02458 Cognitive modelling E19
3rd Assignment

Adrian Atienza (s192780)

Arianna Taormina (arita@fysik.dtu.dk)

Marie Claire Capolei (macca@elektro.dtu.dk)

¶

Homework 1 – Principal components analysis PCA

We are asked to verify how applying PCA is useful to reduce multidimensional data (in this case an image) to lower dimensions while retaining most of the information (variance) of the data.

In Figure 1.1 A we can see that the original data Covariance is linear and positive, which means that there is probably a strong relation between variables.

We use PCA to try convert observations of possibly correlated variables into a set of values of linearly uncorrelated variables, the principal components. The result applying PCA is visible in Figure 1.1 B

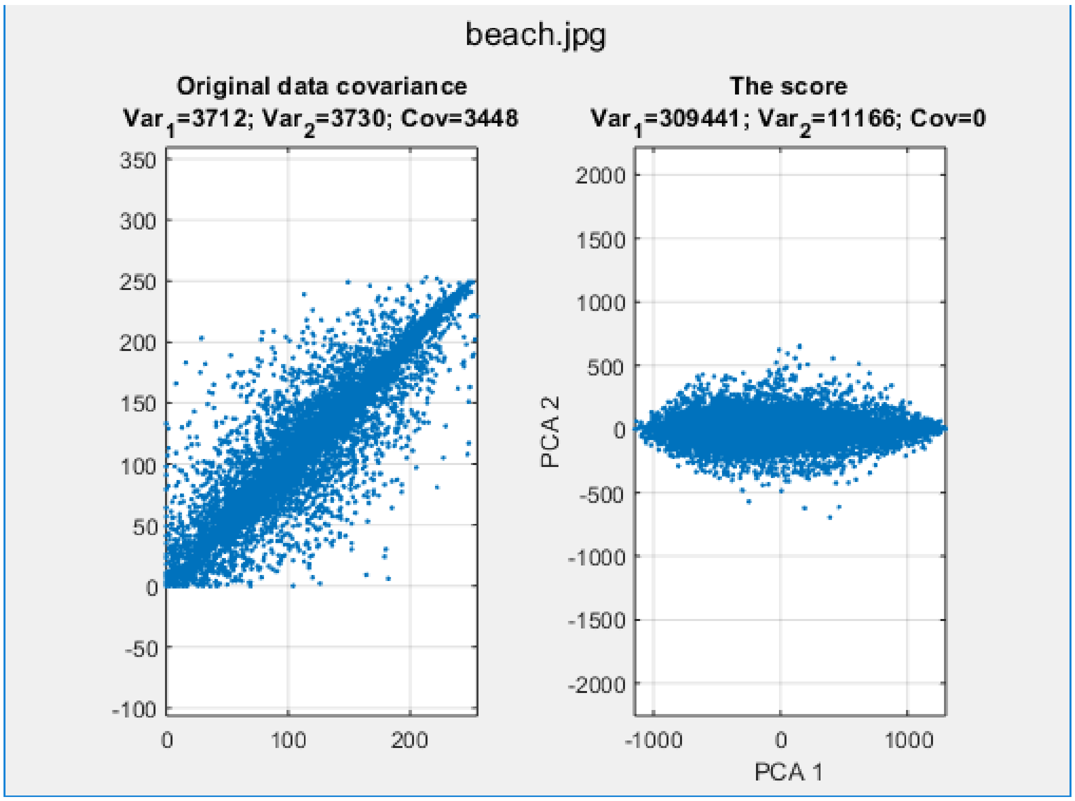

Figure 1.1. A and B : Covariance of the data without and with PCA

We start by fixing the dimensions of the chosen image,

I = image matrix, of shape (1000,1000),

We divide the image in 10000 patches of shape (10x10) and transform each patch in a vector (1x100) Each one of the 10000 vectors in stacked in a transformed matrix S,

S = transformed matrix, of shape (10000,100),

We apply PCA on the transformed Matrix,

[X,W,latent]= pca(S);

Getting:

- X = coefficients matrix, each column contains the coefficients for one principal components, in descending order of component variance. Eigenvectors of covariance matrix.
- W = scores, representations of the data rotated to the new basis of the principal component space.
- Latent = eigenvalues of covariance matrix in descending order of variance.

Exercise 1 – Verify that X*W' is a reconstruction of S

Keeping all the components, we try to reconstruct the original image.

n_pca = 100;

I_pca = W(:,1:n_pca)*X(:,1:n_pca)';

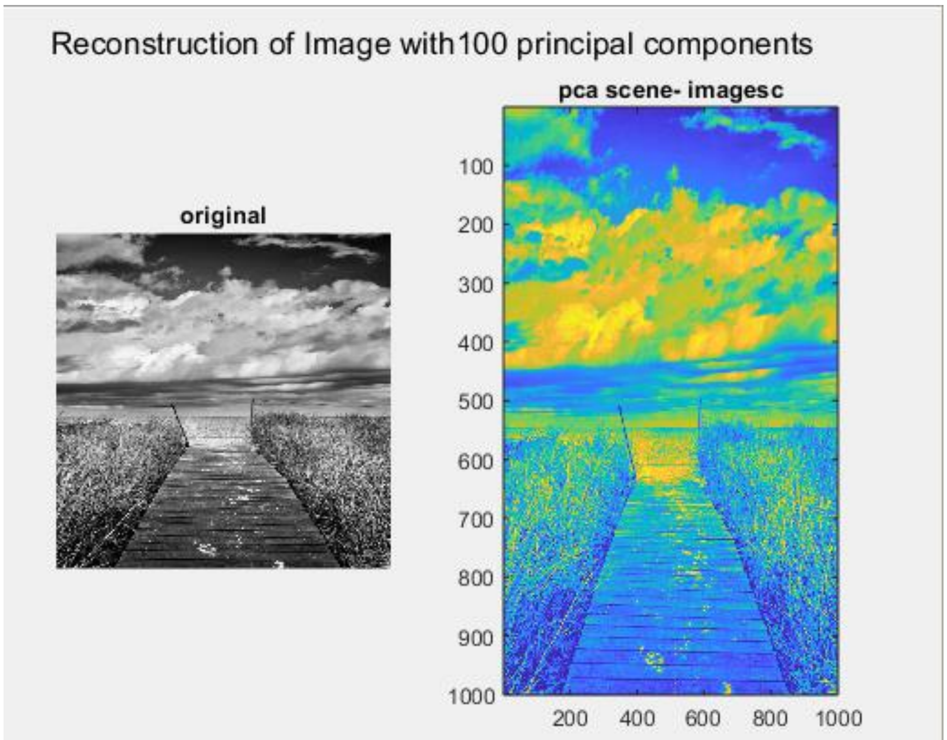

Figure 1.2. I and W*X' reconstructed with 100 principal components ,the difference is negligible

Exercise 2 – How many eigenvalues does it take to explain 95% of the total variance?

We know that the coefficients and eigenvalues stored in the latent variable of PCA are in descending order of variance. From Figure1.3 We can see that the most variance (> 80%) is stored in the very first component.

