Random Matrix Theory for Signal–Noise Decomposition

Ishan Ganguly

December 16, 2025

Abstract

We study signal–noise decomposition in two related settings. First, we consider a low-rank signal matrix X observed in additive white Gaussian noise,

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

and use random matrix theory to justify singular value thresholding and shrinkage rules such as the Gavish–Donoho optimal hard threshold and optimal singular value shrinker. Second, we consider a sparse Fourier signal corrupted by Gaussian noise and observed at a subset of time points, and compare its periodogram- and covariance-based behaviour to the matrix case. We illustrate these ideas on an image denoising experiment and on a synthetic compressed-sensing–style example, and highlight an analogy between the Marchenko–Pastur/Tracy–Widom picture for matrices and exponential extreme-value behaviour for Fourier spectra.

1 Introduction and setup

Suppose we observe data $Y \in \mathbb{R}^{m\times n}$ that we model as

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

where

- X is a low-rank signal matrix of rank $r \ll \min\{m, n\}$;
- Z has i.i.d. entries $Z_{ij} \sim \mathcal{N}(0, 1)$;
- σ^2 is the noise variance.

The task is to recover X from the noisy observation Y.

A natural approach is to search for an estimator $\hat{X}$ such that

$$Y = \hat{X} + (Y - \hat{X}),$$

where the residual $Y-\hat{X}$ is statistically indistinguishable from pure noise at some chosen significance level. Practically, this leads us to spectral methods that partition the singular value decomposition of Y into "signal" and "noise" components.

We write the singular value decomposition (SVD) of Y as

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

with singular values $s_1 \geq s_2 \geq \cdots$ and singular vectors u_i, v_i.

In what follows we first review the random-matrix theory underpinning this model and derive optimal hard-thresholding and shrinkage rules, then apply them to image denoising, and finally compare the matrix setting to a Fourier/periodogram setting for compressed-sensing–style signals.

2 Random-matrix behaviour of singular values and vectors

2.1 Quarter-circle law (square case)

In the noise-only case $X = 0$, the matrix $Y = \sigma Z$ has i.i.d. Gaussian entries. When $m = n$, the empirical distribution of the singular values of $Y/\sqrt{n}$ converges, as $n \to \infty$, to the quarter-circle law supported on $[0, 2]$:

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

and zero outside this interval. In particular, the bulk of singular values lies in

$$s \in \left[\sigma(\sqrt{n} - o(1)),\ 2\sigma\sqrt{n}\right]$$

after appropriate scaling.

This suggests a hard-thresholding strategy: compute the SVD

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

and define

$$\hat{X} = U(\Sigma \odot \mathbb{1}_{\{s_i \ge \tau\}})V^\top,$$

where singular values below a threshold τ are discarded.

2.2 Spiked model and distortion of singular values

We next consider a rank-one "spiked" signal

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

with $\|a\|_2 = 1$, and noise Z with i.i.d. $\mathcal{N}(0, 1)$ entries. We study the top singular value and singular vector of

$$Y = X + Z = \theta aa^\top + Z,$$

in the regime where n is large.

Let s_1 and u_1 denote the top singular value and left singular vector of Y. One can show that s_1 and u_1 are non-trivially distorted relative to the signal parameters (θ, a).

A heuristic derivation proceeds as follows. Consider

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

Since $s_1 I - Z$ is almost surely non-singular, we can write

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

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 we obtain the fixed-point equation

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

If we let $\{(\lambda_i, v_i)\}_{i=1}^n$ denote 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 independence and isotropy assumptions, we have $\langle a, v_i \rangle^2 \to 1/n$ as $n \to \infty$, so

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

In the large-n limit, the trace of the resolvent of Z converges to the Stieltjes transform of the semicircle law, which yields the approximation

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

for $\theta > 1$ (supercritical spike).

