Eigenrange: A Robust Spectral Method for Dimensionality Reduction

Malika Kharouf, Tabea Rebafka, Nataliya Sokolovska

Eigenrange: A Robust Spectral Method for Dimensionality Reduction

Malika Kharouf, Tabea Rebafka, Nataliya Sokolovska

***Abstract*—This paper addresses the problem of dimension reduction of noisy data, more precisely the challenge of determining the dimension of the subspace where the signal lives in. Based on results from random matrix theory, a novel estimator of the signal dimension is developed. Consistency of the estimator is proved in the modern asymptotic regime, where the number of parameters grows proportionally with the sample size. Experimental results show that the novel estimator is robust to noise and, moreover, it gives highly accurate results in settings where standard methods fail. The application of the new dimension estimator on several biomedical data sets in the context of classification illustrates the improvements achieved by the new method compared to the state of the art.**

***Index Terms*—Dimensionality reduction, sample eigenvalues, random matrix theory, spiked population model.**

I. Introduction

DIMENSIONALITY reduction aims at separating signal from noise in order to preserve significant properties of data in a low-dimensional space before analyzing them by further statistical methods. Data representation in a lower dimension is needed in many applications, where the number of observed features or parameters has considerably increased mainly due to technical advances. Nevertheless, increasing the number of features does not automatically increase the dimension of the subspace where the signal lives in. Indeed, in many applications as in speech recognition [1], wireless communications [2], hyperspectral imaging [3], chemometrics [4], medical imaging [5], genomics [6] or mathematical finance [7] the signal space dimension is much lower than the number of observed parameters. Thus, a challenge is to determine the low-dimensional signal space, in order to project the data onto it, for instance by principal component analysis (PCA) [8]. In this context, a fundamental question is how to determine an optimal minimal dimension of a high-dimensional problem. A major difficulty in real data sets is the presence of noise, making the estimation of the signal space rather involved.

A number of methods to determine an optimal dimension have been proposed, an overview is provided in [9]. The most prominent method uses the number of principal components that are necessary to explain a given part of the total variance. Also widely used in practice the scree test or kink method which relies on the detection of an elbow in the scree graph, that is the plot of ordered sample eigenvalues. Both methods are rather heuristical. Further methods are the SURE method [10], [11], model-order selection in a Bayesian framework [5], [12] or maximum-likelihood based approaches [13]. Although reporting reasonable performance, most of them lack theoretical explanations.

Another recent approach is based on eigengaps, that is the distance between consecutive sample eigenvalues (for both white [14] and colored [15] noise). The novelty of this method is that a sound mathematical foundation is provided stemming from results in random matrix theory. Notably consistency in the modern asymptotic regime is proved, when both the sample size *and* the number of parameters tend to infinity. This is most relevant for applications where the number of parameters is comparable with or even larger than the number of observations.

Despite theoretical guarantees, the eigengap method suffers from the fact that it relies on local features of the scree graph. If a single sample eigenvalue is badly estimated, the dimension estimate may be very erroneous. Indeed, in practice high accuracy is only obtained when the signal eigenvalues are well separated.

The purpose of this paper is to propose a method that is robust to the presence of similar or even identical signal eigenvalues, while preserving the strong theoretical properties of the eigengap method. This is achieved by a more global look on the sample eigenvalues. It is noteworthy that the consistency of the proposed eigenrange method does not depend on strong distributional assumptions like normality as it is the case of the maximum-likelihood approaches and others. The eignerange method performs very competitively on real data.

We mention that another important part of dimensionality reduction approaches based on random matrix theory is developed for time-series, what is out of the scope of the current paper [7], [16], [17].

The paper is organised as follows. Section II introduces our approach and provides the theoretical foundations. A numerical comparison of the new method to the state of the art is reported in Section III. Section IV discusses the performance of the eigenrange method in the context of classification on real biomedical challenges. Concluding remarks and perspectives close the paper.

II. Eigenrange method

In this section we introduce the mathematical framework of the new eigenrange method and we show its consistency when the noise level is known. Then the eigenrange method is adapted to unknown noise, which is most relevant for applications.

