Journal of Statistics and Mathematics

Journal of Statistics and Mathematics
ISSN: 0976-8807 & E-ISSN: 0976-8815, Volume 2, Issue 2, 2011, PP-40-48
Online available at : http://www.bioinfo.in/contents.php?id=85

EFFECT OF NOISE IN PRINCIPAL COMPONENT ANALYSIS

KATERINA G. TSAKIRI[1*] AND IGOR G. ZURBENKO[2]
[1*]Division of Math, Science and Technology, Nova Southeastern University, 3301 College Avenue, Fort Lauderdale-Davie, FL 33314-7796, USA
[2]Department of Biometry and Statistics, School of Public Health, State University of New York at Albany, One University Place, Rensselaer, NY 12144, USA
*Corresponding author. E-mail: ktsakiri@gmail.com

Received: November 16, 2010; Accepted: October 19, 2011

Abstract- This paper demonstrates the effect of independent noise in principal components of k normally distributed random variables defined by a population covariance matrix. We prove that the principal components determined by a joint distribution of the original sample affected by noise can be essentially different in comparison with those determined from the original sample. However when the differences between the eigenvalues of the population covariance matrix are sufficiently large compared to the level of the noise, the effect of noise in principal components proved to be negligible. We support the theoretical results by using simulation study and examples. We also compare the results about the eigenvalues and eigenvectors in the two dimensional case with other models examined before. This theory can be applied in any field for the decomposition of the time series in multivariate analysis.
Keywords: principal components; eigenvalue; eigenvector; multivariate analysis

1. Introduction

Several studies have been developed for the investigation of the effect of noise on a population or a sample covariance matrix [1-3, 4-7, 8, 9, 10]. One approach is the analysis of the effect of noise on a sample covariance matrix, given that we have a finite number of observations [1, 9, 11]. Another approach is the analysis of the effect of noise on a population covariance matrix, assuming that we have a virtually infinite collection of observations [2-3, 5-7]. In particular, it has been analyzed the estimation of a covariance matrix of p random variables from n observations by either tapering or banding the sample covariance matrix, or estimated a banded version of the inverse of the covariance [1].

Analyzing the effect of noise on the population covariance matrix is essentially a problem in matrix perturbation theory [2-3, 5-7]. The presence of a small perturbation in a Hermitian matrix or a population covariance matrix has been described by theorems of Davis and Kahan [2-3, 6-7]. Moreover, the comparison of the results about the eigenvalues for the Jacobi's method and QR iteration when small relative errors in the eigenvalues of a matrix occur by small relative errors in its entries has been investigated by Demmel and Veselic [3]. Nadler by using a matrix perturbation approach study how close are the largest eigenvalues and eigenvectors of a sample covariance matrix $\mathbf{S_n}$ with those of a $p \times p$ population matrix $\mathbf{\Sigma}$ [6].

The effect of noise in the covariance matrix has also been analyzed in the field of principal component analysis [4, 9, 10-12]. Martin considers a type of probability based-principal component analysis, in which each of the n observations has a probability distribution in p-dimensional space centered on it [4, 11]. In one example, Martin considers identical and spherical distribution for each observation, so that the underlying covariance matrix has the form $\mathbf{\Sigma} + \sigma^2 \mathbf{I_p}$ [4, 11]. Further, Webb presents an approach to non linear principal components using spherically symmetric kernel functions [9]. Another method is the investigation of the effect of noise on random variables in the two dimensional case described by Zurbenko and Sowizral [10]. The authors show that the principal components can be changed dramatically when noise is present on the random variables in the two dimensional case [10]. However when the data are times series, an appropriate filter can be used to smooth out the noise and evaluate the true parameters of the model [10, 12].

In this paper, we examine the effect of noise in the field of principal component analysis (PCA) defined in more than two dimensions. PCA is a statistical tool in multivariate analysis with main subject the covariance matrix of the observed variables. Unfortunately, statistical models are always related with uncertainties regarding the determination of those matrices. Those uncertainties may be derived by influences into the variables from unobserved sources. Frequently, the observations are affected by different influences from separate scales [13] or the covariance matrices of the observed variables can slowly change in time. All such practical problems deal with covariance matrices which can be affected by certain hidden influences and not necessarily by noise. We may define "noise" what we can not explain directly from the corresponding variables. For clear theoretical formulation of the problem we add true stochastic noise to the original variables, and we investigate the level of the noise which may ruin the standard multivariate analysis.

Principal components (PC's) in different scales are usually different, and may be replaced arbitrarily. In this paper, we want to determine the level of this disturbance. For this reason, we keep the variables with fixed covariance structure and we add independent noise. Such a model allows the determination of the levels of extra influences so as to change completely the PC's. We investigate the PC's of the population covariance matrix defined by the original variables. If the population matrix is affected by the noise, then the sample covariance can only provide worse result for the examination of the PC's. If the PC's are replaced arbitrarily, then any other method of the factor analysis, such as the canonical correlation analysis, will provide false results, as well.

In this paper, we first present the results for the effect of noise on eigenvalues and eigenvectors of a population covariance matrix associated with a random vector **X**, and then we extend the results in principal component analysis [8]. We also compare those results with other models examined before in the two dimensional case [2-3, 6-7]. This theory can be applied when two random vectors are uncorrelated, and therefore the covariance matrices have completely different structures [8, 14].

2. The eigenvalues and eigenvectors of the noisy covariance matrix

For the following sections we denote by **X** a $k \times 1$ random vector associated with the covariance matrix $\mathbf{\Sigma_X}$. We also define by $\mathbf{X}^* = \mathbf{X} + \boldsymbol{\varepsilon}$ a noisy random vector, where $\boldsymbol{\varepsilon}$ is a random vector of the noise associated with the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$. Then the covariance matrix associated with $\mathbf{X}^*$ is denoted by $\mathbf{\Sigma}_{\mathbf{X}^*}$. The spectral decomposition of the covariance matrix $\mathbf{\Sigma_X}$ is given by:

$$\mathbf{\Sigma_X} = \lambda_1 \boldsymbol{e_1 e_1}' + \lambda_2 \boldsymbol{e_2 e_2}' + \ldots + \lambda_k \boldsymbol{e_k e_k}'$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of $\mathbf{\Sigma_X}$ such that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$, and $\boldsymbol{e_1, e_2, \ldots, e_k}$ are the associated normalized eigenvectors [11, 15-16]. For the following sections, we also assume that the eigenvalues are in decreasing order. Similarly it can be defined the spectral decomposition of the covariance matrix $\mathbf{\Sigma}_{\mathbf{X}^*}$.

The norm of the covariance matrix $\mathbf{\Sigma_X}$ can be defined by:

$$\|\mathbf{\Sigma_X}\| = \lambda_1 \qquad (1)$$

where λ_1 is the largest eigenvalue of the covariance matrix, $\mathbf{\Sigma_X}$.

