Electronic Journal of Statistics
Vol. 10 (2016) 74–80
ISSN: 1935-7524
DOI: 10.1214/15-EJS1019

Discussion of "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation"*

Debashis Paul†

Department of Statistics
University of California
Davis, CA 95616, USA
e-mail: debpaul@ucdavis.edu

and

Lili Wang

Department of Mathematics and Statistics
University of Melbourne
Parkville, Vic 3010, Australia
e-mail: liliwang@zju.edu.cn

Abstract: In this discussion, we present a brief overview of recent works on the behavior of summary statistics for high-dimensional observations that are time-dependent, and the inference on parameters associated with high-dimensional time series, with emphasis on covariance and auto-covariance matrices.

AMS 2000 subject classifications: 62H99.
Keywords and phrases: Covariance matrix, principal component analysis, spiked covariance model, Stieltjes transform.

Received March 2015.

1. Introduction

We commend authors Cai, Ren and Zhou for this fine article that provides a comprehensive review of the burgeoning literature on the estimation of covariance and precision matrices from high-dimensional data. The main focus of their article is on the different methods for estimating structured covariance and precision matrices, where the assumed structure is usually sparse in an appropriate sense, and on the analytical tools for deriving the optimal risk bounds for those statistical procedures. In this respect, they were exceptionally thorough and the paper provides a fine technical guide for dealing with these complex problems.

*Main article 10.1214/15-EJS1081.

†Research partially supported by NSF grants DMS-1106690 and DMS-1407530.

In this context, the authors also point out the link between the estimation of the covariance and that of the behavior of the eigenvalues and eigenvectors for high-dimensional sample covariance matrices. Results on the latter have been exploited in detecting number of significant components in a factor model, and in deriving hypothesis tests for covariance matrices. In particular, important insights about the limitations of classical PCA for high-dimensional data is derived from the phase transition results on the behavior of the leading sample eigenvalues and eigenvectors.

One of the scenarios which Cai and coauthors could potentially touch upon is the behavior of summary statistics for high-dimensional observations that are time-dependent, and the inference on quantities such as autocovariance matrices and autoregressive or moving average coefficients for high-dimensional time series. These problems pose challenges associated with both dimensionality and the additional structures governed by temporal dependence. In this discussion, we give a brief overview of some recent works related to these two topics, and present two results on asymptotic behavior of eigenvalues of sample covariance and autocovariance matrices for a class of high-dimensional time series. The latter results demonstrates that there are considerable similarities and some important differences in the phenomena associated with i.i.d. observations and time-dependent data. It is hoped that this discussion will illustrate the scope of further research in the direction of statistical inference for structured high-dimensional time series. It would be interesting to know if Professor Cai and coauthors have any suggestion on formulating inference procedures in any of the regularized estimation schemes mentioned at the end of Section 2 of this discussion.

2. Brief review of inference on high-dimensional time series

Explorations of asymptotic behavior of sample covariance and autocovariance matrices of high-dimensional time series have started relatively recently. Credit for an early work in this direction, in the context of a factor model, goes to Onatski ([11]) who extended the phase transition results for the leading eigenvalues and corresponding eigenvectors of the sample covariance matrix based on i.i.d. observations from a "spiked covariance model" (see, e.g., [2, 4, 12]) to the set up of a time-dependent factor model with "weak factors". He assumed that the factor scores are independent of the isotropic background noise, while the dimension p and sample size n both increase to infinity such that p/n converges to a finite positive constant. A slightly simplified version of this result appears in [8] who assumed that the finite dimensional factor scores have a joint Gaussian distribution with a "weak" correlation across the observations, e.g., when the factor scores form a stationary vector autoregressive process with short range dependence. The assumption that the background noise is independent across time means that the time series characteristic is not dominant in these analyses.

As the analyses in the i.i.d. case by [4] and [12] show, asymptotic behavior of the empirical spectral distribution (ESD), i.e., the empirical distribution of the eigenvalues, form an important component in describing the behavior of leading eigenvalues and eigenvectors of sample covariance matrices. Among the early works in this direction, [9] proved existence of a limiting ESD of the symmetrized sample autocovariance matrix of any given lag order $\tau \geq 0$, defined as

$$\mathbf{S}_\tau := \frac{1}{2(n-\tau)} \sum_{t=1}^{n-\tau} (X_t X_{t+\tau}^* + X_{t+\tau} X_t^*), \tag{2.1}$$

