Random Matrix Theory for Signal–Noise Decomposition

Ishan Ganguly

December 16, 2025

Abstract

This report studies the application of random matrix theory to signal–noise decomposition in two settings. In the first, a low-rank matrix corrupted by additive Gaussian noise is observed, and asymptotic results for the singular values and vectors of spiked random matrices are used to justify data-driven rules for singular value thresholding and shrinkage, following work of Gavish and Donoho. In the second, a sparse Fourier signal is observed with additive Gaussian noise at a subset of time points, and its periodogram and induced autocovariance structure are compared to the classical Marchenko–Pastur and Tracy–Widom picture. In both settings, MSE-optimal truncation and shrinkage rules can be complemented by hypothesis-testing viewpoints in which the complexity of the signal model is increased until the remaining residue is statistically indistinguishable from noise.

1 Introduction

Many high-dimensional data sets can be decomposed into a low-complexity signal component and an unstructured noise component. A simple and widely-studied model for this situation is

$$Y = X + \sigma Z, \tag{1}$$

where $Y \in \mathbb{R}^{n\times n}$ is observed, X is an unknown low-rank signal matrix, and Z has independent standard Gaussian entries. The parameter σ^2 denotes the noise variance. The goal is to recover X from Y as accurately as possible.

A natural class of estimators $\hat{X}$ is obtained by manipulating the singular values of Y. Writing the singular value decomposition (SVD) as

$$Y = U\Sigma V^\top = \sum_{i=1}^{n} s_i\, u_i v_i^\top,$$

one may construct $\hat{X}$ by either discarding singular values below a threshold or shrinking them towards zero while keeping their associated singular vectors.

Random matrix theory (RMT) provides detailed information about the spectrum of Y when $X = 0$, and about how a finite-rank perturbation X affects the leading singular values and vectors. These results allow construction of asymptotically optimal thresholding and shrinkage procedures, and also calibration of hypothesis tests based on the largest singular value via the Tracy–Widom law.

Two types of spectral procedures are emphasized in this report. Sections 2–3 focus on MSE-optimal hard thresholding and singular value shrinkage under an additive Gaussian model, following Gavish and Donoho [4]. Sections 4 and 5 then reinterpret the same experiments from

a hypothesis-testing perspective: after estimating the noise level, the largest remaining singular value (or periodogram peak) is compared to its noise-only distribution (Tracy–Widom in the matrix case, exponential / extreme-value in the Fourier case) to decide whether additional components should be included.

2 Gaussian matrix model and spectral behavior

2.1 Model and basic assumptions

Let us consider the case where $Y \in \mathbb{R}^{n\times n}$ is a square matrix, with the unobserved decomposition

$$Y = X + \sigma Z,$$

where

- X is deterministic (or independent of Z) and has rank $r \ll n$;
- Z has i.i.d. entries $Z_{ij} \sim \mathcal{N}(0,1)$;
- σ^2 is the noise variance.

The SVD of Y is written as

$$Y = U\Sigma V^\top,$$

with singular values $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ and orthonormal singular vectors $u_i, v_i \in \mathbb{R}^n$.

When $X = 0$, the normalized singular values of Y concentrate around a deterministic limiting distribution as $n \to \infty$. These limiting laws give a reference against which the singular values of Y can be compared when $X \neq 0$.

2.2 Quarter-circle law for the noise-only case

Consider the noise-only matrix

$$Y_0 = \sigma Z.$$

Define the normalized singular values

$$\tilde{s}_i = \frac{s_i(Y_0)}{\sigma\sqrt{n}}.$$

When n is large, the empirical distribution of $\{\tilde{s}_i\}_{i=1}^n$ converges almost surely to the quarter-circle law supported on $[0,2]$:

$$\rho_{\mathrm{QC}}(s) = \frac{1}{\pi}\sqrt{4-s^2}, \qquad 0 \leq s \leq 2.$$

Equivalently, the empirical distribution of the squared singular values $s_i^2(Y_0)/(\sigma^2 n)$ converges to the Marchenko–Pastur law with aspect ratio one.

In particular, the bulk of the singular values of Y_0 lies in an interval of the form

$$s_i(Y_0) \approx \sigma\sqrt{n}\, s, \qquad s \in [0,2].$$

The largest singular value concentrates near $2\sigma\sqrt{n}$ up to fluctuations of order $n^{1/6}$, which are described by the Tracy–Widom law.

Figure 1 shows a simulated histogram of singular values of a Gaussian matrix, normalized by $\sigma\sqrt{n}$, together with the quarter-circle density.

This picture suggests that, in the presence of a low-rank signal X, singular values significantly above $2\sigma\sqrt{n}$ are unlikely to have been generated by noise alone and are natural candidates to be treated as signal.

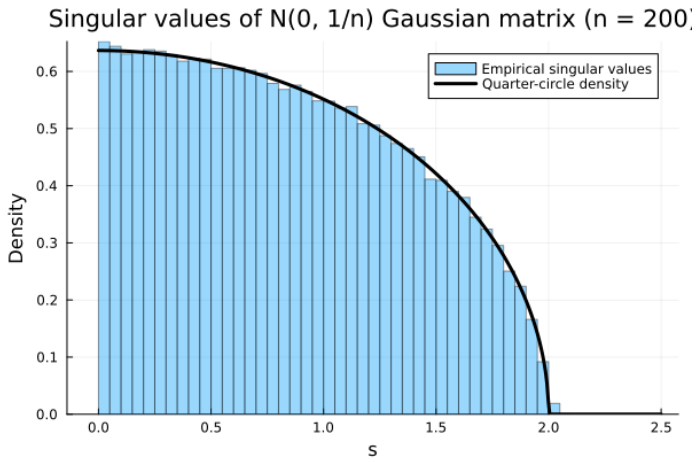


Figure 1: Empirical singular value distribution of a Gaussian noise-only matrix, normalized by $\sigma\sqrt{n}$, overlaid with the quarter-circle density.

