Journal of Multivariate Analysis 111 (2012) 120–135

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

The singular values and vectors of low rank perturbations of large rectangular random matrices

Florent Benaych-Georges [a,b], Raj Rao Nadakuditi [c,*]

[a] LPMA, UPMC Univ Paris 6, Case courier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France
[b] CMAP, École Polytechnique, route de Saclay, 91128 Palaiseau Cedex, France
[c] Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 9 March 2011
Available online 3 May 2012

AMS 2000 subject classifications:
15A52
46L54
60F99

Keywords:
Random matrices
Haar measure
Free probability
Phase transition
Random eigenvalues
Random eigenvectors
Random perturbation
Sample covariance matrices

ABSTRACT

In this paper, we consider the singular values and singular vectors of finite, low rank perturbations of large rectangular random matrices. Specifically, we prove almost sure convergence of the extreme singular values and appropriate projections of the corresponding singular vectors of the perturbed matrix.

As in the prequel, where we considered the eigenvalues of Hermitian matrices, the non-random limiting value is shown to depend explicitly on the limiting singular value distribution of the unperturbed matrix via an integral transform that linearizes rectangular additive convolution in free probability theory. The asymptotic position of the extreme singular values of the perturbed matrix differs from that of the original matrix if and only if the singular values of the perturbing matrix are above a certain critical threshold which depends on this same aforementioned integral transform.

We examine the consequence of this singular value phase transition on the associated left and right singular eigenvectors and discuss the fluctuations of the singular values around these non-random limits.

1. Introduction

In many applications, the $n \times m$ signal-plus-noise data or measurement matrix formed by stacking the m samples or measurements of $n \times 1$ observation vectors alongside each other can be modeled as

$$\widetilde{X} = \sum_{i=1}^{r} \sigma_i u_i v_i^* + X, \tag{1}$$

where u_i and v_i are left and right 'signal' column vectors, σ_i are the associated 'signal' values and X is the noise-only matrix of random noises. This model is ubiquitous in signal processing [51,47], statistics [40,2,34] and machine learning [36] and is known under various guises as a signal subspace model [48], a latent variable statistical model [35], or a probabilistic PCA model [50].

Relative to this model, a common application-driven objective is to estimate the signal subspaces Span$\{u_1, \dots, u_r\}$ and Span$\{v_1, \dots, v_r\}$ that contain signal energy. This is accomplished by computing the *singular value decomposition* (*SVD*,

* Corresponding author.
E-mail addresses: florent.benaych@upmc.fr (F. Benaych-Georges), rajnrao@umich.edu, rajnrao@eecs.umich.edu (R.R. Nadakuditi).
URLs: http://www.cmapx.polytechnique.fr/~benaych/ (F. Benaych-Georges), http://www.eecs.umich.edu/~rajnrao/ (R.R. Nadakuditi).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.04.019

henceforth) of $\widetilde{X}$ and extracting the r largest singular values and the associated singular vectors of $\widetilde{X}$—these are referred to as the r *principal components* [46] and the Eckart–Young–Mirsky theorem states that they provide the best rank-r approximation of the matrix $\widetilde{X}$ for any unitarily invariant norm [24,39]. This theoretical justification along with the fact that these vectors can be efficiently computed using now-standard numerical algorithms for the SVD [28] has led to the ubiquity of the SVD in applications such as array processing [51], genomics [1,52], wireless communications [25], information retrieval [27], to list a few [37,23].

In this paper, motivated by emerging high-dimensional statistical applications [33], we place ourselves in the setting where n and m are large, r is known (or provided by an oracle) and the SVD of $\widetilde{X}$ is used to form estimates of $\{\sigma_i\}$, $\{u_i\}_{i=1}^r$ and $\{v_i\}_{i=1}^r$. We provide a characterization of the relationship between the estimated extreme singular values of $\widetilde{X}$ and the underlying (or latent) 'signal' singular values σ_i (and also the angle between the estimated and true singular vectors).

