Phase transition for eigenvalues and recovery of rank one matrices

Enrico Au-Yeung
Department of Mathematical Sciences
DePaul University, Chicago, IL 60614 USA
Email: eauyeun1@depaul.edu

Greg Zanotti
Department of Mathematical Sciences
DePaul University, Chicago, IL 60614, USA
Email: gregzanotti@gmail.com

***Abstract*—In datasets where the number of parameters is fixed and the number of samples is large, principal component analysis (PCA) is a powerful dimension reduction tool. However, in many contemporary datasets, when the number of parameters is comparable to the sample size, PCA can be misleading. A closely related problem is the following: is it possible to recover a rank-one matrix in the presence of a large amount of noise? In both situations, there is a phase transition in the eigen-structure of the matrix.**

Keywords: principal component analysis, low rank matrix recovery

I. Introduction

The problem of low-rank matrix recovery has received a lot of attention in the signals processing community over the last 10 years. The practical nature of this problem has motivated many researchers to investigate efficient methods to solve this optimization problem. See, e.g., [6], [13], [10]. This list is by no means exhaustive.

In this paper, we take a different direction. We are interested in the following type of situation: Is it possible to recover a rank-one matrix in the presence of a large amount of noise? A useful data model to keep in mind is the following: $X = \lambda \mathbf{x}\mathbf{x}^{\mathbf{T}} + G$, where $\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$. The data matrix X represents our observations, and the Gaussian matrix G represents the noise structure. The challenge is to recover the principal vector $\mathbf{x}$ and the value λ from the data matrix X. We are especially interested in the asymptotic behaviour of the largest eigenvalue and the leading eigenvector of the data matrix X, (as $n \to \infty$), when the operator norm of G is not negligible compared to the operator norm of X. We observe a phase transition phenomenon.

We are also interested in the behaviour of the leading singular vector when the data matrix is a large rectangular matrix, where the number of rows is proportional to the number of columns. Principal component analysis (PCA) is a versatile tool in dimensionality reduction. PCA projects the data onto the principal subspace spanned by the leading eigenvectors of the sample covariance matrix. In theory, these eigenvectors can capture most of the variance in the data. This enables the dimension of the feature space to be reduced, while retaining most of the information. In the contemporary setting, a collection of high-dimensional data can be treated as a low-rank signal with additional noise structure. If the samples of data are organized into a data matrix, then PCA can be used to recover the low-rank signal. It performs well when the number of features, p, is small, and the number of samples n is large. However, in biomedical studies, the number of features p is often comparable to the sample size n. In the biomedical setting, the features are measurements on the expression levels of thousands of genes, and n is the thousands of individuals.

A. Setting and Motivation

Suppose we have a collection of independently and identically distributed random vectors, $x_1, x_2, x_3, \ldots, x_n$ from a p-dimensional real Gaussian distribution with mean zero and covariance $\Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M, 1, 1, \ldots, 1)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M > 1$. Let X be the $p \times n$ matrix with column vectors $x_1, x_2, \ldots, x_n \in \mathbb{R}^p$. Assume that $0 < c < 1$ and $\frac{p}{n} = c$. Let $S = \frac{1}{n} X X^T$ be the sample covariance matrix. A data scientist wants to know how the largest eigenvalues of the matrix S behave as $n \to \infty$. Let us consider a specific scenario, when $p = 500$ and $n = 2000$, is the sample largest eigenvalue $\widehat{\lambda}_1$ of the matrix S a good estimator of the true eigenvalue λ_1? That depends on the true value of the largest eigenvalue λ_1.

The following is a simplified version of a theorem of Baik and Silverstein (see [4], [14]):

Theorem I.1. *Let $\widehat{\lambda}_1$ be the largest eigenvalue of S. (1). Suppose $\lambda_1 \leq 1 + \sqrt{c}$ and $\frac{p}{n} \to c$ as $n \to \infty$. Then, we have*

$$\widehat{\lambda}_1 \to (1 + \sqrt{c})^2 \quad \text{as } n \to \infty.$$

(2). Suppose $\lambda_1 > 1 + \sqrt{c}$ and $\frac{p}{n} \to c$ as $n \to \infty$. Then, we have

$$\widehat{\lambda}_1 \to \lambda_1 (1 + \frac{c}{\lambda_1 - 1}) \text{ as } n \to \infty.$$

What these authors observe is that there is a phase transition in the eigen-structure of a matrix when both the rows and columns are large, i.e. when $\frac{p}{n} \to c, 0 < c < 1$ and $n \to \infty$. The phase transition phenomenon can be quite complicated and this has been analyzed in the seminal paper [3]. For other variations on this theme, see, e.g. [15], [9], and [17].