We also define the norm of a $k \times 1$ random vector $\boldsymbol{x}$ as:

$$\|\boldsymbol{x}\| = (\boldsymbol{x}'\boldsymbol{x})^{1/2} = \left(x_1^2 + x_2^2 + \ldots + x_k^2\right)^{1/2} \qquad (2)$$

Moreover, the eigenvector can be determined with length unity [16]. Then Theorem 1 and Theorem 2 state the change in the noisy eigenvectors and principal components when the differences between the eigenvalues are sufficiently large compared to the maximum noise or disturbance [8].

Theorem 1 Let $(\lambda_1, \boldsymbol{e_1}),\ldots,(\lambda_k, \boldsymbol{e_k})$ be the spectral decomposition of $\mathbf{\Sigma_X}$ such that $|\lambda_i - \lambda_{i+1}| > i\sigma_\varepsilon^2$ for i=1,..,k-1, where $\|\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\| = \sigma_\varepsilon^2$. Then the spectral decomposition of $\mathbf{\Sigma}_{\mathbf{X}^*}$, $(\lambda_1^*, \boldsymbol{e_1^*}),\ldots,(\lambda_k^*, \boldsymbol{e_k^*})$, satisfies the conditions: $|\lambda_i^* - \lambda_i| \le \sigma_\varepsilon^2$ and $\|\boldsymbol{e_i^*} - \boldsymbol{e_i}\| \le i\sigma_\varepsilon$ for i=1,..,k.

Proof Let $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_k \end{pmatrix}'$ be a vector of noise associated with the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$. Assume that $(\lambda_1^*, \boldsymbol{e_1^*})$, $(\lambda_2^*, \boldsymbol{e_2^*}),\ldots,(\lambda_k^*, \boldsymbol{e_k^*})$ are the eigenvalues-eigenvectors pairs of the noisy covariance matrix, $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$. Then by definition (1), we have that:

$$|\lambda_1^* - \lambda_1| = \left| \|\mathbf{\Sigma}_{\mathbf{X}^*}\| - \|\mathbf{\Sigma_X}\| \right| \le \|\mathbf{\Sigma}_{\mathbf{X}^*} - \mathbf{\Sigma_X}\| = \|\mathbf{\Sigma}_\varepsilon\| = \sigma_\varepsilon^2$$

The change between the noisy and non-noisy largest eigenvalue is at most σ_ε^2, and as a result $|\lambda_1^* - \lambda_1| \le \sigma_\varepsilon^2$.

However, $\mathbf{\Sigma_X}$ can be decomposed as: $\mathbf{\Sigma_X} = \lambda_1 \boldsymbol{e_1 e_1'} + \mathbf{\Sigma_{k-1}}$. Similarly, the covariance matrix $\mathbf{\Sigma}_{\mathbf{X}^*}$ can be defined as: $\mathbf{\Sigma}_{\mathbf{X}^*} = \lambda_1^* \boldsymbol{e_1^* e_1^{*\prime}} + \mathbf{\Sigma_{k-1}^*}$. Since $\mathbf{\Sigma_X}$ is considered as an operator, the difference for the second eigenvalue can be estimated by:

$$|\lambda_2^* - \lambda_2| = \left| \|\mathbf{\Sigma_{k-1}^*}\| - \|\mathbf{\Sigma_{k-1}}\| \right| \le \|\mathbf{\Sigma_{k-1}^*} - \mathbf{\Sigma_{k-1}}\| = \|\mathbf{\Sigma}_{\varepsilon,\mathbf{k-1}}\| = \sigma_{\varepsilon,k-1}^2 \le \sigma_\varepsilon^2$$

Thus, we can conclude that $|\lambda_2^* - \lambda_2| \le \sigma_\varepsilon^2$.

Similarly, it can be shown that $\left|\lambda_k^* - \lambda_k\right| \le \sigma_\varepsilon^2$. The covariance matrix $\mathbf{\Sigma_X}$ can be decomposed as: $\mathbf{\Sigma_X} = \lambda_1 \boldsymbol{e_1}\boldsymbol{e_1'} + \ldots + \lambda_{k-1}\boldsymbol{e_{k-1}}\boldsymbol{e_{k-1}'} + \mathbf{\Sigma_1}$. Moreover, the covariance matrix $\mathbf{\Sigma_{X^*}}$ can be defined as: $\mathbf{\Sigma_{X^*}} = \lambda_1^* \boldsymbol{e_1^*}\boldsymbol{e_1^{*\prime}} + \ldots + \lambda_{k-1}^*\boldsymbol{e_{k-1}^*}\boldsymbol{e_{k-1}^*}' + \mathbf{\Sigma_1^*}$. Then the difference between the kth noisy and non-noisy eigenvalue is given by:

$$\left|\lambda_k^* - \lambda_k\right| = \left|\left\|\mathbf{\Sigma_1^*}\right\| - \left\|\mathbf{\Sigma_1}\right\|\right| \le \left\|\mathbf{\Sigma_1^*} - \mathbf{\Sigma_1}\right\| = \left\|\mathbf{\Sigma_{\varepsilon,1}}\right\| = \sigma_{\varepsilon,1}^2 \le \sigma_\varepsilon^2$$

Thus, we can conclude that $\left|\lambda_i^* - \lambda_i\right| \le \sigma_\varepsilon^2$ for all i=1,…,k.

For the determination of the noisy eigenvector, it can be noticed that $\boldsymbol{e_1^*} = \boldsymbol{e_1} + \varepsilon_i$. We can distinguish three cases. The fist case is when the vector of noise has the same direction as $\boldsymbol{e_1}$. Then, the difference of the lengths for the noisy and non-noisy eigenvector is equal with σ_{ε_i}, and as a result $\left|\left\|\boldsymbol{e_1^*}\right\| - \left\|\boldsymbol{e_1}\right\|\right| = \sigma_{\varepsilon_i} \le \sigma_\varepsilon$ where σ_ε is the standard deviation of the component of noise with the maximum length. The second case is when the vector of noise is orthogonal to $\boldsymbol{e_1}$. By the Pythagorean Theorem and definition (2), we can estimate the difference between the noisy and non-noisy eigenvector by sinθ = $\frac{\sigma_{\varepsilon_i}}{\left\|\boldsymbol{e_1^*}\right\|} = \frac{\sigma_{\varepsilon_i}}{\left(1+\sigma_{\varepsilon_i}^2\right)^{1/2}} \le \frac{\sigma_\varepsilon}{\left(1+\sigma_\varepsilon^2\right)^{1/2}}$, where θ is the angle between the noisy and non-noisy eigenvector. Otherwise, sinθ satisfies the following inequality: sinθ < $\frac{\sigma_\varepsilon}{\left(1+\sigma_\varepsilon^2\right)^{1/2}}$. Consequently, the norm of the difference between the noisy and non-noisy eigenvector will not exceed σ_ε, $\left|\left\|\boldsymbol{e_1^*}\right\| - \left\|\boldsymbol{e_1}\right\|\right| \le \sigma_\varepsilon$. Furthermore, it can be observed that $\left\|\boldsymbol{e_1^*} - \boldsymbol{e_1}\right\| = \sigma_{\varepsilon_i}$, where ε_i is an arbitrary component of the vector of noise (Fig. (**1**)). Since we know that $\max\limits_{1\le i\le n}\left\|\varepsilon_i\right\| = \sigma_\varepsilon$, we can conclude that $\left\|\boldsymbol{e_1^*} - \boldsymbol{e_1}\right\| \le \sigma_\varepsilon$.