M. Kharouf is with Institut Charles Delaunay, CNRS UMR 6281, Université de technologie de Troyes, France, e-mail: malika.kharouf@utt.fr

T. Rebafka is with LPMA, CNRS UMR 7599, Université Pierre et Marie Curie, Paris, France, e-mail: tabea.rebafka@upmc.fr

N. Sokolovska is with INSERM UMR_S 1166, NutriOmics Team, Université Pierre et Marie Curie, Paris, France, e-mail: nataliya.sokolovska@upmc.fr

A. Spiked population model

We consider the additive noise model, where the signal vector $\mathbf{s}$ is corrupted by some additive white noise noise $\mathbf{e}$. That is, the observed random vector $\mathbf{y} \in \mathbb{R}^p$ is defined as

$$\mathbf{y} = \mathbf{s} + \mathbf{e}. \tag{1}$$

The random vectors $\mathbf{s}$ and $\mathbf{e}$ are supposed to be independent and the noise $\mathbf{e}$ has zero mean and covariance $\sigma^2\mathbf{I}_p$, $\sigma^2 > 0$. Denoting the signal's covariance matrix by R_s, the covariance of the observation vector $\mathbf{y}$ verifies $R_y = R_s + \sigma^2\mathbf{I}_p$.

Often the signal $\mathbf{s}$ is a linear combination of a relatively small number of predictors, i.e. $\mathbf{s} = \mathbf{B}\mathbf{x}$ with some $(p \times r)$ matrix $\mathbf{B}$ and $r < p$. In other words, the signal lives in a proper subspace of $\mathbb{R}^p$ of dimension r. To separate signal from noise, data may be compressed to this smaller subspace. The signal space dimension r is also given by the rank of the signal's covariance matrix R_s such that $r = \text{rank}(R_s)$.

Denote the non zero eigenvalues of R_s by $\alpha_1 > \cdots > \alpha_r > 0$ and the eigenvalues of R_y by $\lambda_1 \geq \cdots \geq \lambda_p > 0$. In this model, which is also referred to as the spiked population model, the eigenvalues verify

$$\lambda_l = \begin{cases} \alpha_l + \sigma^2, & l = 1, \ldots, r \\ \sigma^2, & l > r \end{cases}$$

The first r eigenvalues $\lambda_1, \ldots, \lambda_r$ are called spikes and they are larger than the nonspiked ones. As a consequence, the eigenvalues of R_y yield important information on the signal dimension r.

Now let the observations $y_1, \ldots y_n$ be n i.i.d. realizations of $\mathbf{y}$. Then the sample covariance matrix is $\hat{R}_y = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^t$ where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. The associated sample eigenvalues are denoted by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and we use them to construct a new consistent estimator of the signal dimension r.

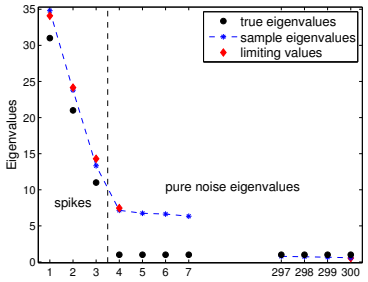
B. Eigenrange method when σ^2 is known

When many features are observed (i.e. p of the order of n), traditional asymptotic results, where the sample size n grows, while the number of features p is fixed, may be bad descriptions of what happens on finite samples. Generally, results in the modern regime, where both n and p tend to infinity, provide much better approximations of the finite sample situation.

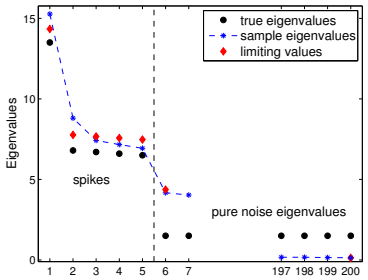
Concerning the estimation of the eigenvalues λ_1 by the sample eigenvalues $\hat{\lambda}_1$, notable works including [18] state consistency in the traditional regime, where only n tends to infinity. However, consistency no longer holds when both n and p grow. Recent advances in random matrix theory provide convergence results in the modern asymptotic regime.

In the pure noise case, where $\mathbf{y} = \mathbf{e}$ and $R_y = \sigma^2\mathbf{I}_p$, the seminal work of Marchenko and Pastur [19] shows that, when $p/n \to c$, all limiting values ϕ_l of the sample eigenvalues $\hat{\lambda}_l$ lie within the interval $[a, b] := [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2]$, which is the support of the so-called Marcenko-Pastur law.