2.3 Rank-one spike: top singular value and alignment

To make this idea more precise, consider a rank-one signal

$$X = \theta\, aa^\top,$$

where $a \in \mathbb{R}^n$ is a unit vector and $\theta > 0$ is the signal strength. The leading singular value and corresponding singular vector of

$$Y = X + \sigma Z$$

are studied in this subsection. For concreteness, take $\sigma = 1$ and focus on the square case.

Let s_1 denote the largest singular value of Y and u_1 the corresponding left singular vector. Random matrix theory gives asymptotic formulas describing how s_1 and u_1 are distorted relative to the signal parameters (θ, a).

One way to analyze s_1 is to consider the characteristic equation

$$\det(s_1 I - Y) = 0.$$

Writing $Y = Z + \theta aa^\top$ and assuming $s_1 I - Z$ is nonsingular (which holds almost surely),

$$\begin{aligned} 0 &= \det(s_1 I - (Z + \theta aa^\top)) \\ &= \det(s_1 I - Z)\, \det\big(I + \theta (s_1 I - Z)^{-1} aa^\top\big). \end{aligned}$$

By the matrix determinant lemma,

$$\det\big(I + \theta (s_1 I - Z)^{-1} aa^\top\big) = 1 + \theta\, a^\top (s_1 I - Z)^{-1} a.$$

Thus s_1 satisfies the fixed-point equation

$$1 + \theta\, a^\top (s_1 I - Z)^{-1} a = 0. \tag{2}$$

Let $\{(\lambda_i, v_i)\}_{i=1}^n$ be the eigenpairs of Z. Then

$$a^\top (s_1 I - Z)^{-1} a = \sum_{i=1}^n \frac{1}{s_1 - \lambda_i} \langle a, v_i \rangle^2.$$

Under the independence and isotropy assumptions, the inner products $\langle a, v_i \rangle$ behave like i.i.d. $\mathcal{N}(0, 1/n)$ random variables, so that $\langle a, v_i \rangle^2 \to 1/n$ in expectation. As $n \to \infty$, the sum in (2) can be approximated by the trace of the resolvent,

$$a^\top (s_1 I - Z)^{-1} a \approx \frac{1}{n} \operatorname{Tr}\big((s_1 I - Z)^{-1}\big),$$

whose limit is the Stieltjes transform of the semicircle law. Evaluating this transform yields the asymptotic relation

$$s_1 \approx \theta + \frac{1}{\theta}$$

for $\theta > 1$, while for $\theta \leq 1$ the spike does not separate from the noise bulk and s_1 remains close to the upper edge 2 of the quarter-circle law.

The corresponding alignment between the spike direction a and the top singular vector u_1 can be analyzed using the resolvent representation of the spectral projector:

$$u_1 u_1^\top = -\frac{1}{2\pi i} \oint_\Gamma (zI - Y)^{-1} \, dz,$$

where Γ is a contour enclosing s_1. Taking the quadratic form with a gives

$$\langle a, u_1 \rangle^2 = -\frac{1}{2\pi i} \oint_\Gamma a^\top (zI - Y)^{-1} a \, dz.$$

Using a rank-one resolvent identity and the limiting spectral distribution of Z, one obtains

$$\langle a, u_1 \rangle^2 \approx 1 - \frac{1}{\theta^2}$$

for $\theta > 1$; see Benaych-Georges and Nadakuditi [3] for precise statements in the rectangular case.

These formulas illustrate two key phenomena:

- *Spectral separation:* For $\theta > 1$, the top singular value separates from the noise bulk and behaves approximately like $\theta + 1/\theta$.
- *Vector alignment:* The leading singular vector u_1 becomes increasingly well-aligned with the signal direction a as θ increases, with squared cosine $1 - 1/\theta^2$.

Figure 2 shows a Monte Carlo illustration of these effects for square matrices.

Both spectral separation and vector alignment are essential for interpreting and calibrating singular value thresholding rules.

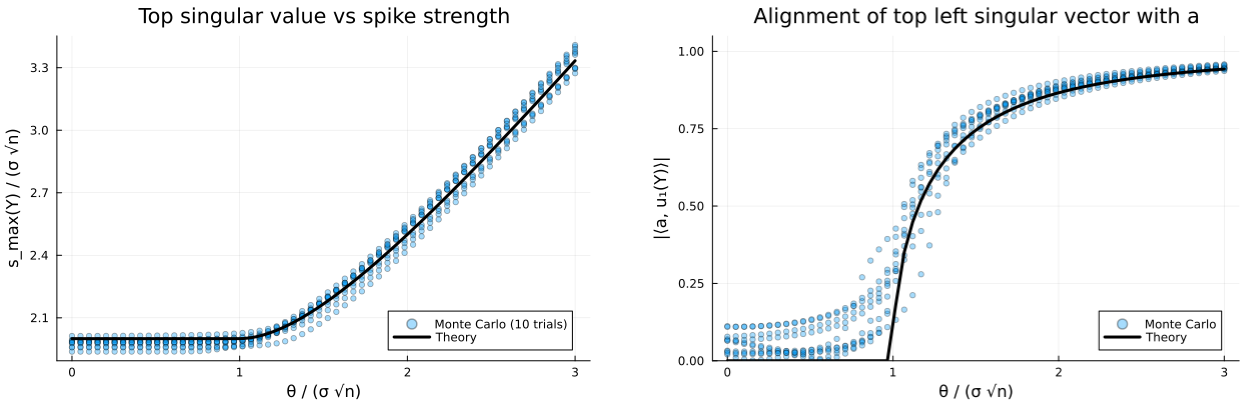
3 Hard thresholding and optimal shrinkage

3.1 Loss-based formulation

This section addresses how to partition the singular components of Y into signal and noise, assuming the square Gaussian model and the asymptotic distortion formulas described above.

Suppose the true signal matrix has rank r and admits a singular value decomposition

$$X = \sum_{i=1}^{r} \theta_i \, a_i b_i^\top.$$



(a) Top singular value vs. spike strength.

(b) Alignment $|\langle a, u_1 \rangle|$ vs. spike strength.

Figure 2: Simulation of the largest singular value and alignment in the rank-one spiked model $Y = \theta a a^\top + Z$ for square matrices.