Given a true signal in the form of an n-dimensional unit vector $\mathbf{x}$ called the *spike*, we can define the spiked Wigner model: observe $Y = \lambda \mathbf{x}\mathbf{x}^{\mathbf{T}} + \frac{\mathbf{1}}{\sqrt{\mathbf{n}}}\mathbf{W}$, where W is an $n \times n$ random symmetric matrix with entries drawn i.i.d. (up to symmetry) from a fixed distribution with mean 0 and variance 1. The parameter λ represents the signal-to-noise ratio (SNR). In the model, the detection problem is the following statistical question: For what values of the SNR is it possible to consistently test, with probability $1 - o(1)$ as $n \to \infty$, between a random matrix drawn from the spiked distribution and one from the un-spiked distribution? The detection problem has been explored by many researchers, see, e.g. [8], [7], [18], [12], [2], [1], [19]. For a recent development in the spiked Wigner model, see [16].

In the data model, $X = \lambda \mathbf{x_1}\mathbf{x_1}^{\mathbf{T}} + G$, suppose we have some additional information about the principal vector $\mathbf{x}$, how can we use that information to recover the vector? We consider the case when each entry of the vector is bounded between 0 and a fixed constant τ. To be precise, we can take the specific value, $\tau = 0.2$. Thus, we know that the vector $\mathbf{x_1}$ lies in a box. The value of τ does not affect the conclusion of the main theorem but it may affect the speed of convergence of an optimization algorithm used to recover the vector $\mathbf{x_1}$.

The leading eigenvector of the matrix X can be computed using the Power Method or more sophisticated variants (see [11]). Numerical experiments show that, when $\lambda = 4$, the leading eigenvector $\mathbf{v_1}$ of the data matrix X is not a good approximation to the desired vector $\mathbf{x_1}$. In fact, the relative error between $\mathbf{v_1}$ and $\mathbf{x_1}$ often exceeds one hundred percent.

The purpose of this paper is to address both the theoretical and practical aspect of this problem. On the theory side, we observe a phase transition in the largest eigenvalue. Moreover, the result shows that, depending on the true value of λ_1, the leading eigenvector of the data matrix X can be nearly orthogonal to the true vector $\mathbf{x_1}$. This means that some caution is warranted: when $n \to \infty$, using principal component analysis as an attempt to retrieve the vector $\mathbf{x_1}$ can give a misleading result. In place of a proof to the theorem, we provide an analytical explanation that gives the main insight to the theorem.

On the practical side, we develop an iterative algorithm to recover the vector $\mathbf{x_1}$ from the matrix X. We view this as a box-constrained optimization problem to find a vector $\mathbf{x}$, where the unknown variable satisfies the constraint, $0 \leq \|\mathbf{x}\|_\infty \leq \tau$. Compared to the leading eigenvector $\mathbf{v_1}$ of X, our algorithm yields a vector that is significantly closer to the desired vector $\mathbf{x_1}$. But first, we need to set some Notations:

- X is a symmetric random matrix, $X \in \mathbb{R}^{n \times n}$
- $\mathbf{x_1}$ is a fixed (non-random) vector, $\|\mathbf{x_1}\|_2 = 1$, and $\mathbf{x_1} \in \mathbb{R}^n$
- G is a Gaussian symmetric matrix, $G \in \mathbb{R}^{n \times n}$ and $G = G^T$, where $G(i,j)$ are independent, normally distributed with mean 0 and variance $\frac{1}{n}$ for $i < j$, and $G(i,i)$ is normally distributed with mean 0 and variance $\frac{2}{n}$

The following theorem is the phase transition phenomenon (for symmetric matrices).

Theorem I.2. *Let $X = \lambda \mathbf{x_1}\mathbf{x_1}^{\mathbf{T}} + G$, where G is Gaussian symmetric matrix. Pick $\tau = 0.2$.*
Suppose $\mathbf{x_1}$ is a fixed vector of length 1, and $0 \leq \mathbf{x_1}(j) \leq \tau$, for $1 \leq j \leq n$.
Let $\widehat{\lambda}_1$ be the largest eigenvalue of the matrix X. Let $\mathbf{v_1}$ be the leading eigenvector of the matrix X, i.e. $\mathbf{v_1}$ is the eigenvector that corresponds to $\widehat{\lambda}_1$.
Then, if $\lambda > 1$, we have