2.3 Alignment of singular vectors

The corresponding alignment between the spike direction a and the top singular vector u_1 can be analysed via the resolvent representation of the spectral projector,

$$u_1 u_1^\top = -\frac{1}{2\pi i} \oint_{\Gamma(s_1)} (zI - Y)^{-1} \, dz,$$

where $\Gamma(s_1)$ is a contour enclosing s_1. Taking the quadratic form with a yields

$$\langle a, u_1 \rangle^2 = -\frac{1}{2\pi i} \oint_{\Gamma(s_1)} a^\top (zI - Y)^{-1} a \, dz.$$

Using the Sherman–Morrison formula, the semicircle law, and the residue theorem, one arrives at the asymptotic formula

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

for supercritical spikes $\theta > 1$.

Precise statements and proofs for rectangular matrices and low-rank perturbations can be found in Benaych-Georges and Nadakuditi (2012), *"The singular values and vectors of low rank perturbations of large rectangular random matrices."*

3 Hard thresholding and optimal shrinkage

3.1 Optimal hard threshold (square case)

Following Gavish and Donoho (2014), we consider the problem of partitioning the singular triplets $\{(u_i, s_i, v_i)\}$ into signal and noise components so that the reconstructed signal best approximates

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

We compare two losses for a given component:

- Loss of *not* including a component:

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

- Loss of including a noisy component:

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

The threshold is chosen so that components are assigned to the signal up to the point where $\ell_+ = \ell_-$.

In the square case ($m = n$), and under the spiked model, we asymptotically have

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

When $\ell_+ = \ell_-$, a short calculation (assuming $u_i \approx v_i$ and $a_i \approx b_i$) gives

$$\theta_i^2 - \left(s_i^2 + \theta_i^2 - 2 s_i \theta_i \langle a_i, u_i \rangle^2 \right) = 0,$$
$$s_i = 2\theta_i \langle a_i, u_i \rangle^2.$$

Substituting the asymptotic expressions,

$$\theta_i + \frac{1}{\theta_i} = 2\left(\theta_i - \frac{1}{\theta_i}\right) \quad \Rightarrow \quad \theta_i^2 = 3 \quad \Rightarrow \quad \theta_i = \sqrt{3},$$

and hence

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

In terms of the unscaled singular values of Y, this yields the asymptotically optimal hard threshold

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

for square matrices, in the sense of minimizing asymptotic mean squared error.

3.2 Optimal singular value shrinkage

Rather than either keeping or discarding each singular value, we may consider a shrunk estimator of the form

$$\hat{X} = \sum_i c_i s_i \, u_i v_i^\top,$$

where $c_i \in [0, 1]$ is a shrinkage factor. Minimizing the same Frobenius loss,

$$\ell_i(c_i) = \left\| c_i s_i u_i v_i^\top - \theta_i a_i b_i^\top \right\|_F^2,$$

and differentiating with respect to c_i yields

$$\frac{d}{dc_i} \ell_i(c_i) = 2 c_i s_i^2 - 2 s_i \theta_i \langle a_i, u_i \rangle^2 = 0,$$

so the optimal shrinkage factor is

$$c_i^\star = \frac{\theta_i}{s_i} \langle a_i, u_i \rangle^2.$$

Substituting asymptotics, Gavish and Donoho (2016) show that, in the square case and after suitable 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.

4 Image denoising experiment

We illustrate these ideas on a simple image denoising task. We represent a colour image as three matrices

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

for the red, green, and blue channels. We add independent Gaussian noise with standard deviation $\sigma = 0.25$ to each channel and seek to recover the original image.

4.1 Data and noise estimation

Figure 1 shows the original (squarified) image of van Gogh's *Starry Night*, and Figure 2 shows the noisy version.

