishing C^3 density function $p(w) > 0$,

(ii) Letting $f(w) = -p'(w)/p(w)$, we have that f and its first two derivatives are polynomially-bounded: there exists $C > 0$ and an even integer $m \geq 2$ such that $|f^{(\ell)}(w)| \leq C + w^m$ for all $0 \leq \ell \leq 2$.

(iii) With m as in (ii), $\mathcal{P}$ has finite moments up to $5m$: $\mathbb{E}|\mathcal{P}|^k < \infty$ for all $1 \leq k \leq 5m$.

An important consequence of assumptions (ii) and (iii) is the following.

Lemma H.1. *$\mathbb{E}|f^{(\ell)}(\mathcal{P})|^q < \infty$ for all $0 \le \ell \le 2$ and $1 \le q \le 5$. Likewise $\mathbb{E}|f^{(\ell)}(\mathcal{P}_d)|^q < \infty$ for all $0 \le \ell \le 2$ and $1 \le q \le 3$.*

Proof. We demonstrate $\mathcal{P}$; then $\mathcal{P}_d$ follows identically. Using $|a+b|^q \le |2a|^q + |2b|^q = 2^q(|a|^q + |b|^q)$ we have

$$\mathbb{E}|f^{(\ell)}(\mathcal{P})|^q \le \mathbb{E}|C + \mathcal{P}^m|^q \le 2^q(C^q + \mathbb{E}|\mathcal{P}|^{mq}) < \infty. \qquad \square$$

The main theorem of this section is the following.

Theorem 4.8. *Let $\lambda \ge 0$ and let $\mathcal{X}, \mathcal{P}$ satisfy Assumption 4.7. Let $\widehat{Y} = \sqrt{n}Y$ where Y is drawn from $\mathrm{Wig}(\lambda, \mathcal{P}, \mathcal{P}_d, \mathcal{X})$. Let $f(\widehat{Y})$ denote entrywise application of the function $f(w) = -p'(w)/p(w)$ to $\widehat{Y}$, except we define the diagonal entries of $f(\widehat{Y})$ to be zero.*

- *If $\lambda \le 1/\sqrt{F_\mathcal{P}}$ then $\frac{1}{\sqrt{n}}\lambda_{\max}(f(\widehat{Y})) \to 2\sqrt{F_\mathcal{P}}$ as $n \to \infty$.*
- *If $\lambda > 1/\sqrt{F_\mathcal{P}}$ then $\frac{1}{\sqrt{n}}\lambda_{\max}(f(\widehat{Y})) \to \lambda F_\mathcal{P} + \frac{1}{\lambda} > 2\sqrt{F_\mathcal{P}}$ as $n \to \infty$ and furthermore the top (unit-norm) eigenvector v of $f(\widehat{Y})$ correlates with the spike: $\langle v, x\rangle^2 \ge (\lambda - 1/\sqrt{F_\mathcal{P}})^2/\lambda^2 - o(1)$ with probability $1 - o(1)$.*

Convergence is in probability. Here $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue.

Note that Lemma H.1 implies that the expectation defining $F_\mathcal{P}$ is finite.

Proof. First we justify a local linear approximation of $f(\widehat{Y}_{ij})$. For $i \ne j$, define the error term $\mathcal{E}_{ij}$ by

$$f(\widehat{Y}_{ij}) = f(W_{ij}) + \lambda\sqrt{n}x_i x_j f'(W_{ij}) + \mathcal{E}_{ij}.$$

(Define $\mathcal{E}_{ii} = 0$.) We will show that the operator norm of $\mathcal{E}$ is small: $\|\mathcal{E}\| = o(\sqrt{n})$ with probability $1 - o(1)$. Apply the mean-value form of the Taylor approximation remainder: $\mathcal{E}_{ij} = \frac{1}{2}f''(W_{ij} + e_{ij})\lambda^2 n x_i^2 x_j^2$ for some $|e_{ij}| \le |\lambda\sqrt{n}x_i x_j|$. Bound the operator norm by the Frobenius norm:

$$\|\mathcal{E}\|^2 \le \|\mathcal{E}\|_F^2 = \frac{\lambda^4 n^2}{4}\sum_{i \ne j} x_i^4 x_j^4 f''(W_{ij} + e_{ij})^2 \le \frac{\lambda^4}{4} n^{8\alpha - 2} \sum_{i \ne j} f''(W_{ij} + e_{ij})^2.$$

Using the polynomial bound on f'' and the fact $|a+b|^k \le 2^k(|a|^k + |b|^k)$, we have

$$\begin{aligned} f''(W_{ij} + e_{ij})^2 &\le (C + (W_{ij} + e_{ij})^m)^2 \le 4C^2 + 4(W_{ij} + e_{ij})^{2m} \\ &\le 4C^2 + 4 \cdot 2^{2m}(W_{ij}^{2m} + e_{ij}^{2m}) \\ &\le 4C^2 + 2^{2m+2}(W_{ij}^{2m} + \lambda^{2m} n^{(4\alpha - 1)m}) \\ &= 4C^2 + 2^{2m+2} W_{ij}^{2m} + o(1). \end{aligned}$$

Using finite moments of $W_{ij} \sim \mathcal{P}$, it follows that $\mathbb{E}\left[\sum_{i \ne j} f''(W_{ij} + e_{ij})^2\right] = \mathcal{O}(n^2)$, and so $\mathbb{E}\|\mathcal{E}\|^2 = \mathcal{O}(n^{8\alpha})$. Since $\alpha < 1/8$, Markov's inequality now gives the desired result: with probability $1 - o(1)$, $\|\mathcal{E}\|^2 = o(n)$ and so $\|\mathcal{E}\| = o(\sqrt{n})$.

Our goal will be to show that $f(\widehat{Y})$ is, up to small error terms, another spiked Wigner matrix. Toward this goal we define another error term: for $i \ne j$, let $\Delta_{ij} = \lambda\sqrt{n}x_i x_j\,(f'(W_{ij}) - \mathbb{E}[f'(W_{ij})])$, so that

$$f(\widehat{Y}_{ij}) = f(W_{ij}) + \lambda\sqrt{n}x_i x_j \mathbb{E}[f'(W_{ij})] + \mathcal{E}_{ij} + \Delta_{ij}. \tag{13}$$