In the additive noise model, the nonspiked sample eigenvalues still tend to lie in the Marchenko-Pastur interval $[a, b]$, while the limits of the spikes are outside [20]. This result holds under the assumption that spikes are sufficiently different



(a) $r = 3$, $\alpha = [30, 20, 10]$, $\sigma^2 = 1$
$c = 3$, $n = 100$, $p = 300$



(b) $r = 5$, $\alpha = [12, 5.3, 5.2, 5.1, 5]$, $\sigma^2 = 1.5$
$c = 0.5$, $n = 400$, $p = 200$

Fig. 1. Illustration of spikes and pure noise eigenvalues, and of the bias of the sample eigenvalues: eigenvalues λ_l (black dots), sample eigenvalues $\hat{\lambda}_l$ (blue stars), theoretical limits of the sample eigenvalues (red diamonds) as $p/n \to c$.

from pure noise eigenvalues. More formally, if $\alpha_l > \sigma^2\sqrt{c}$ for $l = 1, \ldots, r$ and $p/n \to c$, then for $l = 1, \ldots, r$

$$\hat{\lambda}_l \longrightarrow \phi_l = \alpha_l + \sigma^2\left(1 + c + \frac{c\sigma^2}{\alpha_l}\right) \quad a.s. \tag{2}$$

Moreover, the first and last pure noise eigenvalues satisfy

$$\hat{\lambda}_{r+1} \longrightarrow b = \sigma^2(1+\sqrt{c})^2, \ \hat{\lambda}_m \longrightarrow a = \sigma^2(1-\sqrt{c})^2 \quad a.s.$$

where $m = \min(n, p)$. It is clear that any procedure relying on sample eigenvalues must take into consideration their bias, that is the difference between the limiting values of the sample eigenvalues $\hat{\lambda}_l$ and the model eigenvalues λ_l. This bias is illustrated in Figure 1.

As a consequence, the range of the pure noise sample eigenvalues, $\hat{\lambda}_{r+1} - \hat{\lambda}_m$, is about $b - a$, while the distance $\hat{\lambda}_l - \hat{\lambda}_m$ for any $l = 1, \ldots, r$ is significantly larger. From this viewpoint, a natural estimator of the signal dimension r is derived as the number of sample eigenvalues that must be discarded such that the remaining eigenvalues are contained in an interval of approximate length $b-a$. Denote $\delta_l = \hat{\lambda}_{l+1} - \hat{\lambda}_m$ and consider a threshold of the form $b-a+d_n$. A new estimator of the signal space dimension r, referred to as the *eigenrange method*, is defined by

$$\hat{r}^{\text{range}} = \min\{l : \delta_l < b - a + d_n\}. \tag{3}$$

C. Consistency

The eigenrange estimator $\hat{r}^{\text{range}}$ is shown to be consistent for an appropriate choice of sequence d_n in (3). The proof relies

on the rates of convergence of the smallest and the largest pure noise sample eigenvalues when $p/n \to c$. According to [21] and [22], under some moment conditions,

$$n^{2/3}(\hat{\lambda}_{n,r+1} - b) = O_{\mathbb{P}}(1),\ n^{2/3}(\hat{\lambda}_{n,m_n} - a) = O_{\mathbb{P}}(1), \quad (4)$$

where the notation $X_n = O_{\mathbb{P}}(1)$ means that X_n is a stochastically bounded sequence. To ease the understanding of the asymptotics, in this section subscript n is added to all quantities depending on n.

Theorem 1. *Let the sequence d_n be such that $d_n \to 0$ and $n^{2/3}d_n \to \infty$ as $n \to \infty$. Suppose that the signal eigenvalues satisfy $\alpha_k > \sigma^2\sqrt{c}$ for $k = 1, \ldots, r$. Then under some moment conditions on* $\mathbf{s}$ *and* $\mathbf{e}$*, the eigenrange estimator $\hat{r}^{range}$ defined in* (3) *is a consistent estimator of the dimension of the signal space, that is, as $p/n \to c$,*

$$\hat{r}_n^{range} \longrightarrow r \quad \textit{almost surely}.$$

Proof. Without loss of generality let $\sigma^2 = 1$. Denote $\tilde{d}_n = b - a + d_n$. As $\delta_{n,l} > \delta_{n,l-1}$ for all l, we have

$$\{\hat{r}_n^{\text{range}} = r\} = \{\delta_{n,r} < \tilde{d}_n\} \cap \{\delta_{n,r-1} \geq \tilde{d}_n\},$$

implying that

$$\begin{aligned}\mathbb{P}(\hat{r}_n^{\text{range}} = r) &= 1 - \mathbb{P}\left(\{\delta_{n,r} \geq \tilde{d}_n\} \cup \{\delta_{n,r-1} < \tilde{d}_n\}\right) \\ &\geq 1 - \mathbb{P}(\delta_{n,r} \geq \tilde{d}_n) - \mathbb{P}(\delta_{n,r-1} < \tilde{d}_n).\end{aligned}$$