For the determination of the second noisy eigenvector, we project the vector $\boldsymbol{e_2^*}$ to the plane orthogonal to $\boldsymbol{e_1}$, called by proj($\boldsymbol{e_2^*}$) . Then by definition (2), we can conclude that:

$$\left\|\boldsymbol{e}_2^* - \boldsymbol{e}_2\right\| = \left\|\boldsymbol{e}_2^* - \mathrm{proj}(\boldsymbol{e}_2^*) + \mathrm{proj}(\boldsymbol{e}_2^*) - \boldsymbol{e}_2\right\| \le \left\|\boldsymbol{e}_2^* - \mathrm{proj}(\boldsymbol{e}_2^*)\right\| + \left\|\mathrm{proj}(\boldsymbol{e}_2^*) - \boldsymbol{e}_2\right\| \le 2\sigma_\varepsilon$$

The first term is due to $\left\|\boldsymbol{e_1^*} - \boldsymbol{e_1}\right\| \le \sigma_\varepsilon$, and the second term since $\max\limits_{1\le i\le n}\left\|\varepsilon_i\right\| = \sigma_\varepsilon$ (Fig. (**2**)).

Since $\left|\lambda_i - \lambda_{i+1}\right| > i\sigma_\varepsilon^2$ holds for all i=1,…,k-1, it can be shown by induction that $\left\|\boldsymbol{e_i^*} - \boldsymbol{e_i}\right\| \le i\sigma_\varepsilon$ is satisfied for all i=1,...,k. Assume that the condition is satisfied for i equals to n, and therefore the difference between nth noisy and non-noisy eigenvector can be determined by the inequality: $\left\|\boldsymbol{e_n^*} - \boldsymbol{e_n}\right\| \le n\sigma_\varepsilon$. Then:

$$\left\|\boldsymbol{e_{n+1}^*} - \boldsymbol{e_{n+1}}\right\| =$$
$$\left\|\boldsymbol{e_{n+1}^*} - \mathrm{proj}(\boldsymbol{e_{n+1}^*}) + \mathrm{proj}(\boldsymbol{e_{n+1}^*}) - \boldsymbol{e_{n+1}}\right\| \le$$
$$\left\|\boldsymbol{e_{n+1}^*} - \mathrm{proj}(\boldsymbol{e_{n+1}^*})\right\| + \left\|\mathrm{proj}(\boldsymbol{e_{n+1}^*}) - \boldsymbol{e_{n+1}}\right\| \le$$
$$n\sigma_\varepsilon + \sigma_\varepsilon = (n+1)\sigma_\varepsilon$$

where $\mathrm{proj}(\boldsymbol{e_{n+1}^*})$ is the projection of $\boldsymbol{e_{n+1}^*}$ to the plane orthogonal to $\boldsymbol{e_n}$. Therefore, the inequality for difference between the noisy and non-noisy eigenvectors, $\left\|\boldsymbol{e_i^*} - \boldsymbol{e_i}\right\| \le i\sigma_\varepsilon$ is satisfied for all i=1,...,k.

Theorem 2 Let $Y_i = \boldsymbol{e_i'}\mathbf{X}$ be the ith principal component associated with the covariance matrix $\mathbf{\Sigma_X}$ and $\left|\lambda_i - \lambda_{i+1}\right| > i\sigma_\varepsilon^2$ for i=1,…,k-1. Then the length of the noisy principal components Y_i^* can be determined by the inequality: $\left\|Y_i^*\right\| \le \left(\lambda_i\right)^{1/2} + \sigma_\varepsilon$ for i=1,…,k, and the change between Y_i and Y_i^* can be defined by $\left\|\boldsymbol{e_i^*} - \boldsymbol{e_i}\right\| \le i\sigma_\varepsilon$ for i=1,…,k.

Proof Since $\mathrm{Var}(Y_i) = \lambda_i$, the norm of the first noisy principal component is given by:

$$\left\|Y_1^*\right\| = \left(\lambda_1^*\right)^{1/2} = \left\|\mathbf{V_{X^*}^{1/2}}\right\| \le \left\|\mathbf{V_X^{1/2}}\right\| + \left\|\mathbf{V_\varepsilon^{1/2}}\right\| \le \left(\lambda_1\right)^{1/2} + \sigma_\varepsilon$$

where $\mathbf{V_X^{1/2}}$ is the matrix with entries the standard deviations of the non-noisy variables.

Therefore by Theorem 1 and induction, the length of the kth noisy principal component is given by:

$$\left\|Y_k^*\right\| = \left(\lambda_k^*\right)^{1/2} \le \left(\lambda_k\right)^{1/2} + \sigma_\varepsilon.$$

Further, by Theorem 1 and the definition of principal components the change between Y_i and Y_i^* is given by the change between the corresponding noisy and non-noisy eigenvector: $\left\|\boldsymbol{e_i^*} - \boldsymbol{e_i}\right\| \le i\sigma_\varepsilon$ for i=1,..,k.

Corollary 1 Suppose that **X** is a random vector associated with the covariance matrix $\mathbf{\Sigma_X}$ and spectral decomposition ($\lambda_1, \boldsymbol{e_1}$),..,($\lambda_k, \boldsymbol{e_k}$). Let **ε** be a random vector of noise independent of **X** with covariance matrix $\mathbf{\Sigma_\varepsilon}$ such that $\|\mathbf{\Sigma_\varepsilon}\| = \sigma_\varepsilon^2$. Denote by $\mathbf{X}^* = \mathbf{X} + \boldsymbol{\varepsilon}$ a noisy random vector. Assume that the eigenvalues of $\mathbf{\Sigma_X}$ satisfy the condition $|\lambda_i - \lambda_{i+1}| > i\sigma_\varepsilon^2$ for i=1,...,k-1 and the signal-to-noise ratio is equal with $\|\mathbf{\Sigma_X}\| / \|\mathbf{\Sigma_\varepsilon}\|$. Then by Theorem 1, Fig. (**3**) shows the maximum number of the invariant eigenvectors if the signal-to-noise ratio is known (Fig. (**3**)).