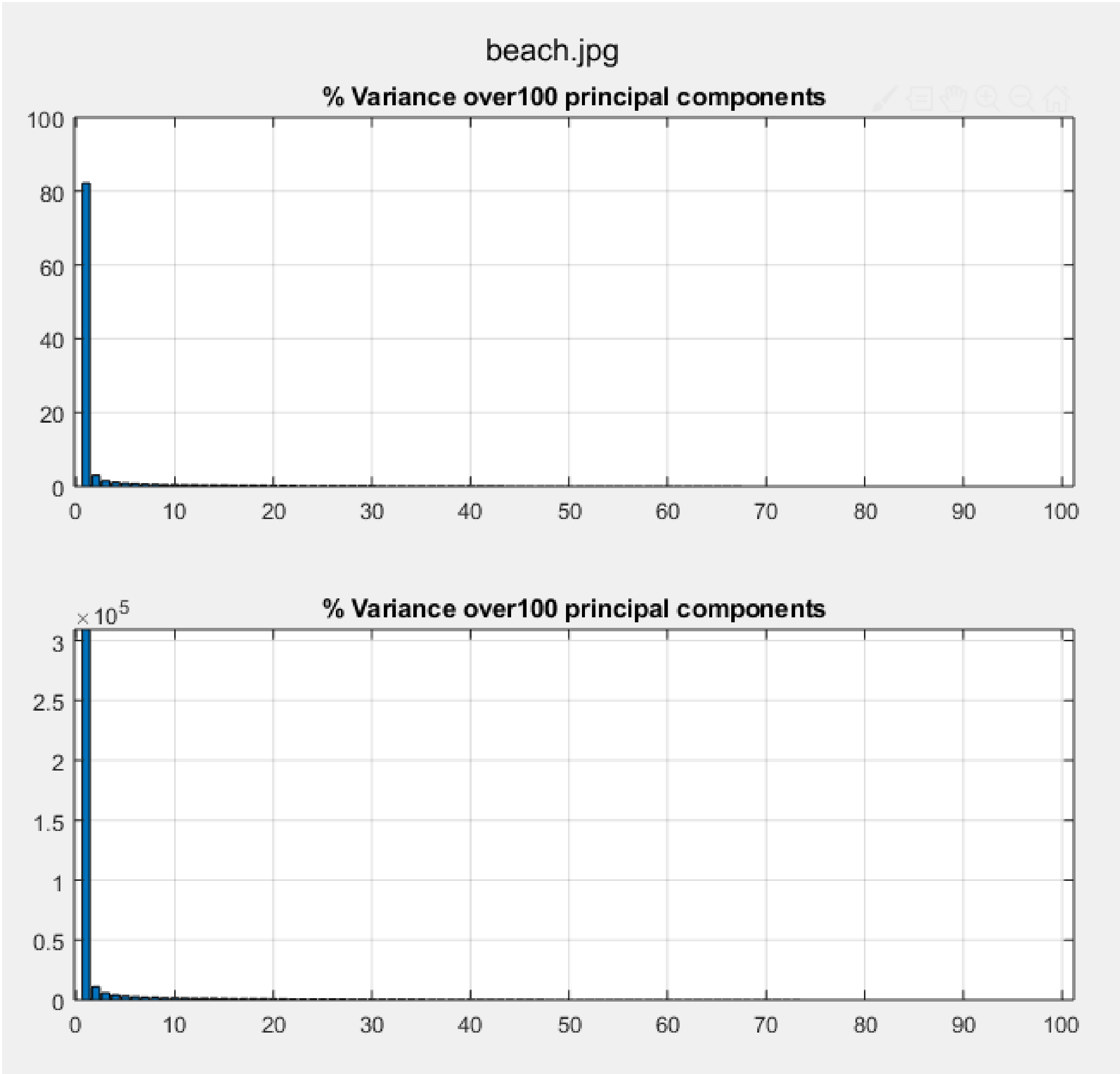

Figure 1.3. Variance % and Variance over 100 principal components

So we can sum eigenvalues one after another until we get to the 95% of the total sum of the latents. It results that the 95% of the variance is explained by the first "i" eigenvalues. Same thing would be looking at the variable "explained" from the full pca output:

[X,W,latent,tsquared, explained, mu] = pca(S);

And we can see the first 'i' values explain the 95% of the variance.

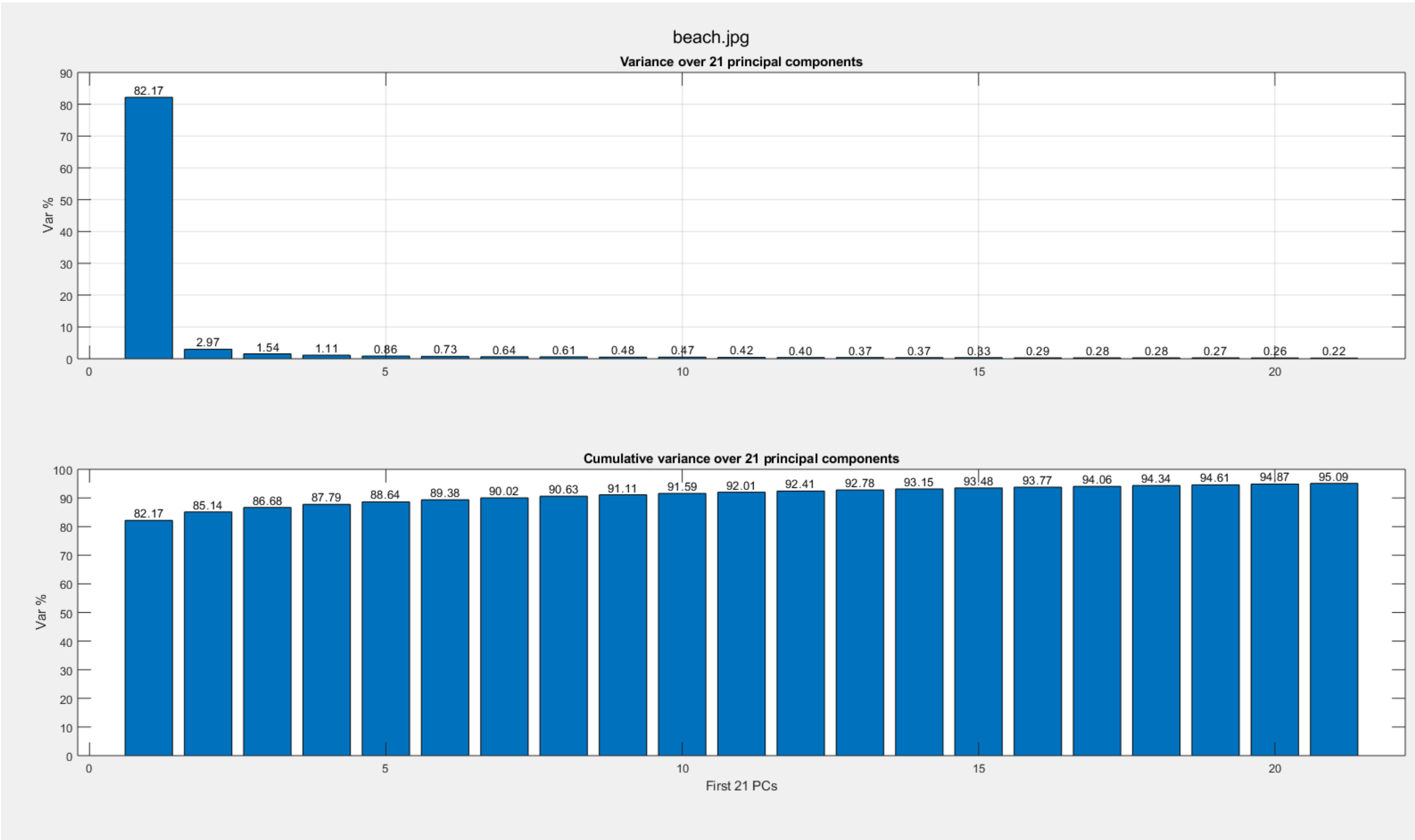

Figure 1.4. Variance and Cum-Variance over 21 principal components

If we reconstruct the image keeping the first i = 21 principal components we get still a very good approximation.

We can see in Figure1.5 how the images appears when is reconstructed with i = 21 components that capture the 95% of the variance of the data.