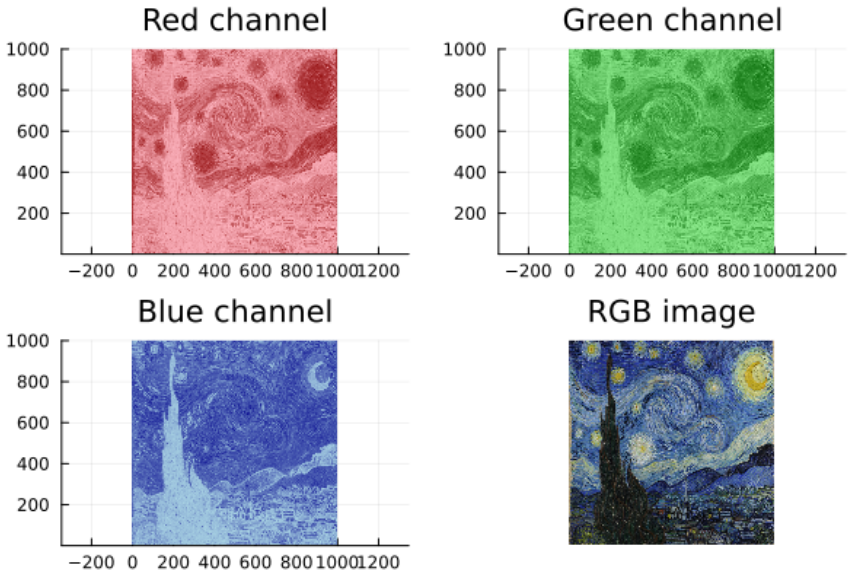

Figure 1: Original squarified *Starry Night*.

We estimate the noise level in each channel using the median singular value and the known median of the Marchenko–Pastur distribution. In one run we obtain, for example,

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

4.2 Optimal hard thresholding

For each channel, we compute the SVD and apply the optimal hard threshold with the estimated $\hat{\sigma}$. We compute the mean squared error (MSE) between the reconstructed and original channels as a function of rank, and identify the optimal rank according to the Gavish–Donoho threshold.

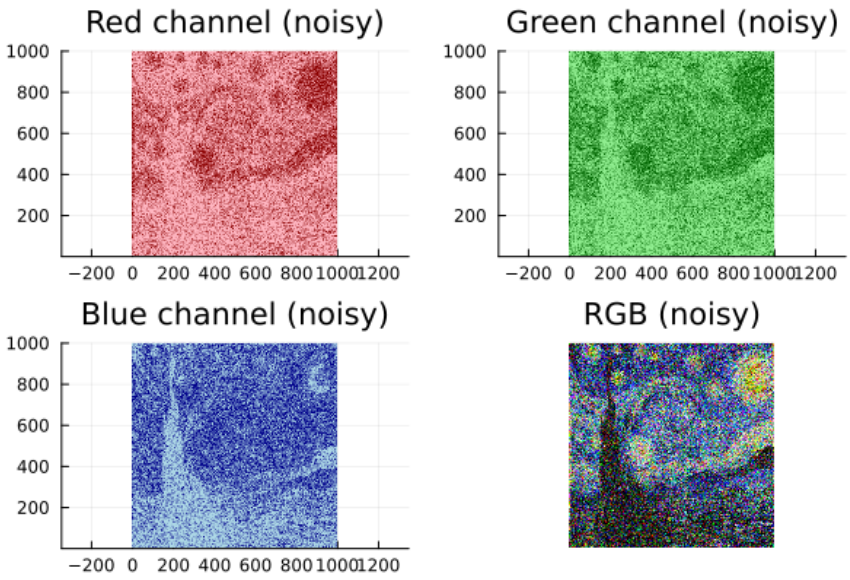

Figure 2: Noisy squarified *Starry Night* with $\sigma = 0.25$.

The resulting hard-threshold reconstruction is shown in Figure 4.

4.3 Optimal shrinkage

We repeat the experiment using the optimal singular value shrinker instead of hard thresholding. Figure 5 compares the singular values and MSEs for hard thresholding and shrinkage, and Figure 6 shows the corresponding reconstruction.

4.4 Observations and a Tracy–Widom test

Empirically, both optimal thresholding and optimal shrinkage turn out to be conservative in this example: the thresholds are relatively high, and the shrinkage is relatively aggressive. In practice, small amounts of residual noise are often less visually salient than loss of fine structure, so slightly less conservative rules might perform better for perceptual quality.