On the one hand, by (4) and as $n^{2/3}d_n \to \infty$,

$$\begin{aligned}&\mathbb{P}\left(\delta_{n,r} \geq \tilde{d}_n\right) \\ &= \mathbb{P}\left(n^{2/3}(\hat{\lambda}_{n,r+1} - b) - n^{2/3}(\hat{\lambda}_{n,m_n} - a) \geq n^{2/3}d_n\right) \longrightarrow 0.\end{aligned}$$

On the other hand, it holds that

$$\begin{aligned}\mathbb{P}(\delta_{n,r-1} < \tilde{d}_n) &= \mathbb{P}\left(\hat{\lambda}_{n,r} - \hat{\lambda}_{n,m_n} < d_n + b - a\right) \\ &= \mathbb{P}\left((\hat{\lambda}_{n,r} - \phi_r) - (\hat{\lambda}_{n,m_n} - a) - d_n < b - \phi_r\right) \\ &\longrightarrow 0,\end{aligned}$$

since $b - \phi_r = -\frac{(\alpha_r - \sqrt{c})^2}{\alpha_r} < 0$ and $(\hat{\lambda}_{n,r} - \phi_r) - (\hat{\lambda}_{n,m_n} - a) - d_n \xrightarrow{\mathbb{P}} 0$ by (2), (4) and as $d_n \to 0$. Combining all arguments yields $\mathbb{P}(\hat{r}_n^{\text{range}} = r) \to 1$. This completes the proof. □

We conducted an extensive simulation study to calibrate the sequence d_n and concluded that the best choice is given by

$$d_n = \sigma^2 \beta n^{-2/3} \log(\log(n)) \quad (5)$$

with $\beta = (1 + \sqrt{c})\,(1 + \sqrt{1/c})^{1/3}$. All experimental results presented in this paper rely on this particular choice.

D. Eigenrange method when σ^2 is unknown

In practice the noise level σ^2 is generally unknown, though required in (5) for the eigenrange estimator. Based on the fact that the mean of the pure noise sample eigenvalues $\hat{\lambda}_l$ is a consistent estimator of σ^2, we propose an iterative procedure (similar to that in [14]) to estimate both r and σ^2. More precisely, we alternate the estimation of σ^2 (using the current value r^{curr} of r) and the estimation of r by the eigenrange method (with σ^2 replaced by its current estimate $\hat{\sigma}^2_{\text{curr}}$) until convergence.

Algorithm 1 Eigenrange method with unknown σ^2

1: Initialize $r^{\text{curr}} = 0$.
2: Estimate σ^2 by $\hat{\sigma}^2_{\text{curr}} = \frac{1}{p - r^{\text{curr}}} \sum_{l=r^{\text{curr}}+1}^{p} \hat{\lambda}_l$.
3: Compute $\hat{r}^{\text{range}}$ where σ^2 is replaced with $\hat{\sigma}^2_{\text{curr}}$ in (5).
4: If $r^{\text{curr}} < \hat{r}^{\text{range}}$, update $r^{\text{curr}} = \hat{r}^{\text{range}}$ and return to 2, otherwise r^{curr} is the final estimate of r.

III. Numerical experiments

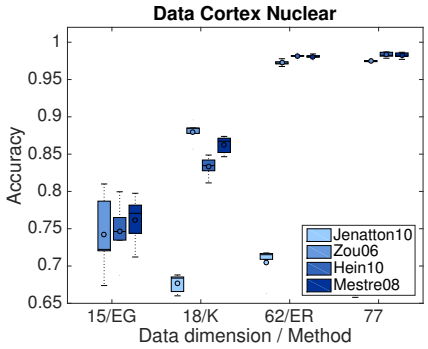
We compared the accuracy of our approach to traditional estimators of the signal dimension. A *standard method* based on PCA uses the number of principal components to explain a given part of the total variance in the data [9], [23]. The *kink method* or *scree test* analyses the scree graph [9]. As the pure noise eigenvalues $\lambda_{r+1}, \ldots, \lambda_m$ are constant, the scree graph has a (more or less pronounced) kink or elbow at the $(r+1)$-th sample eigenvalue, see Figure 1. Another approach relies on so-called eigengaps, that is the distance between consecutive eigenvalues. The *eigengap method*, as formalized in [14], is intended to identify the beginning of the flat part of the scree graph, made up of the pure noise eigenvalues. This method is shown to be consistent when $p/n \to c$.