(Define $\Delta_{ii} = 0$.) We will show that the operator norm of Δ is small: $\|\Delta\| = o(\sqrt{n})$ with probability $1 - o(1)$. Let $A_{ij} = f'(W_{ij}) - \mathbb{E}[f'(W_{ij})]$ so that $\Delta_{ij} = \lambda\sqrt{n}x_i x_j A_{ij}$. (Define $A_{ii} = 0$.) We have $\|\Delta\| \le \lambda n^{-1/2+2\alpha}\|A\|$ because for any unit vector y,

$$\begin{aligned} y^\top \Delta y &= \sum_{i,j} \lambda\sqrt{n}x_i x_j A_{ij} y_i y_j \le \sum_{i,j} \lambda\sqrt{n} z_i A_{ij} z_j \qquad \text{where } z_i = x_i y_i \\ &\le \lambda\sqrt{n}\,\|A\| \cdot \|z\|^2 \le \lambda n^{-1/2+2\alpha}\|A\| \cdot \|y\| = \lambda n^{-1/2+2\alpha}\|A\|. \end{aligned}$$

Note that A is a Wigner matrix (i.e. a symmetric matrix with off-diagonal entries i.i.d.) and so $\|A\| = \mathcal{O}(\sqrt{n})$ with probability $1 - o(1)$. This follows from Pizzo et al. [2013] Theorem 1.1, provided we can check that each entry of A has finite fifth moment. But this follows from Lemma H.1:

$$\mathbb{E}|A_{ij}|^5 \le 2^5 \left(\mathbb{E}|f'(W_{ij})|^5 + |\mathbb{E}[f'(W_{ij})]|^5\right) < \infty.$$

Now we have $\|\Delta\| = \mathcal{O}(n^{2\alpha}) = o(\sqrt{n})$ with probability $1 - o(1)$ as desired.

From (13) we now have that, up to small error terms, $f(\widehat{Y})$ is another spiked Wigner matrix:

$$f(\widehat{Y}) = f(W) + \lambda\sqrt{n}\,\mathbb{E}[f'(\mathcal{P})]\,xx^\top + \mathcal{E} + \Delta - \delta$$

where (to take care of the diagonal) we define $f(W)_{ii} = 0$, $\delta_{ij} = 0$, and $\delta_{ii} = \lambda\sqrt{n}\,\mathbb{E}[f'(\mathcal{P})]x_i^2$. Note that the final error term δ is also small: $\|\delta\| \le \|\delta\|_F = \mathcal{O}(n^{2\alpha}) = o(\sqrt{n})$. We now have

$$\frac{1}{\sqrt{n}}\lambda_{\max}(f(\widehat{Y})) = \lambda_{\max}\left(\frac{1}{\sqrt{n}}f(W) + \lambda\,\mathbb{E}[f'(\mathcal{P})]\,xx^\top\right) + o(1)$$

and so the theorem follows from known results on the spectrum of spiked Wigner matrices, namely Theorem 1.1 from Pizzo et al. [2013]. We need to check the following details. First note that the Wigner matrix $f(W)$ has off-diagonal i.i.d. entries that are centered:

$$\mathbb{E}[f(W_{ij})] = \int_{-\infty}^{\infty} \frac{-p'(w)}{p(w)} p(w)dw = p(-\infty) - p(\infty) = 0.$$

Each off-diagonal entry of $f(W)$ has variance $\mathbb{E}[f(W_{ij})^2] = F_\mathcal{P}$. The rank-1 deformation $\lambda\,\mathbb{E}[f'(\mathcal{P})]\,xx^\top$ has top eigenvalue $\lambda\,\mathbb{E}[f'(\mathcal{P})] \cdot \|x\|^2$. Recall that $\|x\|^2 \to 1$ in probability. Also,

$$f'(w) = \frac{d}{dw}\frac{-p'(w)}{p(w)} = -\frac{p''(w)p(w) - p'(w)^2}{p(w)^2}$$

and so

$$\mathbb{E}[f'(\mathcal{P})] = \int_{-\infty}^{\infty}\left[-p''(w) + \frac{p'(w)^2}{p(w)}\right]dw = \int_{-\infty}^{\infty}\frac{p'(w)^2}{p(w)}dw = F_\mathcal{P}.$$

Therefore the top eigenvalue of the rank-1 deformation converges in probability to $\lambda F_\mathcal{P}$. By Lemma H.1, the entries of $f(W)$ have finite fifth moment.

The desired convergence of the top eigenvalue now follows. It remains to show that when $\lambda > 1/\sqrt{F_\mathcal{P}}$, the top eigenvalue of $f(\widehat{Y})$ correlates with the planted vector x. Let v be the top eigenvector of $f(\widehat{Y})$ with $\|v\| = 1$. From above we have

$$v^\top\left(\frac{1}{\sqrt{n}}f(\widehat{Y})\right)v = v^\top\left(\frac{1}{\sqrt{n}}f(W)\right)v + \lambda F_\mathcal{P}\langle v, x\rangle^2 + o(1).$$

We know $\frac{1}{\sqrt{n}}f(\widehat{Y})$ has top eigenvalue $\lambda F_\mathcal{P} + 1/\lambda + o(1)$ and $\frac{1}{\sqrt{n}}f(W)$ has top eigenvalue $2\sqrt{F_\mathcal{P}} + o(1)$, which yields

$$\langle v, x\rangle^2 \ge \frac{1}{\lambda F_\mathcal{P}}(\lambda F_\mathcal{P} + 1/\lambda - 2\sqrt{F_\mathcal{P}}) - o(1) = \frac{(\lambda - 1/\sqrt{F_\mathcal{P}})^2}{\lambda^2} - o(1). \qquad \square$$

I Proof of Theorem 5.3: MLE for Wishart with finite prior

Note the following well-known Chernoff bound for the χ_k^2 distribution:

Lemma I.1. *For all* $0 < z < 1$,

$$\frac{1}{k}\log\Pr\left[\chi_k^2 < zk\right] \le \frac{1}{2}(1 - z + \log z).$$

Similarly, for all $z > 1$,

$$\frac{1}{k}\log\Pr\left[\chi_k^2 > zk\right] \le \frac{1}{2}(1 - z + \log z).$$

We now prove the following theorem:

Theorem 5.3. *Let* $\beta \in (-1, \infty)$. *Let* $\mathcal{X}_n$ *be a spike prior supported on at most* c^n *points, for some fixed* $c > 0$. *If*

$$2\gamma \log c < \beta - \log(1 + \beta)$$

then there is a (computationally inefficient) procedure that distinguishes between the spiked Wishart model $\mathrm{Wish}(\gamma, \beta, \mathcal{X})$ *and the unspiked model* $\mathrm{Wish}(\gamma)$, *with* $o(1)$ *probability of error.*

Proof. First consider the case $\beta < 0$. Given a matrix Y, consider the test statistic

$$T = \min_{v \in \mathrm{supp}(\mathcal{X}_n)} \frac{v^\top Y v}{\|v\|^2}$$

where $\mathrm{supp}(\mathcal{X}_n)$ denotes the support of $\mathcal{X}_n$. Under $Y \sim \mathrm{Wish}(\gamma, \beta, \mathcal{X})$ with true spike x, we have that $x^\top Y x / \|x\|^2 \sim \frac{1}{N}(1 + \beta\|x\|^2)\chi_N^2$, which converges in probability to $1 + \beta$ (since $\|x\| \to 1$ in probability). Hence, for any $\varepsilon > 0$, we have that $T < 1 + \beta + \varepsilon$ with probability $1 - o(1)$ under the spiked model $\mathrm{Wish}(\gamma, \beta, \mathcal{X})$.