3. Examples and Simulation Study

In this section we prove that the principal components can be replaced arbitrarily when noise is present in a random vector **X** and certain conditions about the noise are satisfied. First we present the results for the two dimensional case (Example 1) and then we generalize the statements for more dimensions (Theorem 3).

Example 1 Let $\mathbf{X} = (X_1 \quad X_2)'$ be a random vector with covariance matrix $\mathbf{\Sigma_X}$ and spectral decomposition ($\lambda_1, \boldsymbol{e_1}$), ($\lambda_2, \boldsymbol{e_2}$). If noise $\boldsymbol{\varepsilon} = (\varepsilon_1 \quad \varepsilon_2)'$ is present in the random vector $\mathbf{X}$ so that $|\lambda_1 - \lambda_2| < \sigma_\varepsilon^2$ where $\sigma_\varepsilon^2 = \max(\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2)$, then the eigenvalues–eigenvectors pairs can be changed dramatically.

Proof Let $\mathbf{X} = (X_1 \quad X_2)'$ be a vector consists of random variables X_1 and X_2 with covariance matrix $\mathbf{\Sigma_X} = \begin{pmatrix} 50 & 0.25 \\ 0.25 & 50 \end{pmatrix}$. Then the eigenvalues and eigenvectors are equal with λ_1=Var(Y_1)=50.25, $\boldsymbol{e_1}' = [0.707 \quad 0.707]$, λ_2=Var(Y_2)=49.75, and $\boldsymbol{e_2}' = [-0.707 \quad 0.707]$. Therefore, the difference between the first and the second eigenvalue is 0.5.

Suppose that noise $\boldsymbol{\varepsilon} = (\varepsilon_1 \quad \varepsilon_2)'$ is present in the random vector **X** with covariance matrix $\mathbf{\Sigma_\varepsilon} = \begin{pmatrix} 3 & 0 \\ 0 & 0.5 \end{pmatrix}$. Then, since the noise is independent of the variables X_1 and X_2, the covariance matrix of the noisy vector $\mathbf{X}^* = \mathbf{X} + \boldsymbol{\varepsilon}$ is given by: $\mathbf{\Sigma_{X^*}} = \mathbf{\Sigma_X} + \mathbf{\Sigma_\varepsilon} = \begin{pmatrix} 53 & 0.25 \\ 0.25 & 50.5 \end{pmatrix}$. The eigenvalues and eigenvectors of the covariance matrix $\mathbf{\Sigma_{X^*}}$ are equal with $\lambda_1^* = \mathrm{Var}(Y_1^*) = 53.02$, $\boldsymbol{e_1^{*\prime}} = [0.99 \quad 0.09]$, $\lambda_2^* = \mathrm{Var}(Y_2^*)$=50.47, and $\boldsymbol{e_2^{*\prime}} = [-0.09 \quad 0.99]$. Thus when noise is present in the random vector $\mathbf{X}$, then the eigenvectors are changed dramatically.

Therefore, the principal components determined by the random vector $\mathbf{X}$ $Y_1 = 0.707X_1 + 0.707X_2$ and $Y_2 = -0.707X_1 + 0.707X_2$, have changed dramatically to $Y_1^* = 0.99X_1^* + 0.09X_2^*$ and $Y_2^* = -0.09X_1^* + 0.99X_2^*$. Fig. (**4**) shows that the principal components of the random vector $\mathbf{X}$ have been rotated by an angle approximately of 45°, when noise is present in the random vector $\mathbf{X}$ (Fig. (**4**)).

3.1 Simulation study

Denote by **X**=$(X_1 \quad X_2 \quad \cdots \quad X_k)'$ a random vector consists of k random variables with covariance matrix $\mathbf{\Sigma_X}$. Suppose that noise **ε** is present in the random vector **X** with covariance matrix $\mathbf{\Sigma_\varepsilon}$, such that $\|\mathbf{\Sigma_\varepsilon}\| = \sigma_\varepsilon^2 = n$ and $n \geq 3$. Then the covariance matrix of the random vector **X** can be defined as follows:

$$\mathbf{\Sigma_X^a} = \begin{pmatrix} m_1 & c_1 & c_2 & c_4 & \cdots & c_j \\ c_1 & m_2 & c_3 & c_5 & \cdots & c_{j+1} \\ c_2 & c_3 & m_3 & c_6 & \cdots & c_{j+2} \\ c_4 & c_5 & c_6 & m_4 & & c_{j+3} \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ c_j & c_{j+1} & c_{j+2} & c_{j+3} & \cdots & m_k \end{pmatrix}.$$

Assume that m_1 = Var(X_1) is an arbitrary integer. Then choose as:

$m_2 = \mathrm{Var}(X_2) = m_1 - q$ such that $0 \leq q < n$ and n, q are integers

$m_3 = \mathrm{Var}(X_3) = m_2 - s$ where s = q + n

$m_4 = \mathrm{Var}(X_4) = m_3$ – q – 2*n and therefore

$m_k = \mathrm{Var}(X_k) = m_{k-1} - q - (k-2)*n$, where n = σ_ε^2.

In the special case where $m_2 = \mathrm{Var}(X_2) = \mathrm{Var}(X_1) = m_1$, choose as $m_3 = \mathrm{Var}(X_3) = m_2 - 1 - n$, and therefore $m_k = \mathrm{Var}(X_k) = m_{k-1} - 1 - (k-2)*n$.

Consider c_1 = Cov(X_1, X_2) and c_2 = Cov(X_1, X_3) such as: c_1 < 1 and c_2 <1. Then, let as c_3 =Cov(X_2, X_3) = $c_1 * c_2$ where c_3 is less than 1. Also, choose as c_4 =Cov(X_1, X_4) = $c_3 * c_2 = c_1 * c_2 * c_2 = c_1 * c_2^2$ where c_1, c_2 are defined above and c_5 = =Cov(X_4, X_2) = $c_4 * c_3 = c_1 * c_2^2 * c_1 * c_2 = c_1^2 * c_2^3$.

Hence, $c_j = c_{j-1} * c_{j-2}$ for $j \geq 3$, where the c_j's are defined similarly. Following the same algorithm, choose as $c_1, c_2 < 0.5$ when n=2, and $c_1, c_2 < 0.25$ for n=1.

Similarly it can be defined the covariance matrix in case the variance is increasing along the diagonal. Then the covariance matrix of the random vector **X** can be defined as:

$$\mathbf{\Sigma_X^b} = \begin{pmatrix} m_1 & c_1 & c_2 & c_4 & \cdots & c_j \\ c_1 & \ddots & c_3 & c_5 & \cdots & c_{j+1} \\ c_2 & c_3 & m_{k-3} & c_6 & \cdots & c_{j+2} \\ c_4 & c_5 & c_6 & m_{k-2} & & c_{j+3} \\ \vdots & \vdots & \vdots & & m_{k-1} & \vdots \\ c_j & c_{j+1} & c_{j+2} & c_{j+3} & \cdots & m_k \end{pmatrix}$$

.