$$\lim_{n \to \infty} |\langle \mathbf{v_1}, \mathbf{x_1} \rangle| = \sqrt{c},$$

where $c = 1 - \frac{1}{\lambda^2}$. Otherwise, if $\lambda \leq 1$, we have

$$\lim_{n \to \infty} |\langle \mathbf{v_1}, \mathbf{x_1} \rangle| = 0.$$

For the largest eigenvalue of the matrix X, the following phase transition occurs. If $\lambda \geq 1$, we have

$$\widehat{\lambda}_1 \to \lambda + \frac{1}{\lambda} \tag{I.1}$$

as $n \to \infty$. Otherwise, if $\lambda \leq 1$, we have $\widehat{\lambda}_1 \to 2$ as $n \to \infty$.

Note: After the initial preparation of an earlier version of this manuscript, we learned that this is a version of a theorem of Florent Benaych-Georges and Raj Rao Nadakuditi, see [5]. We thank the authors of that paper for bringing this to our attention. In their theorem [5], they do not need the hypothesis that the entries of the vector $\mathbf{x_1}$ are between 0 and τ. We include this additional condition, since we can use it to improve convergence in our numerical optimization algorithm.

II. Background for Main Theorem

The symmetric Gaussian random matrix in our first result is an example of a Wigner matrix. We summarize here some background and standard facts regarding the Wigner Semicircular Law for symmetric random matri-

ces. Given any probability measure μ on the real line, the Stieltjes transform is defined by

$$S_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(t)}{z-t},$$

where z is any complex number in the upper half of the complex plane. For any $n \times n$ symmetric matrix M_n, we can work with the normalized matrix $\frac{1}{\sqrt{n}} M_n$ and form its empirical spectral distribution (ESD),

$$\mu_{\frac{1}{\sqrt{n}} M_n}(x) = \frac{1}{n} \sum_{j=1}^{n} \delta\left(x - \frac{\lambda_j(M_n)}{\sqrt{n}}\right)$$

of M_n, where $\lambda_j(M)$ are the eigenvalues of M_n. The ESD is a probability measure, also known as the spectral measure for the matrix. For the square matrix M_n with spectral measure $\mu(x) = \mu_{M_n}(x)$, we can define its corresponding Stieltjes transform. We have the following useful identity,

$$S_n(z) = S_{\mu_{\frac{1}{\sqrt{n}} M_n}}(z) = \frac{1}{n} Tr\left[\left(\frac{1}{\sqrt{n}} M_n - z I_n\right)^{-1}\right]$$

where Tr denotes the trace of a matrix, and I_n is the $n \times n$ identity matrix. We define the semicircular distribution $\mu_{sc}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$. The Wigner semicircular law states that the sequence of ESDs $\mu_{\frac{1}{\sqrt{n}} M_n}(x)$ converges almost surely to $\mu_{sc}(x)$. The Stieltjes transform for the spectral measure μ_{sc} is

$$S_{\mu_{sc}}(z) = \int_{\mathbb{R}} \frac{d\mu_{sc}(x)}{x - z} = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

III. Main insight for Main Theorem

We now give the analytical explanation for the quantity $\lambda + \frac{1}{\lambda}$ that appears in equation (I.1) in the phase transition phenomenon of Theorem I.2.

Recall that $X = \lambda \mathbf{x_1 x_1}^T + G$, where G is Gaussian symmetric matrix. Since G is symmetric, we can write $G = U^T D U$, where U is an orthogonal matrix and $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix. Instead of the matrix X, we can consider the matrix $UXU^T = D + \lambda U \mathbf{x_1 x_1}^T U^T$, i.e. a diagonal matrix D plus a rank one positive definite matrix $P \equiv \lambda U \mathbf{x_1 x_1}^T U^T$. The random orthogonal matrix U rotates the fixed vector $\mathbf{x}$ of length one to a random vector $\mathbf{u}$ of length one. The intuition is that when n is large, then with high probability, the vector $\mathbf{u}$ is uniformly distributed on the unit sphere $\{x \in \mathbb{R}^n \colon \|x\|_2 = 1\}$. Hence, each entry $u(k)$ of the unit-length vector $\mathbf{u}$ is approximately equal to the square root of $1/n$. Fix z and suppose the matrix $(D - zI_n)$ is invertible. Then, we have the relation,

$$\det(zI_n - (D+P)) = \det(zI_n - D) \cdot \det(I_n - (zI_n - D)^{-1} P). \tag{III.1}$$