where X_t are i.i.d. p-dimensional observations with independent entries having zero mean, unit variance and bounded fourth moments, and $p/n \to c \in (0, \infty)$ as $p, n \to \infty$. They used this result to test for the presence of factors in a dynamic factor model. [10] extended this result to the setting of a linear process of the form

$$X_t = Z_t + \sum_{l=1}^{\infty} \mathbf{A}_l Z_{t-l} \tag{2.2}$$

under the following conditions. (**A0**) $\{Z_t\}$ is a sequence of p-dimensional vectors with i.i.d. coordinates that have zero mean, unit variance and finite fourth moment; the moving average coefficients $\{\mathbf{A}_l\}$ satisfy (**A1**) $\mathbf{A}_l = \mathbf{A}_l^* = \mathbf{U} D_l \mathbf{U}^*$, for all $l \geq 1$, for a unitary/orthogonal matrix $\mathbf{U}$ and diagonal matrices D_l with j-th diagonal element $f_l(\alpha_j)$; where (**A2**) the functions $f_l : \mathbb{R}^m \to \mathbb{R}$ are bounded and continuous; (**A3**) the empirical distribution of $\{\alpha_j\}_{j=1}^p$ converge to a distribution in $\mathbb{R}^m$, denoted by $F^{\mathcal{A}}$; and (**A4**) $\sum_{l=1}^{\infty} l^2 \parallel f_l \parallel_\infty < \infty$.

In moderately large dimensional settings, i.e., when $p, n \to \infty$ such that $p/n \to 0$, Wang [15] explored the fluctuations of the sample autocovariances from their population counterparts. Specifically, she proved the existence of the limiting ESD of the matrices of the form $\sqrt{n/p}(\mathbf{S}_\tau - \Gamma_\tau)$ under the model studied by [10], where Γ_τ is the lag-τ autocovariance matrix of the process. We state the main result in [15] below, which is described in terms of Stieltjes transforms. Note that, for a measure μ on $\mathbb{R}$, the Stieltjes transform of μ is defined as $s_\mu(z) = \int \frac{d\mu(x)}{x-z}$, where $z \in \mathbb{C}^+$, where $\mathbb{C}^+$ denotes the upper half of the complex plane. They are used extensively in random matrix theory to characterize the convergence of ESDs, see, for example [1] for an overview.

Theorem 2.1. *Suppose that the process $\{X_t\}$ is given by (2.2), and satisfy (**A0**)–(**A3**) and (**A4′**) $\sum_{l=1}^{\infty} l^4 \parallel f_l \parallel_\infty < \infty$ and (**A5**) the functions $\{f_l\}$ are uniformly Lipschitz. Further, assume that $n, p \to \infty$ such that $p/n \to 0$. Then, for each integer $\tau \geq 0$, with $\mathbf{S}_\tau$ defined through (2.1) and Γ_τ denoting lag-τ population autocovariance matrix, the ESD of $\sqrt{n/p}(\mathbf{S}_\tau - \Gamma_\tau)$ converges almost surely to a nonrandom distribution, whose Stieltjes transform is given by s_τ and satisfies the equation*

$$s_\tau(z) = -\int \frac{dF^{\mathcal{A}}(a)}{z + \beta_\tau(z,a)}, \qquad z \in \mathbb{C}^+ \tag{2.3}$$

where the Stieltjes kernel $\beta_\tau(z,a)$ is determined by

$$\beta_\tau(z,a) = -\int \frac{\mathcal{R}_\tau(a,b) dF^{\mathcal{A}}(b)}{z + \beta_\tau(z,b)}, \qquad z \in \mathbb{C}^+, a \in \mathbb{R}^m \tag{2.4}$$

in which the kernel $\mathcal{R}_\tau(a,b) := \frac{1}{2\pi}\int_0^{2\pi}\cos^2(\tau\theta)\psi(a,\theta)\psi(b,\theta)d\theta$*, where* $\psi(a,\theta) = |1+\sum_{l=1}^{\infty} f_l(a)e^{il\theta}|^2$*,* $\theta \in [0,2\pi]$*.*

Theorem 2.1, like the result by [10], demonstrates the complex interplay between the dimensional and temporal correlation structures, and also serves to highlight some inherent difficulties in extending results from the i.i.d. to the time-dependent settings. The limiting distribution in Theorem 2.1 can be seen as a certain generalization of the well-known semi-circle law. This result can be used to test hypotheses about certain specific restrictions on the time series. Details will be provided in a forthcoming paper ([16]).