The observed matrix admits

$$Y = \sum_{i=1}^{n} s_i \, u_i v_i^\top,$$

and the task is to estimate X by either including or discarding each observed singular triplet (u_i, s_i, v_i).

Following Gavish and Donoho [1], consider the Frobenius risk

$$\|\hat{X} - X\|_F^2$$

and compare two possibilities for a given index i:

- If the i-th component is excluded, the loss is

$$\ell_- = \left\|0 - \theta_i a_i b_i^\top\right\|_F^2 = \theta_i^2.$$

- If the noisy component $s_i u_i v_i^\top$ is included, the loss is

$$\ell_+ = \left\|s_i u_i v_i^\top - \theta_i a_i b_i^\top\right\|_F^2.$$

Inclusion is preferred when $\ell_+ < \ell_-$ and exclusion otherwise.

In the square case, for large n, assume

$$s_i \approx \theta_i + \frac{1}{\theta_i}, \qquad \langle a_i, u_i \rangle^2 \approx 1 - \frac{1}{\theta_i^2},$$

and similarly for b_i, v_i. A short calculation under these approximations leads to a critical signal strength $\theta_i = \sqrt{3}$ at which $\ell_+ = \ell_-$. The corresponding singular value threshold is

$$s_i = \theta_i + \frac{1}{\theta_i} = \frac{4}{\sqrt{3}}.$$

Reintroducing the noise variance and dimensional scaling yields the asymptotically optimal hard threshold

$$s_i \geq \frac{4}{\sqrt{3}} \, \hat{\sigma} \sqrt{n}, \tag{3}$$

where $\hat{\sigma}$ is an estimate of the noise standard deviation.

This analysis assumes the true signal is low-rank, which is not often the case in many real-life examples. Nevertheless, these results remain asymptotically optimal when the noise level is known. This is because any component of the signal with a smaller singular value than the threshold is unrecoverable from noise.

3.2 Estimating the noise level

The threshold (3) depends on the noise level σ, which is typically unknown. Gavish and Donoho propose estimating σ from the median singular value of Y.

Let m_n denote the median of the quarter-circle distribution. If there is no signal, then

$$\text{median}(s_i(Y)) \approx \sigma\sqrt{n}\, m_n.$$

In the presence of a low-rank signal, the lower portion of the singular values remains close to the noise-only distribution. Thus the estimator

$$\hat{\sigma} = \frac{\text{median}(s_i(Y))}{m_n\sqrt{n}} \tag{4}$$

is natural. Plugging $\hat{\sigma}$ into (3) yields a fully data-driven threshold.

3.3 Optimal singular value shrinkage

Instead of deciding whether to keep or discard each singular value, one can consider a more general estimator of the form

$$\hat{X} = \sum_{i=1}^{n} c(s_i)\, u_i v_i^\top,$$

where $c(\cdot)$ is a shrinkage function applied to the singular values. Under the same Gaussian model and asymptotic regime, Gavish and Donoho [2] derive an optimal shrinker that minimizes Frobenius risk.

In the square case and with appropriate normalization, the optimal shrinker takes the form

$$c(y) = \begin{cases} \sqrt{y^2 - 4}, & y > 2, \\ 0, & y \le 2, \end{cases}$$

where $y = s_i/(\sigma\sqrt{n})$ is the normalized singular value. Thus singular values below the edge of the noise bulk are discarded, while those above the edge are shrunk in a way that compensates for their distortion.

4 Image denoising experiment

4.1 Experimental setup

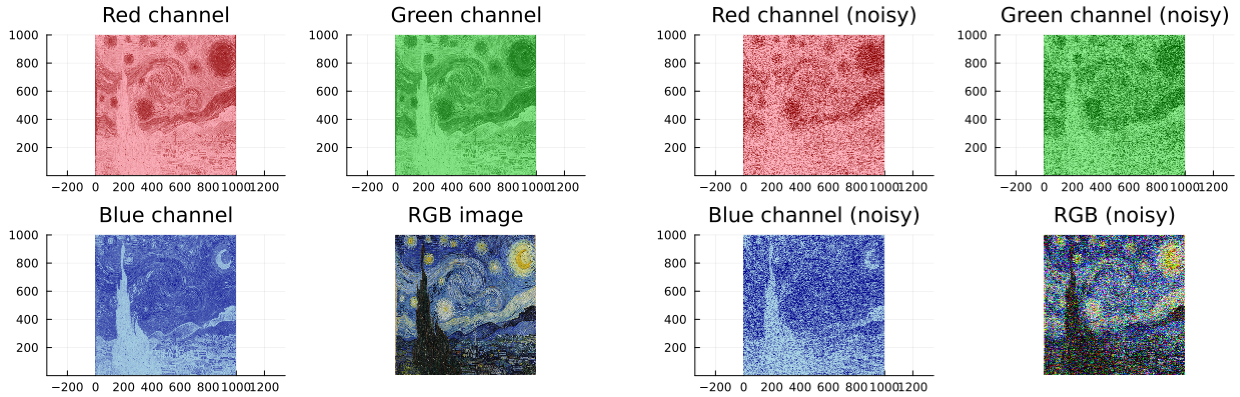
To illustrate the behavior of hard thresholding and optimal shrinkage, we look at the task of image denoising. A color image is represented as three matrices

$$R, G, B \in [0,1]^{n\times n},$$

corresponding to the red, green, and blue channels. Each channel is treated separately as an $n \times n$ matrix and independent Gaussian noise with standard deviation $\sigma = 0.25$ is added to each pixel.

The goal is to reconstruct the original image from the noisy observation using the matrix denoising methods of Section 3. No additional structural constraints (such as non-negativity or spatial smoothness) are imposed beyond the low-rank approximation implicit in the spectral methods.

Figure 3 shows the squarified original image and its noisy version.



(a) Original image.

(b) Noisy observation ($\sigma = 0.25$).

Figure 3: Squarified "Starry Night" before and after adding Gaussian noise.

