First Essay: Application of Principal Component Analysis to Image Compression

Author: Gabriel Harrison Fidelis Teotonio
e-mail: ghft1@de.ufpe.br
Department of Statistics - Nature and Exact Sciences Center
Federal University of Pernambuco

April 19, 2019

Abstract

This essay is a report of the paper *Application of Principal Component Analysis to Image Compression* (Hernandez and Mendez, 2018) with the aim to highlight the main points of theory and application with image compression. It will be evaluated as part of the first grade of the Multivariate Analysis 2 course, under the guidance of Professor Abraão D. C. Nascimento.

1 Introduction

Principal components analysis (PCA) is a multivarite analysis technique that has the objective to reduce the dimensionality in the data, in relation to the number of variables collected by creating new orthogonal variables that shall maintain the maximum of the variance in the data. This new variables are called by principal components. This method is also classified as a unsupervised machine learning algorithm, once we have only a set of variables $X_1, X_2, ..., X_p$ measured on n observations. In this case, we are not interested in prediction, because we do not have an associated response variable Y.

PCA can be used and applied in a lot of problems where the high dimension number is a problem or it is something that has an expensive cost in computations and interpretation. On section 3, there is an application of PCA in image compression. Suppose an image, say .jpeg or .png, and this image has a number of p variables that describe it and there is an amount of storage needed to save it in relation to the dimension of this image. Can we storage this image using less space on disk? That is, storage it with a lower number of p original variables and keep the image identifiable by human look. PCA fits perfectly on this problem. A few number of principal components can be found in order to maintain most part of the variance and so keep the image visible. In order the evaluate this *visibility* or identification of the image by using principal components, consider some measures as *Peak Signal-to-Noise Ration* (PSNR) and *Structural Similarity Index* (SSIM) to see the quality of reconstruction.

2 Theory Recapitulation

This section is a compiled of the theory background required to perform and interpret PCA.

2.1 Probability

Let $\mathbf{X} = (X_1, X_2, ..., X_p)^t$ be a random vector of dimension p. The expected values, variances, and covariances can be defined as

$$\begin{aligned} \boldsymbol{\mu} = E[\boldsymbol{X}] &= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \\ \boldsymbol{\Sigma} = Cov[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t] &= \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}, \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}$ is the population mean vector and $\boldsymbol{\Sigma}$ is the population covariance matrix.
The population correlation matrix is given by $\boldsymbol{\rho}$, where $\rho_{ij} = \dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$.
Consider $\mathbf{X}$ a simple random sample of a p-dimensional random variable. The estimators of the previous population quantities:

$$\begin{aligned} \bar{\boldsymbol{X}} &= \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \\ \boldsymbol{S} &= \frac{n}{n-1}(\boldsymbol{X} - \bar{\boldsymbol{X}})(\boldsymbol{X} - \bar{\boldsymbol{X}})^t \end{aligned} \tag{2}$$

.

$\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The sample correlation matrix is $\boldsymbol{R}$, where $r_{ij} = \dfrac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$

2.2 Linear Algebra

Some considerations about linear algebra:

- Consider $\boldsymbol{A}$ a square matrix. If $\boldsymbol{v^t A v} \geq 0$ for any vector $\boldsymbol{v}, \mathbf{A}$ is a nonnegative definite matrix;
- If $\boldsymbol{v^t A} = \lambda \boldsymbol{v}$, with $\boldsymbol{v} \neq 0, \lambda$ is an eigenvalue associated with the eigenvector $\boldsymbol{v}$;
- Let $\boldsymbol{A}$ be a symmetric $p \times p$ matrix with real-valued entries. $\boldsymbol{A}$ has p pairs of eigenvalues and eigenvectors, $(\lambda_1, \boldsymbol{e}_1), ..., (\lambda_p, \boldsymbol{e}_p)$;
- The spectral decomposition of $\boldsymbol{A}$ is $\boldsymbol{A} = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^{\boldsymbol{t}} + ... + \lambda_p \boldsymbol{e}_p \boldsymbol{e}_p^{\boldsymbol{t}}$;