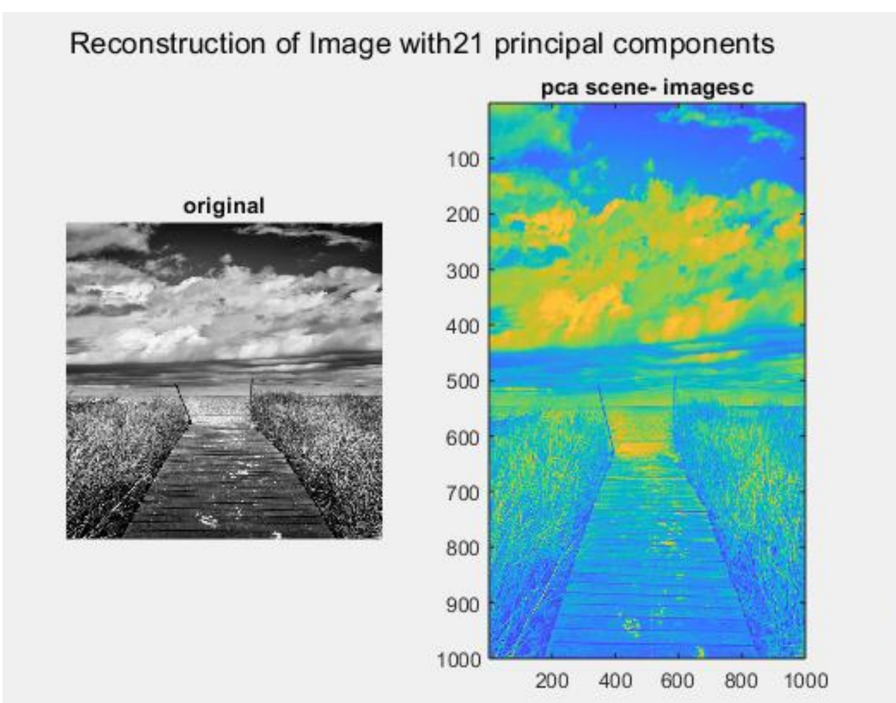

Figure 1.5. Reconstruction of image with 21 principal components.

Exercise 3 – Reconstruct S using only the first 6 PCs and compare it to the original

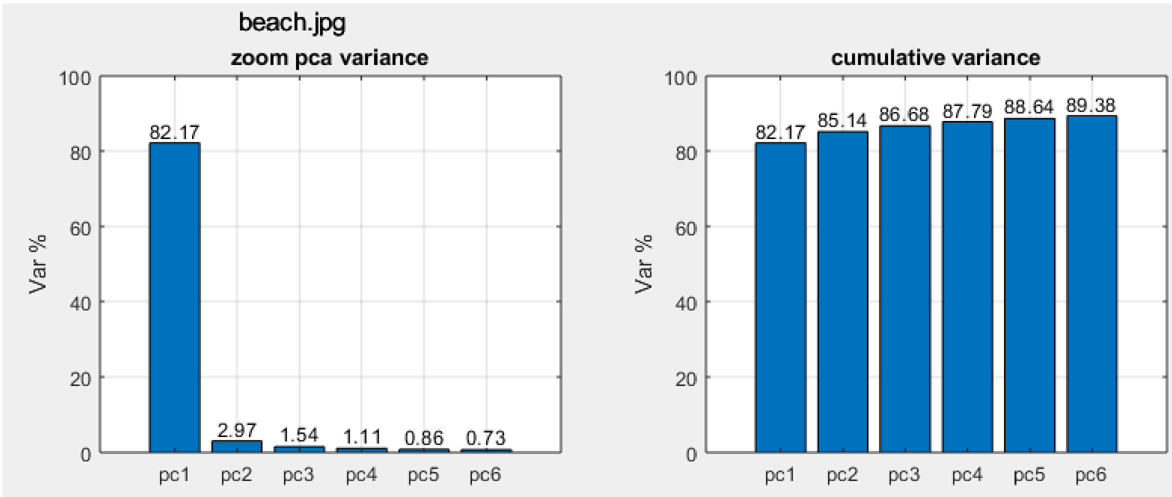

Figure 1.6. Histogram of variance and cumulative variance of the first 6 principal components.

From the Histogram we can see how the first PC is responsible for the 82,17% of the variance and that the first 6 PCs retain the 89,38% of the original information.

We reconstruct the Image, as previously done:

n_pca = 6,

I_pca = W(:,1:n_pca)*X(:,1:n_pca)',

We can see in Figure 1.7 that the reconstructed image appears smoother than the original image. This makes sense, since part of the background information (mostly noise) has been lost.

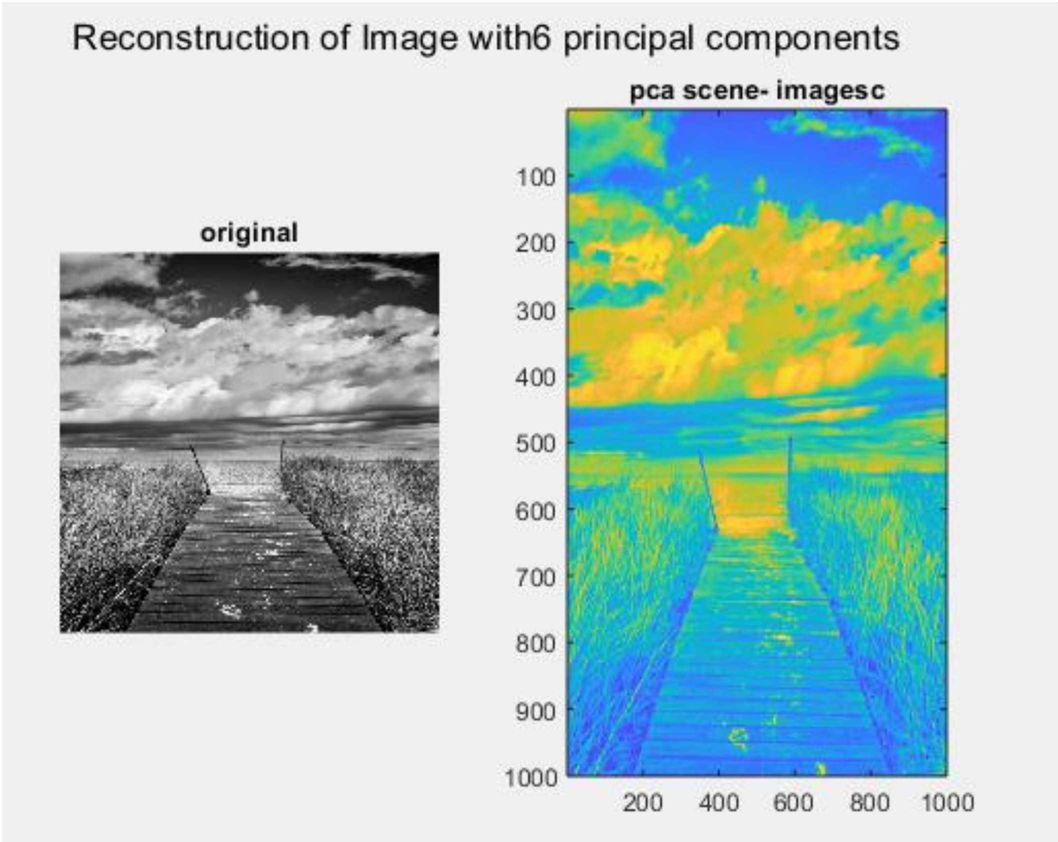

Figure 1.7. Reconstruction of image with 6 principal components.