4.2 Noise estimation and hard thresholding

For each channel, the SVD of the noisy matrix is computed and the noise level is estimated using the median singular value estimator (4). In one run, the estimates are

$$\hat{\sigma}_R \approx 0.226, \quad \hat{\sigma}_G \approx 0.234, \quad \hat{\sigma}_B \approx 0.237,$$

which are reasonably close to the true $\sigma = 0.25$.

The hard threshold (3) is then applied to each channel separately and the mean squared error (MSE) between the reconstructed and original channel matrices is computed as a function of the retained rank. The theoretical threshold corresponds to a particular rank that is near the minimum of the empirical MSE curve.

Figure 4 shows the channel-wise MSE as a function of rank together with the Gavish–Donoho thresholds, and Figure 5 shows the resulting hard-threshold reconstruction.

Qualitatively, the hard-threshold reconstruction preserves the dominant structure of the image while removing much of the high-frequency noise. Fine texture and small-scale details are partially suppressed, reflecting the conservative nature of the threshold.

4.3 Optimal shrinkage

The experiment is repeated using the optimal singular value shrinker instead of hard thresholding. Conceptually, this applies the nonlinear map $c(y) = \sqrt{y^2 - 4}$ (for $y > 2$) to the normalized singular values and rescales back to the original units. As in the hard-threshold case, the noise level is estimated from the median singular value.

Figure 6 compares the empirical singular values with the shrinkage map, Figure 7 shows the corresponding MSE curves, and Figure 8 shows the reconstructed image.

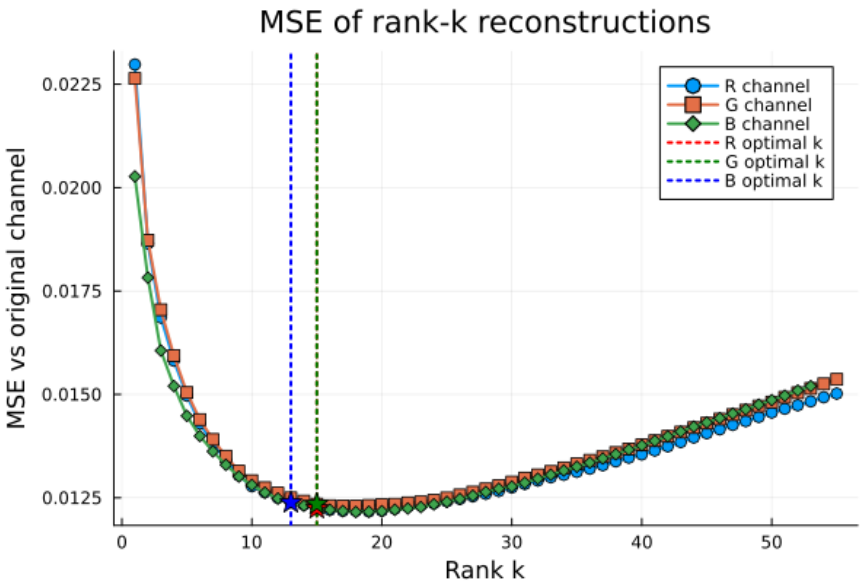


Figure 4: Channel-wise reconstruction error as a function of rank. Vertical lines indicate the Gavish–Donoho hard thresholds.

The shrinkage estimator typically yields slightly lower MSE than hard thresholding and often preserves more subtle structure, since it does not discard large singular values entirely but corrects for their distortion. At the same time, shrinkage can remain conservative in the sense that small singular values near the noise edge are strongly attenuated.

4.4 Statistical method for signal recovery

The hard threshold and shrinker in Section 3 are designed to optimize mean squared error under the spiked Gaussian model. In many applications, however, it is also useful to quantify how much apparent structure in Y can still be explained by noise alone. Rather than fixing a single "optimal" rank, one would like a family of reconstructions indexed by how aggressive the decomposition is allowed to be before the signal is indistinguishable from noise.

One way to obtain such a calibration is to compare the largest singular values of the residual matrix to a noise-only benchmark. For a candidate rank k, define the residual

$$\hat{Z}_k := \sum_{i=k+1}^{n} s_i u_i v_i^\top,$$

and consider the question: is the leading singular value of $\hat{Z}_k$ still compatible with a matrix whose entries are Gaussian noise with variance $\hat{\sigma}^2$? If it is, then all components $i > k$ can reasonably be treated as noise. If not, at least one additional singular component must be attributed to signal.

In the square Gaussian case, the largest eigenvalue of $YY^\top$ under pure noise exhibits a universal edge fluctuation: writing $\lambda_{\max}$ for the largest eigenvalue of $YY^\top$,

$$\frac{\lambda_{\max}/\sigma^2 - 4n}{2^{4/3} n^{1/3}} \Rightarrow \mathrm{TW}_1,$$

where TW_1 is the Tracy–Widom law for real symmetric ensembles. This limit provides exactly the reference distribution needed to judge whether an observed top singular value is unusually large

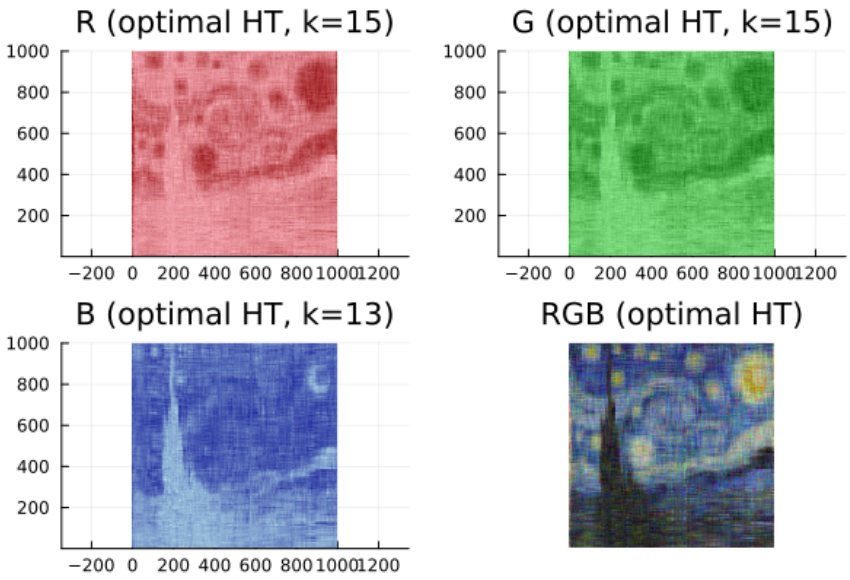


Figure 5: Reconstruction obtained by applying optimal hard thresholding channelwise.

under a noise model. After replacing σ by the plug-in estimator $\hat{\sigma}$, the standardized top eigenvalue of $\hat{Z}_k \hat{Z}_k^\top$ can be compared to Tracy–Widom quantiles, in the same spirit as eigenvalue-based tests in high-dimensional covariance problems; see, for example, recent work on Tracy–Widom-calibrated detection in multivariate settings [4].