- If $\boldsymbol{P} = (\boldsymbol{e_1}, ..., \boldsymbol{e_p})$ is an orthogonal matrix and Λ is a diagonal matrix with main diagonal entries $(\lambda_1, ..., \lambda_p)$, the spectral decomposition of $\boldsymbol{A}$ can be given by $\boldsymbol{A} = \boldsymbol{P\Lambda P^t}$. Therefore, $\boldsymbol{A^{-1}} = \boldsymbol{P\Lambda^{-1}P^t} = \sum_{i=1}^{p} \frac{1}{\lambda_i} \boldsymbol{e_i e_i^t}$.

2.3 Population Principal Component

Let $\boldsymbol{X} = (X_1, X_2, ..., X_p)^t$ be a p-dimensional random vector with covariance matrix $\boldsymbol{\Sigma}$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Consider the following p linear combinations:

$$\begin{aligned} Y_1 &= l_1^t \boldsymbol{X} = l_{11} X_1 + ... + l_{p1} X_p \\ &\vdots \\ Y_p &= l_p^t \boldsymbol{X} = l_{1p} X_1 + ... + l_{pp} X_p \end{aligned} \tag{3}$$

These new random variables verify the following equalities:

$$\begin{aligned} &V[Y_i] = l_i^t \boldsymbol{\Sigma} l_i \qquad i = 1, ..., p \\ &Cov[Y_i, Y_j] = l_i^t \boldsymbol{\Sigma} l_j \qquad i, j = 1, ..., p \quad i \neq j \end{aligned} \tag{4}$$

Principal components are those linear combinations that, being uncorrelated among them, have the greatest possible variance. Thus, the first principal component is the linear combinations with the greatest variance, that is, $V[Y_1] = l_1^t \boldsymbol{\Sigma} l_1$ is the maximum. Since if we multiply l_1 by some constant the previous variance rows, we will restrict our attention to vectors of norm one with the aforementioned indeterminacy disappears. The second principal component is the linear combination that maximizes the variance and in uncorrelated with the first on, and the norm of the coefficient vector is equal to 1.

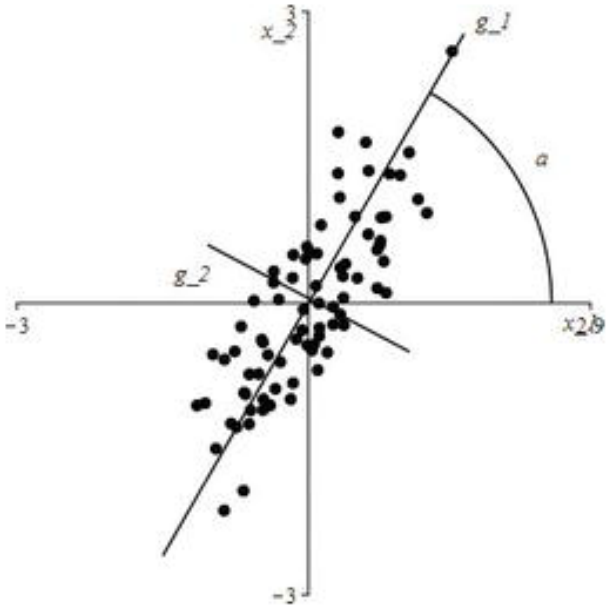


Figure 1: Scatter plot with two original variables and two new variables in relation to a rotation with angles a to achieve the maximum variance for the first new variable and the second one as orthogonal.