Homework 2 – Independent components analysis ICA

Exercise 1 – Create and linearly combine two Laplacian distributions

Given the Laplace distribution,

$$Lp(k) = \mu - \sigma \cdot sign(c) \cdot \log(1 - 2 \cdot |c - 0.5|),$$

where c is a random value, μ and σ are the mean and standard deviation of the distribution respectively, we define M=2 Laplace distributions of N=10000 samples each (Fig.1.**A**),

$$\mathbf{d}_1 = [d_1(1), \ldots, d_1(k), \ldots, d_1(N)] \quad \text{where } d_1(k) = Lp(k),$$

$$\mathbf{d}_2 = [d_2(1), \ldots, d_2(k), \ldots, d_2(N)] \quad \text{where } d_2(k) = Lp(k),$$

$$D_{2xN} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

We then create M linear combinations,

$$\mathbf{S}_{NxM} = \mathbf{A}_{MxM} \cdot \mathbf{D}_{MxN},$$

Where **A** is a MxM matrix of steady-state scalar variables,

$$\mathbf{A}_{\mathrm{MxM}} = \begin{bmatrix} a_{11} & \dots & a_{1M} \\ & \vdots & \\ a_{M1} & \dots & a_{MM} \end{bmatrix},$$

Which in the considered case,

$$\mathbf{A}_{2\mathrm{x}2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{2} \end{bmatrix} = \begin{bmatrix} 10 & 20 \\ -2 & 10 \end{bmatrix}$$

And **S** is the MxN matrix containing the resulting M linear combinations (Fig.2.1.**B**),

$$\mathbf{S}_{\mathrm{MxN}} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix},$$

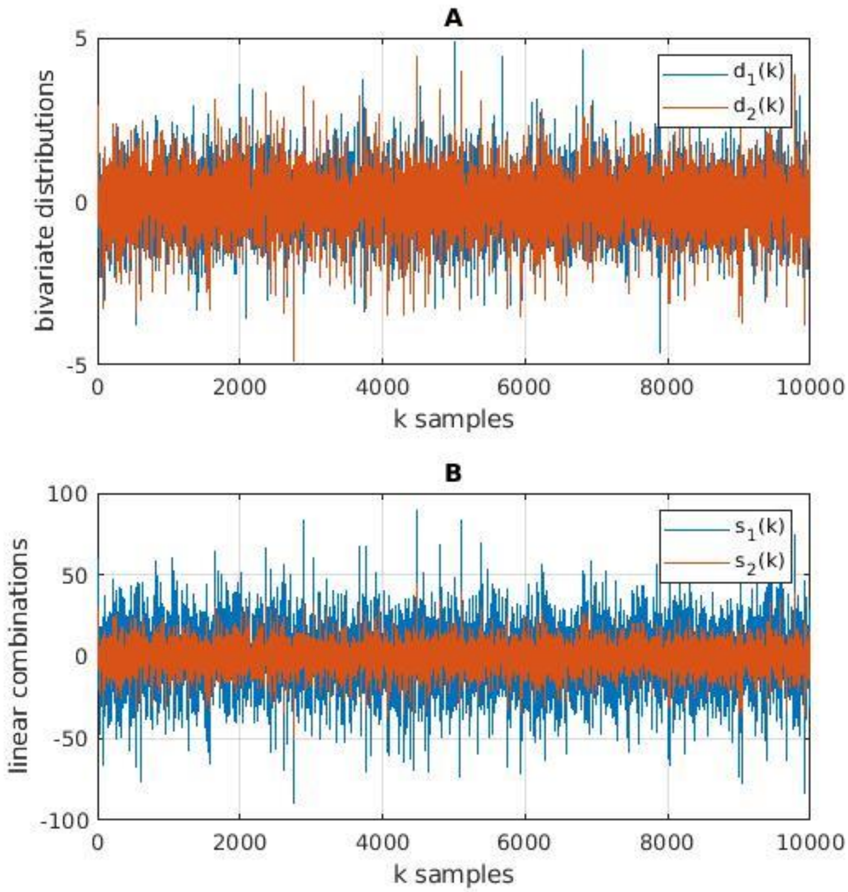

Figure 2.1. A the two laplacian distributions and **B** their linear combinations.

In Figure 2.2, the scatterplot shows the relationship existing between the linear combinations $s_1(k)$ and $s_2(k)$. We can see a strong linear positive correlation between the two.

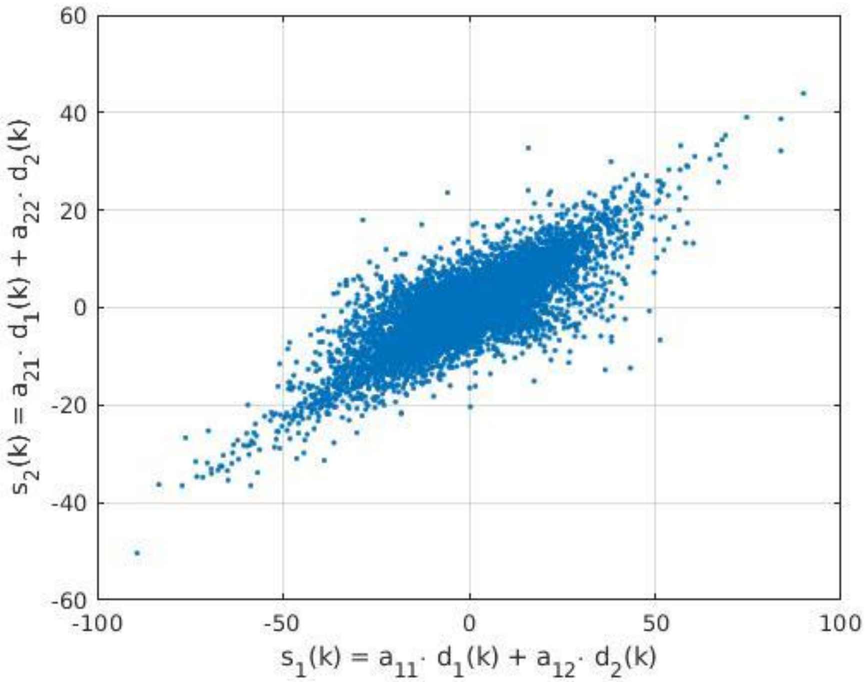

Figure 2.2. The two linear combinations plot one against the other.

Exercise 2 – PCA of the bivariate distribution

The PCA analysis on the **S** matrix returns a set of new variables, the $\mathbf{X}_{MxM}$ PC principal components,

$$\mathbf{X}_{2x2} = [\mathbf{x_1}_{2x1}, \mathbf{x_2}_{2x1}] = \begin{bmatrix} 0.93 & -0.36 \\ 0.36 & 0.93 \end{bmatrix},$$

and the transformed data, or scores, in a matrix **W** with the same dimension of $\mathbf{S}^T$. The PC scores are the representations of **S** in the principal component space. Rows of **W** correspond to the **N** observations, and columns correspond to the **M** components.

The total variance of original and reconstructed data remains the same. Although, it is distributed among the PC in the most unbalanced way. In Fig.2.3, it is represented the amount of variance each PC can possibly explain, this value is computed by finding the eigenvalues of the covariance matrix of $\mathbf{S}^T$. The first PC contains the 94,39% of the entire variance, or rather it is the PC that encodes most of the information, while the second PC represent the 5.60% of the total variance. In order to efficiently reconstruct the original data we need to select a number of PC with final total variance of at least 95% ot of the total, I.e., both PC are needed in this case.