Practically, this leads to a simple edge-based calibration procedure. Fix a significance level α and, for each k, standardize the largest singular value of $\hat{Z}_k$ using the Tracy–Widom scaling. If it exceeds the $(1-\alpha)$-quantile, k is deemed too small (the residual still contains signal); if it does not, the residual is indistinguishable from noise at level α and k is an acceptable rank. Scanning k and recording the smallest value for which the residual passes this check yields a decomposition tuned not to asymptotic MSE, but to a user-chosen tolerance for mistaking noise for signal.

In the image experiment of Section 4, the Gavish–Donoho hard threshold corresponds to an extremely small tail probability under this noise benchmark, confirming that the retained components are far outside the range expected under a noise-only model. Moving down the spectrum and using more moderate values such as $\alpha = 1\%, 5\%, 10\%$ produces reconstructions with progressively smaller rank and more aggressive denoising. Figure 9 illustrates several such reconstructions obtained by applying this edge-based calibration to the residual singular values.

5 Compressed sensing experiment

5.1 From matrix spectrum to frequency spectrum

In Section 4, we used the singular value decomposition to isolate signal from noise in matrix data. We now demonstrate that this spectral viewpoint is not limited to matrices. The same philosophy—separating a deterministic "spike" from a random "noise bulk"—applies to one-dimensional signals, provided we substitute the Singular Value Decomposition (SVD) with the Discrete Fourier Transform (DFT).

While the noise geometry changes—shifting from the interacting eigenvalues of the Gaussian Orthogonal Ensemble to the independent coefficients of white noise—the recovery strategy remains

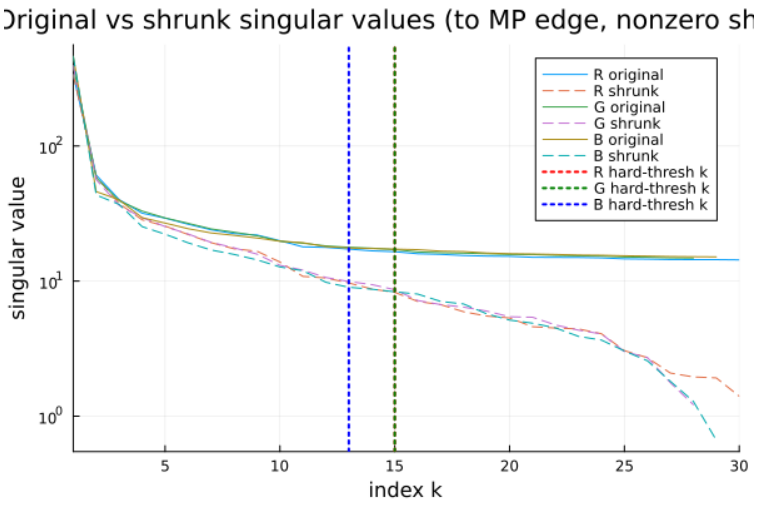


Figure 6: Empirical singular values and optimal shrinkage map for one channel.

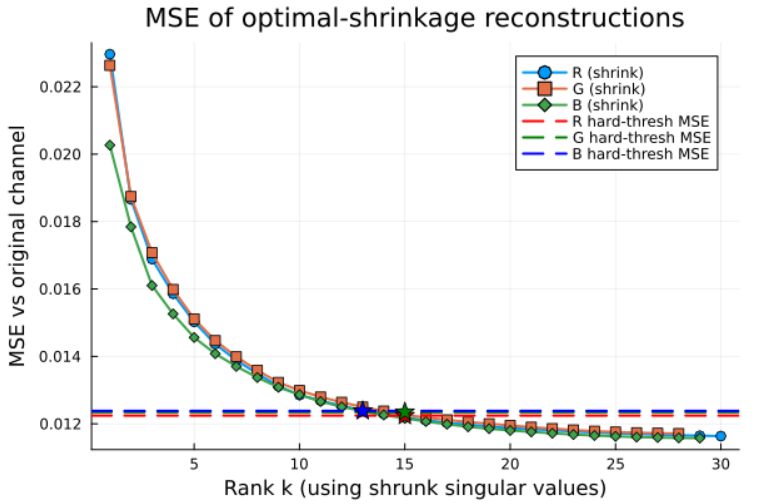


Figure 7: Channel-wise MSE for hard thresholding and optimal singular value shrinkage.

identical: we construct a null hypothesis for the "noise edge" and reject components that exceed it.