Let $\mathbf{\Sigma}$ be the covariance matrix of the random vector $\boldsymbol{X} = (X_1, X_2, ..., X_p)^t$. Let us assume that $\mathbf{\Sigma}$ has p pairs of the eigenvalues and eigenvectors, $(\lambda_1, \boldsymbol{e}_1), ..., (\lambda_p, \boldsymbol{e}_p)$, with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Then, the ith principal component is given by

$$Y_i = \boldsymbol{e_i^t X} = e_{1i}X_1 + ... + e_{pi}X_p \qquad i = 1, ..., p \tag{5}$$

With this in mind we can verify that:

- $V[Y_i] = \boldsymbol{e_i^t \Sigma e_i} = \lambda_i, \qquad i = 1, ..., p;$
- $Cov[Y_i, Y_j] = 0 \qquad i, j = 1, ..., p \quad i \neq j;$
- $\sigma_{11} + ... + \sigma_{pp} = \sum_{i=1}^{p} V[X_i] = \lambda_1 + ... + \lambda_p = \sum_{j=1}^{p} V[Y_j].$

3 Application to Image Compression

In this section, we consider the famous and known image of Lena, highly used as research application in field of image processing, in the application of image compression by using PCA.

Figure 2: Black and white photograph of Lena

3.1 Compression Process

The black and white photograph shown in Figure 2 was considered. PCA assumes that the data is a matrix $\boldsymbol{X}$ which has n observations and p variables. Taking it into account, we need to do some processing in the image before compute PCA. First, the image in .jpg format was converted into he

numerical matrix **Image** of dimension 512 by 512. Second, to obtain the observations vectors, the matrix was divided into blocks of dimension 64 by 64, $\boldsymbol{A_{ij}}$, with which 4096 blocks were obtained, and each of them was a vector of observations.

$$\textbf{Image} = \begin{pmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,64} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{64,1} & \cdots & \boldsymbol{A}_{64,64} \end{pmatrix} \tag{6}$$

Third, each matrix $\boldsymbol{A_{ij}}$ was stored in a vector of dimension 64, $\boldsymbol{x}$, which contained the elements of the matrix by rows, that is, $\boldsymbol{x} = (a_{i,1}, ..., a_{i,8}, a_{i+1,1}, , ..., a_{i+1,8}, a_{i+8,8})$. This way, we had the observations $\{\boldsymbol{x}_k \in \Re^{64} | k = 1, ..., 4096\}$, which were grouped in the observation matrix $\boldsymbol{x} \in \mathbf{M}_{4096,64}$.
Fourth, the average of each column, $\bar{\boldsymbol{x}} = (\bar{x}_1, ..., \bar{x}_{64})$ was calculated obtaining the vector of means, and from each observation x_{ij}, its corresponding mean $\bar{x}_j$ was subtracted. Thus, the matrix of centered observations $\boldsymbol{U}$ was obtained. The covariance matrix of $\boldsymbol{x}$ was $\boldsymbol{S} = \boldsymbol{U^t U} \in \mathbf{M}_{64,64}$.
Fifth, the 64 pairs of eigenvalues and eigenvectors of $\boldsymbol{S}, (\hat{\lambda}_i, \hat{\boldsymbol{e}}_{\boldsymbol{i}})$, were found, and they were ordered according to the eigenvalues from the highest to lowest.