Let $\widehat{\gamma} = n/N$ so that $\widehat{\gamma} \to \gamma$. Under the unspiked model, we have

$$\begin{aligned}
\Pr[T \le 1 + \beta + \varepsilon] &\le \sum_{v \in \mathrm{supp}(\mathcal{X})} \Pr[v^\top Y v / \|v\|^2 \le 1 + \beta + \varepsilon] \\
&\le c^n \Pr\left[\chi_N^2 \le (1 + \beta + \varepsilon)N\right] \\
&= \exp\left[N\left(\widehat{\gamma}\log c + \frac{1}{N}\log\Pr\left[\chi_N^2 \le (1 + \beta + \varepsilon)N\right]\right)\right] \\
&\le \exp\left[N\left(\widehat{\gamma}\log c + \frac{1}{2}(1 - (1 + \beta + \varepsilon) + \log(1 + \beta + \varepsilon))\right)\right] \quad \text{by Lemma I.1.}
\end{aligned}$$

This is $o(1)$ so long as

$$2\gamma\log c - \beta - \varepsilon + \log(1 + \beta + \varepsilon) < 0.$$

We can choose such $\varepsilon > 0$ precisely under the hypothesis of this theorem.

Hence, by thresholding the statistic T at $1 + \beta + \varepsilon$, we obtain a hypothesis test that distinguishes $Y \sim \mathrm{Wish}(\gamma, \beta, \mathcal{X})$ from $Y \sim \mathrm{Wish}(\gamma)$, with probability $o(1)$ of error of either type.

The proof for the case $\beta > 0$ is similar, using instead the test statistic $T = \max_{v \in \mathrm{supp}(\mathcal{X}_n)} v^\top Y v / \|v\|^2$ along with the upper tail bound for χ_k^2. ◻

J Basic properties of Wishart lower bound

In this section we give basic properties of the condition on γ, β required by Theorem 5.7. Recall that this condition is $\gamma > \gamma^*$ where

$$\gamma^* f_{\mathcal{X}}(t) \ge F(\beta, t) \quad \forall t \in (0, 1) \tag{14}$$

where

$$F(\beta, t) \triangleq (1 + \beta)\frac{t(w - t)}{1 - t^2} + \frac{1}{2}\log\left(\frac{1 - w^2}{1 - t^2}\right)$$

and

$$w = \sqrt{A^2+1} - A \quad \text{with} \quad A = \frac{1-t^2}{2t(\beta+1)}.$$

We have the following properties of $F(\beta, t)$, which can be shown using basic calculus.

- The $t \to 0^+$ and $t \to 1^-$ limits of $F(\beta, t)$ exist and so $F(\beta, t)$ is defined and continuous in both variables on the domain $\beta \in (-1, \infty)$, $t \in [0, 1]$. The boundary values are $F(\beta, 0) = 0$ and $F(\beta, 1) = \frac{1}{2}(\beta - \log(1 + \beta))$.
- For any $\beta \in (-1, \infty) \setminus \{0\}$, $F(\beta, t)$ is a strictly increasing function of t. In particular, $F(\beta, t) \geq 0$ with equality only at $t = 0$.
- For any $\beta \in (-1, \infty)$, $\lim_{t\to 0^+} \frac{\partial}{\partial t} F(\beta, t) = 0$ and $\lim_{t\to 0^+} \frac{\partial^2}{\partial t^2} F(\beta, t) = \beta^2$.

We now give some lemmas that allow for a tradeoff between certain variables while keeping (14) true. The first allows the rate function to be weakened slightly at the expense of increasing γ^* slightly.

Lemma J.1. *Let $\gamma^* > 0$, $\beta \in (-1, \infty) \setminus \{0\}$, and $\varepsilon > 0$. Let $f(t)$ be a function on $(0, 1)$. If $\gamma^* f(t) \geq F(\beta, t)$ $\forall t \in (0, 1)$ then there exists $\delta > 0$ such that $(\gamma^* + \varepsilon) f(t(1 - \delta)^2) \geq F(\beta, t)$ $\forall t \in (0, 1)$.*

Proof. We have

$$(\gamma^* + \varepsilon) f(t(1-\delta)^2)) \geq \frac{\gamma^*+\varepsilon}{\gamma^*} F(\beta, t(1-\delta)^2)$$

so it is sufficient to show

$$\frac{F(\beta, t)}{F(\beta, t(1-\delta)^2)} \leq \frac{\gamma^*+\varepsilon}{\gamma^*} \quad \forall t \in (0, 1]. \tag{15}$$

For each $t \in (0, 1]$ there exists a maximal $\delta = \delta(t) > 0$ such that (15) holds, and $\delta(t)$ is a continuous function of t. We want to show that $\delta(t)$ is bounded above 0, so we only need to check the limit $t \to 0$.

Since $\lim_{t\to 0} F(\beta, t) = \lim_{t\to 0} \frac{\partial}{\partial t} F(\beta, t) = 0$ and $\lim_{t\to 0} \frac{\partial^2}{\partial t^2} F(\beta, t) = \beta^2 > 0$ we have, using L'Hôpital's rule,

$$\lim_{t\to 0} \frac{F(\beta, t)}{F(\beta, t(1-\delta)^2)} = \frac{1}{(1-\delta)^4}$$

which can be made smaller than $(\gamma^* + \varepsilon)/\gamma^*$ by taking $\delta > 0$ small enough. □

The next lemma allows β to be increased slightly at the expense of increasing γ^* slightly.