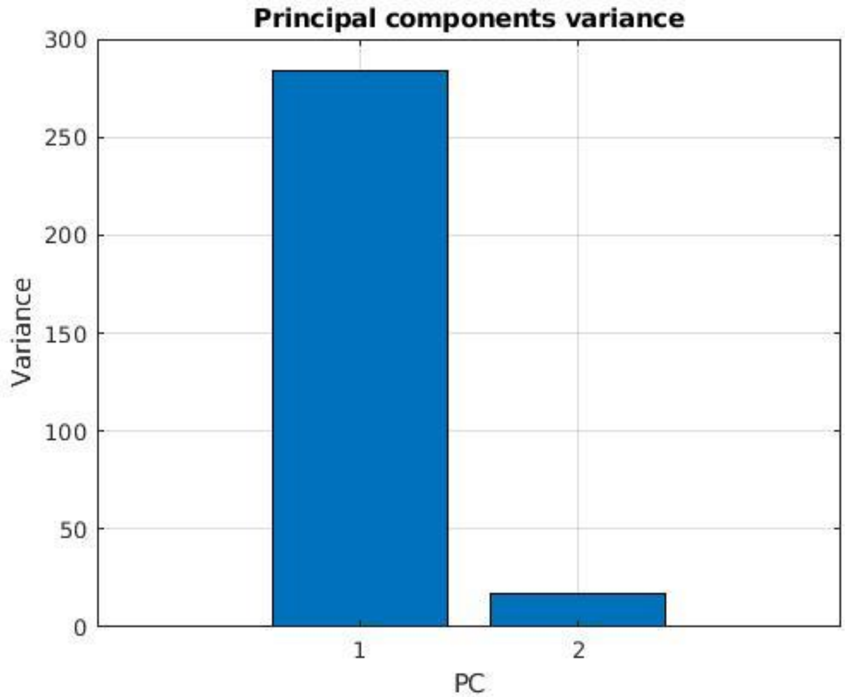

Figure 2.3. Principal components variance.

From the analysis of the score (Fig.2.4.**B**), the two distribution results still dependent, the scores radiate from the origin towards two main directions that not completely follow the two cartesian axis. In Fig.2.4.**C**, the reconstructed data

$$\mathbf{S'} = \mathbf{XW}^{T.}$$

are overlapped by the two PC, it is evident how the new variables are not following the two main direction of the data. One of the main drawbacks of using PCA with bivariate distribution, such as S, is the orthogonality assumption. As matter of fact, PCA works fine with Gaussian distribution, but it is not optimal for minimizing the dependency of non-orthogonal components. In general, PCA works better for normal distributions then for skewed or asymmetric distributions, where it can fail because it retains and maximize the variance only of the projected dimensions. PC method rotates the axis to decorrelate the data, in order to eliminate second order dependencies. However, if the distribution is not Gaussian there are higher order dependencies that can not be removed (as seen in Fig.2.4.**B**).

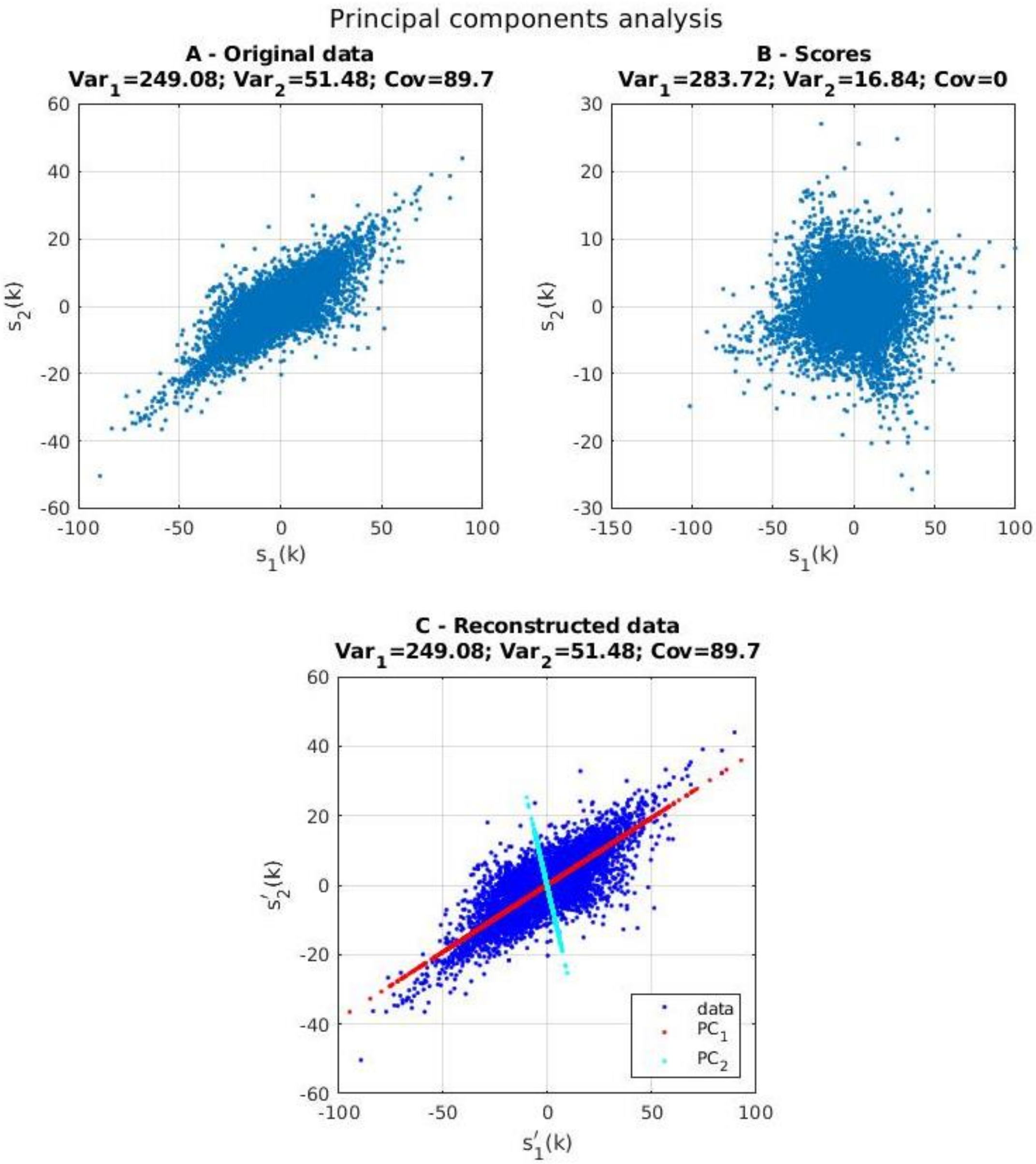

Figure 2.4. Principal components analysis: (**A)** the original data **S**; (**B)** the scores, i.e., the reconstructed data **S** in PC space; (**C)** the reconstructed data **S'**.

Exercise 3 – ICA of the bivariate distribution

The Independent Component Analysis (ICA) is a more efficient method than PCA when analyzing non-gaussian distribution such as **S**. We employed this method in order to separate the data into two independent non-gaussian distribution and the results demonstrate the efficacy of this method. In Fig.2.5.**B**-**C**, we compare the PCA and the ICA scores respectively, the latest one is independent and radiate from the origin in directions of the axis. The ICA score has equal unit variance, zero covariance, and different scale with respect to PCA score and the original data.

