Fundamental Limits of Singular Value Based Signal Detection from Randomly Compressed Signal-plus-Noise Matrices

Nicholas Asendorf
Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: asendorf@umich.edu

Raj Rao Nadakuditi
Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: rajnrao@umich.edu

***Abstract*—The singular value spectrum of a data matrix is commonly used to detect high-dimensional signals. However, as the size of this data matrix grows, taking its SVD becomes intractable. We consider projecting the data matrix into a lower dimensional space and using the resulting singular value spectrum for signal detection. We derive the almost sure limit of the top singular values of the resulting projected matrix both when using a Gaussian and unitary projection matrix. We highlight our prediction accuracy and discuss the benefits and drawbacks of each projection matrix using numerical simulations.**

I. Introduction

In many classical signal processing applications, we stack observations in a data matrix that is assumed low-rank plus noise, modeled as

$$\widetilde{X}_n = \sum_{i=1}^{r} \theta_i u_i v_i^T + X_n. \tag{1}$$

In the above equation, for $i = 1, \ldots, r$, $u_i \in \mathbb{C}^{n\times 1}$ and $v_i \in \mathbb{C}^{N\times 1}$ are independent unit norm signal vectors, $\theta_i > 0$ are the associated signal values and X_n is a noise-only matrix. Assume that $u_i^H u_j = \delta_{\{i=j\}}$ and $v_i^H v_j = \delta_{\{i=j\}}$. Let $X_n \in \mathbb{C}^{n\times N}$ be a real or complex random matrix. Let $\sigma_1, \ldots, \sigma_{\min(n,N)}$ be the singular values of X_n. Let μ_{X_n} be the empirical singular value distribution, i.e, the probability measure defined as

$$\mu_{X_n}(x) = \frac{1}{\min(n,N)} \sum_{i=1}^{\min(n,N)} \delta_{\sigma_i}(x).$$

In many signal processing applications, we treat the columns of $\widetilde{X}_n$ as noisy observations of a desired target signal lying in the span of $\{u_1, \ldots, u_r\}$. In this light, we treat θ_i as the signal-to-noise ratio (SNR) for its corresponding subspace component, n as the intrinsic dimension of the problem, and N as the number of samples (or snapshots or observations) we have at our disposal. To recover the underlying signal subspace, $\mathbf{span}\{u_1, \ldots, u_r\}$, it is common to take the left singular vectors of $\widetilde{X}_n$ corresponding to the largest r singular values. The accuracy of this estimate is well studied (see [1]–[4]). Specifically, when X_n has independent $\mathcal{CN}(0,1)$ entries, the individual subspace component estimates are known to have a non-random estimate when $\theta_i > \left(\frac{n}{N}\right)^{1/4}$.

However, in many such applications the intrinsic dimension, n, of the system is so large that taking the SVD of $\widetilde{X}_n$ may not be tractable. In this paper, we explore the performance of signal detection when randomly projecting $\widetilde{X}$ into a lower dimensional space using either a Gaussian or unitary projection. Specifically for $m < n$, let $G_n \in \mathbb{C}^{n\times m}$ be a random matrix with independent $\mathcal{CN}(0,1)$ entries and let $Q_n \in \mathbb{C}^{n\times m}$ be a unitary matrix such that $Q_n^H Q_n = I_m$. Define the $m \times N$ complex matrices

$$Y_n^G = G_n^H \widetilde{X}_n, \quad Y_n^Q = Q_n^H \widetilde{X}_n \tag{2}$$

Since $m < n$, taking the SVD of Y_n^G and Y_n^Q is more tractable than taking the SVD of $\widetilde{X}_n$. Such compressed sensing strategies for both unitary [5]–[7] and Gaussian [8]–[10] sensing matrices have been extensively studied. These algorithms have been extended to include Gaussian-like strategies that employ matrices with partially observed entries [11], [12] as well as unitary-like strategies that use a discrete Fourier transform matrix [13] or discrete cosine transform [14]. For excellent reviews of such compressed sensing algorithms please see [15]–[17], for example.