Lemma J.2. *Let $\gamma^* > 0$, $\beta \in (-1, \infty) \setminus \{0\}$, and $\varepsilon > 0$. Let $f(t)$ be a function on $(0, 1)$. If $\gamma^* f(t) \geq F(\beta, t)$ $\forall t \in (0, 1)$ then there exists $\delta > 0$ such that $(\gamma^* + \varepsilon) f(t) \geq F(\beta(1 + \delta)^2, t)$ $\forall t \in (0, 1)$.*

Proof. We have

$$(\gamma^* + \varepsilon) f(t) \geq \frac{\gamma^*+\varepsilon}{\gamma^*} F(\beta, t)$$

so it is sufficient to show

$$\frac{F(\beta(1+\delta)^2, t)}{F(\beta, t)} \leq \frac{\gamma^*+\varepsilon}{\gamma^*} \quad \forall t \in (0, 1],$$

or equivalently,

$$\log F(\beta(1+\delta)^2, t) - \log F(\beta, t) \leq \log\left(\frac{\gamma^*+\varepsilon}{\gamma^*}\right).$$

It is sufficient to have, for any fixed compact interval $\mathcal{I} \subseteq (-1, \infty)$ not containing zero, that $|\frac{\partial}{\partial \beta} \log F(\beta, t)|$ is bounded by a constant, uniformly over all $t \in (0, 1]$ and $\beta \in \mathcal{I}$. Since $\frac{\partial}{\partial \beta} \log F(\beta, t)$ is defined and continuous in both variables (on the domain $t \in (0, 1]$ and $\beta > -1$), we only need to check the limit $t \to 0$. We have $\lim_{t\to 0} \frac{\partial}{\partial \beta} \log F(\beta, t) = 2/\beta$.

□

K Proof of Lemma 5.13

Here we show how to use the local Chernoff bound to bound the small deviations of the Wishart second moment. Letting $\widehat{\gamma} = n/N$ so that $\widehat{\gamma} \to \gamma$ we have

$$\begin{aligned} S(\varepsilon) &= \mathop{\mathbb{E}}_{x,x'\sim\mathcal{X}} \exp\left(\frac{-n}{2\widehat{\gamma}}\log(1-\beta^2\langle x,x'\rangle^2)\right) \mathbb{1}_{\langle x,x'\rangle^2\le\varepsilon} \\ &\le \mathop{\mathbb{E}}_{x,x'\sim\mathcal{X}} \exp\left(\frac{-n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\langle x,x'\rangle^2\right) \mathbb{1}_{\langle x,x'\rangle^2\le\varepsilon} \end{aligned}$$

where we used the convexity of $t \mapsto -\log(1-\beta^2 t)$

$$\begin{aligned} &= \int_0^\infty \Pr\left[\exp\left(\frac{-n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\langle x,x'\rangle^2\right)\mathbb{1}_{\langle x,x'\rangle^2\le\varepsilon} \ge u\right]\mathrm{d}u \\ &= \int_0^\infty \Pr\left[\langle x,x'\rangle^2 \le \varepsilon \quad \text{and} \quad \exp\left(\frac{-n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\langle x,x'\rangle^2\right) \ge u\right]\mathrm{d}u \\ &= \int_0^\infty \Pr\left[\langle x,x'\rangle^2 \le \varepsilon \text{ and } \langle x,x'\rangle^2 \ge t\right]\frac{-n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\exp\left(-\frac{n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)t\right)\mathrm{d}t \\ &= \int_0^\varepsilon \Pr\left[\langle x,x'\rangle^2 \ge t\right]\frac{-n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\exp\left(-\frac{n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)t\right)\mathrm{d}t \end{aligned}$$

where t is defined by $\exp(-\frac{n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)t) = u$. If ε is sufficiently small we can apply the local Chernoff bound:

$$\le \int_0^\varepsilon \frac{-Cn}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)\exp\left(-nf_{\mathcal{X}}(\sqrt{t}) - \frac{n}{2\widehat{\gamma}\varepsilon^2}\log(1-\varepsilon^2\beta^2)t\right)\mathrm{d}t.$$

Using the identity $\int_0^\infty n\exp(-n\alpha t)\,\mathrm{d}t = 1/\alpha$ (for $\alpha > 0$), the above is bounded provided we have $\varepsilon > 0$ and $\alpha > 0$ such that

$$f_{\mathcal{X}}(\sqrt{t}) \ge -\frac{1}{2\gamma\varepsilon^2}\log(1-\varepsilon^2\beta^2)t + \alpha t \quad \forall t \in [0,\varepsilon).$$

Using the bound $\log t \ge 1 - 1/t$ we have $-\frac{1}{\varepsilon^2}\log(1-\varepsilon^2\beta^2) \le \frac{\beta^2}{1-\varepsilon^2\beta^2}$ and so it is sufficient to show $f_{\mathcal{X}}(\sqrt{t}) \ge \left(\frac{\beta^2}{2\gamma}+\eta\right)t$ for all $t \le \varepsilon$, for some $\eta > 0$. But this can be derived from (14) as follows. With $\gamma > \gamma^*$ we have $\gamma^* f_{\mathcal{X}}(t) \ge F(\beta,t)$ for all $t \in (1,0)$. Rewrite this as $f_{\mathcal{X}}(\sqrt{t}) \ge F(\beta,\sqrt{t})/\gamma^*$ and compute $\lim_{t\to 0}\frac{\partial}{\partial t}F(\beta,\sqrt{t})/\gamma^* = \beta^2/2\gamma^* > \beta^2/2\gamma$.

L Comparison of priors and general case of Wishart lower bound

In the main text we have proven Theorem 5.7 in the special case that $\mathcal{X}$ is supported on unit vectors. Here we extend the proof to the general case where $\|x\| \to 1$ in probability. The same argument also yields a result for comparison of priors (Proposition L.1), similar to Proposition 3.13 for the Gaussian Wigner model.

Suppose that $\mathcal{X}, \beta, \gamma^*$ satisfy the assumptions of Theorem 5.7 and let $\gamma > \gamma^*$. Our goal is to show $\mathrm{Wish}(\gamma,\beta,\mathcal{X}) \lhd \mathrm{Wish}(\gamma)$. Let $\delta > 0$ and let $\tilde{\mathcal{X}}$ be the conditional distribution of $x \sim \mathcal{X}$ given $1-\delta \le \|x\| \le 1+\delta$. Let $M(\gamma,\beta,\tilde{\mathcal{X}})$ denote the conditional Wishart second moment defined in Section 5.6. We will show that for δ small enough, $M(\gamma,\beta,\tilde{\mathcal{X}})$ is bounded, implying the desired result (via Lemma 2.4).

Let $\overline{\mathcal{X}}$ be the distribution of $\bar{x} \triangleq \tilde{x}/\|\tilde{x}\|$ with $\tilde{x} \sim \tilde{\mathcal{X}}$. The idea of the proof is to show that the assumptions of Theorem 5.7 are satisfied for $\overline{\mathcal{X}}$ so that we can apply the basic ($\|x\| = 1$) version of the theorem (which we have already proven). Note that $f_{\overline{\mathcal{X}}}(t) \triangleq f_{\mathcal{X}}(t(1-\delta)^2)$ is a valid rate function for $\overline{\mathcal{X}}$. This follows from

$$\Pr[|\langle\bar{x},\bar{x}'\rangle| \ge t] \le \Pr[|\langle\tilde{x},\tilde{x}'\rangle| \ge t(1-\delta)^2] \le c\cdot\Pr[|\langle x,x'\rangle| \ge t(1-\delta)^2]$$

where $c = 1 + o(1)$. We can take the lower bound (in Definition 5.4) to be $b_{n,\overline{\mathcal{X}}} = -\frac{1}{n}\log c + b_{n,\mathcal{X}}(t(1-\delta)^2)$. If $f_{\mathcal{X}}$ admits a local Chernoff bound (condition (ii) of Theorem 5.7) then so does $f_{\overline{\mathcal{X}}}$.

