Problem 4.2

Chen Bo Calvin Zhang

26/10/2020

In this problem, we are going to explore the five natural whitening procedures seen in the lectures using a sample of generated data. In particular, we are going to look at ZCA, ZCA-cor, PCA, PCA-cor, and Cholesky.

We will need the following libraries for this problem.

```
library("mnormt")
library("corpcor")
library("whitening")
```

First, we want to generate a synthetic bivariate normal data set with sample size $n = 500$ and with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$ and mean zero.

```
mu = c(0, 0)
Sigma = matrix(c(2, 2, 2, 4), nrow = 2, ncol = 2)
X = rmnorm(500, mu, Sigma)
```

We will then use the command whiten() in the R package "whitening" produce 5 different whitened versions of the simulated data set, corresponding to ZCA, ZCA.cor, PCA, PCA.cor and Cholesky whitening.

```
Z.ZCA = whiten(X, method = "ZCA")
Z.ZCA.cor = whiten(X, method = "ZCA-cor")
Z.PCA = whiten(X, method = "PCA")
Z.PCA.cor = whiten(X, method = "PCA-cor")
Z.Chol = whiten(X, method = "Cholesky")
```

Now, we have to verify quantitatively that the five methods produce uncorrelated transformed variable. This means that we need to check that the covariance matrices are identities.

```
print(cov(Z.ZCA))
```

```
##               [,1]          [,2]
## [1,]  1.000000e+00 -1.707814e-16
## [2,] -1.707814e-16  1.000000e+00
```

```
print(cov(Z.ZCA.cor))
```

```
##              [,1]         [,2]
## [1,] 1.000000e+00 5.975562e-17
## [2,] 5.975562e-17 1.000000e+00
```

```r
print(cov(Z.PCA))
```

```
##               [,1]          [,2]
## [1,]  1.000000e+00 -1.871316e-16
## [2,] -1.871316e-16  1.000000e+00
```

```r
print(cov(Z.PCA.cor))
```

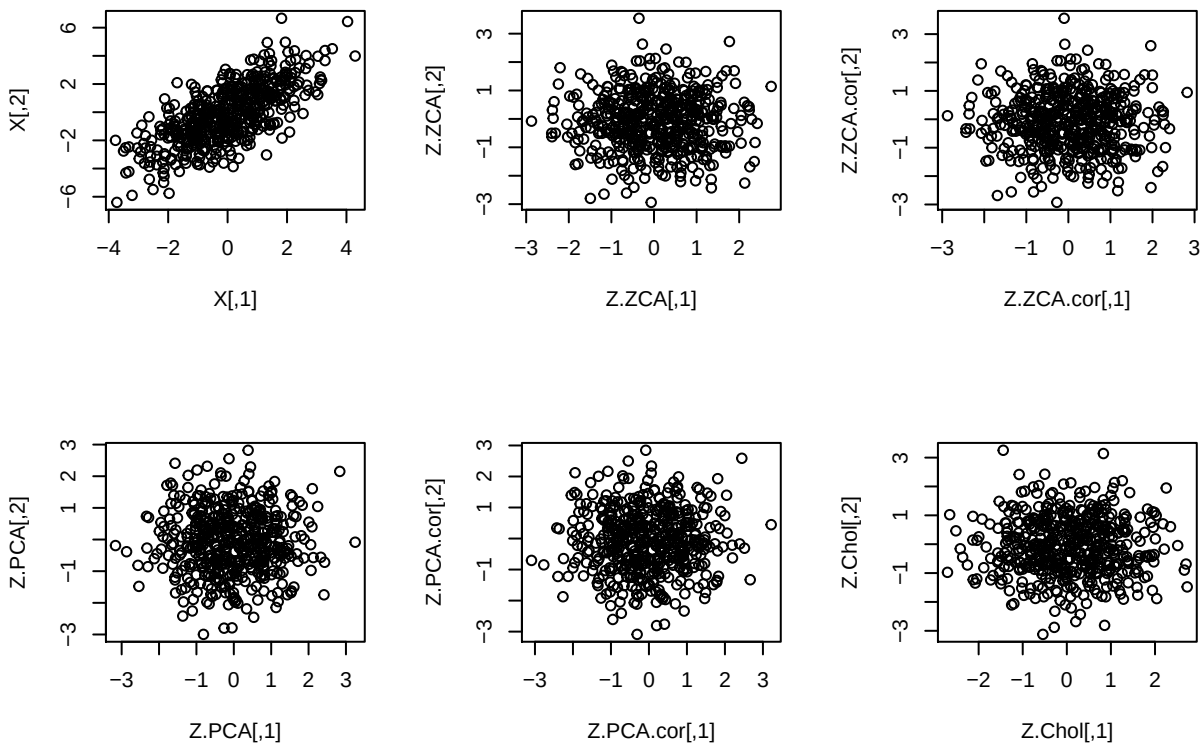
```
##              [,1]         [,2]
## [1,] 1.000000e+00 6.594546e-17
## [2,] 6.594546e-17 1.000000e+00
```

```r
print(cov(Z.Chol))
```

```
##               [,1]          [,2]
## [1,]  1.000000e+00 -1.122434e-16
## [2,] -1.122434e-16  1.000000e+00
```