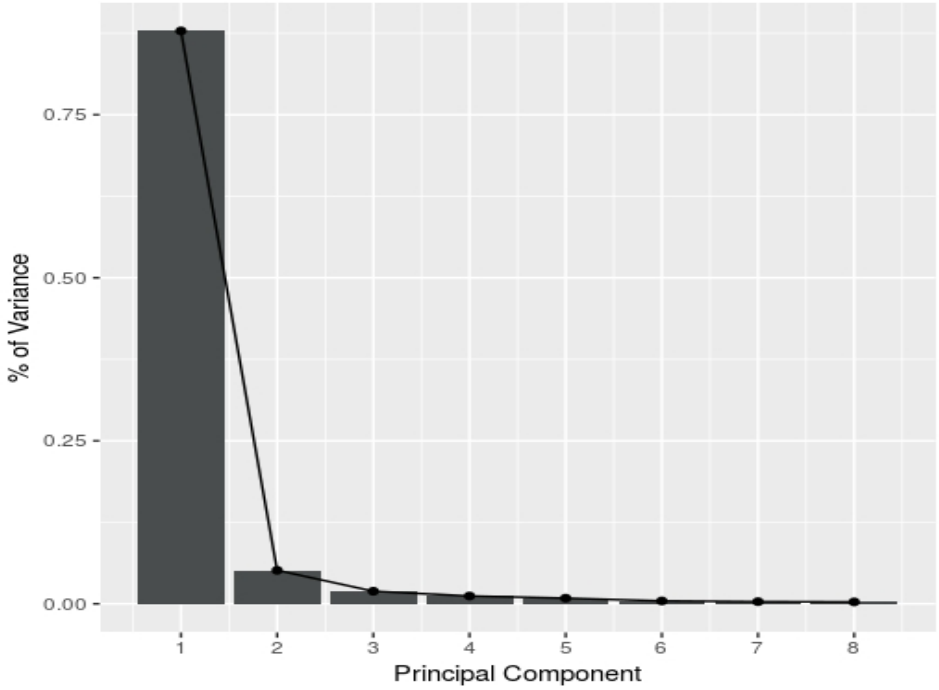


Figure 3: Variation proportion maintained by principal component.

In the Figure 3 we have the 8 largest eigenvalues. As can be seen, the first eigenvalue is much larger than the rest. Thus, the first principal component completely dominates the total variability, about 87%.
Sixth, we compute the 64 principal components, $\boldsymbol{y}$, from the eigenvalues and eigenvectors.
Figure 4 shows the image matrices of some principal components. We can observe that the image of fourth principal component captures much more boundaries if comparing with the 35th and 64th component. Each eigenvector are associated with a eigenvalue, and we know they (λ_i) are ordered, so the

eigenvectors associated with the greatest eigenvalues must have images with more identification of limits and boundaries. Seventh, now we calculate the inverse of the transformation $\boldsymbol{y} = \boldsymbol{x}.(\hat{\boldsymbol{e}}_{\boldsymbol{1}}^{\boldsymbol{t}}, ..., \hat{\boldsymbol{e}}_{\boldsymbol{64}}^{\boldsymbol{t}}) \Rightarrow \boldsymbol{x} = \boldsymbol{y}.(\hat{\boldsymbol{e}}_{\boldsymbol{1}}^{\boldsymbol{t}}, ..., \hat{\boldsymbol{e}}_{\boldsymbol{64}}^{\boldsymbol{t}})^t$
Eighth, to reconstruct the compressed image, each row of the result of last item was regrouped in an 8 by 8 matrix and then into a numerical matrix of dimension 512 by 512.