As in the proof for the $\|x\| = 1$ case, we treat the small and large deviations separately. The parameter α that separates the small ($|\alpha| \in [0, \varepsilon]$) and large ($|\alpha| \in (\varepsilon, 1]$) deviations is now defined with normalization: $\alpha \triangleq \langle x, x' \rangle / (\|x\| \cdot \|x'\|)$.

L.1 Small deviations

We have

$$\mathop{\mathbb{E}}_{\tilde{x},\tilde{x}'\sim\tilde{\mathcal{X}}}(1-\beta^2\langle\tilde{x},\tilde{x}'\rangle^2)^{-N/2}\mathbb{1}_{\langle\tilde{x},\tilde{x}'\rangle^2/(\|\tilde{x}\|\cdot\|\tilde{x}'\|)^2\le\varepsilon} \le \mathop{\mathbb{E}}_{\bar{x},\bar{x}'\sim\overline{\mathcal{X}}}(1-\beta^2(1+\delta)^4\langle\bar{x},\bar{x}'\rangle^2)^{-N/2}\mathbb{1}_{\langle\bar{x},\bar{x}'\rangle^2\le\varepsilon},$$

i.e. the small deviations of $M(\gamma, \beta, \tilde{\mathcal{X}})$ are bounded by the small deviations of $M(\gamma, \beta(1+\delta)^2, \overline{\mathcal{X}})$. Therefore it is sufficient to verify the conditions of Theorem 5.7 for $\gamma, \beta(1+\delta)^2, \overline{\mathcal{X}}$.

First we show that if condition (i) ($\beta^2/\gamma^* \le (\lambda^*_{\mathcal{X}})^2$) in Theorem 5.7 was satisfied for $\gamma, \beta, \mathcal{X}$ then it is still satisfied for $\gamma, \beta(1+\delta)^2, \overline{\mathcal{X}}$ (provided we allow an arbitrarily-small increase in γ^*). Since conditioning on a $(1-o(1))$-probability event can only increase the Wigner second moment by a $(1 + o(1))$ factor, we have $\lambda^*_{\tilde{\mathcal{X}}} \ge \lambda^*_{\mathcal{X}}$. We also have

$$\mathop{\mathbb{E}}_{\bar{x},\bar{x}'\sim\overline{\mathcal{X}}}\exp\left(\frac{n\lambda^2}{2}\langle\bar{x},\bar{x}'\rangle^2\right) \le \mathop{\mathbb{E}}_{\tilde{x},\tilde{x}'\sim\tilde{\mathcal{X}}}\exp\left(\frac{n\lambda^2}{2(1-\delta)^2}\langle\tilde{x},\tilde{x}'\rangle^2\right)$$

and so $\lambda^*_{\overline{\mathcal{X}}} \ge (1-\delta)\lambda^*_{\tilde{\mathcal{X}}} \ge (1-\delta)\lambda^*_{\mathcal{X}}$. Therefore by choosing δ small enough we can find $\overline{\gamma}^*$ with $\gamma^* < \overline{\gamma}^* < \gamma$ such that $\beta^2(1+\delta)^4/\overline{\gamma}^* \le (\lambda^*_{\overline{\mathcal{X}}})^2$ as desired.

Now we check that (14) is satisfied for $\gamma, \beta(1+\delta)^2, \overline{\mathcal{X}}$. We are guaranteed $\gamma > \gamma^*$ with

$$\gamma^* f_{\mathcal{X}}(t) \ge F(\beta, t) \quad \forall t \in (0,1). \tag{16}$$

Our goal is to show (for sufficiently small δ) $\gamma > \overline{\gamma}^*$ with

$$\overline{\gamma}^* f_{\overline{\mathcal{X}}}(t) \ge F(\beta(1+\delta)^2, t) \quad \forall t \in (0,1). \tag{17}$$

The proof of (17) follows from (16) by Lemmas J.1 and J.2. The first allows us to replace $f_{\mathcal{X}}$ by $f_{\overline{\mathcal{X}}}$ and the second allows us to increase β to $\beta(1+\delta)^2$. Each of these changes comes at the expensive of increasing γ^* (to $\overline{\gamma}^*$) by an arbitrarily-small amount (which can be done such that $\gamma > \overline{\gamma}^*$).

L.2 Large deviations

We now consider the contribution to $M(\gamma, \beta, \tilde{\mathcal{X}})$ from $|\alpha| \in [\varepsilon, 1-\varepsilon]$. The contribution from $|\alpha| \in (1-\varepsilon, 1]$ can be handled similarly. We have

$$\mathop{\mathbb{E}}_{\tilde{x},\tilde{x}'\sim\tilde{\mathcal{X}}}[\mathbb{1}_{|\alpha|\in[\varepsilon,1-\varepsilon]}\,\tilde{m}(\tilde{x},\tilde{x}')]$$

where

$$\begin{aligned}\tilde{m}(\tilde{x},\tilde{x}') &\triangleq \mathop{\mathbb{E}}_{Y\sim P_n}(1+\beta\|\tilde{x}\|^2)^{-N/2}(1+\beta\|\tilde{x}'\|^2)^{-N/2}\exp\left(\frac{N}{2}\left(\frac{\beta}{1+\beta\|\tilde{x}\|^2}\tilde{x}^\top Y\tilde{x} + \frac{\beta}{1+\beta\|\tilde{x}'\|^2}\tilde{x}'^\top Y\tilde{x}'\right)\right)\mathbb{1}_{\Omega(\tilde{x},Y)}\mathbb{1}_{\Omega(\tilde{x}',Y)} \\ &\le \mathop{\mathbb{E}}_{Y\sim P_n}(1+\beta(1-\delta)^2)^{-N}\exp\left(\frac{N}{2}\left(\frac{\beta\|\tilde{x}\|^2}{1+\beta\|\tilde{x}\|^2}\bar{x}^\top Y\bar{x} + \frac{\beta\|\tilde{x}'\|^2}{1+\beta\|\tilde{x}'\|^2}\bar{x}'^\top Y\bar{x}'\right)\right)\mathbb{1}_{\Omega(\tilde{x},Y)}\mathbb{1}_{\Omega(\tilde{x}',Y)}\end{aligned}$$

