Signal, Image, and Data Processing
(236200)

Tutorial 10

Principle Component Analysis
Random Vectors Basics: Exercise #1
Random Vector under Linear Transformation

The $N$-length column random vector $\mathbf{x}$ is characterized by its second-order statistics:

Mean: $\mu_\mathbf{x} \triangleq E\{\mathbf{x}\}$

Covariance matrix: $C_\mathbf{x} \triangleq E\{(\mathbf{x} - \mu_\mathbf{x})(\mathbf{x} - \mu_\mathbf{x})^T\}$

Consider the real-valued $M \times N$ matrix $\mathbf{A}$, and the transformed random variable:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

What are the second-order statistics of $\mathbf{y}$?
Random Vectors Basics: Exercise #1
Random Vector under Linear Transformation

First, note that $\mathbf{y} = \mathbf{A}\mathbf{x}$ is a $M$-length column vector.

Its mean is $\mathbf{\mu}_y = E\{\mathbf{y}\} = E\{\mathbf{A}\mathbf{x}\} = \mathbf{A}E\{\mathbf{x}\} = \mathbf{A}\mathbf{\mu}_x$ (which is a $M$-length column vector)

and its Covariance matrix is

$$
C_y = E \left\{ (\mathbf{y} - \mathbf{\mu}_y)(\mathbf{y} - \mathbf{\mu}_y)^T \right\} \\
= E\{(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{\mu}_x)(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{\mu}_x)^T\} \\
= E\{\mathbf{A}(\mathbf{x} - \mathbf{\mu}_x)(\mathbf{x} - \mathbf{\mu}_x)^T\mathbf{A}^T\} \\
= \mathbf{A}E\{\mathbf{(x} - \mathbf{\mu}_x)(\mathbf{x} - \mathbf{\mu}_x)^T\}\mathbf{A}^T \\
= \mathbf{A}\mathbf{C}_x\mathbf{A}^T
$$

(which is a $M \times M$ matrix)
Random Vectors Basics: Exercise #2
Random Vector under Unitary Transformation

The $N$-length column random vector $\mathbf{x}$ is characterized by its second-order statistics:

Mean: $\mathbf{\mu}_x \triangleq E\{\mathbf{x}\}$

Covariance matrix: $\mathbf{C}_x \triangleq E\{(\mathbf{x} - \mathbf{\mu}_x)(\mathbf{x} - \mathbf{\mu}_x)^T\}$

Consider the $N \times N$ unitary matrix $\mathbf{U}$, and the transformed random variable:

$$\mathbf{y} = \mathbf{U}^*\mathbf{x}$$

where $^*$ is the conjugate transpose operator.

What are the second-order statistics of $\mathbf{y}$?

Note that $\mathbf{U}$ can be a complex-valued matrix (for example, the DFT matrix).
Random Vectors Basics: Exercise #2
Random Vector under Unitary Transformation

From the results we got for a general linear transformation we get:

Mean: \( \mu_y = U^*\mu_x \)

Covariance matrix: \( C_y = U^*C_xU \)

Note that the transpose is extended here to be a conjugate transpose (since \( U \) can be complex valued).

1. Prove that \( Trace[C_y] = Trace[C_x] \) :

236200, CS Department, Technion
Random Vectors Basics: Exercise #2
Random Vector under Unitary Transformation

1. Prove that \( \text{Trace}[C_y] = \text{Trace}[C_x] \):

The covariance matrix of the transformed signal is \( C_y = U^*C_xU \)

Then,

\[
\text{Trace}[C_y] = \text{Trace}[U^*C_xU]
\]

\[
= \text{Trace}[UU^*C_x] \quad \text{Due to the cyclic property of the trace}
\]

\[
= \text{Trace}[C_x] \quad \text{Since \( U \) is unitary, i.e., \( UU^* = I \)}
\]
Random Vectors Basics: Exercise #2
Random Vector under Unitary Transformation

2. What is the meaning of the above equality?

The trace of a covariance matrix is the sum of the variances of all the vector components, which is the **expected energy of the signal**.