As an alternative, we can formulate a sequence of hypothesis tests of the form

$$H_{0,k}: \quad \hat{Z}_k := \sum_{i=k+1}^{n} s_i u_i v_i^\top \text{ is generated from noise of level } \hat{\sigma},$$
$$H_{1,k}: \quad \hat{Z}_k \text{ contains a residual signal component.}$$

We can sequentially test $H_{0,k}$ as k decreases (i.e. as we move down the spectrum) and stop when we fail to reject the null at significance level α.

In the square case, the largest singular value of a pure-noise matrix σZ has Tracy–Widom fluctuations. For the $(k+1)$-st singular value s_{k+1}, we may consider the test statistic

$$t = \frac{(s_{k+1}^2/\sigma_{\text{noise}}^2) - 4n}{2^{4/3} n^{1/3}},$$

compare t to the TW_1 distribution, and compute the p-value

$$p = 1 - F_{\text{TW}_1}(t).$$

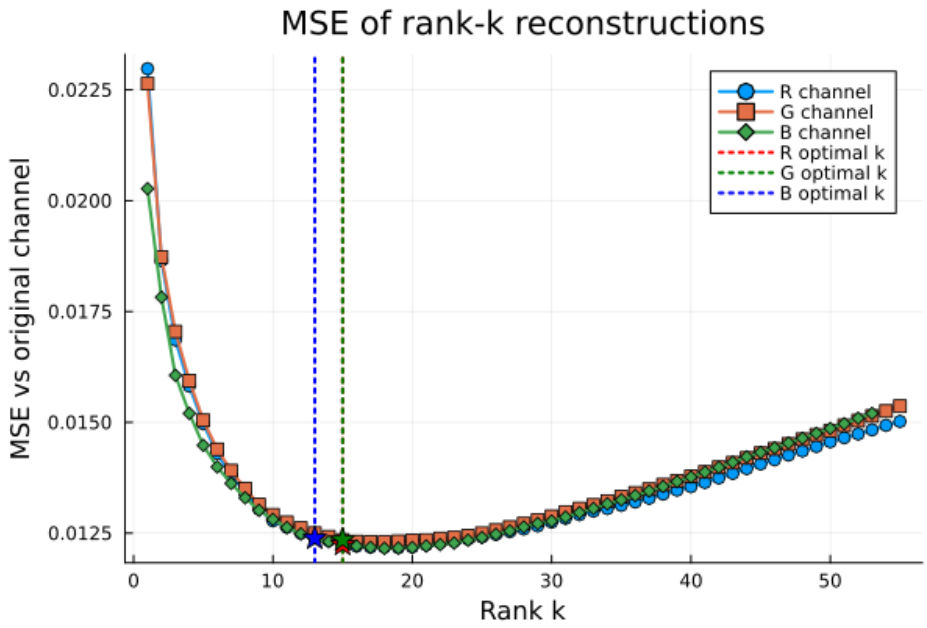

Figure 3: MSE vs. rank and optimal hard thresholds for each channel.

We reject $H_{0,k}$ if $p < \alpha$.

In the image experiment, the hard threshold corresponds to an *extremely* small significance level (e.g. of order 10^{-207}), indicating that the top retained components are overwhelmingly unlikely to be generated by noise alone. Figure 7 shows reconstructions corresponding to different significance levels.

5 Compressed sensing–style Fourier experiment

5.1 Noisy observations of a Fourier-sparse signal

We now move to a one-dimensional compressed-sensing–style setting, where the signal is sparse in the Fourier basis and the noise does *not* correspond to a classical Gaussian matrix ensemble.

We consider a real-valued signal

$$x_t = s_t + \varepsilon_t, \qquad t = 0, \dots, N-1,$$

where s_t is sparse in the Fourier basis and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, and we only observe x_t at a subset of $n \ll N$ time points. Classical compressed sensing (e.g. Tao et al., 2005) tells us that such signals can be recovered under suitable conditions.

Figure 8 shows an example of a noisy, partially observed Fourier-sparse signal.