where $\bar{x} = \tilde{x}/\|\tilde{x}\|$ and $\bar{x}' = \tilde{x}'/\|\tilde{x}'\|$. Note that $\Omega(\tilde{x}, Y)$ can be written as $\bar{x}^\top Y\bar{x} \in [(1+\beta\|\tilde{x}\|^2)(1-\delta), (1+\beta\|\tilde{x}\|^2)(1+\delta)]$. We can upper bound the resulting expression by replacing each instance of $\|\tilde{x}\|^2$ by either $1+\delta$ or $1-\delta$. Since only $\bar{x}, \bar{x}'$ (and not $\tilde{x}, \tilde{x}'$) now appear, we have reduced to the original case of the proof (since $\|\bar{x}\| = \|\bar{x}'\| = 1$) but with the β's replaced by slightly different constants; carrying through the proof as before yields the sufficient condition $\gamma > \bar{\gamma}^*$ with $\bar{\gamma}^* f_{\overline{\mathcal{X}}}(t) \ge F_\delta(\beta, t)$ $\forall t \in [\varepsilon, 1-\varepsilon]$ where for each t, $F_\delta(\beta,t) \to F(\beta,t)$ as $\delta \to 0^+$. Since F, F_δ are continuous and $[\varepsilon, 1-\varepsilon]$ is compact, the convergence $F_\delta(\beta,t) \to F(\beta,t)$ is uniform over $t \in [\varepsilon, 1-\varepsilon]$. Let $\gamma^* < \widehat{\gamma}^* < \bar{\gamma}^* < \gamma$. $F(\beta,t)$ is positive and increasing in t for $t \in [\varepsilon, 1-\varepsilon]$ (see Appendix J), so provided δ is small enough, it is sufficient to show $\widehat{\gamma}^* f_{\overline{\mathcal{X}}}(t) \ge F(\beta,t)$ $\forall t \in [\varepsilon, 1-\varepsilon]$. This follows from the assumption $\gamma^* f_{\mathcal{X}}(t) \ge F(\beta,t)$ along with Lemma J.1. The proof of Theorem 5.7 in full generality is now complete.

L.3 Comparison of similar priors

The same argument used above implies the following which may be of independent interest.

Proposition L.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be spike priors. Suppose that $x \sim \mathcal{X}_n$ and $y \sim \mathcal{Y}_n$ can be coupled such that $y = \alpha x$ where $\alpha = \alpha_n$ is a random variable with $\alpha_n \to 1$ in probability as $n \to \infty$. Suppose that the conditions of Theorem 5.7 are satisfied for $\mathcal{X}, \beta, \gamma^*$. Then for any $\gamma > \gamma^*$, $\mathrm{Wish}(\beta, \gamma, \mathcal{Y}) \lhd \mathrm{Wish}(0)$.*

Proof. The proof is similar to the arguments above so we only give a sketch. We define modified priors $\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \overline{\mathcal{X}}, \overline{\mathcal{Y}}$ as above and note that $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ are the same. Since the conditions of Theorem 5.7 are satisfied for $\mathcal{X}$, they are also satisfied for $\overline{\mathcal{X}}$ at the expense of an arbitrarily-small increase in γ^*. We can then control the Wishart conditional second moment $M(\gamma, \beta, \tilde{\mathcal{Y}})$ by comparison to $\overline{\mathcal{X}}$ (i.e. $\overline{\mathcal{Y}}$). □

M Monotonicity of Wishart lower bound

In this section we prove various properties of the condition (14), implying certain monotonicity properties of the Wishart lower bound (Theorem 5.7). Informally speaking, we will show the following.

- If the PCA threshold is optimal for some $\bar{\beta} \in (-1, \infty) \setminus \{0\}$, it is also optimal for all $\beta > \bar{\beta}$ (Proposition 5.8).
- If the PCA threshold is optimal for the Wigner model, it is also optimal for the positively-spiked ($\beta > 0$) Wishart model (Corollary 5.9). Conversely, if PCA is optimal for Wishart for all $\beta > 0$ then it is optimal for Wigner (Proposition M.2).
- For any reasonable i.i.d. prior, if β is sufficiently large then the PCA threshold is optimal (Proposition 5.10).

The statements above are informal; the true results we prove are of the form e.g. "if *our methods* show a Wigner lower bound then they also show a Wishart lower bound."

Proposition 5.8. *Let $\mathcal{X}$ be a spike prior. Fix $\lambda > 0$ and $\bar{\beta} \in (-1, \infty) \setminus \{0\}$. If (14) holds for $\bar{\beta}$ and $\gamma^* = \bar{\beta}^2/\lambda^2$ then it also holds for any $\beta > \bar{\beta}$ and $\gamma^* = \beta^2/\lambda^2$.*

Proof. The condition (14) takes the form $\gamma^* f_{\mathcal{X}}(t) \ge F(\beta,t)$ $\forall t \in (0,1)$. With λ fixed and $\gamma^* = \beta^2/\lambda^2$, this is equivalent to $f_{\mathcal{X}}(t) \ge \lambda^2 F(\beta,t)/\beta^2$ $\forall t \in (0,1)$. It is therefore sufficient to show the following lemma.

Lemma M.1. *For any fixed $t \in (0,1)$, $F(\beta,t)/\beta^2$ is a decreasing function of β on the domain $\beta \in (-1,\infty)$. (When $\beta = 0$ we define $F(\beta,t)/\beta^2$ by its limit value.)*

Proof. We wish to show that $\frac{\mathrm{d}}{\mathrm{d}\beta} \frac{F(\beta,t)}{\beta^2} < 0$. One computes that $\lim_{t \to 0} \frac{\mathrm{d}}{\mathrm{d}\beta} \frac{F(\beta,t)}{\beta^2} = 0$, so it suffices to show that $\frac{\partial^2}{\partial\beta\,\partial t} \frac{F(\beta,t)}{\beta^2} \le 0$ for $0 < t < 1$. We compute:

$$\frac{\partial^2}{\partial\beta\,\partial t} \frac{F(\beta,t)}{\beta^2} = \frac{U_1 - U_2}{\beta^3\, S\, t(1-t^2)^2}, \quad \text{where}$$

$S = \sqrt{1 + t^2(2 + 8\beta + 4\beta^2) + t^4}$, $\quad U_1 = S(1+2(1+\beta)t^2+t^4) \ge 0$, $\quad U_2 = (1+t^2)(1+t^2(2+6\beta+2\beta^2)+t^4) \ge 0.$

The denominator evidently has sign matching β, so it suffices to see that the numerator has sign matching $-\beta$. As $U_1 \geq 0$ and $U_2 \geq 0$, the sign of $U_1 - U_2$ will match that of $U_1^2 - U_2^2$, and we compute:

$$U_1^2 - U_2^2 = -4\beta^3(2+\beta)t^4(1-t^2)^2$$

which has sign matching that of $-\beta$, as desired. □

Corollary 5.9. *Suppose $\langle x, x' \rangle$ is (σ^2/n)-subgaussian, where x and x' are drawn independently from $\mathcal{X}_n$. Then for any $\beta > 0$ and any $\gamma > \beta^2\sigma^2$ we have* $\mathrm{Wish}(\gamma, \beta, \mathcal{X}) \lhd \mathrm{Wish}(\gamma)$.