Accordingly, the result $Trace[C_y] = Trace[C_x]$ means that the expected energy of the signal is preserved under unitary transformation.
Principle Component Analysis (PCA)

Consider a class of signals that is described by a random vector $\overline{\phi}_\omega$ with zero mean $E_\omega\{\overline{\phi}_\omega\} = \overline{0}$ and an autocorrelation matrix $R_{\overline{\phi}} = E_\omega\{\overline{\phi}_\omega \overline{\phi}_\omega^*\}$.

The diagonalization of the autocorrelation matrix $R_{\overline{\phi}}$ is obtained using the unitary matrix $U$ via

$$R_{\overline{\phi}} = U \Lambda U^*$$

where $\Lambda$ is a diagonal matrix, composed of $R_{\overline{\phi}}$’s eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$$
Principle Component Analysis (PCA)

We consider the matrix $\mathbf{U} = [\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_N]$ by its columns, and the direct transform to be applied via

$$\overline{\varphi}(\mathbf{U}) = \mathbf{U}^* \overline{\varphi}$$

where the $k^{th}$ component of $\overline{\varphi}(\mathbf{U})$ is the scalar $\langle \overline{\varphi}, \bar{u}_k \rangle = \bar{u}_k^* \overline{\varphi}$.

The corresponding $K$-term approximation, for any $K \in \{1, \ldots, N\}$, is

$$\hat{\varphi}^{(K)} = \sum_{k=1}^{K} \langle \overline{\varphi}, \bar{u}_k \rangle \bar{u}_k$$
Principle Component Analysis (PCA)

The above matrix $U$ is the PCA transformation, providing the optimal $K$-term approximation in the sense of minimal expected squared-error

$$E\{\mathcal{E}_{\omega}^2(K)\} = E \left\{ \left( \overline{\varphi}_\omega - \widehat{\varphi}_\omega^{(K)} \right)^* \left( \overline{\varphi}_\omega - \widehat{\varphi}_\omega^{(K)} \right) \right\}$$

for any $K \in \{1, \ldots, N\}$.

Specifically, the minimal expected error is

$$E\{\mathcal{E}_{\omega}^2(K)\} = \sum_{k=K+1}^{N} \lambda_k$$
PCA: An Example for A Simple Class

A class of one-dimensional discrete signals is defined as follows.

A signal is the column vector of N samples of the form:

$$\overline{\phi} = [M, \ldots, M, M + L, M, \ldots, M]^T$$

i.e., all the vector components have the value $M$ except for the $K^{th}$ component that has the value $M + L$.

The vector components are indexed starting at 1, i.e., the vector can be generally formulated as $\overline{\phi} = [\phi_1, \ldots, \phi_N]^T$
PCA: An Example for A Simple Class

A class of one-dimensional discrete signals is defined as follows.

A signal is the column vector of N samples of the form:

\[ \bar{\phi} = [M, \ldots M, M + L, M, \ldots M]^T \]

\[ K^{th} \text{ component} \]

\( M, L, \text{ and } K \) are independent random variables:

\( K \) is a uniform random variable over the integers \( \{1, \ldots, N\} \).

\( M \) obeys \( E\{M\} = 0 \) and \( E\{M^2\} = c \)

\( L \) obeys \( E\{L\} = 0 \) and \( E\{L^2\} = N(1 - c) \)

where \( 0 < c < 1 \) is a (deterministic) constant.
PCA: An Example for A Simple Class

a. Show that the random vector $\bar{\varphi}$ has a zero mean.

**Solution:**

The expected value of the $i^{th}$ component is

$$E\{\varphi_i\} = E\{\varphi_i|K = i\}P(K = i) + E\{\varphi_i|K \neq i\}P(K \neq i)$$

$$= E\{M + L\}P(K = i) + E\{M\}P(K \neq i) = 0$$

Hence, the mean signal (vector) is $E\{\bar{\varphi}\} = \bar{0}$.
b. Calculate the autocorrelation matrix of $\vec{\varphi}$, denoted as $R_{\vec{\varphi}}$, and show it is circulant.