5.2 Periodogram and autocovariance matrices

Let $\hat{x}_k$ denote the discrete Fourier transform (DFT) coefficients and $I_k = |\hat{x}_k|^2$ the periodogram. For white noise, I_k/σ^2 behaves approximately like i.i.d. Exp(1) variables. We can form an empirical autocovariance matrix at the observation times by treating

$$C = \Phi \, \mathrm{diag}(I) \, \Phi^\top$$

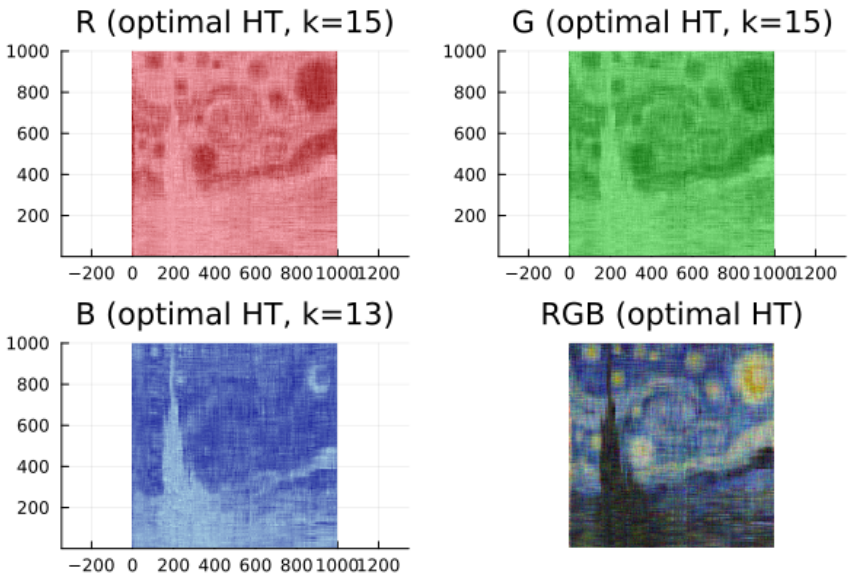

Figure 4: Hard-threshold reconstruction of *Starry Night*.

as a circulant covariance and restricting to the observed indices.

Figure 10 shows the magnitudes of the first 100 Fourier basis elements and the reconstructed autocovariance matrices at the observation times.

5.3 Matrix vs Fourier comparison

We can now explicitly compare the matrix model $Y = X + \sigma Z$ and the Fourier model $x = s + \varepsilon$ at the level of their autocovariance-like objects S and C:

	Matrix model (TW)	**Fourier model (Exp)**
Target	Data $Y \in \mathbb{R}^{p\times n}$, 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 $\mathcal{N}(0,\sigma^2 I_p)$	$\hat{x}_k \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$
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 $A_k \sim \sigma\,$Rayleigh
Maximum magnitude of noise	Largest eigenvalue $\lambda_{\max}$ (or $s_{\max}$) has Tracy–Widom edge fluctuations	$M = \max_k I_k$ with $\Pr(M \le z) = (1 - e^{-z/\sigma^2})^K$ for K frequencies

Table 1: Comparison between the matrix and Fourier models.

Figure 12 shows the empirical distribution of Fourier amplitudes and the MSE of reconstructions as a function of the number of frequencies retained.

Finally, Figure 13 shows reconstructions at different significance levels based on exponential extreme-value thresholds for the periodogram.