These works examine the ability of such matrices to approximate the original data matrix as low rank. In this paper, we consider the fundamental limits of the resulting singular values when used to detect low-rank signals. We quantify how the dimensions of our matrices, m, n, N, and the SNR θ affect the behavior of the largest singular values of Y_n^G and Y_n^Q. Finally, we compare the detection performance of these two specific choices of the projection matrix and show that a unitary projection matrix can more reliably detect low-rank signals than a Gaussian projection matrix.

This paper is organized as follows. In Section II, we provide the main results of this paper including the almost sure limit of the top singular values of the projection matrices in (2). We then provide corollaries to the main result that highlight a phase transition below which signal detection is impossible and a closed form expression of our main theorem for unitary

projections. In Section III, we verify our asymptotic results on finite sized systems and highlight the accuracy of our predictions. We provide concluding remarks in Section IV.

II. Main Results

Our main result characterizes the asymptotic behavior of the largest singular values of the projection matrices defined in (2). We begin with the following assumptions and definitions.

Assumption II.1. *The probability measures μ_{X_n}, μ_{G_n}, and μ_{Q_n} converge almost surely weakly to non-random compactly supported probability measures μ_X,μ_G, and μ_Q, respectively.*

Definition II.1. *Let $M_n^G = G_n^H X_n$ be the product of the random matrices G_n and X_n and let $M_n^Q = Q_n^H X_n$ be the product of the random matrices Q_n and X_n.*

Assumption II.2. *The probability measure $\mu_{M_n^G}$ converges almost surely weakly to a non-random compactly supported probability measure μ_{M_G}. The probability measure $\mu_{M_n^Q}$ converges almost surely weakly to a non-random compactly supported probability measure μ_{M_Q}.*

Assumption II.3. *Let a_G be the infimum of the support μ_{M_G}. The smallest singular value of M_n^G converges almost surely to a_G. Let a_Q be the infimum of the support μ_{M_Q}. The smallest singular value of M_n^Q converges almost surely to a_Q.*

Assumption II.4. *Let b_G be the supremum of the support μ_{M_G}. The largest singular value of M_n^G converges almost surely to b_G. Let b_Q be the supremum of the support μ_{M_Q}. The largest singular value of M_n^Q converges almost surely to b_Q.*

With these assumptions and definitions, we now state our main results about the almost sure limit of the largest singular values of the projection matrices defined in (2).

Theorem II.1. *Let Y_n be the projection of $\widetilde{X}_n$ onto either G_n or Q_n as in (2). The largest r singular values of the $m \times N$ matrix Y_n exhibit the following behavior as $n, m, N \to \infty$ with $n/N \to c_1$ and $m/n \to c_2$. We have that for each fixed $1 \leq i \leq r$, $\sigma_i(Y_n)$ solves*

$$\sigma_i^2 \varphi_F(\sigma_i)\varphi_H(\sigma_i) = \frac{1}{\theta_i^2}, \tag{3}$$

where

$$\varphi_F(\sigma_i) \xrightarrow{a.s.} -\mathbb{E}\left[x m_{\mu_{RS|R}}\left(\sigma_i^2, x\right)\right]_{\mu_R}$$
$$\varphi_H(\sigma_i) \xrightarrow{a.s.} -\frac{n}{N} m_{M_3}(\sigma_i^2) - \frac{1}{\sigma_i^2}\frac{n-N}{N}$$

where m_{μ_M} is the Stieltjes transform of a matrix M defined as

$$m_{\mu_M}(z) \int \frac{1}{x - z} \mu_M(x),$$

and μ_R is the limiting eigenvalue density of either $G_n G_n^H$ or $Q_n Q_n^H$, μ_S is the limiting eigenvalue density of $X_n X_n^H$, $m_{\mu_{RS|S}}$ is the Stieltjes transform of the limiting conditional density and $m_{\mu_{M_3}}$ is the Stieltjes transform of $G_n G_n^H X_n X_n^H$ or $Q_n Q_n^H X_n X_n^H$. When using G_n, $m_{\mu_{RS|S}}(z, x)$ solves the following equation