Assume that m_k =Var(X_k) is an arbitrary integer. Consider as

$$m_{k-1} = \mathrm{Var}(X_{k-1}) = m_k - q, \text{ where } 0 \leq q < n$$

$$m_{k-2} = \mathrm{Var}(X_{k-2}) = m_{k-1} - q - n$$

$$m_{k-3} = \mathrm{Var}(X_{k-3}) = m_{k-2} - q - 2*n$$

$$\vdots$$

$$m_1 = \mathrm{Var}(X_1) = m_2 - q - (k-2)*n.$$

The remaining elements of the matrix are defined as described above.

Then, it can be shown that the eigenvalues of either the covariance matrix $\mathbf{\Sigma_X^a}$ or $\mathbf{\Sigma_X^b}$ satisfy the condition: $(i-1)\sigma_\varepsilon^2 < |\lambda_i - \lambda_{i+1}| < i\sigma_\varepsilon^2$, for i=1,...,k–1 and $\|\mathbf{\Sigma_\varepsilon}\| = \sigma_\varepsilon^2$. Moreover, when noise **ε** is present in the random vector **X**, then the eigenvectors of the noisy random vector $\mathbf{X}^* = \mathbf{X} + \boldsymbol{\varepsilon}$ can be changed arbitrarily.

Theorem 3 Let **X** be a random vector with covariance matrix $\mathbf{\Sigma_X}$ and let $\Delta\lambda_i = |\lambda_i - \lambda_{i+1}|$ for i=1,..,k-1. If random vector of noise **ε** with covariance matrix $\mathbf{\Sigma_\varepsilon}$ such that $\|\mathbf{\Sigma_\varepsilon}\| = \sigma_\varepsilon^2$ is present in the random vector **X** so that $(i-1)\sigma_\varepsilon^2 < |\lambda_i - \lambda_{i+1}| < i\sigma_\varepsilon^2$ for i=1,...,k-1 (3), then the eigenspace is asymptotically invariant, while the eigenvectors within the space can be replaced arbitrarily. In addition, the conditions for $\Delta\lambda_i$ described in (3) cannot be improved.

4. Comparison of the models for the spectral decomposition of the noisy covariance matrix

In this section we compare the results for the noisy eigenvalues and eigenvectors described by Theorem 1 in the two dimensional case with other models examined before [2-3, 6-7]. We show that our results for certain conditions satisfied by the noise are more accurate compared to other models [2-3, 6-7].

Lemma 1 Let **A** be a $p \times p$ symmetric matrix and let **B** be a symmetric perturbation. Let $(\lambda, \boldsymbol{v})$ be the eigenvalue-eigenvector pair of **A**+**B** corresponding to $(\lambda_i, \boldsymbol{v_i})$ of **A** and $\delta = \min_{j \neq i} |\lambda - \lambda_j|$, where $\{\lambda_j\}_{j=1}^p$ are the eigenvalues of **A**; then

$$\sin\theta(\boldsymbol{v}, \boldsymbol{v_i}) \leq \frac{\|\mathbf{B}\|}{\delta}.$$

Lemma 1 follows from classical results in matrix or operator perturbation theory [2, 6-7]. The next statement (Statement 1) compares the results for the angle between the noise and non-noisy eigenvectors derived by Lemma 1 with the results by Theorem 1.

Statement 1 Suppose that $\mathbf{\Sigma_X}$ is a 2×2 covariance matrix associated with the vector **X** and $\mathbf{\Sigma_\varepsilon}$ a 2×2 covariance matrix of the noise, such that $\sigma_\varepsilon^2 < |\lambda_1 - \lambda_2| < 2\sigma_\varepsilon^2$ with $\sigma_\varepsilon < (1/8)^{1/2}$. Then the angle between the noisy and non-noisy eigenvector can be estimated more accurately by Theorem 1 than by Lemma 1.

Proof In Lemma 1, let $\|\mathbf{B}\| = \|\mathbf{\Sigma_\varepsilon}\| = \sigma_\varepsilon^2$ and $\delta = |\lambda_1^* - \lambda_2|$. Then by Lemma 1, $\sin\theta(\boldsymbol{e_1^*}, \boldsymbol{e_1}) \leq \frac{\sigma_\varepsilon^2}{|\lambda_1^* - \lambda_2|}$. By Theorem 1, the angle between the first noisy and non-noisy eigenvector can be determined by: $\sin\theta \leq \frac{\sigma_\varepsilon}{(1+\sigma_\varepsilon^2)^{1/2}}$. If $2\sigma_\varepsilon^2 > |\lambda_1 - \lambda_2| > \sigma_\varepsilon^2$ and $\sigma_\varepsilon < (1/8)^{1/2}$ then we can prove that $\frac{\sigma_\varepsilon}{(1+\sigma_\varepsilon^2)^{1/2}} < \frac{\sigma_\varepsilon^2}{|\lambda_1^* - \lambda_2|}$.

By Theorem 1 we know that: $|\lambda_1^* - \lambda_2| \leq |\lambda_1 - \lambda_2| + \sigma_\varepsilon^2$. Since $\sigma_\varepsilon < (1/8)^{1/2}$, it can be observed that

$\left|\lambda_1^* - \lambda_2\right| < 3\sigma_\varepsilon^2 < \sigma_\varepsilon(1+\sigma_\varepsilon^2)^{1/2}$.

Therefore we can conclude that: $\frac{\sigma_\varepsilon}{(1+\sigma_\varepsilon^2)^{1/2}} < \frac{\sigma_\varepsilon^2}{\left|\lambda_1^* - \lambda_2\right|}$ for the first noisy eigenvector.

For the second eigenvector it follows by Lemma 1 that: $\sin\theta(\boldsymbol{e}_2^*, \boldsymbol{e}_2) \le \frac{\sigma_\varepsilon^2}{\left|\lambda_2^* - \lambda_1\right|}$. But by Theorem 1, we have that: $\left|\lambda_2^* - \lambda_1\right| \le \left|\lambda_2^* - \lambda_2\right| + \left|\lambda_2 - \lambda_1\right| \le 3\sigma_\varepsilon^2$ $< \sigma_\varepsilon(1+\sigma_\varepsilon^2)^{1/2}$. Consequently, we show that $\frac{\sigma_\varepsilon}{(1+\sigma_\varepsilon^2)^{1/2}} < \frac{\sigma_\varepsilon^2}{\left|\lambda_2^* - \lambda_1\right|}$.

Statement 1 can also be extended when $\sigma_\varepsilon^2 < \left|\lambda_1 - \lambda_2\right| < k\sigma_\varepsilon^2$ and $\sigma_\varepsilon < [1/(k^2+2k)]^{1/2}$ for k>1.