**Solution:**

The variance of the $i^{th}$ component is the $(i,i)$ entry in $R_{\vec{\varphi}}$:

$$r_{ii} = E\{\varphi_i^2\} = E\{\varphi_i^2 | K = i\}P(K = i) + E\{\varphi_i^2 | K \neq i\}P(K \neq i)$$

$$= (E\{L^2\} + E\{M^2\})P(K = i) + E\{M^2\}P(K \neq i)$$

Using the independence of $K$ with $M$ and $L$

$$= E\{L^2\}P(K = i) + E\{M^2\} =$$

$$= N(1 - c) \frac{1}{N} + c = 1$$
b. Calculate the autocorrelation matrix of $\bar{\varphi}$, denoted as $R_{\bar{\varphi}}$, and show it is circulant.

**Solution:**

The correlation between the $i^{th}$ and the $j^{th}$ components of $\bar{\varphi}$, which is the $(i, j)$ entry in $R_{\bar{\varphi}}$:

$r_{ij} = E\{\varphi_i \varphi_j\}$

$= E\{\varphi_i \varphi_j | K = i\}P(K = i) + E\{\varphi_i \varphi_j | K = j\}P(K = j) + E\{\varphi_i \varphi_j | K \neq i, j\}P(K \neq i, j)$

$= E\{(M + L)M | K = i\}P(K = i) + E\{M(M + L) | K = j\}P(K = j) + E\{M^2 | K \neq i, j\}P(K \neq i, j)$

$= E\{M^2\}P(K = i) + E\{M^2\}P(K = j) + E\{M^2\}P(K \neq i)$

$= E\{M^2\}$

$= c$

Where we used above the independence of $M$, $L$, and $K$ in the calculation of $E\{(M + L)M | K = i\} = E\{(M + L)M\} = E\{M^2 + LM\} = E\{M^2\} + E\{L\}E\{M\} = E\{M^2\}$
b. Calculate the autocorrelation matrix of $\bar{\phi}$, denoted as $R_{\bar{\phi}}$.

**Solution:**

To conclude the structure of $R_{\bar{\phi}}$ is

$$R_{\bar{\phi}} = \begin{bmatrix}
1 & c & c & \cdots & c \\
c & 1 & c & \ddots & \vdots \\
c & c & \ddots & \ddots & c \\
\vdots & \ddots & \ddots & 1 & c \\
c & \cdots & c & c & 1
\end{bmatrix}$$

exhibiting a **circulant** form.
PCA: An Example for A Simple Class

c. Is the random signal is cyclo-stationary?

**Solution:**

Since the autocorrelation matrix is circulant, all the \((i, j)\) entries that obey \(j - i \equiv a \pmod{N}\) have the same value (for \(a = 0, \ldots, N - 1\)).

This means that \(E\{\varphi_i \varphi_{i+a} \pmod{N}\} = E\{\varphi_l \varphi_{l+a} \pmod{N}\}\) for any \(i, l = 1, \ldots, N\). Then, by definition, the cyclo-stationarity property is satisfied.
PCA: An Example for A Simple Class

d. What is the PCA matrix corresponding to the autocorrelation matrix $R_{\phi}$?

**Solution:**

Since the autocorrelation matrix is circulant and of size $N \times N$, it is diagonalized by the $[DFT]^*$ matrix of size $N \times N$.

Hence, the PCA transform matrix is the $[DFT]^*$ matrix!
PCA: An Example for A Simple Class

e. Explain how the eigenvalues of $\mathbf{R}_{\varphi}$ can be computed in a way that is simpler than an explicit eigendecomposition procedure.

**Solution:**

Since $\mathbf{R}_{\varphi}$ is a circulant matrix, it is diagonalized by the $[DFT]^*$ matrix, and the eigenvalues can be obtained by applying the $[DFT]^*$ matrix on the column-ordering of the first row of $\mathbf{R}_{\varphi}$. 
The Discrete Cosine Transform (DCT)

The $N \times N$ DCT matrix, $[DCT]$, is defined by its component formulations:

$$[DCT]_{n,k} = \begin{cases} 
\frac{1}{\sqrt{N}} & \text{for } k = 1 \\
\frac{2}{\sqrt{N}} \cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right) & \text{for } k = 2, \ldots, N 
\end{cases}$$

where $[DCT]_{n,k}$ is the value of the matrix component at the $k^{th}$ column and $n^{th}$ row.