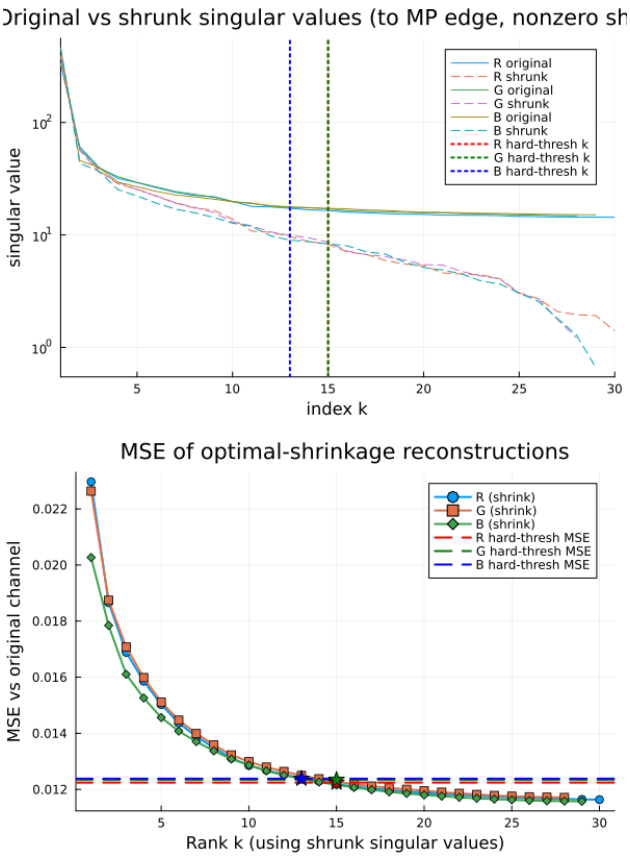

Figure 5: Comparison of singular values and MSE for hard thresholding and optimal shrinkage.

6 Questions and extensions

We close with some questions suggested by this comparison:

- **Structured noise.** Can these frameworks be extended to other structured noise models, such as Toeplitz or ARMA covariance in time, or correlated matrix noise?
- **Distortion and misspecification.** How can we incorporate knowledge of windowing, leakage, irregular sampling, or non-circulant structure into the hypothesis tests, so that the resulting thresholds remain valid?
- **Asymptotic distortion.** What is the analogue, in the Fourier setting, of the asymptotic distortion of top singular values and singular vectors (e.g. $s_1 \approx \theta + 1/\theta$, alignment $\approx 1 - 1/\theta^2$)? Can we characterize the distortion of top periodogram peaks and amplitudes under sparse signals plus noise?

References

- D. Donoho and M. Gavish, "The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$," *IEEE Transactions on Information Theory*, 2014.
- D. Donoho and M. Gavish, "Optimal Shrinkage of Singular Values," *IEEE Transactions on Information Theory*, 2016.

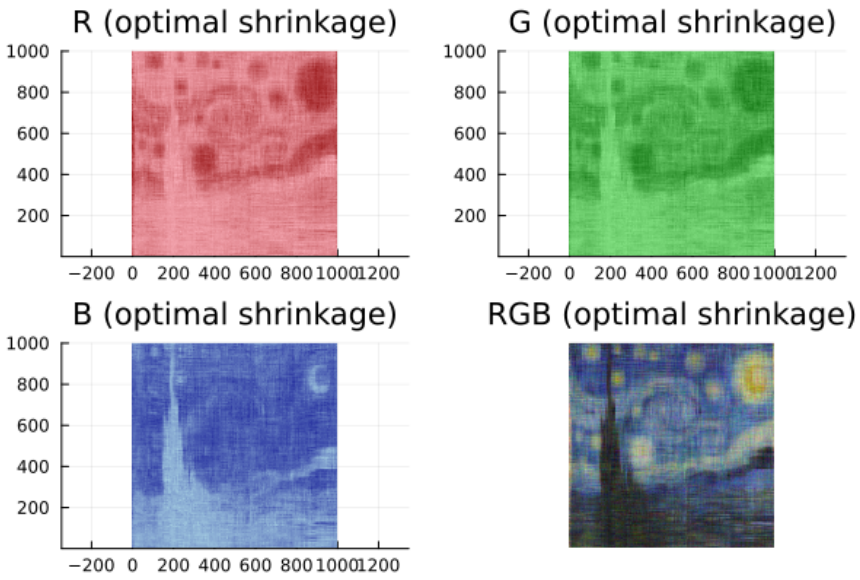

Figure 6: Shrinkage reconstruction of *Starry Night*.

- 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*, 2012.
- T. Tao, E. Candès, J. Romberg, and others, work on compressed sensing (mid-2000s).