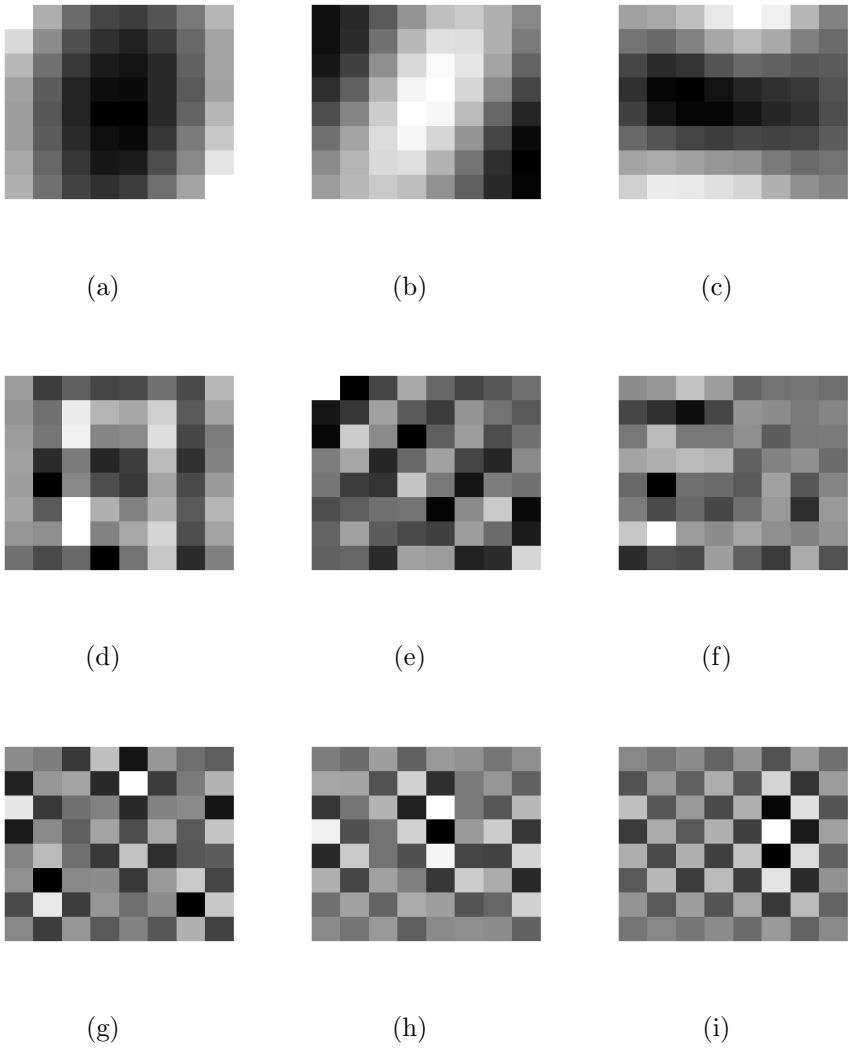
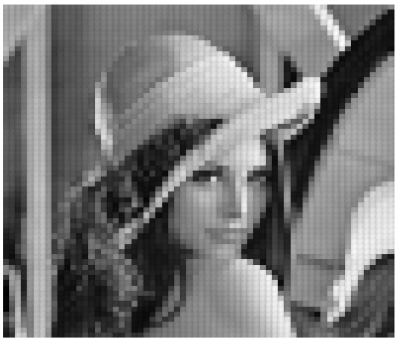


Figure 4: Images of the matrices of the 1st, 4th, 8th, 25th, 30th, 35th, 55th, 60th, and 64th principal components. (a) 1st. (b) 4th. (c) 8th. (d) 25th. (e)30th. (f) 35th. (g) 55th. (h) 60th. (i) 64th.

Figure 5 shows the original and some compressed images varying the number of principal components. As the graphic in Figure 3 indicated, the first principal component captures about 87% of the variability. We can see it in the item (a), Figure 5. The first seven principal components captures about 97% of the variability and the image (c) already display a really good approximation to the original photograph.

3.2 Measures of The Quality of Reconstructions

Here, we will use three methods to evaluate the quality of the reconstructions. Consider the methods *Mean Square Error* (MSE), *Peak Signal-to-Noise Ratio*



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5: Original and compressed image with one, three, seven, ten, and thirteen principal components. (a) Compression with one component. (b) Compression with three components. (c) Compression with seven components. (d) Compression with ten components. (e) Compression with thirteen components. (f) Original image.

(PSNR) and *Structural Similarity Index* (SSIM). The MSE and PSNR measures evaluate the quality in terms of deviations between the process and the original image.

Let N be the number of rows by the number of columns in the image. Let $\{x_n | n = 1, ..., N\}$ be the set of pixels of the original image. Let $\{y_n | n = 1, ..., N\}$ be the set of reconstruction pixels. Let $\{r_n = x_n - y_n | n = 1, ..., N\}$ be the error. The MSE is

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^{N} r_n^2 \tag{7}$$

Let the images under study be the 8 bit images. The PSNR of the reconstruction is

$$\text{PSNR} = 10 log_{10} \left(\frac{(2^8 - 1)^2}{\text{MSE}} \right) \tag{8}$$