$$\begin{aligned} 0 = &\left(-n^2 z^2\right)\left(m_{\mu_{RS|S}}(z,x)\right)^3 + Nm \\ &+ \left(Nnz + mnz - 2n^2 z\right)\left(m_{\mu_{RS|S}}(z,x)\right)^2 \\ &+ \left(Nn + mn + Nmz - n^2 - Nm\right) m_{\mu_{RS|S}}(z,x). \end{aligned} \tag{4}$$

When using the Gaussian projection matrix, G_n, we do not get a closed form of the top singular values. Solving (4) for $m_{\mu_{RS|S}}(z, x)$ is unwieldy as we must solve a cubic polynomial. Furthermore, we must take the expectation of the resulting solution with respect to the distribution μ_R. To solve the expressions φ_F and φ_H when using a Gaussian projection matrix, we use RMTool [18]. We note that this process still yields an analytic solution, although not closed form. However, when using a unitary projection matrix, we do get a closed form expression for the largest singular values.

Corollary II.1. *When Y_n is a generated using a unitary matrix Q_n, we have that for each fixed $1 \leq i \leq r$,*

$$\sigma_i \xrightarrow{a.s.} \begin{cases} \sqrt{\frac{c_1}{\theta_i^2} + c_2\theta_i^2 + 1 + c_1 c_2} & \text{if } \theta_i \geq \left(\frac{c_1}{c_2}\right)^{1/4} \\ \sqrt{c_1 c_2} + 1 & \text{if } \theta_i < \left(\frac{c_1}{c_2}\right)^{1/4} \end{cases}.$$

This corollary nicely gives the almost sure limit of the top singular values as a function of the system parameters n, m, N, and θ_i. This corollary also makes contact with a natural phase transition. When the SNR of a component is below a critical value depending only on n, m, N, the corresponding top singular value behaves as if Y_n is a noise only matrix. Such phase transitions appear in other matrix analyses (see [1]–[4]). Similarly, we may solve for the phase transition when using a Gaussian matrix, although we do not get a closed form expression as we do in the unitary case. We state a more general version of this phase transition in the form of the following corollary.

Corollary II.2. *When*

$$\theta_i \leq \theta_{crit} = \frac{1}{b\sqrt{\varphi_F(b)\varphi_H(b)}}$$

then

$$\sigma_i \xrightarrow{a.s.} b,$$

where b is either b_Q or b_G depending on our projection matrix.

When a component's SNR is below a critical SNR (given in the corollary), then the corresponding singular value of Y_n does not separate from its noise distribution. In this scenario, this signal is not detectable.

III. Empirical Results

In this section we verify the singular value prediction given in (3) that relies on the asymptotic approximations φ_F and φ_H. We consider two different types of projection matrices. In the first setting, we use a matrix G_n with independent $\mathcal{N}(0, 1)$ entries. In the second setting, we use a unitary matrix Q_n such that $Q_n^H Q_n = I_m$. In both settings, we let the noise matrix

X_n be an appropriately scaled random Gaussian matrix whose entries are independent standard Gaussian random variables.

A. Gaussian Projection, G

In this setting, $R = GG^H$ and $S = XX^H$ are independent Wishart random matrices with parameters $c_2 = \frac{n}{m}$ and $c_1 = \frac{n}{N}$. To compute φ_F and φ_H, we use RMTool [18]. We numerically approximate the expected value using RMTool to approximate the density of R and to compute the Stieltjes transform of the kernel. We use 2500 points in the approximation. To compute φ_H, We consider the matrix $M_2 = GG^H XX^H = RS$, which is a product of Wishart matrices. This is desirable as we can use RMTool to compute the Stieltjes transform as above for φ_H.