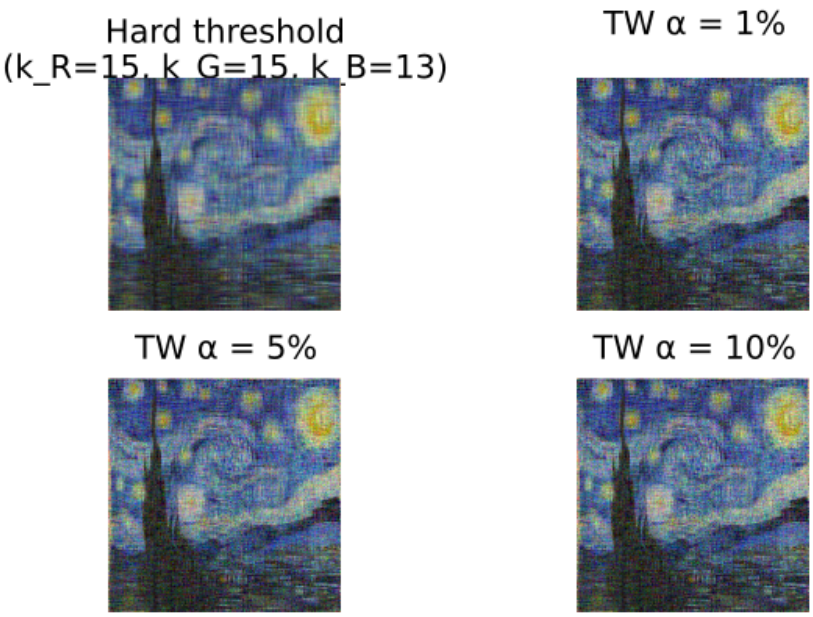

Figure 7: Reconstructions at various significance levels based on Tracy–Widom p-values.

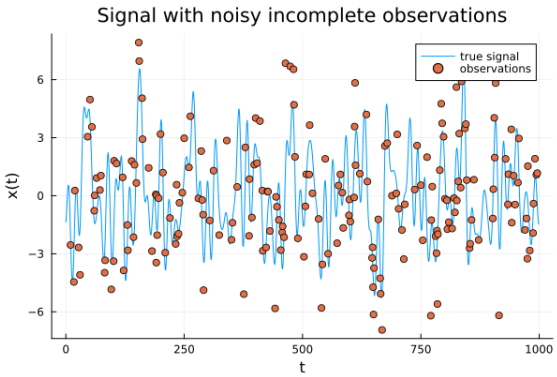

Figure 8: A noisy, partially-observed Fourier-sparse signal.

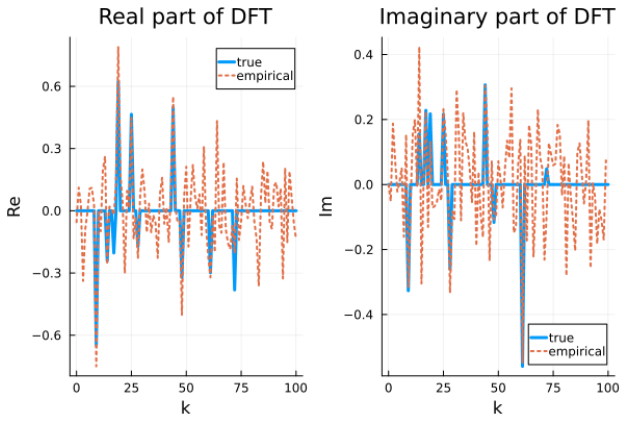

Figure 9: Magnitudes of the first 100 Fourier coefficients.

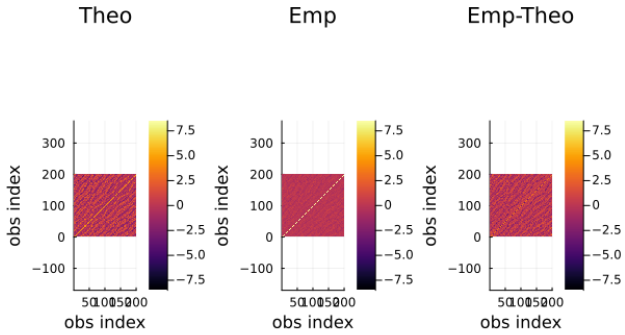

Figure 10: Reconstructed autocovariance matrices at observation times.