When the goal is measure the human perception, the SSIM comes out to evaluate this quantity, see (Wang, Bovik, Sheikh, and Simoncelli, 2004)
We have in Figure 6 the MSE for different numbers of PC. One can observe as the number of PC increases, the MSE has decreased. Until nº PC = 20, the curve decreases really fast, after nº PC = 20 the curve goes to 0 slower. This is related to the variance maintained by first principal components. The Figure 7 shows the curve of the SSIM. The SSIM is between 0 and 1. The closer to 0, the fewer the images resemble each other. And the closer to 1, the images resemble each other. We can identify the same behaviour if comparing with the MSE plot. After the nº PC = 20, the SSIM goes slower to 1, once the first principal component already gives us a good similarity to the original image in relation to our perception.
Figure 8 gives us the quality measures in relation to the number of PC evaluated on Figure 5. The inversely proportional relation between MSE and (PSNR;SSIM) is explicit and we can see that a few number of PC already give us a good image reconstruction. Also, all the compressed images had a smaller storage size.

4 Conclusion

Principal component analysis as a reduction dimensionality technique has showed a great performance on image compression. It was developed a computational implementation to evaluate and analyze the performance of this compression in different contexts. The analysis was realized taking account a block matrix, 8 by 8. We could confirm the behaviour of PCA in capture the variance in the data by a few new variables in the application to the Lena photograph. Excellent compression was developed using only seven variables (principal components), which captured about 97% of total variance and and SSIM equal to 0.93, approximately. That is, beyond maintain great part of the variation in the data, it also keep the similarity and identification by observation. A further investigation could evaluate the performance of PCA in image compression when we vary the dimension of the block matrix and vary the image too. To compare the results by using data with different variances.

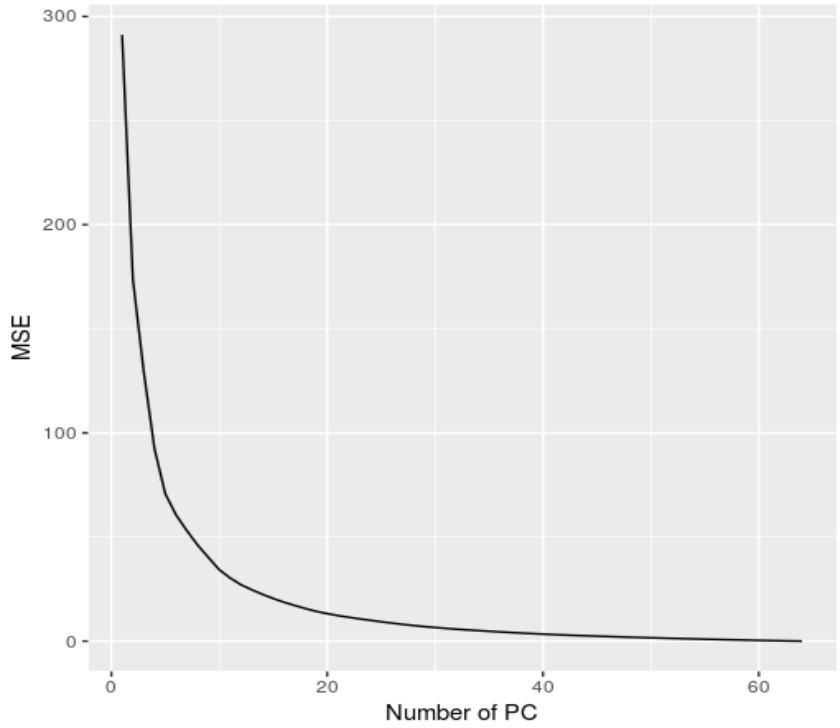


Figure 6: MSE for different numbers of PC.