Figure 1(a) shows the performance of our theoretical prediction when using a Gaussian projection matrix, G, for a rank-1 setting with a fixed $n = 1000$, $N = 1220$, $m = 100$. In our empirical setup, we generate 500 matrices from (1) and 500 noise only matrices. We then generate a random G selected as above. The figure plots the empirical and theoretically predicted top singular value for a number of $\theta_1 = \theta$. As evident in the figure, the theoretical prediction does a good job except.

In Figure 1(b) we consider a Gaussian-like projection matrix for the same rank-1 setting as Figure 1(a). For this figure, the entries of G are

$$G_{ij} = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases} \tag{5}$$

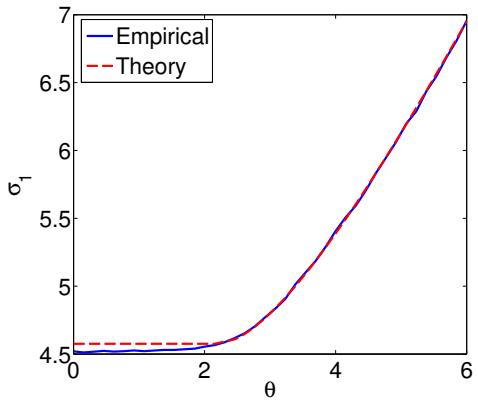
so that they have zero mean and unit variance. We see that the theoretical prediction from (3) for the Gaussian setting is still valid for this Gaussian-like projection matrix.

We then explore the accuracy of the phase transition boundary for the Gaussian setting in Figure 2 by plotting the KS-statistic between the largest singular values from these 500 signal and noise only matrices for $n = 1000$. Figure 2(a) sweeps over θ and N while Figure 2(b) sweeps over θ and m. In both figures, we plot our theoretical phase transition prediction in solid white line. Using a dashed white line, we plot the theoretical phase transition when no projection is used; this is $\theta = \left(\frac{n}{N}\right)^{1/4}$.

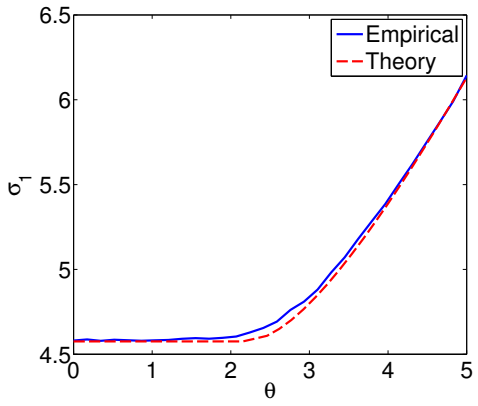
From this figure, we observe that the phase transition prediction is very accurate. Similarly we notice that the phase transition when using the Gaussian projection is significantly worse that that when not projecting. Figure 2(a) sets $m = 100$ so that we reduce our SVD dimension by one order of magnitude. Interestingly, and perhaps most importantly, when using a Gaussian projection matrix, setting $m = n = 1000$ results in worse performance than the non-projecting case even though we aren't reducing the dimension of the problem. This is evident in Figure 2(a).

B. Unitary Projection, Q

In this setting, we use a unitary projection matrix Q using a QR decomposition of a random matrix. Figure 3(a) plots the performance of our theoretical prediction for a rank-1



(a) Gaussian G



(b) Gaussian-like G

Fig. 1. Singular value prediction for Gaussian G and X for a rank-1 setting with fixed $n = 1000$, $N = 1220$ and $m = 100$. The theoretical prediction uses (3) with approximations using RMTool. Empirical results are averaged over 500 trials. (a) Gaussian G (b) Entries of G are from (5).