This connects the Wigner model to the Wishart model because the subgaussian condition above implies a *Wigner* lower bound for all $\lambda < 1/\sigma$ (Proposition 3.8).

Proof. The subgaussian tail bound $\Pr[|\langle x, x' \rangle| \geq t] \leq 2\exp(-nt^2/2\sigma^2)$ implies that we have the rate function $f_{\mathcal{X}}(t) = t^2/2\sigma^2$.

Let $F(\beta, t)$ be defined as in (14). For any fixed $t \in (0, 1)$, we have

$$\lim_{\beta \to 0} \frac{F(\beta, t)}{\beta^2} = \frac{t^2}{2(1+t^2)}. \tag{18}$$

This can be shown by computing the Taylor series of $F(\beta, t)$ at $\beta = 0$. From Lemma M.1 above, we know that $F(\beta, t)/\beta^2$ is a decreasing function of β. Therefore, for any $t \in (0, 1)$ and any $\beta > 0$ we have

$$\frac{F(\beta, t)}{\beta^2} \leq \frac{t^2}{2(1+t^2)} \leq \frac{t^2}{2}.$$

By combining the above results it follows that (14) holds with $\gamma^* = \beta^2\sigma^2$. Proposition 3.8 implies that the Wigner threshold is $\lambda^*_{\mathcal{X}} \geq 1/\sigma$ and so condition (i) of Theorem 5.7 is satisfied. The result now follows from Theorem 5.7.

We remark that (14) would follow from the weaker condition $f_{\mathcal{X}}(t) \geq \frac{t^2}{2\sigma^2(1+t^2)}$ (instead of $f_{\mathcal{X}}(t) \geq \frac{t^2}{2\sigma^2}$). This is exactly the condition for the Wigner lower bound of Perry et al. [2016] with $\lambda^* = 1/\sigma$. □

Proposition M.2. *Fix $\lambda^* > 0$. Suppose that for each $\beta > 0$, the assumptions of Theorem 5.7 are satisfied for $\mathcal{X}$ and $\gamma^* = \beta^2/(\lambda^*)^2$. Then* $\mathrm{GWig}(\lambda, \mathcal{X}) \lhd \mathrm{GWig}(0)$ *for any $\lambda < \lambda^*$.*

Proof. If condition (i) of Theorem 5.7 is satisfied for some β then we are done immediately, so assume condition (ii) holds for all $\beta > 0$. For all $\beta > 0$ and all $t \in (0, 1)$ we have $\gamma^* f_{\mathcal{X}}(t) \geq F(\beta, t)$, i.e. $f_{\mathcal{X}}(t) \geq (\lambda^*)^2 F(\beta, t)/\beta^2$. Using (18) this implies $f_{\mathcal{X}}(t) \geq \frac{(\lambda^*)^2}{2}\frac{t^2}{1+t^2}$ and so the result follows from Perry et al. [2016]. (Although Perry et al. [2016] assume that the spike has exactly unit norm, arguments similar to Appendix L can be used.) □

Proposition 5.10. *Suppose $\mathcal{X} = \mathrm{iid}(\pi/\sqrt{n})$ where π is a mean-zero unit-variance distribution for which $\pi\pi'$ (product of two independent copies of π) has a moment-generating function $M(\theta) \triangleq \mathbb{E}\exp(\theta\pi\pi')$ which is finite on an open interval containing zero. Then there exists $\overline{\beta}$ such that for any $\beta \geq \overline{\beta}$ and any $\gamma > \beta^2$ we have* $\mathrm{Wish}(\gamma, \beta, \mathcal{X}) \lhd \mathrm{Wish}(\gamma)$.

In other words, for sufficiently large β, detection is impossible below the spectral threshold.

Proof. We will show that for sufficiently large β, the assumptions of Theorem 5.7 hold with $\gamma^* = \beta^2$ so that the result follows. The usual Chernoff bound yields

$$\Pr[\langle x, x' \rangle \geq t] \leq \exp[-n(t\theta - \log M(\theta))] \quad \forall \theta \in \mathbb{R}$$

and so, letting $\theta = t$,

$$\Pr[\langle x, x' \rangle \geq t] \leq \exp[-n(t^2 - \log M(t))].$$

Similarly,

$$\Pr[\langle x, x' \rangle \le -t] \le \exp[-n(t^2 - \log M(-t))].$$

Therefore, $f_{\mathcal{X}}(t) \triangleq \min\{f_1(t), f_2(t)\}$ is a valid rate function for $\mathcal{X}$ with a local Chernoff bound, where $f_1(t) \triangleq t^2 - \log M(t)$ and $f_2(t) \triangleq t^2 - \log M(-t)$. It remains to show that for sufficiently large β, (14) holds with $\gamma^* = \beta^2$, i.e. $f_{\mathcal{X}}(t) \ge F(\beta,t)/\beta^2 \; \forall t \in (0,1)$. We will show

$$f_1(t) \ge F(\beta,t)/\beta^2 \quad \forall t \in (0,1) \tag{19}$$

but the proof for f_2 is similar.

First we show that (19) holds for $t \in (0,\varepsilon]$ for some $\varepsilon > 0$. Using the well-known identity $\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} M(\theta)|_{\theta=0} = \mathbb{E}[(\pi\pi')^k]$ we have derivatives $M(0) = 0$, $M'(0) = 0$, $M''(0) = 1$, $|M'''(0)| < \infty$, $|M''''(0)| < \infty$. We can use these to compute $f_1(0) = 0$, $\lim_{t\to 0^+} f_1'(t) = 0$, $\lim_{t\to 0^+} f_1''(t) = 1$, $\lim_{t\to 0^+} f_1'''(t) = 0$, $|\lim_{t\to 0^+} f_1''''(t)| < \infty$. We can also compute $F(\beta,0)/\beta^2 = 0$, $\lim_{t\to 0^+} \frac{\partial}{\partial t} F(\beta,t)/\beta^2 = 0$, $\lim_{t\to 0^+} \frac{\partial^2}{\partial t^2} F(\beta,t)/\beta^2 = 1$, $\lim_{t\to 0^+} \frac{\partial^3}{\partial t^3} F(\beta,t)/\beta^2 = 0$, $\lim_{t\to 0^+} \frac{\partial^4}{\partial t^4} F(\beta,t)/\beta^2 = -6(b^2+4b+2)$. Note that $f_1(t)$ and $F(\beta,t)/\beta^2$ have matching derivatives (at $t = 0$) up to third order and that the fourth derivative of $F(\beta,t)/\beta^2$ goes to $-\infty$ as $\beta \to \infty$. Therefore we can find β and $\varepsilon > 0$ such that $f_1(t) \ge F(\beta,t)/\beta^2$ for all $t \in (0,\varepsilon]$. Since $F(\beta,t)/\beta^2$ is a decreasing function of β (Lemma M.1), this remains true for any larger β.