There have been several works extending the scope of existing approaches and deriving rates of convergence of estimators for covariance and precision matrices under time dependence of the observations, while others deal with regularized estimation of parameters directly linked to the time series structure. For example, [7] investigated the banded and tapered estimates of high-dimensional covariance matrices when the observations are weakly dependent and showed that these estimators remain consistent in the operator norm with appropriate rates of convergence under suitable classes of models. [17] investigated the behavior of the sparse PCA estimator proposed by [14] when the observations follow a stationary first order vector autoregressive process. [13] considered the problem of jointly estimating multiple graphical models, through estimation of corresponding precision matrices, when data are collected from n subjects, each of which consists of observations forming a first order autoregressive process whose parameters change smoothly across subjects. In related works, [5] and [6] considered sparse estimation of the autoregression coefficients of a high-dimensional vector autoregressive model under different forms of sparsity constraints.

3. Phase transition phenomena in a VAR model

We present our analysis of a problem where the observations constitute a stationary time series with a first order vector autoregressive process, with i.i.d. and isotropic innovations. The key assumption is that the autoregression coefficient is symmetric and of low rank, while the dimensionality p of the process increases with sample size n in such a way that $p/n \to c \in (0,1)$. We are interested in studying the behavior of the leading eigenvalues of the sample covariance matrix. In this respect, the problem is closely related to the one studied by [17]. Our result, a simplified version of which is presented here, demonstrates the precise interplay of signal-to-noise and p/n ratios in determining the detection limit of the signal, i.e., the presence of the autoregressive term.

Let X_t be the p-dimensional process be defined by

$$X_t = \mathbf{A}X_{t-1} + Z_t, \qquad t \in \mathbb{Z} \tag{3.1}$$

where $\mathbf{A} = \mathbf{A}^T$ and $\{Z_t\}$ is a sequence of real-valued i.i.d. random variables having $N(0, I_p)$ distribution. We assume that $\|\mathbf{A}\| < 1$ to ensure that the process $\{X_t\}$ is stationary with short range dependence. Suppose further that $\text{rank}(\mathbf{A}) = M$, a fixed nonnegative integer ($<p$). Indeed, $M = 0$ means that

the process $\{X_t\}$ is i.i.d. white noise with covariance equal to I_p. Note that, the assumptions on $\mathbf{A}$ imply that there exists a $p \times p$ orthogonal matrix $\mathbf{U}$ such that the process $Y_t := \mathbf{U}^T X_t$ can be expressed as

$$Y_t = \begin{bmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Theta Y_{t-1}^{(1)} + \epsilon_t^{(1)} \\ \epsilon_t^{(2)} \end{bmatrix} \tag{3.2}$$

where $\{\epsilon_t^{(1)}\}$ are i.i.d. $N(0, I_M)$, $\{\epsilon_t^{(2)}\}$ are i.i.d. $N(0, I_{p-M})$, and Θ is an $M \times M$ diagonal matrix with diagonal elements $\theta_1 \geq \cdots \geq \theta_M$ satisfying the constraint $\max_{1 \leq j \leq M} |\theta_j| < 1$. It should be noted that $\Sigma := \mathrm{Var}(X_t) = (I - \mathbf{A}^2)^{-1}$, and that $\mathrm{Var}(Y_t^{(1)}) = (I - \Theta^2)^{-1}$. Thus, Σ can be seen as a rank M perturbation of the identity matrix, which falls in the class of "spiked" covariance matrices studied by [4] and [12], though under i.i.d. setting. We suppose that we have a sequence of realizations $X_1, \ldots, X_n$ from such processes such that $p, n \to \infty$ and $p/n \to c \in (0,1)$. Let $\mathbf{S}$ denote the sample covariance matrix $\mathbf{S} := \frac{1}{n}\sum_{t=1}^n X_t X_t^T$, and let $\hat{\lambda}_j$ denote the j-th largest eigenvalue of $\mathbf{S}$. Then, we have the following eigenvalue phase transition result.

Theorem 3.1. *Under the assumptions above, as $n, p \to \infty$ such that $p/n \to c \in (0,1)$, the following holds.*

(i) For all $1 \leq j \leq M$ such that $|\theta_j| \leq (\frac{\sqrt{c}}{1+\sqrt{c}})^{1/2}$, $\hat{\lambda}_j \to (1+\sqrt{c})^2$ almost surely.