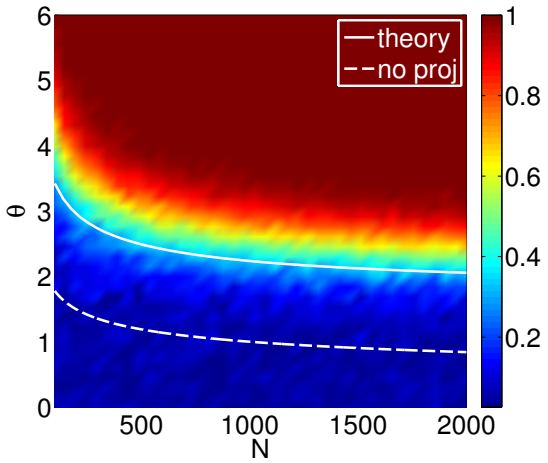
setting with a fixed $n = 1000$, $N = 1220$, $m = 100$. In our empirical setup, we generate 500 matrices from (1) and 500 noise only matrices. We then generate a random Q selected as above. The figure plots the empirical and theoretically predicted top singular value for a number of $\theta_1 = \theta$. The theoretical prediction uses the result from Corollary II.1 and does an excellent job at the singular value prediction.

In Figure 3(b), we consider a specific choice of unitary matrix. Here, we randomly select columns from the $n \times n$ discrete Fourier matrix F with entries

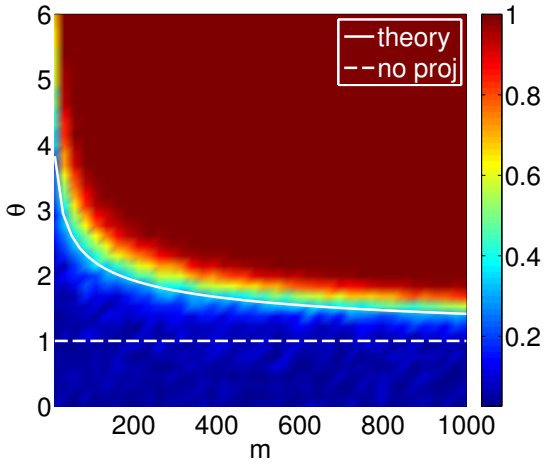
$$F_{kj} = \frac{1}{\sqrt{n}} \exp\left\{\frac{-2\pi i(k-1)(j-1)}{n}\right\} \tag{6}$$

for $k = 1 \ldots, n$ and $j = 1, \ldots, n$. To generate Q we then select m columns from F. We see that the theoretical prediction from Corollary II.1 still does an excellent job at the singular value prediction for this specific choice of unitary matrix.

Figure 4 plots the performance of our theoretical prediction when using a unitary projection matrix, Q. Our parameter sweep is the same as described for Figure 2, except that here we use the phase transition prediction given in Corollary II.1.