Lemma 2 Let **H** be a positive definite symmetric matrix and **δH** a small relative perturbation of **H** in the sense of $\|\boldsymbol{\delta H}\| \le \eta\|\mathbf{H}\|$. Let λ_i, $\boldsymbol{v_i}$ and λ_i^*, $\boldsymbol{v_i^*}$ be the i-th eigenvalue-eigenvector of **H** and **H+δH**, respectively. Then $\left\|\boldsymbol{v_i} - \boldsymbol{v_i^*}\right\| \le \frac{\eta}{\text{absgap}_{\lambda_i}} + O(\eta^2)$

where the absolute gap of eigenvalues is defined as

$$\text{absgap}_{\lambda_i} = \min_{j \ne i} \frac{\left|\lambda_i - \lambda_j\right|}{\|\mathbf{H}\|}.$$

Lemma 2 is a result of the standard perturbation theory for the eigenvectors [3, 7]. Next, we compare our results derived by Theorem 1 for the angle between the noisy and non-noisy eigenvectors with the results by Lemma 2.

Statement 2 Let $\boldsymbol{\Sigma_X}$ be a 2×2 covariance matrix associated with the random vector **X** and $\boldsymbol{\Sigma_\varepsilon}$ a 2×2 covariance matrix of the noise, so that $\sigma_\varepsilon^2 < \left|\lambda_1 - \lambda_2\right| < 2\sigma_\varepsilon^2$ with $\sigma_\varepsilon < 1/4$. Then the difference between the noisy and non-noisy eigenvector can be estimated more accurately by Theorem 1 than by Lemma 2.

Proof In Lemma 2, consider as **δH** $\equiv \boldsymbol{\Sigma_\varepsilon}$ and **H** $\equiv \boldsymbol{\Sigma_X}$. Then for the first eigenvector, it follows that:

$\left\|\boldsymbol{e_1^*} - \boldsymbol{e_1}\right\| \le \frac{\eta}{\frac{\left|\lambda_1 - \lambda_2\right|}{\lambda_1}} + O(\eta^2)$, where $\sigma_\varepsilon^2 \le \eta\lambda_1$.

However, $\frac{\eta\lambda_1}{\left|\lambda_1 - \lambda_2\right|} \ge \frac{\sigma_\varepsilon^2}{\left|\lambda_1 - \lambda_2\right|} > \frac{\sigma_\varepsilon^2}{2\sigma_\varepsilon^2} = 1/2$.

Since $\sigma_\varepsilon < 1/4$, we have that $\sigma_\varepsilon < \frac{\eta\lambda_1}{\left|\lambda_1 - \lambda_2\right|}$.

For the second eigenvector, it follows by Lemma 2 that:

$\left\|\boldsymbol{e_2^*} - \boldsymbol{e_2}\right\| \le \frac{\eta}{\frac{\left|\lambda_2 - \lambda_1\right|}{\lambda_1}} + O(\eta^2)$.

But, $\frac{\eta\lambda_1}{\left|\lambda_2 - \lambda_1\right|} \ge \frac{\sigma_\varepsilon^2}{\left|\lambda_2 - \lambda_1\right|} > 1/2$. Therefore since $\sigma_\varepsilon < 1/4$, we can conclude that $2\sigma_\varepsilon < \frac{\eta\lambda_1}{\left|\lambda_1 - \lambda_2\right|}$.

It can also be noted that Statement 2 can be extended when $\sigma_\varepsilon^2 < \left|\lambda_1 - \lambda_2\right| < k\sigma_\varepsilon^2$ and $\sigma_\varepsilon < 1/(2k)$ for k >1.

Therefore we proved that the bounds for the eigenvectors described by Theorem 1 are more accurately compared to Lemma 1 and Lemma 2. Those results have been proved for the two dimensional case. Since we have proved that the results for the simplest case as described by Theorem 1 are more accurately compared with Lemma 1 and 2, the investigation for higher dimensions is not indispensable.

5. Discussion

In this paper, we show that the presence of noise or disturbance in a random vector **X** can change the principal components dramatically. In particular, when the differences between the eigenvalues of the covariance matrix defined by the random vector **X** are small compared to the level of the noise then by Theorem 3 the principal components can be changed arbitrarily when noise is present in the random vector **X**. On the other hand, when the differences between the eigenvalues are sufficiently large compared to the noise then by Theorem 1 and 2 the difference between the noisy and non-noisy eigenvalues and eigenvectors can be estimated by using the covariance matrix of the noise. We also compare the results with other models studied before [2-3, 6-7] in the two dimensional case and we prove that the noisy eigenvalues and eigenvectors can be determined more accurate by using our model when certain conditions of the noise are satisfied. The results of the theorems can be applied when two random vectors are uncorrelated and associated with covariance matrices with completely different structures [8, 14].

In general, the principal component analysis is the key issue in multivariate analysis. However uncertainties in basic variables or contribution from different scales can easily ruin the analysis and make inferences erroneous [8]. One application of the theory described here, is the necessity of the decomposition of a time series into components associated with different covariance or correlation structures. In this case, we proved by Theorem 3 that we cannot keep the components together because we may obtain inconclusive results.

An example is the determination of the main atmospheric factor for the explanation and prediction of ozone concentrations [14, 17]. Many authors study the main atmospheric factor on ozone concentrations without using the decomposition of the time series of ozone and the remaining atmospheric variables [16, 18-20]. In particular, raw data of ozone, solar radiation and wind are analyzed by using the principal component analysis but no prominent conclusion has been obtained because of the presence of two different components (global and synoptic scale component) associated with completely different correlation structures in the time series of ozone and the atmospheric variables [16].
However, when the Kolmogorov-Zurbenko filter [21,22] is used to decompose the time series of ozone and the atmospheric variables into low frequency (global component) and high frequency (synoptic scale component), the explanation of ozone concentrations has been improved essentially [12, 14, 17]. The global term components are highly correlated, while the synoptic scale component show a low correlation. Since different scales provide different covariance structures, by Theorem 3 they must be separated. Thus, the decomposition of the time series is necessary for the analysis and improves the explanation and prediction for the ozone time series approximately two times [14, 17].
In particular, it is shown that solar radiation is the main atmospheric factor for the explanation and prediction of ozone concentrations when the decomposition of the time series and the canonical correlation analysis are applied in the model [14, 17]. The results regarding the study of ozone are possible only with the decomposition of the time series, which can remove contradictory effects in similar variables considered in different scales. Without such a separation, the multivariate analysis will be strongly compromised as the current paper proves. Thus, it is essential before any analysis to verify whether a random vector consists of vectors associated with different covariance or correlation structures. In this case, we need to study the random vectors separately for avoiding erroneous results in the principal components. This theory described here can be extended for other methods of the factor analysis, such as the canonical correlation analysis.