5.2 Signal model and partial observation

We consider a one-dimensional setting in which the signal is sparse in the Fourier basis. Let N be a positive integer and consider a real-valued signal defined on the discrete grid $t = 0, \dots, N-1$:

$$x_t = s_t + \varepsilon_t.$$

The signal component s_t is taken to be a sparse sum of Fourier modes,

$$s_t = \sum_{k \in S} A_k \cos\left(\frac{2\pi k t}{N} + \phi_k\right),$$

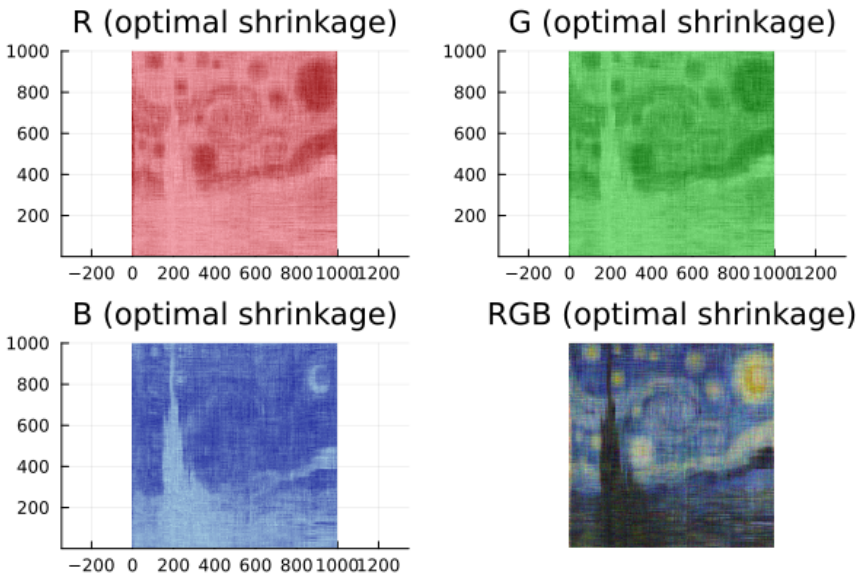


Figure 8: Reconstruction obtained by optimal singular value shrinkage.

where $S \subset \{0, \dots, K_{\max}\}$ is a set of active frequencies with $|S| \ll K_{\max}$, $A_k > 0$ are amplitudes, and ϕ_k are phases. The noise component ε_t consists of i.i.d. Gaussian variables with distribution $\mathcal{N}(0, \sigma^2)$.

The entire vector $x \in \mathbb{R}^N$ is not observed; instead, entries at a subset of indices

$$\mathcal{T}_{\text{obs}} \subset \{0, \dots, N-1\}$$

of size $n \ll N$ are observed. At these indices,

$$y_t = x_t, \quad t \in \mathcal{T}_{\text{obs}},$$

and the unobserved entries are treated as missing.

Figure 10 shows a typical realization of such a signal together with noisy partial observations.

5.3 Periodogram and empirical autocovariance

Let $\hat{x}_k$ denote the discrete Fourier transform (DFT) coefficients of x at frequencies $k = 0, \dots, N-1$, and let

$$I_k = |\hat{x}_k|^2$$

denote the periodogram. In the noise-only case $x_t = \varepsilon_t$, the coefficients $\hat{x}_k$ are complex Gaussian with variance proportional to σ^2, and I_k/σ^2 is approximately exponential with mean one. Therefore $|\hat{x}_k|$ has approximately Rayleigh distribution.

In the experiment, an empirical DFT and periodogram are constructed using only the observed entries y_t at the times $t \in \mathcal{T}_{\text{obs}}$ by treating the missing entries as absent and averaging only over the observed time points. Given the empirical periodogram $\{I_k\}$, the autocovariance function is approximated via an inverse discrete Fourier transform,

$$\hat{\gamma}(h) \approx \frac{1}{N} \sum_{k=0}^{N-1} I_k e^{2\pi i k h/N},$$

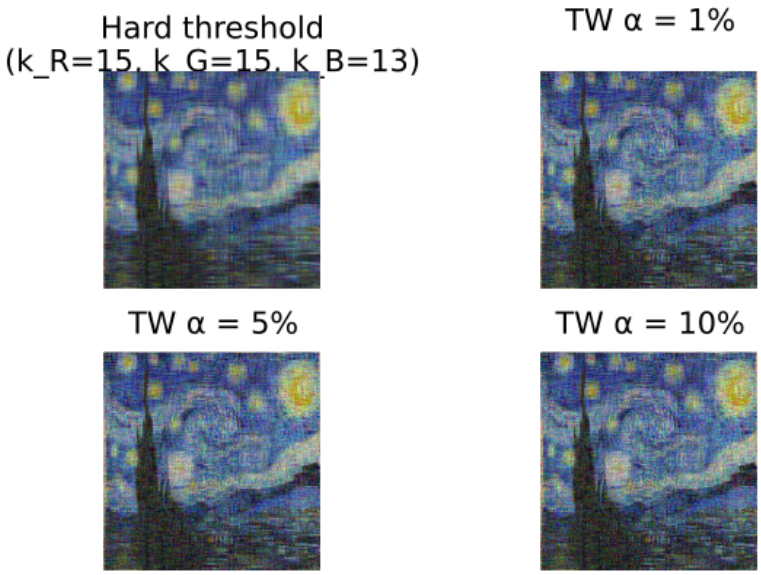


Figure 9: Reconstructions at various significance levels obtained by edge-based calibration of the residual singular values against a noise benchmark.

and a Gram autocovariance matrix at the observed times is constructed as

$$C_{\mathrm{emp},ij} = \hat{\gamma}(|t_i - t_j|), \qquad t_i, t_j \in \mathcal{T}_{\mathrm{obs}}.$$

For comparison, the theoretical autocovariance of the full signal x (with known s_t and σ^2) can be computed and restricted to the same observed indices, yielding a "true" Gram matrix C_{true}.

Figure 11 illustrates the empirical and true Fourier coefficients on a low-frequency band, together with the corresponding Gram autocovariance matrices at the observation times.

5.4 Noise behaviour and extreme values

In the Fourier setting, the role of the Marchenko–Pastur bulk is played by the exponential distribution of the squared Fourier amplitudes under pure noise. If $\hat{\varepsilon}_k$ denotes the DFT of ε, then

$$|\hat{\varepsilon}_k|^2/\sigma^2$$

is approximately Exp(1). The magnitudes $|\hat{\varepsilon}_k|$ are therefore approximately Rayleigh distributed.

The distribution of the maximum

$$M = \max_{1 \le k \le K} |\hat{\varepsilon}_k|^2$$

over a set of K frequencies is given exactly (under the independence approximation) by

$$\mathbb{P}(M \le z) = \left(1 - e^{-z/\sigma^2}\right)^K.$$

Figure 12 shows the empirical distribution of Fourier amplitudes together with the Rayleigh model, and the effect of ordering frequencies by decreasing periodogram magnitude on train/test MSE.