We conducted simulations in the settings of Figure 1. The standard method completely fails to estimate r. In setting (a) the other three methods give all excellent results and their performances are very similar. Slight differences in accuracy only occur on very small data sets. However, in setting (b) the eigenrange method clearly outperforms the eigengap and the kink method. Indeed, in this setting the methods that focus on local features of the scree graph underestimate r, while the global approach of the eigenrange method works correctly. In this sense the eigenrange method is robust to the presence of similar spiked eigenvalues, whereas standard methods fail. For more details see the supplementary file.

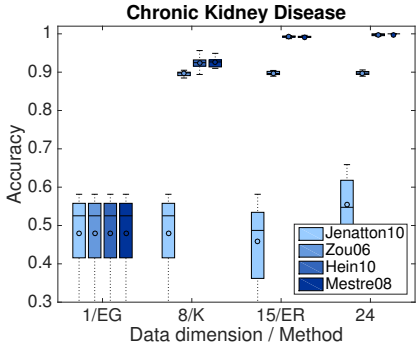
IV. Application to classification

To illustrate the importance of the accurate estimation of the signal space dimension, we show how the eigenrange method improves results in the context of the classification task. It is sometimes practical to first reduce the dimension of the data prior to any further data analysis. For instance, PCA can be used, where the number of principal components is chosen by an appropriate selection method [24], [9]. Then a supervised learning method, e.g. logistic regression or SVM, can be applied to the reduced data to perform classification.

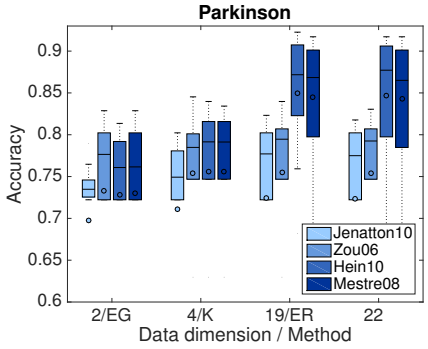
PCA is known to be problematic [25], mainly because sample eigenvectors are not consistent when the number of parameters p grows proportionally with the sample size n. Much effort has been spent to remedy this problem and there are mainly two ways to regain consistency. Several versions of PCA have been proposed that rely on sparsity assumptions, namely the structured sparse PCA [26] to which we refer as *Jenatton10*, sparse PCA that we call *Zou06* method [27], and the inverse power method applied to sparse PCA called *Hein10* [28]. A different approach, here referred to as *Mestre08*, relies on new consistent estimators of the eigenvectors when $p/n \to c$ with $c < 1$ [29]. The data used in our experiments are recent cohorts



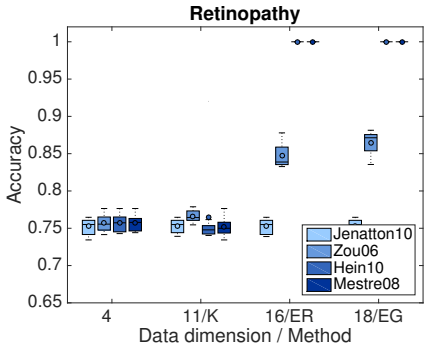
a) Mice protein expression data set



b) Chronic kidney disease data set



c) Parkinson data set



d) Diabetic retinopathy Debrecen cohort

Fig. 2. Classification accuracy on real data sets.

from the UCI machine learning repository. We have considered four life sciences data sets, namely a) mice protein expression data set [30] which is dedicated to mice classification into control or trisomic groups; b) chronic kidney disease data set[1] which classifies patients into ill and healthy; c) the Parkinson data set [31] which has also 2 classes, healthy or ill; and d) the diabetic retinopathy Debrecen cohort [32], where classes indicate whether an image contains signs of diabetic retinopathy or not. In our experiments, we compare the following: for the estimation of the signal dimension we apply the kink, the eigengap and the eigenrange method (K, EG, and ER on Figure 2). We also show results for a bigger dimension (the dimensions are provided on the plots) to illustrate the optimal accuracy, but for the Debrecen data set d) the eigengap method estimated the optimal rank to be very close to the full dimension of the original problem. So, in this case, we decided to also try a small dimension (which we fixed to 4). For the projection of the data on a subspace of smaller dimension, we compare the methods *Mestre08*, *Jenatton10*, *Zou06* and *Hein10*. We apply an SVM as a learning approach, and the optimal hyper-parameters are chosen by cross-validation. Rank and projections are learned from training data. From Figure 2 which shows the 10-fold cross validation test accuracy, we see that the eigenrange method achieves the optimal performance on all tested data sets.