In the limit of large matrices, the extreme singular values only depend on integral transforms of the distribution of the singular values of the noise-only matrix X in (1) and exhibit a phase transition about a critical value; this critical value depends on integral transforms which arise from rectangular free probability theory [10,11]. The phase transition in the singular value is a new manifestation of the so-called *BBP phase transition*, named after the authors of the seminal paper [5] that first brought into focus this phenomenon for the eigenvalues of a special class of 'spiked' Wishart or sample covariance matrices. In this paper, we also characterize the fluctuations of the singular values about these asymptotic limits. The results obtained are precise in the large matrix limit and, akin to our results in [17], go beyond answers that might be obtained using matrix perturbation theory [49].

Our results are very general in terms of possible distributions for the noise model X, in a sense that which will be made more precise shortly; consequently, our theorems yield as a special case, results found in the literature for the eigenvalues [5,6] and eigenvectors [32,44,42] of $\widetilde{X}\widetilde{X}^*$ in the setting where X in (1) is Gaussian. For the Gaussian setting, we provide a new characterization for the right singular vectors, or equivalently, the eigenvectors of $\widetilde{X}^*\widetilde{X}$.

Such results had already been proved in the particular case where X is a Gaussian matrix, but our approach brings to light a general principle, which can be applied beyond the Gaussian case. Roughly speaking, this principle says that for X a $n \times m$ matrix (with $n, m \gg 1$), if one adds an independent small rank perturbation $\sum_{i=1}^r \sigma_i u_i v_i^*$ to X, then the extreme singular values will move to positions which are approximately the solutions z of the equations

$$\frac{1}{n}\operatorname{Tr}\frac{z}{z^2I - XX^*} \times \frac{1}{m}\operatorname{Tr}\frac{z}{z^2I - X^*X} = \frac{1}{\theta_i^2}, \quad (1 \le i \le r).$$

In the case where these equations have no solutions (which means that the θ_i's are below a certain threshold), then the extreme singular values of X will not move significantly. We also provide similar results for the associated left and right singular vectors and give limit theorems for the fluctuations. These expressions provide the basis for the parameter estimation algorithm developed by Hachem et al. in [29].

The papers [17,15] considered the eigenvalues of finite rank perturbations of Hermitian matrices. We employ the strategy developed in these papers for our proofs in this paper. Specifically, we derive master equation representations that implicitly encode the relationship between the singular values and singular vectors of X and $\widetilde{X}$ in terms of the low-rank perturbing matrix. We then employ concentration results to obtain the stated analytical expressions. Of course, because of these similarities in the proofs, we chose to focus, in the present paper, in what differs from [17,15].

At a certain level, our proof also present analogies with the ones of other papers devoted to other occurrences of the BBP phase transition, such as [45,26,20–22,41]. We mention that the approach of the paper [16] could also be used to consider large deviations of the extreme singular values of $\widetilde{X}$.

This paper is organized as follows. We state our main results in Section 2 and provide some examples in Section 3. The proofs are provided in Sections 4–7 with some technical details relegated to the Appendix.

2. Main results

2.1. Definitions and hypotheses

Let X_n be a $n \times m$ real or complex random matrix. Throughout this paper we assume that $n \le m$ so that we may simplify the exposition of the proofs. We may do so without loss of generality because in the setting where $n > m$, the expressions derived will hold for X_n^*. Let the $n \le m$ singular values[1] of X_n be $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$. Let μ_{X_n} be the empirical singular value distribution, *i.e.*, the probability measure defined as

$$\mu_{X_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\sigma_i}.$$

Let m depend on n—we denote this dependence explicitly by m_n which we will sometimes omit for brevity by substituting m for m_n. Assume that as $n \longrightarrow \infty$, $n/m_n \longrightarrow c \in [0, 1]$. In the following, we shall need some of the following hypotheses.

[1] Recall that for $n \le m$, the *singular values* of an $n \times m$ matrix X are the eigenvalues of the $n \times n$ matrix $\sqrt{XX^*}$.