Consider the matrix $M \equiv (zI_n - D)^{-1} P$. Then, 1 is an eigenvalue of the matrix M if and only if z is not an eigenvalue of D and z is an eigenvalue of $D + P$. Since the matrix $M = (zI_n - D)^{-1} \lambda \mathbf{u u^T}$ has rank one, we know that the trace of M is equal to the only nonzero eigenvalue of M. On the other hand, we have

$$Tr(M) = \lambda \sum_{k=1}^{n} \frac{|u(k)|^2}{z - \lambda_k}.$$

This implies that z is not an eigenvalue of D and z is an eigenvalue of $D + P$ if and only if

$$\lambda \sum_{k=1}^{n} \frac{|u(k)|^2}{z - \lambda_k} = 1. \tag{III.2}$$

Here, $u(k)$ are the entries of the vector $\mathbf{u}$. The left hand side of (III.2) is $\lambda S_{\mu_n}(z)$, where μ_n represents a weighted spectral measure associated to the diagonal matrix D,

$$\mu_n(x) = \sum_{k=1}^{n} |u(k)|^2 \cdot \delta(x - \lambda_k).$$

Recall that when n is large, the square of each entry $u(k)$ of the vector $\mathbf{u}$ is about $1/n$, with high probability. Thus, we replace the previous relation (III.2) with

$$\frac{1}{n} \sum_{k=1}^{n} \frac{1}{z - \lambda_k} = \frac{1}{\lambda}. \tag{III.3}$$

But the left hand side of equation (III.3) converges to the Stieltjes transform of the semicircular distribution, so we conclude that

$$S_{\mu_{sc}}(z) = \frac{1}{\lambda}.$$

Inverting the Stieltjes transform, we have $z = S^{-1}_{\mu_{sc}}\left(\frac{1}{\lambda}\right)$. By direct calculation, we can verify that

$$S^{-1}_{\mu_{sc}}\left(\frac{1}{\lambda}\right) = \lambda + \frac{1}{\lambda}.$$

Finally, since z is an eigenvalue of $D + P$, we have shown that $\lambda + \frac{1}{\lambda}$ is indeed an eigenvalue of $D + P$. This completes our analytical explanation for the appearance of $\lambda + \frac{1}{\lambda}$ in the phase transition of eigenvalues for symmetric matrices.

IV. Optimization Algorithm

Our optimization algorithm is based on gradient descent, but includes an additional projection to satisfy our assumptions reduce the increased variance caused by the additive Gaussian noise. We want to solve the following optimization problem, where $\mathbf{X}$ is the observed, noise corrupted matrix:

$$\begin{aligned} \min_{\mathbf{x}} \quad & ||\mathbf{X} - \mathbf{x}\mathbf{x}^T||_F + \text{Tr}(\mathbf{x}\mathbf{x}^T) \\ \text{s.t.} \quad & 0 \le \mathbf{x}(i) \le \tau \quad \forall i \in \{1, 2, ..., N\} \\ & ||\mathbf{x}||_2 = 1. \end{aligned} \tag{IV.1}$$

(Here, $\tau = 0.2$ is a fixed parameter). As the additive Gaussian noise increases the magnitude of the eigenvalues of the observed matrix $\mathbf{X}$, we perform gradient descent on an estimate of the true gradient (i.e. for $\mathbf{x_1 x_1^T}$ instead of $\mathbf{X}$) by penalizing the magnitude of the eigenvalues of our recovered matrix: that is, we penalize the trace of $\mathbf{x}\mathbf{x}^T$ by adding as a penalty term the L_2 norm of $\mathbf{x}$. Our update equation is as follows:

$$\mathbf{x_{k+1}} = \mathbf{x_k} - \alpha[\mathbf{x_k^T}(\mathbf{x_k x_k^T} - \mathbf{X}) + \gamma(\mathbf{x_k^T x_k})\mathbf{1^T}]^\mathbf{T} \tag{IV.2}$$

Above, α is the usual step size, which we experimentally observe to work best when set in the range $[10^{-3}, 10^{-1}]$, with slightly better recovery results toward the lower end of the range. The second parameter γ is the regularization parameter for the L_2 norm of $\mathbf{x}$, which we set to 10^{-1} and do not change. Our gradient descent continues until the following termination condition is met, which is usually satisfied within roughly 50 iterations when $\alpha = 0.1$:

$$\frac{||\mathbf{x_{k+1}} - \mathbf{x_k}||_2}{||\mathbf{x_{k+1}}||_2} \leq 10^{-5}. \tag{IV.3}$$

We initialize $\mathbf{x}$ by generating a length N vector of uniform random numbers in $[0, \tau]$ and then dividing it by its L_2 norm.