The DCT matrix is **unitary** and **real valued**.

Note that this is **not** the real part of the DFT matrix.
The Discrete Cosine Transform (DCT)

Is the DCT matrix the real part of the DFT matrix?

No.

The \((n, k)\) component of the DFT matrix is

\[
[DFT]_{n,k} = \frac{1}{\sqrt{N}} (W^*)^{(n-1)(k-1)} = \frac{1}{\sqrt{N}} e^{-\frac{i2\pi(n-1)(k-1)}{N}}
\]

and its real part is \(Re\{[DFT]_{n,k}\} = \frac{1}{\sqrt{N}} \cos \left(\frac{2\pi(n-1)(k-1)}{N}\right)\).

that clearly differs from the \((n, k)\) component of the DCT matrix defined in the question (for \(k > 1\)) as

\[
[DCT]_{n,k} = \sqrt{\frac{2}{N}} \cos \left(\frac{\pi(2n-1)(k-1)}{2N}\right)
\]
Another PCA Example

Consider a class of signals described by a random vector $\bar{\phi}_\omega \in \mathbb{R}^3$ with zero mean and an autocorrelation matrix

$$R_{\bar{\phi}} = \begin{bmatrix}
1 - \alpha & -\alpha & 0 \\
-\alpha & 1 & -\alpha \\
0 & -\alpha & 1 - \alpha
\end{bmatrix}$$

where $0 < \alpha < 1$ is some scalar value.

What is the PCA matrix for this class?
Another PCA Example

**Solution:**
The PCA matrix $\mathbf{U}$ diagonalizes the autocorrelation matrix $\mathbf{R}_\varphi$, i.e.,

$$\mathbf{U}^* \mathbf{R}_\varphi \mathbf{U} = \Lambda$$

where $\Lambda$ is a $3 \times 3$ diagonal matrix.

In this case where the matrix $\mathbf{R}_\varphi$ is not circulant we need to explicitly calculate its eigenvectors (that, of course, form the diagonalizing matrix $\mathbf{U}$).
Another PCA Example

\[
\det \left( R_\varphi - \lambda I \right) = \det \left( \begin{bmatrix}
1 - \alpha - \lambda & -\alpha & 0 \\
-\alpha & 1 - \lambda & -\alpha \\
0 & -\alpha & 1 - \alpha - \lambda
\end{bmatrix} \right)
\]

\[
= (1 - \alpha - \lambda) \left[ (1 - \lambda)(1 - \alpha - \lambda) - \alpha^2 \right] - \alpha^2 (1 - \alpha - \lambda)
\]

\[
= (1 - \alpha - \lambda) \left[ (1 - \lambda)(1 - \alpha - \lambda) - 2\alpha^2 \right]
\]

\[
= (1 - \alpha - \lambda)(1 - \lambda - 2\alpha)(1 - \lambda + \alpha)
\]

\[
(1 - \alpha - \lambda)(1 - \lambda - 2\alpha)(1 - \lambda + \alpha) = 0
\]

\[
\Rightarrow \lambda_1 = 1 + \alpha, \quad \lambda_2 = 1 - \alpha, \quad \lambda_3 = 1 - 2\alpha \quad \Rightarrow \quad \Lambda = \begin{bmatrix}
1 + \alpha & 0 \\
0 & 1 - \alpha \\
0 & 1 - 2\alpha
\end{bmatrix}
\]
Another PCA Example

Finding eigenvector #1:

$$\begin{bmatrix} 1-\alpha & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & -\alpha & 1-\alpha \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = (1+\alpha) \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$(1-\alpha)u_1 - \alpha u_2 = (1+\alpha)u_1 \rightarrow u_1 = -\frac{1}{2}u_2$$

$$-\alpha u_1 + u_2 - \alpha u_3 = (1+\alpha)u_2$$

$$-\alpha u_2 + (1-\alpha)u_3 = (1+\alpha)u_3 \rightarrow u_3 = -\frac{1}{2}u_2 = u_1$$

$$\Rightarrow \begin{bmatrix} \beta \\ -2\beta \\ \beta \end{bmatrix} \Rightarrow u = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