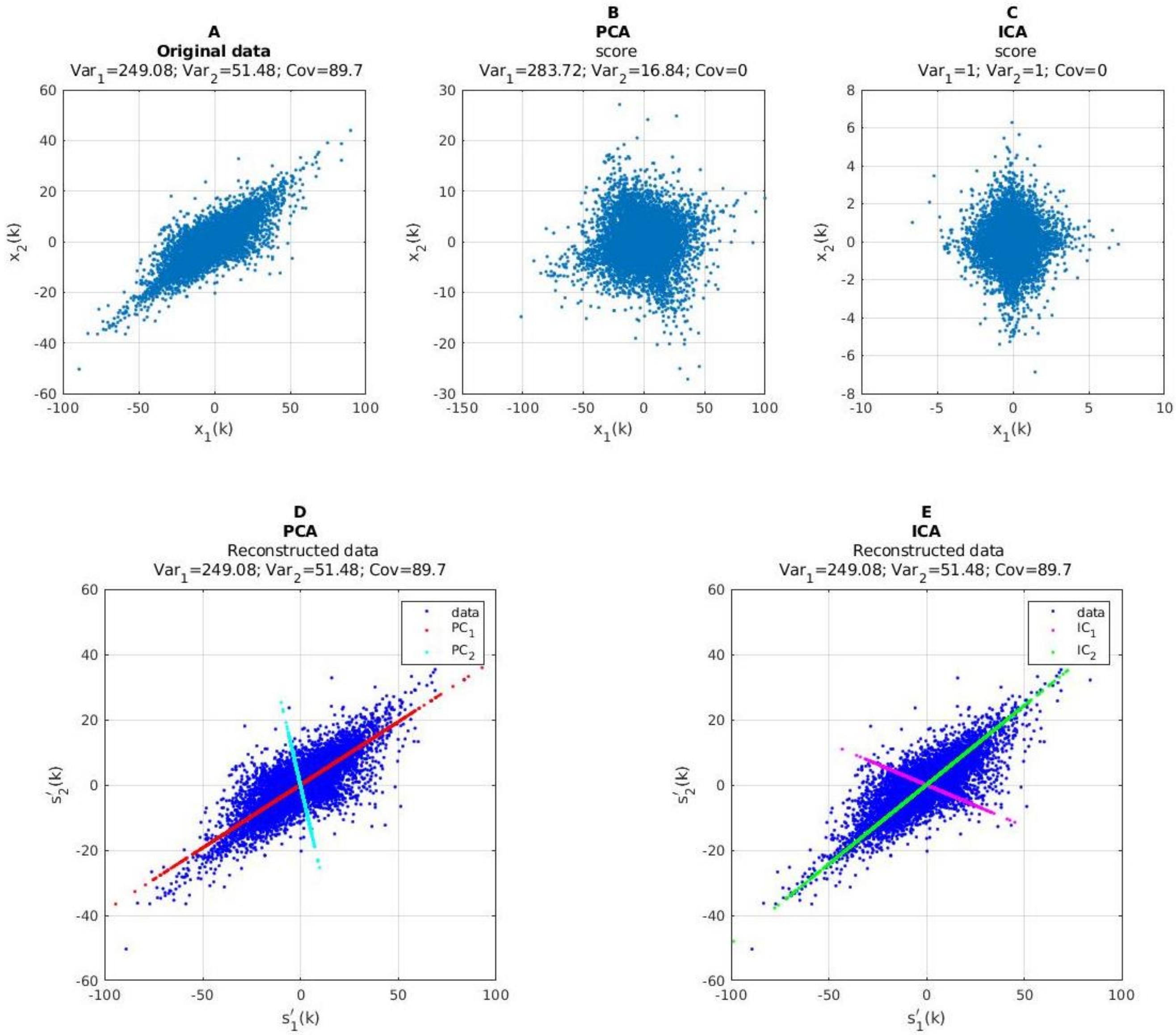

Figure 2.5. Principal vs Independent components analysis: **(A)** the original data S; **(B)** the PCA scores, i.e., the reconstructed data **S'** in PC space; **(C)** the ICA scores, i.e., the reconstructed data **S'** in IC space; (**D**) the data **S** reconstructed with PCs (in red and cyan); **(E)** the data S reconstructed with ICs (in magenta and green)**.**

Exercise 4 – Compare ICA results with the original samples

The ICA score has different scale with respect to the original data and the variance that each independent component IC can explain is equal, this could lead to a substantial loss of information or possible overfitting. As matter of fact, assigning equal variance to a variable that explain less information can lead to over fitting a specific feature, especially when working with high-dimensional space this approach could lead to a wrong reconstruction of the data.

$$\mathbf{W}_{ica} = A_{ica} \bullet \mathbf{S},$$

$$S'_{ica} = \mathbf{X}_{ica} \bullet \mathbf{W}_{ica}^{T} = \mathbf{X}_{ica} \bullet (A_{ica} \bullet \mathbf{S})^{T}.$$

Exercise 5 – ICA of linearly combined time series

Given two guitar soundtracks, we store N=10.000 samples of the 1-channel tracks with a sampling frequency,

$$F_s = 44{,}100 \text{ Hz},$$

into a matrix MxN,

$$\mathbf{D}_{\mathrm{MxN}} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \quad \mathrm{M}=2,$$

where,

$$\mathbf{d}_i = [d_i(1), \ldots, d_i(k), \ldots, d_i(N)],$$

We then create M linear combinations,

$$\mathbf{S}_{\mathrm{NxM}} = \mathbf{A}_{\mathrm{MxM}} \bullet \mathbf{D}_{\mathrm{MxN}} ,$$

where **A** is a MxM matrix of steady-state scalar variables,

$$\mathbf{A}_{\mathrm{MxM}} = \begin{bmatrix} 18 & 20 \\ -30 & 10 \end{bmatrix},$$

And **S** is the MxN matrix containing the resulting M linear combinations (Fig.2.6.**C**),

$$\mathbf{S}_{\mathrm{MxN}} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} .$$

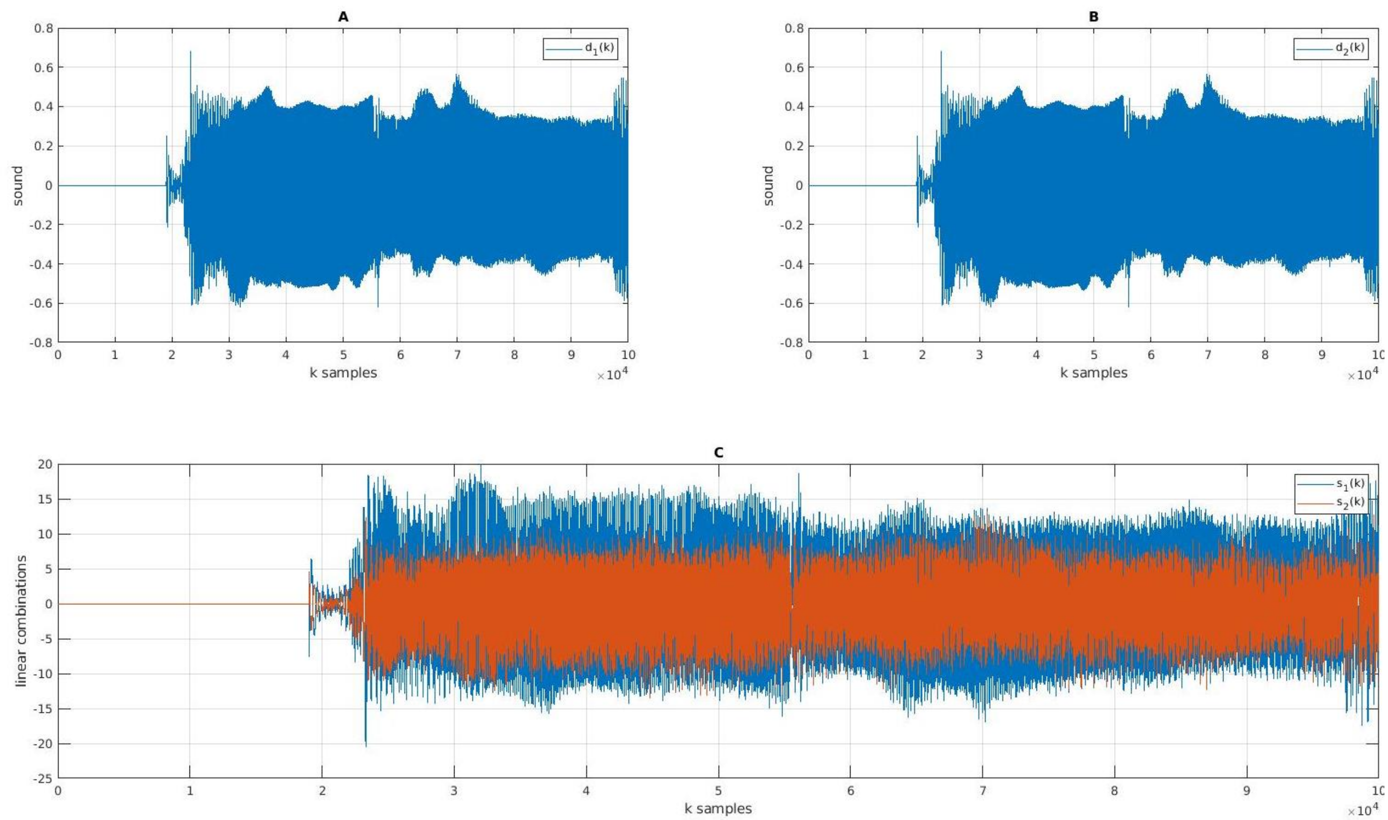

Figure 2.6. A-B the two original time series and **C** their linear combinations.