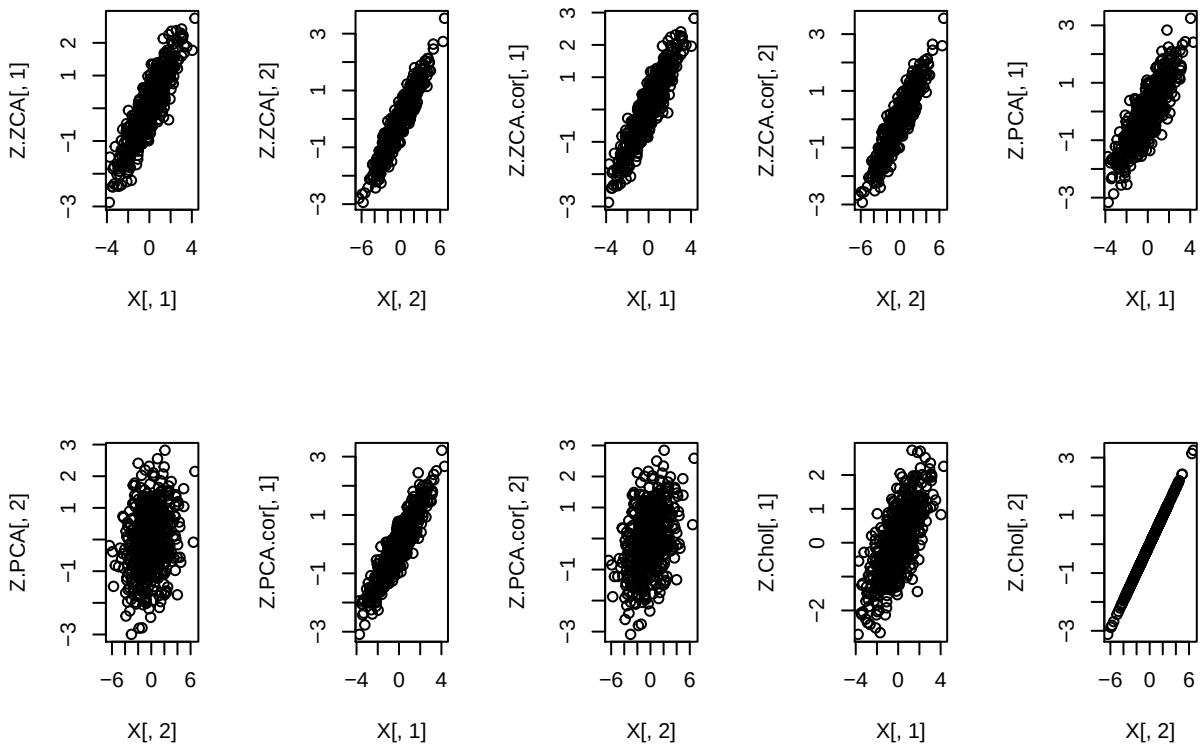
We also want to verify the same thing graphically.

```r
par(mfrow = c(2, 3))
plot(X)
plot(Z.ZCA)
plot(Z.ZCA.cor)
plot(Z.PCA)
plot(Z.PCA.cor)
plot(Z.Chol)
```



Let us inspect the pairwise scatterplots between corresponding components of the original and the transformed data.