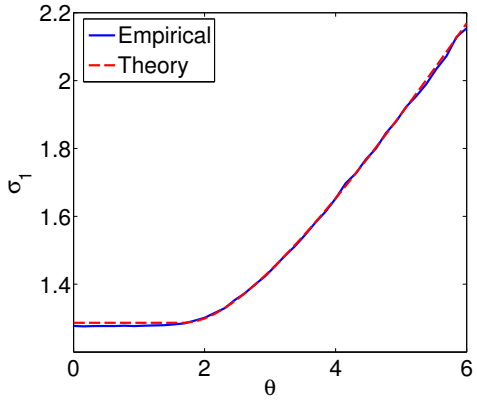


(a) KS Statistic - N,θ sweep

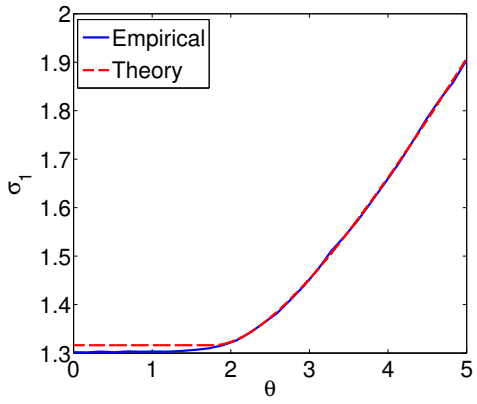


(b) KS Statistic - m,θ sweep

Fig. 2. Accuracy of theoretical phase transition prediction for Gaussian G and X for a rank-1 setting with fixed $n = 1000$. The theoretical prediction uses (3) with approximations using RMTool. Figures plot the KS statistic between singular values generated from 500 signal bearing and 500 noise only matrices. (a) Sweeps over both θ and N for a fixed $m = 100$. (b) Sweeps over θ and m for a fixed $N = 1000$.



(a) Unitary Q



(b) Fourier Q

Fig. 3. Singular value prediction for unitary projection matrix Q and Gaussian noise matrix X for a rank-1 setting with fixed $n = 1000$, $N = 1220$ and $m = 100$. The theoretical prediction uses Corollary II.1. Empirical results are averaged over 500 trials. (a) Columns of Q are from the QR decomposition of a random matrix. (b) Columns of Q are sampled from the $n \times n$ discrete Fourier matrix defined in (6).

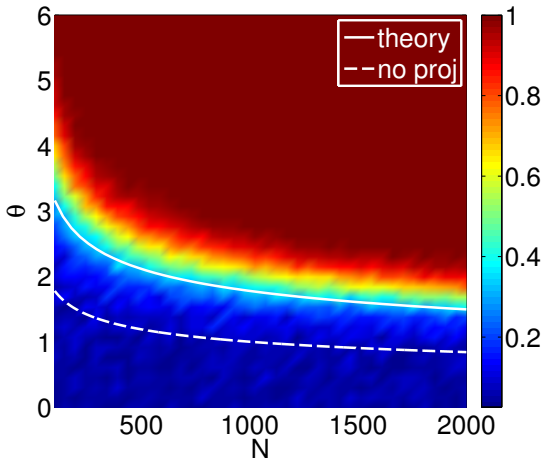
Again, we notice that our phase transition prediction is very accurate. A key observation is that for a unitary projection matrix, as $m \to n$, the phase transition approaches that of not using a projection matrix. This is very desirable as we don't want to suffer much performance loss for only slightly reducing the dimension of the problem.

C. Comparison of Projection Matrices

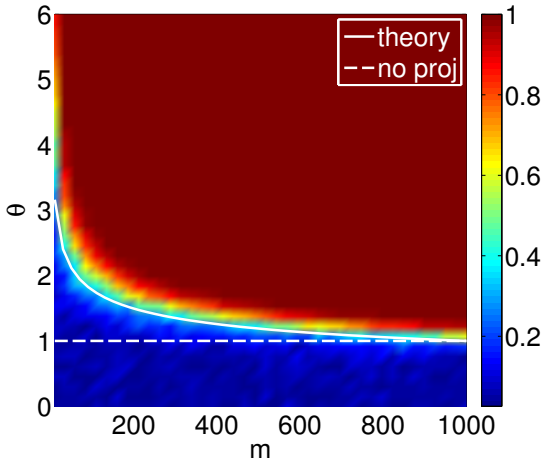
We see that the unitary projection matrix performs uniformly better than the Gaussian projection matrix above the phase transition. Importantly, even when setting $m = n$ so that the projection doesn't reduce the dimension, the unitary projection keeps the same phase transition while the Gaussian projection changes the phase transition so that it is harder to detect the presence of a signal. In terms of detection performance, the unitary projection matrix is preferred over the Gaussian projection matrix. However, we must consider the cost for generating these projection matrices, particularly for large dimensions. Generating the Gaussian projection matrix is very easy as every entry is an independent Gaussian random variable. However, generating a $n \times m$ unitary matrix for high dimensions may be prohibitive. The analysis in this paper gives the practitioner the ability to choose the projection matrix that best fits his or her needs. Given system parameters, the practitioner can select the projection dimension m to achieve a certain detection ability. The decision may be driven by the ease of creating each projection matrix.