References

[1] Bickel P.J., Levina E. (2008) *Ann. of Stat.*, 36, 199-227.
[2] Davis C. and Kahan W.M. (1970) *III SIAM J. of Numerical Anal.*, 70, 1-47.
[3] Demmel J., Veselic K. (1992) *SIAM J. of Matrix Anal. and Applications*, 13, 1204-1245.
[4] Martin J.F. (1988) *Component and Correspondence Analysis. Dimension Reduction by Functional Approximation, J.L.A. van Rijckevorsel and J. de Leeuw, eds., Wiley, Chichester*, 103-114.
[5] Mathias R. (1997) *SIAM J. of Matrix Anal. and Applications*, 18, 959-980.
[6] Nadler B. (2008) *Ann. of Stat.*, 36, 2791-2817.
[7] Parlett B.N. (1980) *The symmetric eigenvalue problem, Prentice-Hall, New Jersey.*
[8] Tsakiri K.G. and Zurbenko I.G. (2008) *JSM proceedings, Stat. Comput. Section, American Statistical Association, Alexandria, VA: American*, 569-576.
[9] Webb A.R. (1996) *Stat. and Comput.*, 6, 159-168.
[10] Zurbenko I.G. and Sowizral M. (1999) *Far East J. of Theor. Stat.*, 3, 139-157.
[11] Jolliffe I.T. (2002) *Principal Component Analysis, Springer-Verlang, New York*.
[12] Zurbenko I.G. (1986) *The Spectral Anal. of Time Series, North Holland Series in Statistics and Probability, Amsterdam.*
[13] Yang W. and Zurbenko I. (2010) *Wiley Interdisciplinary Revives: In Comput. Stat., Wiley*, 2, 107-115.
[14] Tsakiri K.G. and Zurbenko I.G. (2010) *J. of Air Qual., Atmos. and Health*, DOI: 10.1007/s11869-010-0084-5.
[15] Shores T.S. (2007) *Appl. Linear Algebra and Matrix Anal., Springer, New York.*
[16] Johnson R.A., Wichern D.W. (2002) *Appl. Multivariate Stat. Anal., Prentice Hall, New Jersey*.
[17] Tsakiri K.G. and Zurbenko I.G. (2010) *Meteorol. and Atmos. Phys.,* 109, 129-137.
[18] Kelly N.A., Ferman M.A., and Wolff G.T. (1986) *J. of the Air Pollut. Control Assoc.*, 36, 150-158.
[19] Korsog P.E. and Wolff G.T. (1991) *Atmos. Environ.*, 25B, 47-57.
[20] Salazar-Ruiz E., Ordieres J.B., Vergara E.P., Capuz-Rizo S.F. (2008) *J. of Environ. Model. and Soft.,* 23, 1056-1069.
[21] Yang W. and Zurbenko I. (2010) *Wiley interdisciplinary revives in comput. Stat., WIREs in Comput. Stat.*, 2, 340-351.
[22] Kolmogorov-Zurbenko Filters (2011) *In Wikipedia, http://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Zurbenko_filter*

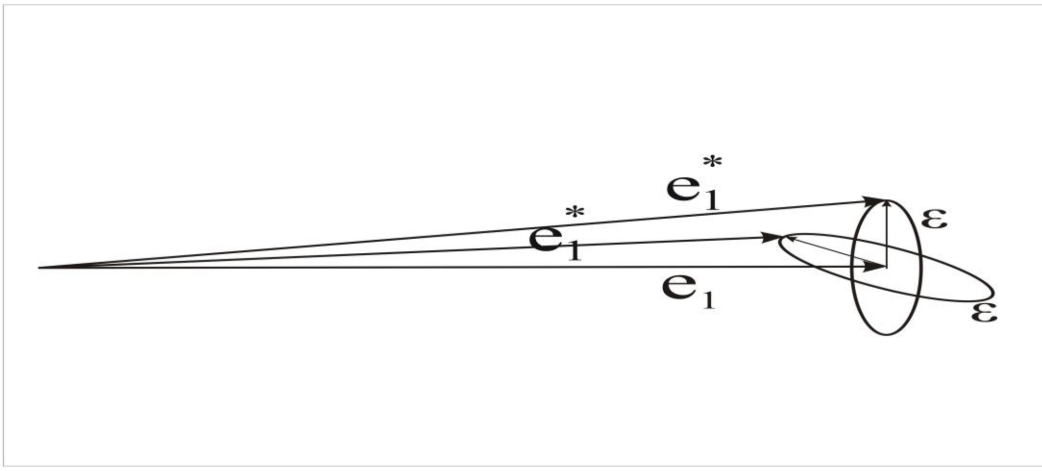

Fig. **1-**Representation of the first noisy and non-noisy eigenvector when the differences between the eigenvalues are large compared to the noise. The first noisy eigenvector, e_1^*, may fluctuate depending on the value of the noise.