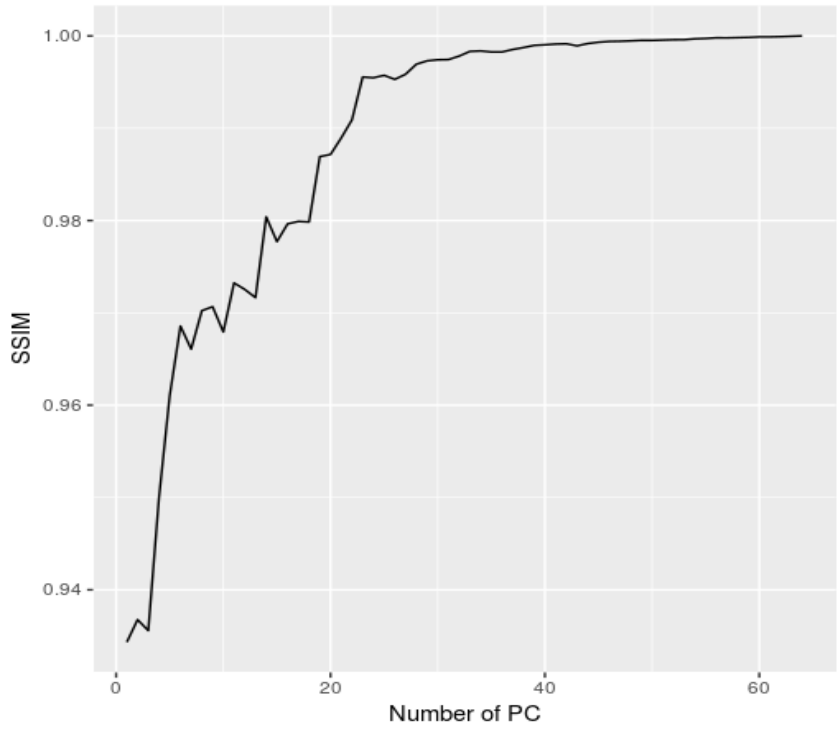
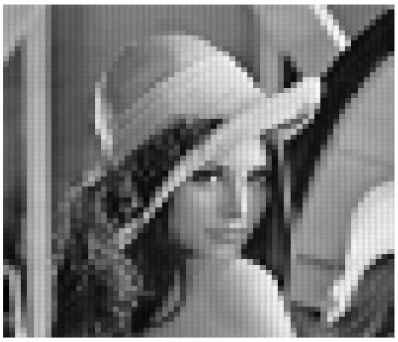


Figure 7: SSIM for different numbers of PC.



(a) MSE = 291; PSNR = 23; SSIM = 0.9343

(b) MSE = 129; PSNR = 26; SSIM = 0.9356

(c) MSE = 53; PSNR = 30; SSIM = 0.9661

(d) MSE = 34; PSNR = 32; SSIM = 0.9679

(e) MSE = 24; PSNR = 33; SSIM = 0.9716

Figure 8: Original and compressed image with one, three, seven, ten, and thirteen principal components evaluating quality measures. (a) Compression with one component. (b) Compression with three components. (c) Compression with seven components. (d) Compression with ten components. (e) Compression with thirteen components.

5 References

D. KI, K. SY. *Principal Component Neural Networks: Theory and Applications*. John Wiley Sons; 1996.

D. Wichern, R. Johnson. *Applied Multivariate Statistical Analysis*. 6th ed. Pearson Education Limited; 2014.

D. Salomon, G. Motta, and D. Bryant. *Data Compression: The Complete Reference*, ser. Molecular biology intelligence unit. Springer, 2007.

G. James, D. Witten, T. Hastie, R. Tibishirani. 7th ed. *An Introduction Statistical Learning*. Springer, 2017.

R, Oliveira. *Aproximações para a DCT Baseadas em Medida Angular: Baixa Complexidade e Compressão de Imagens*. Trabalho de Conclusão de Curso da Graduação em Estatística na Univesidade Federal de Pernambuco, 2016.

V. Britanak, P. Yip, and K. R. Rao. *Discrete Cosine and Sine Transforms*. Academic Press, 2007.

W. Hernandez, A. Mendez. *Application of Principal Component Analysis to Image Compression*. IntechOpen; 2018.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. *Image quality assessment: from error visibility to structural similarity*. IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600– 612, Apr. 2004.