Using the normality requirement:

$$u_1^2 + u_2^2 + u_3^2 = 1$$
Another PCA Example

Finding eigenvector #2:

\[
\begin{pmatrix}
1-\alpha & -\alpha & 0 \\
-\alpha & 1 & -\alpha \\
0 & -\alpha & 1-\alpha \\
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{pmatrix}
= (1-\alpha)
\begin{pmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{pmatrix}
\]

\[(1-\alpha)u_1 - \alpha u_2 = (1-\alpha)u_1 \rightarrow u_2 = 0\]

\[-\alpha u_1 + u_2 - \alpha u_3 = (1-\alpha)u_2 \rightarrow u_1 = -u_3 \Rightarrow u = \begin{pmatrix} \beta \\ 0 \\ -\beta \end{pmatrix} \Rightarrow u = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}\]

Using the normality requirement:

\[u_1^2 + u_2^2 + u_3^2 = 1\]
Another PCA Example

Finding eigenvector #3:

\[
\begin{bmatrix}
1 - \alpha & -\alpha & 0 \\
-\alpha & 1 & -\alpha \\
0 & -\alpha & 1 - \alpha
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
= (1 - 2\alpha)
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
\]

\[(1 - \alpha)u_1 - \alpha u_2 = (1 - 2\alpha)u_1 \rightarrow u_1 = u_2
\]

\[-\alpha u_1 + u_2 - \alpha u_3 = (1 - 2\alpha)u_2
\]

\[-\alpha u_2 + (1 - \alpha)u_3 = (1 - 2\alpha)u_3 \rightarrow u_3 = u_2 \Rightarrow u = \begin{bmatrix}
\beta \\
\beta \\
\beta
\end{bmatrix} \Rightarrow u = \begin{bmatrix}
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}}
\end{bmatrix}
\]

Using the normality requirement:
\[u_1^2 + u_2^2 + u_3^2 = 1\]
Another PCA Example

We got that the PCA matrix corresponding to

\[ \mathbf{R}_\phi = \begin{bmatrix} 1 - \alpha & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & -\alpha & 1 - \alpha \end{bmatrix} \]

is

\[ \mathbf{U} = \begin{bmatrix} 1 & 1 & 1 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & 0 & \sqrt{3} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \]
Another PCA Example

Is the PCA matrix $\mathbf{U}$ related to the $3 \times 3$ DCT matrix?

By the definition given earlier, the $3 \times 3$ DCT matrix is

$$ [DCT] = \begin{bmatrix} 1 & 1 & 1 \\ \sqrt{3} & \sqrt{2} & \sqrt{6} \\ 1 & 0 & 2 \\ \sqrt{3} & 0 & -\sqrt{6} \\ 1 & -1 & 1 \\ \sqrt{3} & -\sqrt{2} & \sqrt{6} \end{bmatrix} $$

By swapping the first and the third columns of $[DCT]$ we get the PCA matrix $\mathbf{U}$ that we found above!
Principle Component Analysis (PCA) for A Dataset

We considered the PCA for a class with a probabilistic definition.

The PCA approach can be employed also for a given set of data vectors \( \{ \phi_1, \phi_2, \ldots, \phi_M \} \) that, for a large \( M \), are assumed to empirically represent the class.

Then, one can **empirically estimate** the class mean via

\[
\hat{\mu}_\phi = \frac{1}{M} \sum_{m=1}^{M} \phi_m
\]

and the corresponding autocorrelation matrix

\[
\hat{R}_\phi = \frac{1}{M} \sum_{m=1}^{M} (\phi_m - \hat{\mu}_\phi)(\phi_m - \hat{\mu}_\phi)^*
\]

and the PCA can be computed based on \( \hat{R}_\phi \).