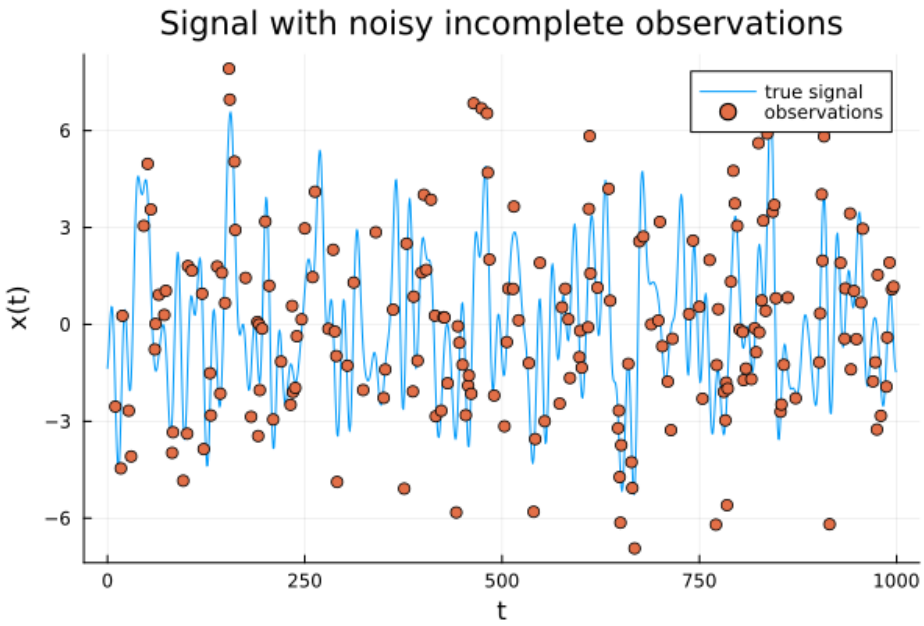


Figure 10: Example Fourier-sparse signal on $t = 0, \dots, N-1$ (solid line) and noisy partial observations (points).

As in the matrix case, this suggests a sequential testing framework. The periodogram values $\{I_k\}$ are sorted in descending order, and frequencies are added one by one to the signal model. For a given number of included frequencies r, the largest remaining periodogram value is compared to the exponential or extreme-value distribution under the noise-only model. For a chosen significance level α, one stops adding frequencies once the maximum remaining I_k falls below the corresponding critical value. This provides a Fourier-domain analogue of the Tracy–Widom-based sequential test on residual singular values.

Figure 13 shows reconstructions obtained at significance levels derived from this exponential/extreme-value viewpoint.

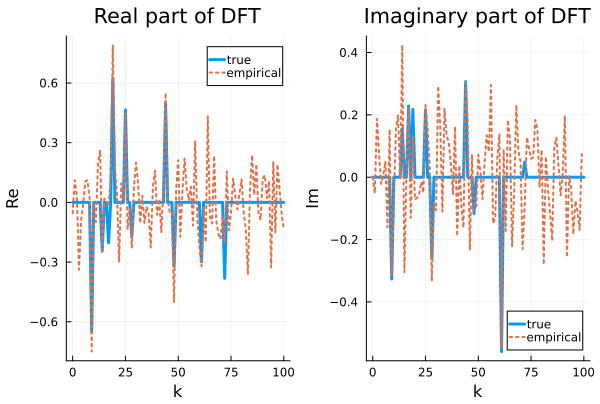
5.5 Comparison with the matrix model

Table 1 summarizes the analogy between the matrix model and the Fourier model at the level of their natural autocovariance-like objects.

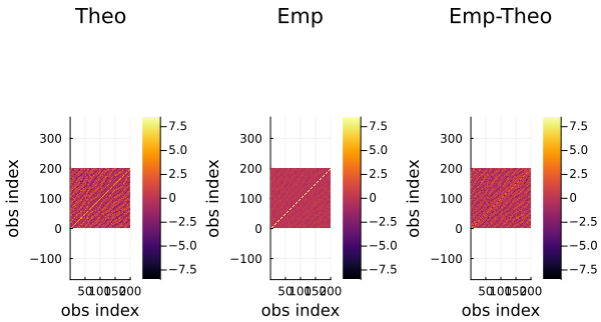
In both settings there is a bulk region generated by noise and an edge region where statistically significant deviations from the noise-only distribution appear. In the matrix case the relevant limiting distributions are Marchenko–Pastur and Tracy–Widom; in the Fourier case they are exponential, Rayleigh, and Gumbel. In both cases, hypothesis tests can be formulated that sequentially add components to the signal model until the remaining residue is consistent with the bulk noise distribution.

6 Discussion and open questions

The two experiments discussed above illustrate how random matrix theory and related probabilistic tools can be used to design and calibrate signal–noise decompositions in different domains. Several questions arise naturally from this perspective:

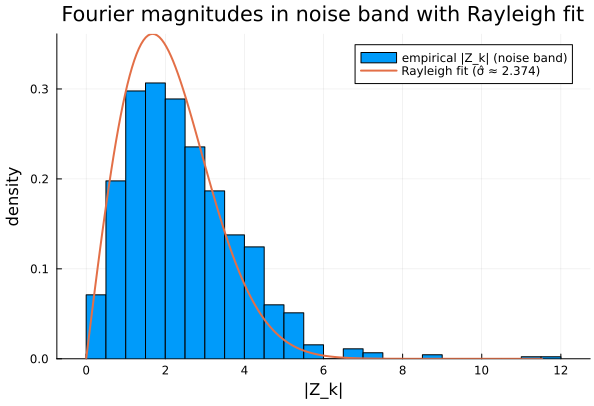


(a) Empirical vs. true DFT on the first 100 frequencies.

