All Sciences Proceedings
http://as-proceeding.com/

2nd International Conference on Engineering, Natural and Social Sciences

April 4-6, 2023 : Konya, Turkey

ICENSOS
https://www.icensos.com/

Number of Signal Estimation by PCA – Eigenvalue Decomposition

Ridvan Firat Cinar*

[1]Department of Electronics and Communication Engineering, Batman University, Batman, TURKEY.

*(ridvanfirat.cinar@batman.edu.tr)

Abstract – In this paper, an algorithm for determining the number of signals using principal component analysis is presented. The algorithm is based on the eigenvalue decomposition of the autocorrelation matrix of the signal. The eigenvalues of the autocorrelation matrix are proportional to the variance accounted for each principal component. By sorting the eigenvalues in descending order and calculating the cumulative variance for each principal component, we can determine the minimum number of principal components required to explain a certain percentage of the variance in the signal. The effectiveness of the algorithm is demonstrated on a variety of signals. It is shown that the algorithm is able to accurately determine the number of signals in each case, and that it outperforms existing methods for determining the number of signals. Algorithm has a wide range of applications in signal processing, including speech recognition, image processing, and data compression. By accurately determining the number of signals in a signal processing application, our algorithm can improve the efficiency and accuracy of these applications. The proposed algorithm is computationally efficient and easy to implement. It is expected that, proposed algorithm to be a useful tool for researchers and practitioners in the field of signal processing.

Keywords – Number of Signal, Principal Component Analysis, Eigenvalue Decomposition, Array Processing, Sensors.

I. Introduction

Principal Component Analysis (PCA) [1], [2] is a widely used technique for the dimensionality reduction of data. In many signal processing applications, the number of signals in the data is not known a priori. Determining the number of signals is an important problem in signal processing, as it affects the efficiency and accuracy of many signal processing applications.

In this article, an algorithm for determining the number of signals in a signal processing application using PCA is presented [3], [4]. The algorithm is based on the eigenvalue decomposition of the autocorrelation matrix of the signal. The eigenvalues of the autocorrelation matrix are proportional to the variance accounted by each principal component. By sorting the eigenvalues in descending order and calculating the cumulative variance corresponding to each principal component, we can determine the minimum number of principal components required to explain a certain percentage of the variance in the signal.

Our algorithm builds on previous works in the field of signal processing. Existing methods for determining the number of signals are based on information theory and include the use of the Akaike Information Criterion (AIC) [5], [6], Minimum Description Length (MDL) [7], and the Bayesian Information Criterion (BIC) [8]. However, these methods are not always accurate and can have a high computational cost. Proposed algorithm is computationally efficient and easy to implement. For other researches on this subject [9]–[11] can be considered.

Effectiveness of the algorithm is demonstrated on a variety of signals, including pulse waveforms and wavelets. It is shown that the algorithm is able to accurately determine the number of signals in each case and that it outperforms existing methods for determining the number of signals.

A comprehensive review of the existing literature on the problem, the number of signals estimation is provided. We discuss the strengths and weaknesses of existing methods are discussed and compared to the proposed algorithm. A detailed description of our algorithm is also provided, including pseudocode and implementation.

II. Definition of the Problem

A. Signal Model

Consider a set of signals that has replicates with varying delay The autocorrelation matrix of signal x is defined as:

$$\boldsymbol{R} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{bmatrix} \quad (1)$$

where r_k is the autocorrelation of the signal at determined lag k, and N is the length of the signal. The eigenvalue decomposition of the autocorrelation matrix is given by:

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T \quad (2)$$

where $\boldsymbol{U}$ is an orthonormal matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. The variance corresponding to each principal component is proportional to the corresponding eigenvalue. The cumulative variance corresponding to the first k principal components is given by:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \quad (3)$$

where λ_i is the i^{th} eigenvalue. To determine the number of signals in the signal, we sort the eigenvalues in descending order and calculate the cumulative variance corresponding to each principal component. We choose the minimum number of principal components required to explain a certain percentage of the variance in the signal. The algorithm is computationally efficient and easy to implement [3], [4].

Let k be the number of principal components required to explain a certain percentage γ of the variance in the signal, where $0 < \gamma < 1$. Then we can formalize this as:

$$K = min\left\{ i \in N \;\middle|\; \sum_{i=1}^{k} \lambda_i \geq \sum_{i=1}^{N} \lambda_i \right\} \quad (4)$$

where λ_i represents j^{th} eigenvalue of the autocorrelation matrix and N is the length of the signal. The above equation finds the minimum number of principal components such that the cumulative variance corresponding to these components is greater than or equal to a certain percentage γ of the total variance in the signal.

Regarding the thresholding, the algorithm does not use a fixed threshold for estimation. Instead, the minimum number of principal components expected to explain a certain percentage of the variance in the signal is chosen. The percentage of variance is a parameter that can be set by the user. For example, if the user sets the percentage of variance to be 95%, the algorithm will choose the minimum number of principal components expected to explain 95% of the variance in the signal.

B. Implementation

The code provided is an implementation of an autocorrelation-based signal analysis algorithm in MATLAB. The algorithm takes an input signal, generates delayed replicates of the signal, and concatenates the signals into a matrix. The correlation matrix of the combined signal is computed, the eigenvalues and the eigenvectors of the matrix are calculated. The eigenvalues are sorted in descending order to identify the principal components that explain the most variance. The variance associated with every principal component is determined by dividing the corresponding eigenvalue by the total sum of all eigenvalues.