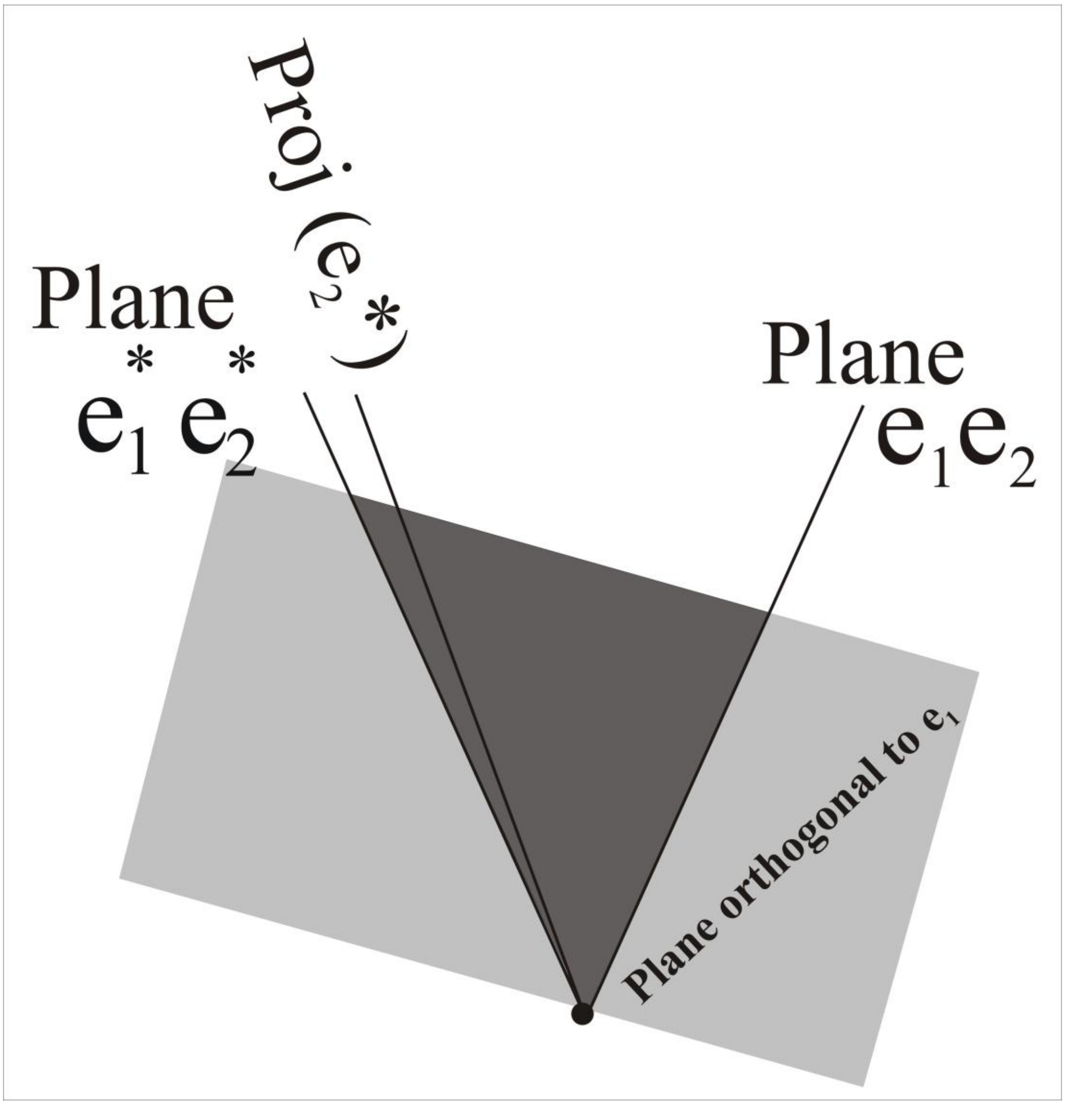

Fig.2-Representation of the second noisy eigenvector when the differences between the eigenvalues are large compared to the noise. The second noisy eigenvector, e_2^*, may fluctuate depending on the value of the noise.

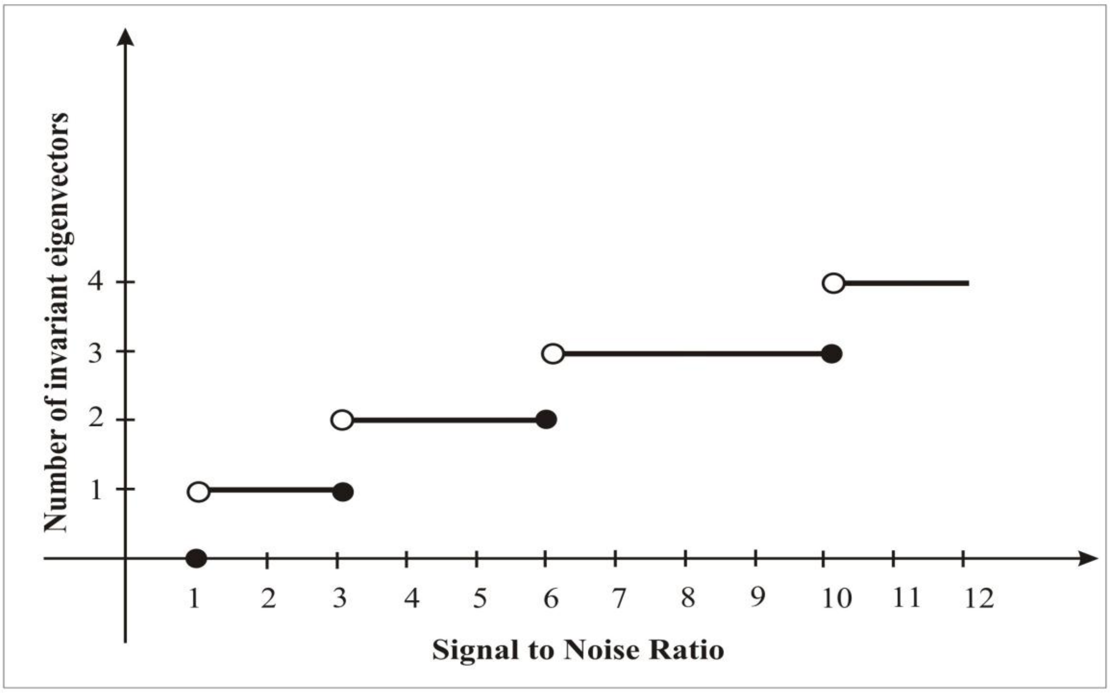

Fig. 3-Maximum number of eigenvectors which can be detected if the signal-to-noise ratio is known.

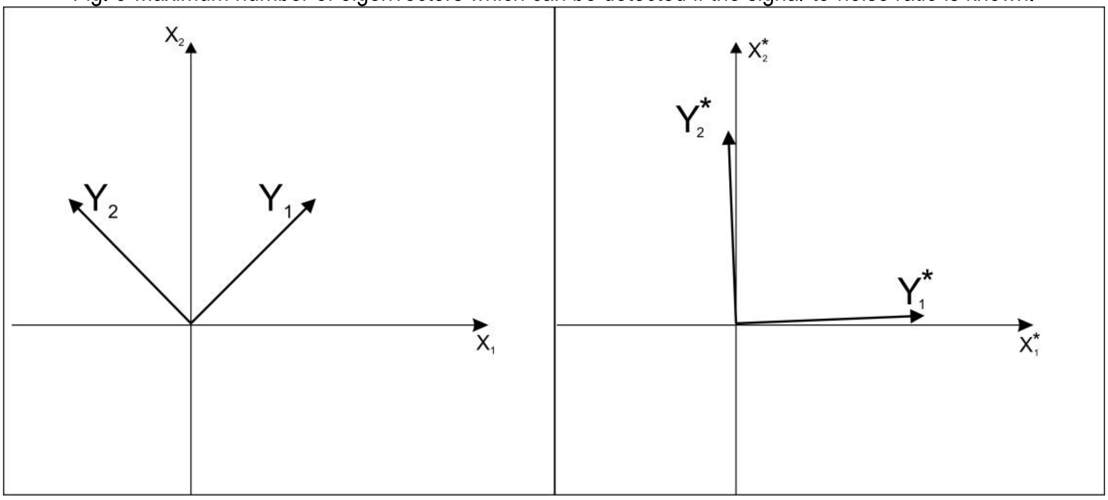

Fig. 4-Representation of the change in the angle of non-noisy (left hand figure) and noisy (right hand figure) principal components when the differences between the eigenvalues of the variables are small compared to the noise.

List of abbreviations:
Principal Component Analysis: PCA
Principal Components: PC's