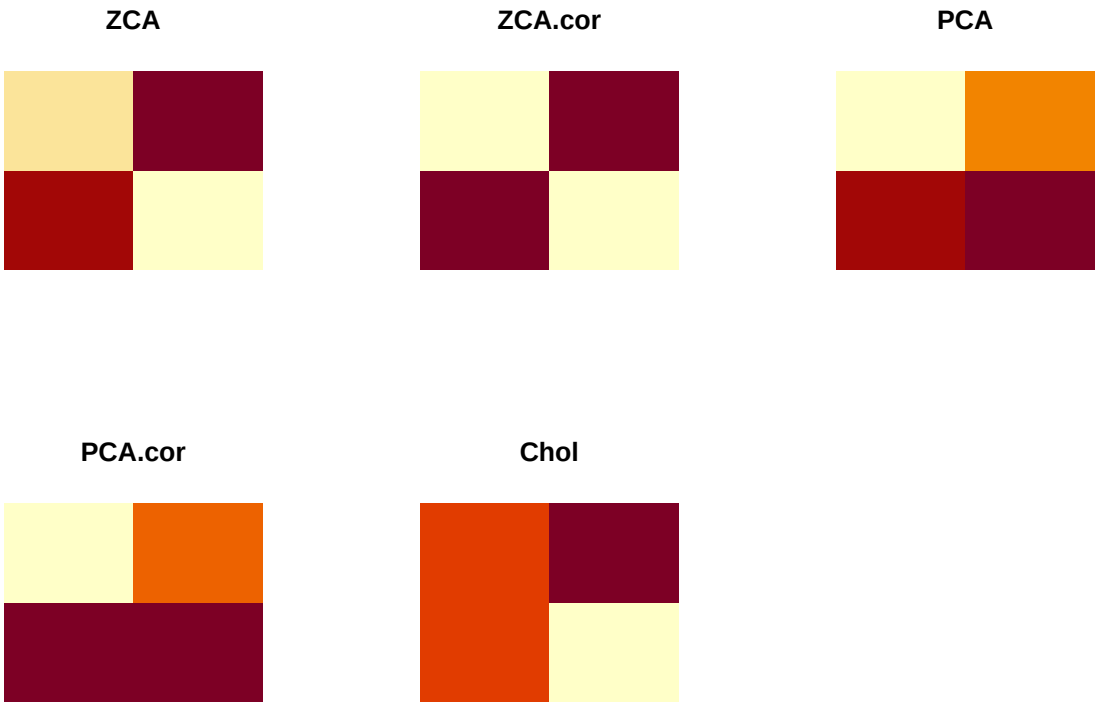
```
par(mfrow = c(2, 5))
plot(X[, 1], Z.ZCA[, 1])
plot(X[, 2], Z.ZCA[, 2])
plot(X[, 1], Z.ZCA.cor[, 1])
plot(X[, 2], Z.ZCA.cor[, 2])
plot(X[, 1], Z.PCA[, 1])
plot(X[, 2], Z.PCA[, 2])
plot(X[, 1], Z.PCA.cor[, 1])
plot(X[, 2], Z.PCA.cor[, 2])
plot(X[, 1], Z.Chol[, 1])
plot(X[, 2], Z.Chol[, 2])
```



We also need to compute and plot the cross-correlation matrix for all five approaches.

```
Psi.ZCA = cor(Z.ZCA, X)
Psi.ZCA.cor = cor(Z.ZCA.cor, X)
Psi.PCA = cor(Z.PCA, X)
Psi.PCA.cor = cor(Z.PCA.cor, X)
Psi.Chol = cor(Z.Chol, X)

par(mfrow = c(2, 3))
image(t(Psi.ZCA), axes = FALSE, main = "ZCA")
image(t(Psi.ZCA.cor), axes = FALSE, main = "ZCA.cor")
image(t(Psi.PCA), axes = FALSE, main = "PCA")
image(t(Psi.PCA.cor), axes = FALSE, main = "PCA.cor")
image(t(Psi.Chol), axes = FALSE, main = "Chol")
```



Lastly, we want to demonstrate that ZCA.cor whitening produces the optimal component-wise cross-correlation $Tr(\mathbf{\Psi})$ and that PCA.cor whitening produces the optimal compression in terms of squared cross-correlations $Diag(\mathbf{\Psi}\mathbf{\Psi}^T)$.

```
# demonstrate that ZCA.cor whitening produces
# the optimal average component-wise cross-correlations
print(sum(diag(Psi.ZCA)))
```

```
## [1] 1.845943
```

```
print(sum(diag(Psi.ZCA.cor)))
```

```
## [1] 1.850577
```

```
print(sum(diag(Psi.PCA)))
```

```
## [1] 1.076937
```

```
print(sum(diag(Psi.PCA.cor)))
```

```
## [1] 1.308556
```

```r
print(sum(diag(Psi.Chol)))
```

```
## [1] 1.712318
```

```r
# demonstrate that PCA.cor whitening produces
# the optimal compression in terms of squared cross-correlations
print(rowSums(Psi.ZCA^2))
```

```
## [1] 0.9009728 1.0990272
```

```r
print(rowSums(Psi.ZCA.cor^2))
```

```
## [1] 1 1
```

```r
print(rowSums(Psi.PCA^2))
```

```
## [1] 1.664314 0.335686
```

```r
print(rowSums(Psi.PCA.cor^2))
```

```
## [1] 1.7018567 0.2981433
```

```r
print(rowSums(Psi.Chol^2))
```

```
## [1] 0.5073972 1.4926028
```