We employed the ICA method to separate the two linear combinations. The resulting 2 independent components are,

$$\mathbf{X}_{2\mathrm{x}2} = [\mathbf{x_1}_{2\mathrm{x}1}, \mathbf{x_2}_{2\mathrm{x}1}] = \begin{bmatrix} 1.0 & 4.05 \\ 1.27 & -6.6 \end{bmatrix}$$

Exercise 5.a - Reconstruction error

Figure 2.7 compares the **S** original data to the **S'** reconstructed and **W** IC-space (or score) data. From Figure 2.7.**E** the data seems correctly reconstructed and the IC follows the two main directions. However, if we calculate the maximum root mean square error between **S** and **S'** we will notice a large deviation,

$$\text{Max(RMSE)} = \max\ (\Sigma(N^{-1} \bullet (\mathbf{S}(k)-\mathbf{S}'(k))^2))^{0.5} = 6.71.$$

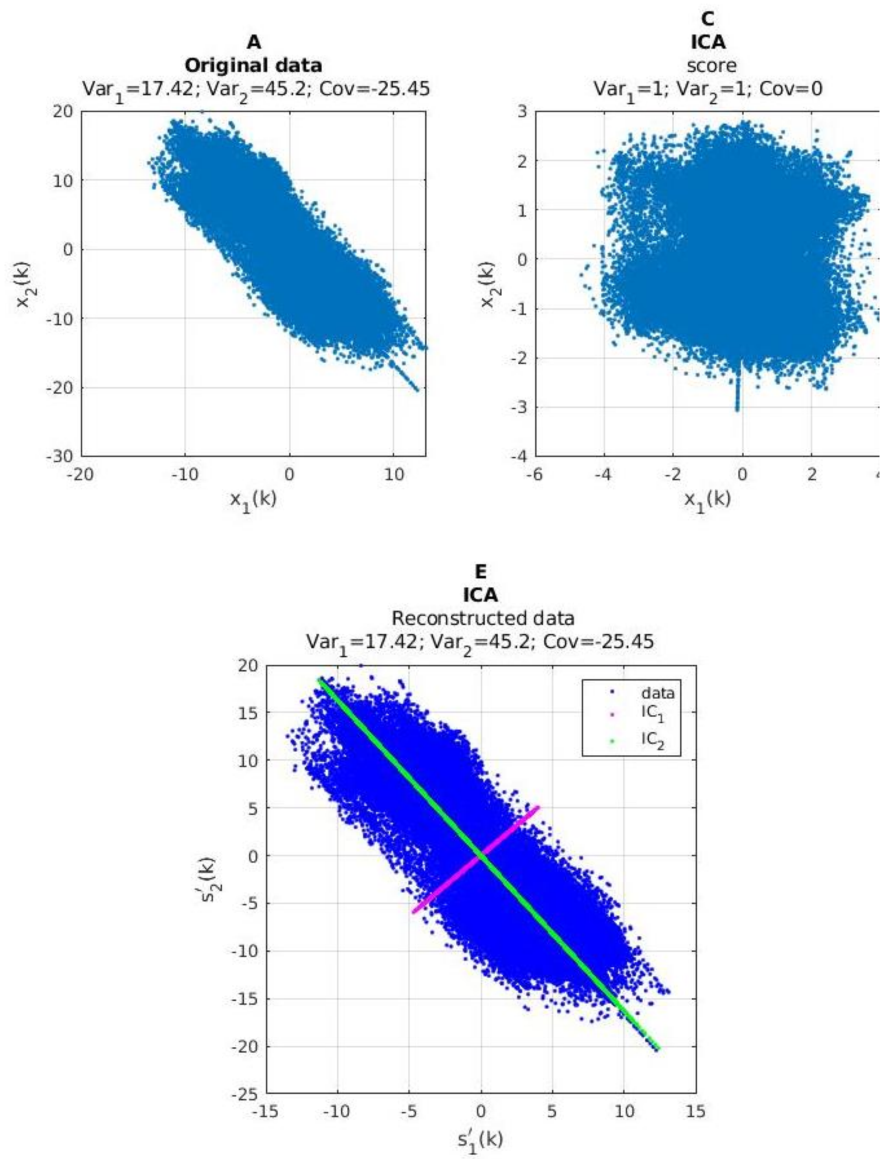

Figure 2.7. Independent components analysis: **(A)** the original data S; (**B)** the scores, i.e., the reconstructed data **S** in IC space; (**C**) the reconstructed data **S'**.

Exercise 5.b - Subjective perception evaluation

The two audio samples sound similar. Although, the reconstruction is noisy it is still possible to recognize the original sound.

Exercise 5.c - Compare the evaluation methods

Although the RMSE gives as a general evaluation about the deviation from the original data, the listener can still recognize the sound. In our opinion, the efficiency of the evaluation method depends on the application. For instance, such large error could be a problem if present in assistive device for people with hearing impairment, and the presence of noise could lead to severe discomfort.

Homework 3 – Bayesian multisensory integration

In this exercise, we have data from three experiments, one with only visual tasks, one with auditory tasks and one with a mixture of both. Each one collects 50 trials and their answers.

Our task is to obtain information about the signals of both stimuli using the data collected by estimating the mean values and the variance of each normal distribution. If only we had the results of the visual and auditory experiments, the task would be trivial, it would be enough to look for the optimal values of the mean and the variance that minimize each function neg log likelihood of both tasks.

What happens is that in addition to these data, we have the data from the joint experiment that also gives us information about the signals separately, and therefore should be considered to obtain more accurate results.

Knowing that the log likelihood functions are independent of each other, we construct a joint neg log likelihood function of the three corresponding to the signals of the three experiments, we substitute the mean and variance of the function of the audiovisual stimuli for their equivalence with the means and variances of the stimuli separately, in order to have only four variables to estimate, and we look for the optimal values that minimize this function.