This gradient descent procedure, however, does not account for the main constraint in our optimization problem: the box constraint. To satisfy this constraint, after gradient descent reaches the termination condition above, we apply a projection step that mitigates the effect of the additional additive noise in the off-diagonal entries of the observed matrix $\mathbf{X}$. First, we divide $\mathbf{x}$ by its L_2 norm. Then, we project $\mathbf{x}$ onto the box $[0, \tau]^N$ by setting each $\mathbf{x}(i) = \min(\max(\mathbf{x}(i), 0), \tau)$. We repeat this alternating projection until the constraints are satisfied to a precision of 10^{-5}. As both the unit circle and the $[0, \tau]$ box are convex, this procedure is guaranteed to converge.

In our experiments, we initialize the true vector $\mathbf{x_1}$ by setting a block of 2% of the entries to $1/\sqrt{2(10^{-2})N}$ and dividing it by its L_2 norm. We keep $\alpha = 10^{-1}$. The observed data $\mathbf{X}$ is the rank-one matrix $\lambda \mathbf{x_1 x_1^T}$, plus a Gaussian random matrix G, as described in the previous section, with $\lambda = 4$. We define the relative error as:

$$\mathrm{E}(\mathbf{x}) = 100 \cdot \frac{||\mathbf{x_1} - \mathbf{x}||_2}{||\mathbf{x_1}||_2} \tag{IV.4}$$

where $\mathbf{x}$ is our recovered vector. Regardless of our selection of α in the range above, the standard deviations of the relative error at each N are consistently below 2% for $N \geq 500$. For sizes of $\mathbf{x_1}$ ranging from 500 to 5000, we observe that the average relative error for the recovered vector using our optimization is substantially lower in comparison to that of using the leading eigenvector. Average relative error is computed over 200 trials (i.e. draws of G and optimization procedures) per N. The results of our experiment are shown in Table I.

TABLE I
MEAN RELATIVE ERROR: "OPT" IS OUR OPTIMIZATION PROCEDURE AND "EIG" IS THE TOP EIGENVECTOR PROCEDURE.

n	Opt Mean(E)	Eig Mean(E)
500	15.4%	113.5%
1000	14.1%	107.4%
2500	12.4%	123.9%
5000	10.3%	111.8%

REFERENCES

[1] Ery Arias-Castro, Sebestien Bubeck, and Gabor Lugosi. Detection of correlations. *Ann. Statist.*, 40(1):412–435, 2012.

[2] Ery Arias-Castro, Emmanuel J. Candes, and Arnaud Durand. Detection of an anomalous cluster in a network. *Ann. Statist.*, 39(1):278–304, 2011.

[3] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[4] Jinho Baik and Jack Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006.

[5] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. in Math.*, 227:494–521, 2011.

[6] J.F. Cai, Emmanuel J. Candes, and Z.W. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20:1956–1982, 2010.

[7] T. Tony Cai, Jiashun Jin, and Mark G. Low. Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.*, 35(6):2421–2449, 2007.

[8] David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.

[9] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Comm. Math. Phys.*, 272(1):185–228, 2007.

[10] Massimo Fornasier, Holger Rauhut, and Rachel Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. Optim.*, 21(4):1614–1640, 2011.

[11] G. Golub and C. van Loan. *Matrix computations (4. ed.)*. Johns Hopkins University Press, 2013.

[12] Yuri I. Ingster, Alexandre B. Tsybakov, and Nicholas Verzelen. Detection boundary in sparse regression. *Electron. J. Stat.*, (4):1476–1526, 2010.

[13] Karthik Mohan and Maryam Fazel. Iterative reweighted algorithms for matrix rank minimization. *J. Mach. Learn. Res.*, 13:3441–3473, 2012.

[14] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642, 2007.

[15] Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006.

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

[17] Alessandro Pizzo, David Renfrew, and Alexander Soshnikov. On finite rank deformations of wigner matrices. *Ann. Inst. Henri Poincare Probab. Stat.*, 49(1):64–94, 2013.

[18] Xing Sun and Andrew B. Nobel. On the size and recovery of submatrices of ones in a random binary matrix. *J. Mach. Learn. Res.*, (9):2431–2453, 2008.

[19] Xing Sun and Andrew B. Nobel. On the maximal size of large-average and anova-fit submatrices in a gaussian random matrix. *Bernoulli*, 19(1):275–294, 2013.