IV. Conclusion

In this paper we considered detecting low-rank multidimensional signals given multiple noisy snapshots. Motivated by the computational cost of taking SVDs on large scale datasets, we explored using the SVD of a projection of our data matrix to smaller dimensional space. We analyzed the fundamental detection limits of two specific choices of projection matrix, unitary and Gaussian. Through numerical simulations, we



(a) KS Statistic - N,θ sweep



(b) KS Statistic - m,θ sweep

Fig. 4. Accuracy of theoretical phase transition in Corollary II.1 for unitary projection matrix Q and Gaussian noise matrix X for a rank-1 setting with fixed $n = 1000$. Figure plots the KS statistic between singular values generated from 500 signal bearing and 500 noise only matrices. (a) Sweeps over both θ and N for a fixed $m = 100$ (b) Sweeps over θ and m for a fixed $N = 1000$.

verified our theoretical predictions that identified a phase transition in detection ability. Below a critical SNR threshold, the largest singular value of the resulting projected matrix cannot be used to infer the presence of a signal in the dataset. Importantly, we observed that this critical threshold is lower for the unitary projection matrix than the Gaussian projection matrix, implying that the unitary projection matrix may be used to detect signal at lower SNRs than the Gaussian projection matrix.

ACKNOWLEDGMENT

This work was supported by ONR Young Investigator Award N000141110660, ONR Award N00014-15-1-2141, AFOSR Young Investigator Award FA9550-12-1-0266, NSF award CCF-1116115, ARfO MURI grant W911NF-11-1-0391, and ARO MURI grant W911NF-15-1-0479.

REFERENCES

[1] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, p. 1617, 2007.

[2] F. Benaych-Georges and R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Adv. in Math.*, 2011.

[3] N. Asendorf and R. Nadakuditi, "The performance of a matched subspace detector that uses subspaces estimated from finite, noisy, training data," *Signal Processing, IEEE Trans. on*, vol. 61, no. 8, pp. 1972–1985, 2013.

[4] 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*, vol. 111, pp. 120–135, 2012.

[5] M.-A. Belabbas and P. J. Wolfe, "Fast low-rank approximation for covariance matrices," in *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*. IEEE, 2007, pp. 293–296.

[6] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing qr factorization," *SIAM Journal on Scientific Computing*, vol. 17, no. 4, pp. 848–869, 1996.

[7] M. Rudelson and R. Vershynin, "Sampling from large matrices: An approach through geometric functional analysis," *Journal of the ACM (JACM)*, vol. 54, no. 4, p. 21, 2007.

[8] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation."

[9] V. Rokhlin, A. Szlam, and M. Tygert, "A randomized algorithm for principal component analysis," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1100–1124, 2009.

[10] N. Halko, P.-G. Martinsson, Y. Shkolnisky, and M. Tygert, "An algorithm for the principal component analysis of large data sets," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2580–2594, 2011.

[11] D. Achlioptas and F. Mcsherry, "Fast computation of low-rank matrix approximations," *Journal of the ACM (JACM)*, vol. 54, no. 2, p. 9, 2007.

[12] S. Arora, E. Hazan, and S. Kale, "A fast random sampling algorithm for sparsifying matrices," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2006, pp. 272–279.

[13] E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, "Randomized algorithms for the low-rank approximation of matrices," *Proceedings of the National Academy of Sciences*, vol. 104, no. 51, pp. 20 167–20 172, 2007.

[14] P. Ramachandra and M. Sartipi, "Compressive sensing based imaging via belief propagation," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 254–256.

[15] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[16] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[17] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[18] N. R. Rao and A. Edelman, "The polynomial method for random matrices," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 649–702, 2008.