[1]http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease#

V. Conclusion

Dimensionality reduction is a challenge, especially in applications where data are noisy. We proposed a highly accurate estimator of the signal dimension, based on global statistics involving sample eigenvalues, that outperforms state-of-the-art methods as kink and eigengap that focus on local features of the scree graph. Based on recent advances of random matrix theory, we formulated our main theoretical result summarized in Theorem 1. The simulated and real data experiments confirm our theoretical findings and, moreover, show the robustness in situations where the state-of-the-art methods fail, that is when spiked eigenvalues are close to each other. The novel eigenrange method is of great interest for applications where assumptions on the spiked eigenvalues or on the distribution of the observations are difficult to assert. A further advantage of the proposed method is that it is straightforward to implement, and it is not computationally expensive. We also underline that we do not waste any effort on potentially expensive optimisation.

Our method is not directly connected to classification, since the proposed dimensionality reduction procedure is completely unsupervised. However, we have tested the algorithm to reduce data before applying a prediction method, and we observed a clearly beneficial effect. It is promising to adapt the eigenrange approach to other methods that rely on the eigenstructure of a data matrix. In particular, we are currently investigating its extension to the problem of spectral clustering.

References

[1] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics*, vol. 23, no. 1, pp. 73–102, 02 1995.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *Special Issue of the IEEE Signal Processing Magazine*, vol. 19, pp. 17–28, 2002.

[4] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, no. 1, pp. 19 – 32, 2008.

[5] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Med. Imaging*, vol. 23, no. 2, pp. 137–152, 2004.

[6] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, no. 8, pp. 904–909, Aug. 2006.

[7] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 3, pp. 391–397, 2000.

[8] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.

[9] R. Cangelosi and A. Goriely, "Component retention in principal component analysis with application to cDNA microarray data," *Biology Direct*, vol. 2, no. 2, 2007.

[10] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with SURE and random matrix theory," *IEEE Trans. Signal Processing*, vol. 56, no. 12, pp. 5804–5816, 2008.

[11] ——, "Selecting the number of principal components with SURE," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 239–243, 2015.

[12] A. Seghouane and A. Cichocki, "Bayesian estimation of the number of principal components," *Signal Processing*, vol. 87, no. 3, pp. 562–568, 2007.

[13] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[14] D. Passemier and J. F. Yao, "On determining the number of spikes in a high-dimensional spiked population model," *Random Matrices : Theory and Applications*, vol. 1, 2012.

[15] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J.-Y. Tourneret, "Estimating the Intrinsic Dimension of Hyperspectral Images Using a Noise-Whitened Eigengap Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3811–3821, 2016.

[16] B. Rosenow, "Determining the optimal dimensionality of multivariate volatility models with tools from random matrix theory," *Journal of Economic Dynamics and Control*, vol. 32, no. 1, pp. 279–302, Jan. 2008.

[17] A. Onatski, "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economics and Statistics*, vol. 92, no. 4, pp. 1004–1016, November 2010.

[18] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122–148, 1963.

[19] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–486, 1967.

[20] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, pp. 1382–1408, 2006.

[21] R. Couillet and W. Hachem, "Fluctuations of spiked random matrix models and failure diagnosis in sensor networks," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 509–525, 2013.

[22] F. Benaych-Georges, A. Guionnet, and M. Maïda, "Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices," *Electronic Journal of Probability*, vol. 16, 2011.

[23] J. Harrington and S. Cassidy, *Techniques in Speech Acoustics*. New York: Springer Science and Business Media, New York, 2012.

[24] M. Ringnér, "What is principal component analysis?" *Nature Biotechnology*, vol. 26, no. 3, 2008.

[25] I. Johnstone and A. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 682–693, 2009.

[26] R. Jenatton, G. Obozinski, and F. Bach, "Structured sparse principal component analysis," in *AISTATS*, 2010.

[27] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[28] M. Hein and T. Buehler, "An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA," in *NIPS*, 2010.

[29] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, 2008.

[30] C. Higuera, K. J. Gerdiner, and K. J. Cios, "Self-organizing feature maps identify proteins critical to learning in a mouse model of down syndrome," *PloS One*, vol. 10, no. 6, 2015.

[31] M. A. Little et al., "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed Eng Online*, vol. 6, no. 23, 2007.

[32] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-based systems*, vol. 60, pp. 20–27, 2014.