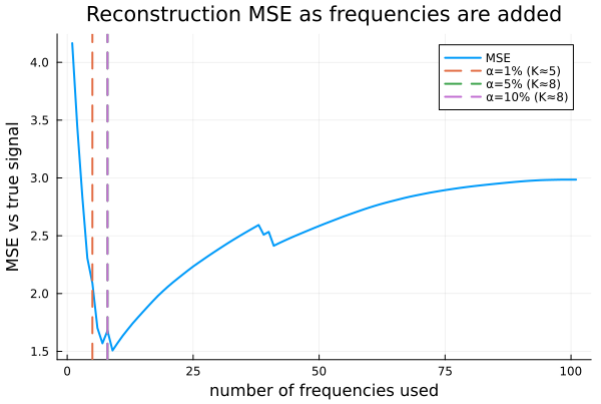


(b) Empirical vs. true Gram autocovariance at observation times.

Figure 11: Fourier and covariance structure for the partially observed Fourier-sparse signal.



(a) Distribution of Fourier amplitudes.



(b) Train/test MSE as frequencies are added in order of decreasing I_k.

Figure 12: Noise behaviour in the Fourier model and its impact on reconstruction error.

	Matrix model	**Fourier model**
Target	Data $Y \in \mathbb{R}^{n\times n}$, autocovariance $S = \frac{1}{n}YY^\top$	Signal $x \in \mathbb{R}^N$, DFT $\hat{x} = \Phi^* x$, periodogram $I_k = \lvert\hat{x}_k\rvert^2$, autocovariance $C = \Phi\,\mathrm{diag}(I)\,\Phi^*$
Model	$Y = X + \sigma Z, \quad Z_{ij} \sim \mathcal{N}(0,1)$	$x_t = s_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,\sigma^2)$
Noise distribution	Columns of Y are $\mathcal{N}(0,\sigma^2 I_n)$	$\hat{x}_k \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$
Components	Singular vectors u_j, singular values s_j of Y	Frequencies k, amplitudes $\hat{x}_k = A_k e^{i\phi_k}$, magnitudes $A_k = \lvert\hat{x}_k\rvert$
Square magnitude of noise	Bulk s_j^2 follows the Marchenko–Pastur law	Bulk $A_k^2 \sim \sigma^2\mathrm{Exp}(1)$
Magnitude of noise	Bulk s_j follows a quarter-circle law	Bulk $A_k \sim \sigma\,\mathrm{Rayleigh}$
Maximum magnitude of noise	Largest eigenvalue $\lambda_{\max}$ (or $s_{\max}$) has Tracy–Widom edge fluctuations	$M = \max_k I_k$ with $\mathbb{P}(M \le z) = (1-e^{-z/\sigma^2})^K$

Table 1: Comparison between the Gaussian matrix model and the Fourier model.

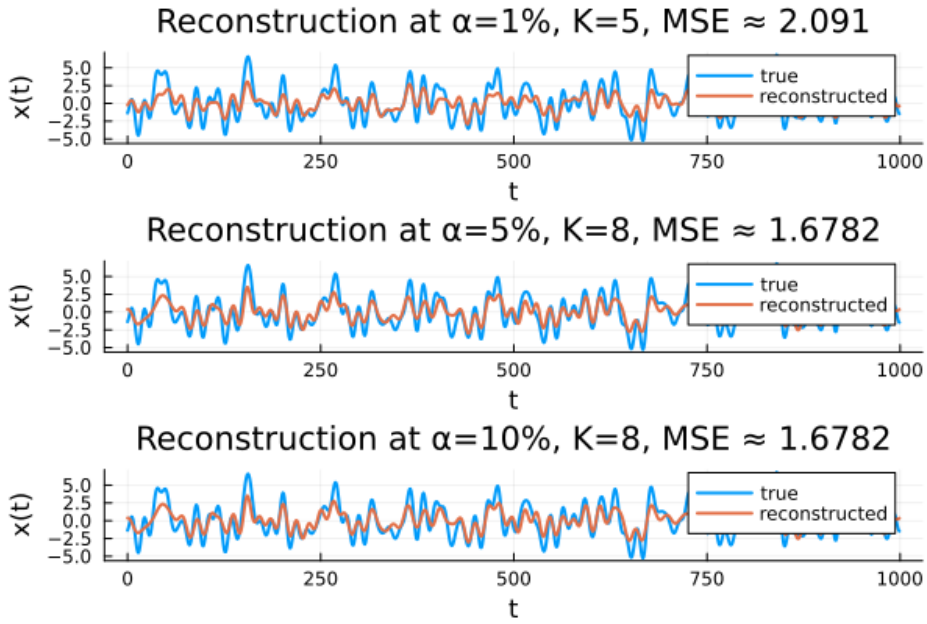


Figure 13: Reconstructions at different significance levels based on the distribution of maximum periodogram values.

- **Structured noise.** The Gaussian i.i.d. noise model is convenient but often unrealistic. In practice, noise may have temporal or spatial structure (for example, ARMA models). Extending the matrix and Fourier frameworks to structured noise models requires understanding the corresponding spectral limits and edge behaviour.
- **Model misspecification and distortion.** In both experiments, there are sources of distortion such as finite-sample effects, windowing, or irregular sampling. An important question is how to incorporate approximate knowledge of these distortions into the hypothesis tests so that the resulting thresholds remain valid but not overly conservative.
- **Asymptotic distortion of top components.** In the matrix model, the asymptotic distortion of top singular values and singular vectors is well understood through spiked random matrix theory. A similar characterization of the distortion of top periodogram peaks and Fourier amplitudes under sparse signals plus noise would help formalize the analogy between the two settings and guide the design of optimal thresholding rules in the Fourier domain.

References

[1] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.

[2] M. Gavish and D. L. Donoho. Optimal shrinkage of singular values. *IEEE Transactions on Information Theory*, 63(4):2137–2152, 2017.

[3] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 102(4):363–375, 2011.

[4] Z. T. Ke, J. Fan, and Y. Wu. Homogeneity Pursuit in High-dimensional Linear Models: A Spectral Approach. *Journal of the American Statistical Association*, 2021.

[5] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.