(ii) For all $1 \leq j \leq M$ such that $|\theta_j| > (\frac{\sqrt{c}}{1+\sqrt{c}})^{1/2}$, $\hat{\lambda}_j \to \rho_j$ almost surely, where $\rho_j = \beta_j(1 + \frac{c}{\beta_j - 1})$ with $\beta_j = (1-\theta_j^2)^{-1}$. Moreover, if in addition, $\sqrt{n}|p/n - c| \to 0$, and θ_j has multiplicity 1, then $\sqrt{n}(\hat{\lambda}_j - \rho_j)$ converges in distribution to $N(0, \sigma^2(\theta_j, c))$, for an appropriate $\sigma^2(\theta_j, c) > 0$.

Theorem 3.1 parallels the main result in [12] and can be proved by making use of the same technique, while taking into account the fact that $Y_t^{(1)}$ is a stationary Gaussian process. The key feature is that, only if the time dependence in a coordinate of $\{Y_t^{(1)}\}$ is sufficiently strong, as is indicated by the magnitude of θ_j exceeding the threshold $(\sqrt{c}/(1+\sqrt{c}))^{1/2}$, the corresponding eigenvalue of $\mathbf{S}$ can be distinguished from the largest noise eigenvalue. It should be noted further that it is the structure of the process as given by (3.2), namely, up to a rotation, the large dimensional, i.i.d., isotropic noise and the (possibly non-isotropic and time-dependent) finite dimensional signal are independent, that determines the phenomena, rather than the precise nature of the time dependence. This result can be extended to the setting of stationary linear processes (2.2) with symmetric, commuting coefficients $\{\mathbf{A}_l\}$, as studied by [10], if we assume further that the j-th diagonal elements of the matrices $D_l = \mathbf{U}^T \mathbf{A}_l \mathbf{U}$ are zero for $j > M$, for some finite M, where $\mathbf{U}$ is a $p \times p$ orthogonal matrix. The results can be generalized further to the setting of linear processes with non-symmetric coefficients, provided the time dependent component of the process belongs to a finite dimensional subspace and is independent of the corresponding orthogonal component which forms an i.i.d. process. In the case of the AR(1) model, it is conjectured that the when $|\theta_j| < (\sqrt{c}/(1+\sqrt{c}))^{1/2}$, after appropriate normalization, and scaling of $n^{2/3}$, $\hat{\lambda}_j$ will have a Tracy-Widom type limiting distribution, as is observed in the spiked covariance model with i.i.d. observations (see, e.g., [3] for corresponding results in the complex-valued case).

References

[1] Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer. MR2567175

[2] Bai, Z. D. and Yao, J.-F. (2011). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, **106**, 167–177. MR2887686

[3] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, **33**, 1643–1697. MR2165575

[4] Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, **97**, 1382–1408. MR2279680

[5] Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, **43**(4), 1535–1567. *arXiv:1311.4175*. MR3357870

[6] Basu, S., Shojaie, A. and Michailidis G. (2015). Network Granger causality with inherent grouping structure. *Journal of Machine Learning Research*, **16**, 417–453. MR3335801

[7] Bhattacharjee, M. and Bose, A. (2014). Consistency of large dimensional sample covariance matrix under weak dependence. *Statistical Methodology*, **20**, 11–26. MR3205718

[8] Chen, K., Paul, D. and Wang, J.-L. (2009). Properties of principal component analysis for correlated data. Technical Report, University of California, Davis.

[9] Jin, B., Wang, C., Bai, Z. D., Nair, K. K. and Harding, M. C. (2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *The Annals of Applied Probability*, **24**, 1199–1225. MR3199984

[10] Liu, H., Aue, A. and Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *The Annals of Statistics* (to appear). MR3319140

[11] Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**, 244–258. MR2923766

[12] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, **17**, 1617–1642. MR2399865

[13] Qiu, H., Han, F., Liu, H. and Caffo, B. (2013). Joint estimation of multiple graphical models from high dimensional time series. *arXiv:1311.0219*.

[14] Vu, V. and Lei, J. (2012). Minimax rates of estimation for sparse pca

in high dimensions. *International Conference on Artificial Intelligence and Statistics*.

[15] WANG, L. (2014). *Topics on Spectral Analysis of Covariance Matrices and Its Application*. Ph.D. Thesis, Zhejiang University.

[16] WANG, L., AUE, A. and PAUL, D. (2015). LSD of normalized sample autocovariance matrices of linear time series with commutating coefficients when $p/n \to 0$. Manuscript under preparation.

[17] WANG, Z., HAN, F. and LIU, H. (2013). Sparse PCA for high dimensional vector autoregressive models. *arXiv:1307.0164*.