Now we show that (19) holds for $t \in (\varepsilon, 1)$. For any $t \in (\varepsilon,1)$ we have $f_1(t) \ge f_1(\varepsilon) > 0$ (using the derivatives above and the fact that rate functions are increasing). Also, for any $t \in [0,1]$ we have $\lim_{\beta\to\infty} F(\beta,t)/\beta^2 = 0$ and by compactness this convergence is uniform over t. Therefore if β is sufficiently large, (19) holds for $t \in (\varepsilon,1)$, completing the proof.

□

N Wishart results for specific priors

N.1 Spherical prior

Our lower bound for the spherical prior is obtained by combining Theorem 5.7 with the rate function of Proposition 5.5, along with the fact that $\lambda^*_{\mathcal{X}} = 1$ for the spherical prior (Corollary 3.14). The result is that the PCA threshold is optimal (i.e. we have contiguity for all $\gamma > \beta^2$) for all $\beta \in (-1,\infty)$. To show this, we need to check (14) with $\gamma^* = \beta^2$. This follows from $\lim_{\beta\to -1^+} F(\beta,t)/\beta^2 = -\frac{1}{2}\log(1-t^2)$ (which is precisely the spherical rate function of Proposition 5.5) along with the fact that $F(\beta,t)/\beta^2$ is a decreasing function of β (Lemma M.1).

N.2 Rademacher prior

Our lower bound for the Rademacher prior is obtained by combining Theorem 5.7 with the rate function of Proposition 5.5, along with the fact that $\lambda^*_{\mathcal{X}} = 1$ for the Rademacher prior (Corollary 3.12). We also obtain an upper bound from Theorem 5.3, taking $c = 2$.

N.3 Sparse Rademacher prior

First consider the variant of the sparse Rademacher prior where the spike has exactly ρn nonzero entries, which are i.i.d. $\pm 1/\sqrt{\rho n}$ (and we restrict to n for which ρn is an integer). In this case the rate function f_ρ stated in Proposition 5.5 has been proven to be valid, and furthermore to admit a local Chernoff bound [Perry et al., 2016]. This yields a lower bound via Theorem 5.7. We show that the same lower bound holds for the i.i.d. sparse Rademacher prior:

Proposition N.1. *Let $\mathcal{X}_\rho$ be the i.i.d. sparse Rademacher prior with sparsity ρ (as defined in Section 3.7). Let f_ρ be the rate function defined in Proposition 5.5. Let $F(\beta,t)$ be defined as in (14). If*

$$\gamma^* f_\rho(t) \ge F(\beta,t) \quad \forall t \in (0,1) \tag{20}$$

then $\mathrm{Wish}(\gamma, \beta, \mathcal{X}_\rho) \lhd \mathrm{Wish}(\gamma)$ *for all* $\gamma > \gamma^*$

In proving this we will not quite show that f_ρ is a rate function for $\mathcal{X}_\rho$ (but we will show that something arbitrarily-close is).

Proof. First we prove the result for the variant $\overline{\mathcal{X}}_\rho$ of the sparse Rademacher prior where the number K of nonzeros satisfies $K/n \to \rho$ in probability, and the nonzero entries are i.i.d. $\pm 1/\sqrt{K}$. Suppose we have some γ^*, β, ρ for which (20) holds. At the expense of increasing γ^* by an arbitrarily-small constant, we can find a small compact interval $[\rho_1, \rho_2]$ with ρ in its interior such that (20) holds on the entire interval. Condition on the $(1-o(1))$-probability event that K/n lies in this interval. The function $f(t) = \min_{\widehat{\rho} \in [\rho_1, \rho_2]} f_{\widehat{\rho}}(t)$ is a valid rate function for $\overline{\mathcal{X}}_\rho$ and furthermore has a local Chernoff bound. This follows from the sparse Rademacher tail bounds of Perry et al. [2016] (Propositions 4.8 and 4.9 of Perry et al. [2016]). The proof for $\overline{\mathcal{X}}$ now follows from Theorem 5.7 because f satisfies (14). The same lower bound holds for $\mathcal{X}_\rho$ by Proposition L.1 (comparison of priors).

□

For the upper bound, we will apply Theorem 5.3. However, instead of $\mathcal{X}_\rho$ we consider the conditional distribution $\tilde{\mathcal{X}}_\rho$ of $\mathcal{X}_\rho$ given the $(1-o(1))$-probability event that the number K of nonzero entries of x satisfies $\rho n - \sqrt{n} \log n < K < \rho n + \sqrt{n} \log n$. The support size of $\tilde{\mathcal{X}}_\rho$ is at most

$$2\sqrt{n} \log n \cdot 2^{(\rho + o(1))n} \binom{n}{(\rho \pm o(1))n}.$$

By Stirling's approximation, $\log \binom{n}{\rho n} = nH(\rho) + o(n)$ where $H(\rho) = -\rho \log \rho - (1-\rho) \log(1-\rho)$ is the binary entropy. We can therefore apply Theorem 5.3 with any $c > 2^\rho \exp(H(\rho))$.

References

D. Achlioptas and A. Naor. The two possible values of the chromatic number of a random graph. In *Proceedings of the thirty-sixth annual ACM Symposium on the Theory of Computing*, pages 587–593. ACM, 2004.

J. Banks, C. Moore, J. Neeman, and P. Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *29th Annual Conference on Learning Theory*, pages 383–416, june 2016.

DLMF. NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.11 of 2016-06-08. URL `http://dlmf.nist.gov/`. Online companion to Olver et al. [2010].

S. Elder. Bayesian adaptive data analysis guarantees from subgaussianity. *arXiv preprint arXiv:1611.00065*, 2016.

I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *arXiv:1509.07269*, 2015.

F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. Print companion to DLMF.

A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231, 2013.

A. Perry, A. S. Wein, and A. S. Bandeira. Statistical limits of spiked tensor models. *arXiv:1612.07728*, 2016.

A. Pizzo, D. Renfrew, and A. Soshnikov. On finite rank deformations of Wigner matrices. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 49, pages 64–94. Institut Henri Poincaré, 2013.