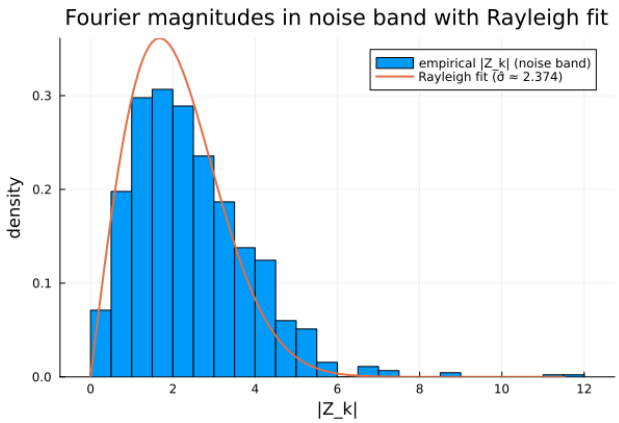

Figure 11: Distribution of Fourier amplitudes.

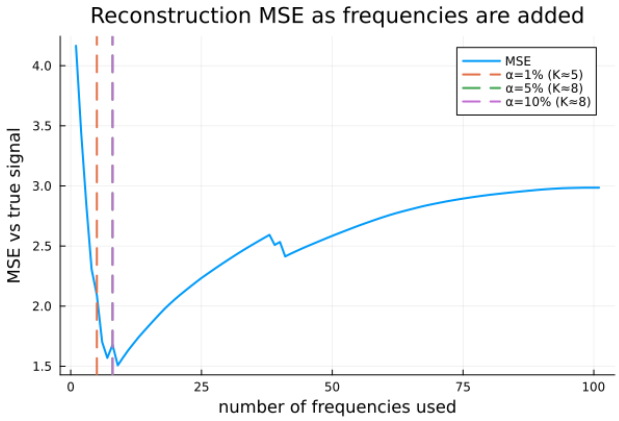

Figure 12: MSE of Fourier-based reconstructions vs. number of frequencies.

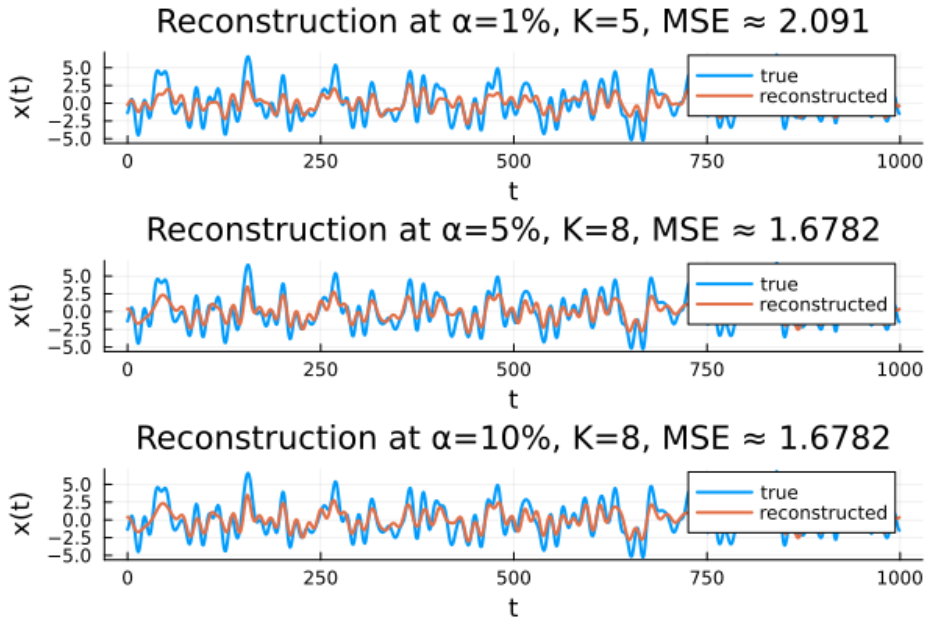

Figure 13: Reconstructions at different significance levels in the Fourier experiment.