The function neg log likelihood to minimize is the following:

$$-LogLikelihood = -\left(LogLikelihood_{Visual} + LogLikelihood_{Audio} + LogLikelihood_{Audio\,Visual}\right)$$

Where,

$$\text{LogLikelihood}(\mu, \sigma^2, x_1, \ldots, x_n) = (2\pi\sigma 2)^{-n/2} \exp(-(2\sigma 2)^{-1}\, \Sigma(x_j-\mu)^2)$$

The estimates are as follows:

Variables	Real	Calculated
'm_a'	-15.62523364	-15.3917948865441
's_a'	2.00197196121263	1.76029883124763
'm_v'	22.86545656	22.5307075121455
's_v'	2.39030759784214	2.53526702658564
'm_av'	-4.0469457068	-3.05652414282423
's_av'	1.82852413036846	2.09073560273964

Figure 3.1 Results of estimates

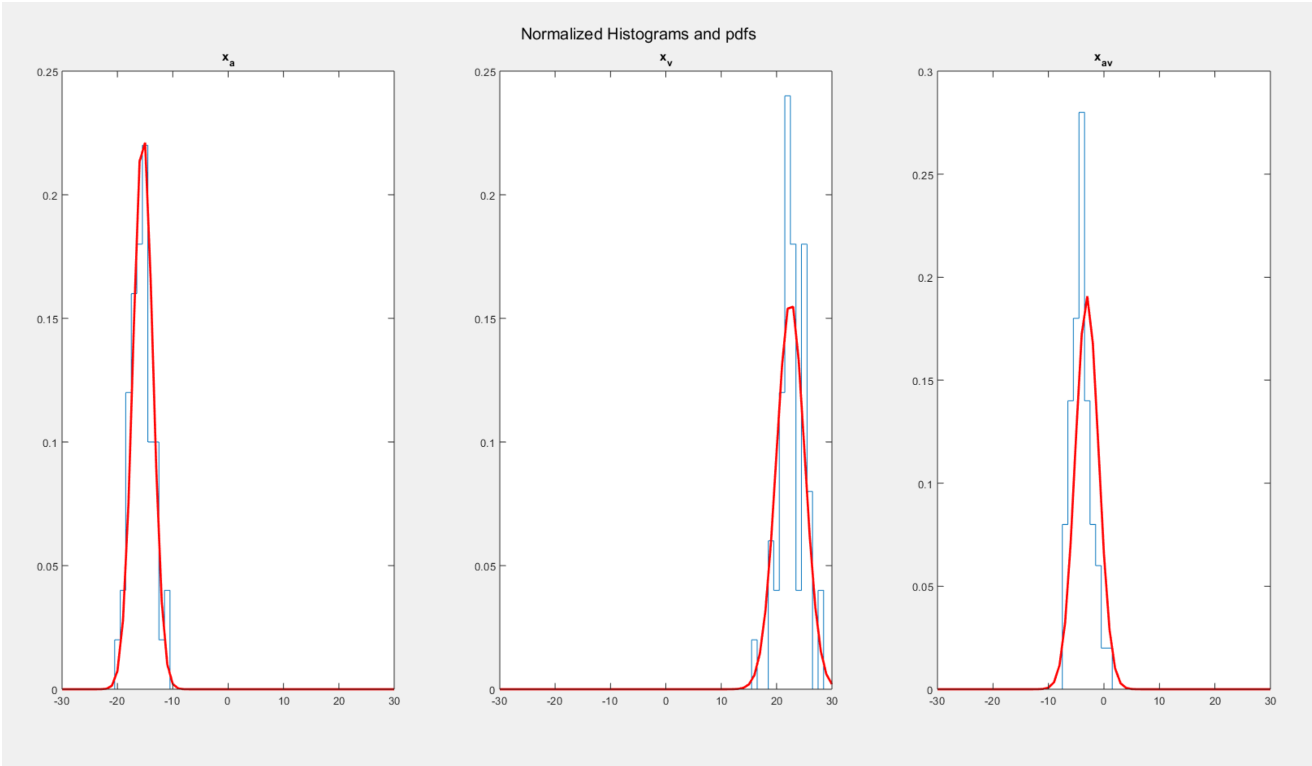

Figure 3.2 Goodness of fit, histogram of the real data, overlapped by pdf (estimated_variables)

Variables	Real	Calculated
'm_a'	-15.832550462	-16.0344807504864
's_a'	4.69142465642395	2.98992586944464
'm_v'	22.460033	22.2578545657044
's_v'	3.69968448661963	4.02608524684125
'm_av'	-3.584196146	-2.42280347103478
's_av'	3.37070735453587	5.76190008456044

Figure 3.3 Results of estimates with the wandering Observer (input lapse series)

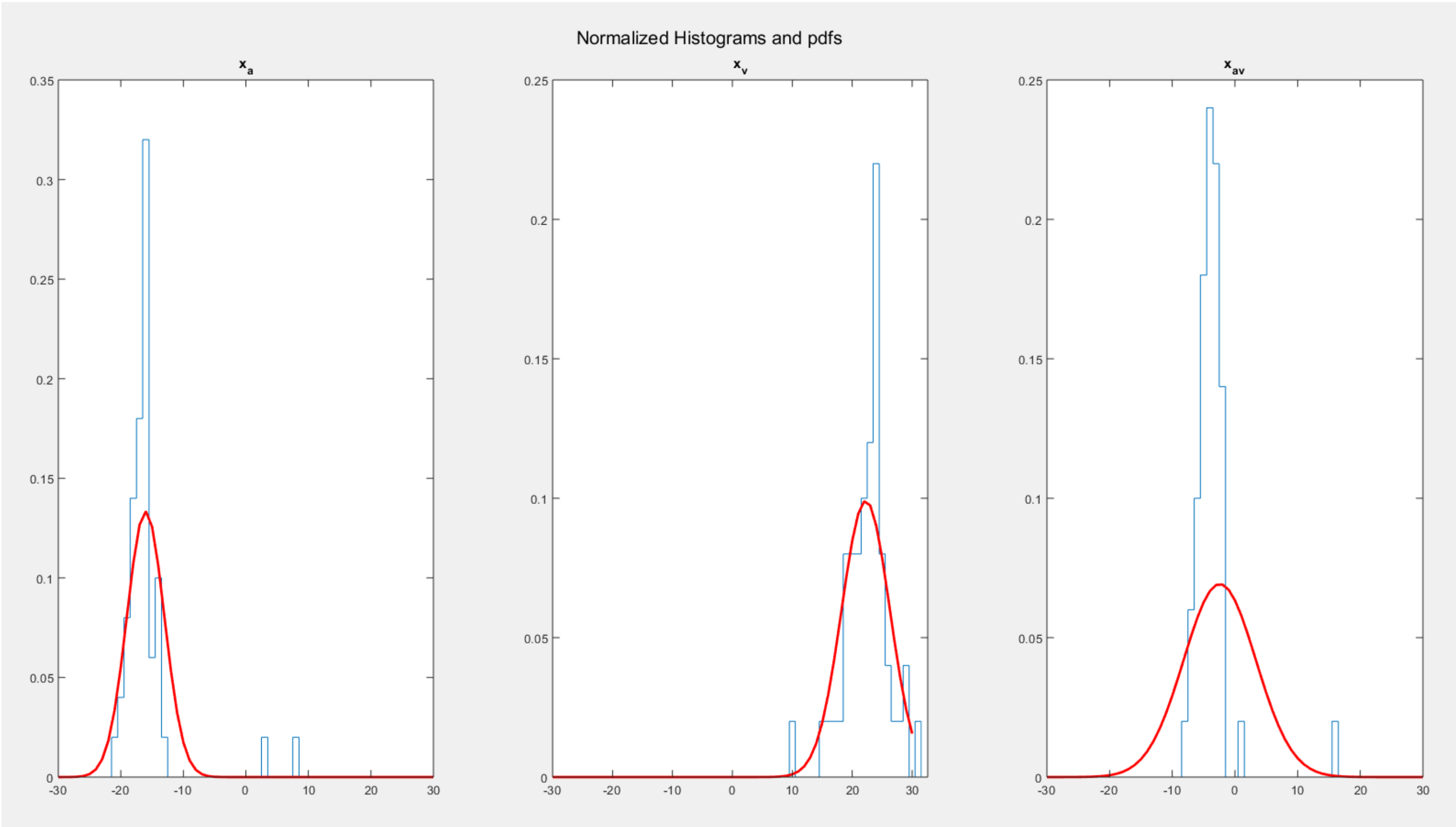

Figure 3.4 Goodness of fit, histogram of the real-modified (lapse) data, overlapped by pdf (new_estimated_variables)

As expected, the randomness of 4% impacts negatively on the estimates of the parameters and on the overall goodness of fit because there are outliers.