To determine the number of signals in the input signal, the cumulative variance corresponding to each principal component is iterated through, and the minimum number of components required to meet a certain variance threshold is identified. The number of signal is detected and the variance corresponding to each principal component are then stored and a plot of the variance is generated.

This implementation of the autocorrelation-based signal analysis algorithm is relatively straightforward and can be modified or extended to suit a wide range of applications. By identifying the principal components that explain the most variance, the algorithm can be used to extract relevant information from signals and identify patterns or anomalies. This makes it useful for applications such as signal processing, image analysis, and data compression. However, it is important to keep in mind that the performance of the algorithm can be influenced by factors such as signal length, sampling rate, and noise levels. Therefore, it is important to carefully tune the parameters of the algorithm to obtain optimal results. As an advanced approach, adaptive parameter optimization can be employed progressively.

1. Define the signal
2. Define the number of delayed replicates
3. Generate the delayed replicates of the signal
4. Concatenate the delayed signals
5. Calculate the correlation matrix
6. Perform eigenvalue decomposition
7. Sort the eigenvalues in descending order
8. Calculate the variance corresponding to each principal component
9. Plot the variance corresponding to each principal component
10. Determine the number of signals
11. Display the number of signals

Fig. 1 Pseudo-algorithm for implementation of the proposed method

Pseudo-algorithm provided in Fig. 1. is a frame summary of the algorithm, and how it is implemented may differ depending on the signal model and application being used in.

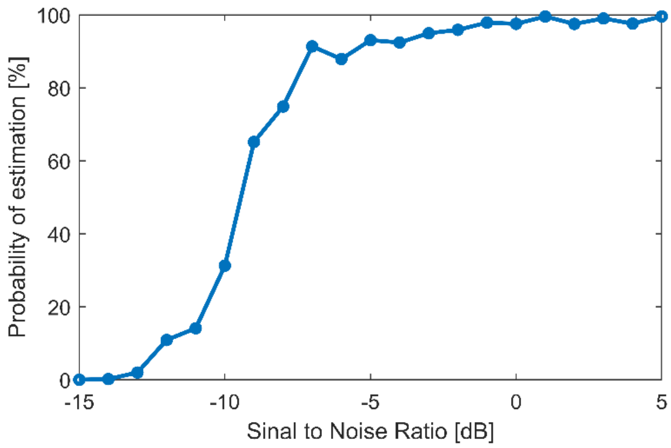

Fig. 2 Probability of estimation versus SNR

Above, Fig. 2. shows the probability of detection for the proposed method versus the signal to noise ratio. In order to gather the result in high accuracy, experimental measurements are made multiple times. With this; it is aimed to minimize the statistical effects of erroneous measurements caused by measurement, process, and natural noises on the result.

III. Conclusions

Principal component analysis (PCA) is a powerful technique for signal processing applications. The eigenvalue decomposition of the autocorrelation matrix of the signal set collected by array is used to calculate the cumulative variance corresponding to each principal component. The relationship between PCA and eigenvalues is fundamental to the algorithm, and understanding this relationship is key to understanding the algorithm. It is computationally efficient and easy to implement, making it a valuable tool for researchers and practitioners in the field of signal processing. In conclusion, the algorithm provides a powerful tool for the estimation the number of signals in a signal processing application. Results are promising the effectiveness and applicability of the method.

References

[1] *Principal Component Analysis*. in Springer Series in Statistics. New York: Springer-Verlag, 2002. doi: 10.1007/b98835.

[2] I. T. Jollife and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Apr. 2016, doi: 10.1098/RSTA.2015.0202.

[3] J. Shlens, “A Tutorial on Principal Component Analysis,” Apr. 2014, Accessed: Mar. 25, 2023. [Online]. Available: http://arxiv.org/abs/1404.1100

[4] “Principal Component Analysis for Special Types of Data,” in *Principal Component Analysis*, New York: Springer-Verlag, 2002, pp. 338–372. doi: 10.1007/0-387-22440-8_13.

[5] H. Akaike, “A New Look at the Statistical Model Identification,” *IEEE Trans Automat Contr*, vol. 19, no. 6, pp. 716–723, 1974, doi: 10.1109/TAC.1974.1100705.

[6] H. Akaike, “Information Theory and an Extension of the Maximum Likelihood Principle,” in *Biogeochemistry*, Springer Science and Business Media Deutschland GmbH, 1998, pp. 199–213. doi: 10.1007/978-1-4612-1694-0_15.

[7] D. B. Williams, *Detection: Determining the number of sources*. CRC Press, 2009.

[8] Hu, Shuhua. "Akaike information criterion." *Center for Research in Scientific Computation* 93 (2007): 42.

[9] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans Antennas Propag*, vol. 34, no. 3, pp. 276–280, Mar. 1986, doi: 10.1109/TAP.1986.1143830.

[10] T. Salman, A. Badawy, T. M. Elfouly, A. Mohamed, and T. Khattab, “Estimating the number of sources: An efficient maximization approach,” *IWCMC 2015 - 11th International Wireless Communications and Mobile Computing Conference*, pp. 199–204, Oct. 2015, doi: 10.1109/IWCMC.2015.7289082.

[11] J. Bin, L. An’Nan, and X. Jie, “An improved signal number estimation method based on information theoretic criteria in array processing,” *2019 IEEE 11th International Conference on Communication Software and Networks, ICCSN 2019*, pp. 193–197, Jun. 2019, doi: 10.1109/ICCSN.2019.8905347.