# Principal Component Transform

## Multivariate Random Signals

A real time signal $x(t)$ can be considered as a random process and its samples $x_m \;\; (m = 0, \cdots, N-1)$ a random vector:

$$X = [x_0, \cdots, x_{N-1}]^T$$

The *mean vector* of $X$ is

$$M_X \stackrel{\triangle}{=} E(X) = [E(x_0), \cdots, E(x_{N-1})]^T = [\mu_0, \cdots, \mu_{N-1}]^T$$

The *covariance matrix* of $X$ is

$$\Sigma_X \stackrel{\triangle}{=} E[(X - M_X)(X - M_X)^T] = E(XX^T) - MM^T = \begin{bmatrix} .. & .. & .. \\ .. & \sigma_{ij}^2 & .. \\ .. & .. & .. \end{bmatrix}$$

where $\sigma_{ij}^2 \stackrel{\triangle}{=} E(x_i x_j) - \mu_i \mu_j$ is the covariance of two random variables $x_i$ and $x_j$. When $i = j$, $\sigma_{ij}$ becomes the variance of $x_i$, $\sigma_i^2 \stackrel{\triangle}{=} E(x_i^2) - \mu_i^2$.

The *correlation matrix* of $X$ is

$$R_X \stackrel{\triangle}{=} E(XX^T) = \begin{bmatrix} .. & .. & .. \\ .. & r_{ij} & .. \\ .. & .. & .. \end{bmatrix}$$

where $r_{ij} = \sigma_{ij}^2 + \mu_i \mu_j$. Both $\Sigma_X$ and $R_X$ are symmetric matrices (Hermitian if $X$ is complex).

A signal vector $X$ can always be easily converted into a zero-mean vector $X' = X - M_X$ with all of its information (or dynamic energy) conserved. In the following, without loss of generality, we will assume $M_X = 0$ and therefore $\Sigma_X = R_X$.

# The Principal Component Transform

The *Principal Component Transform* is also called *Karhunen-Loeve Transform (KLT)*, *Hotelling Transform*, or *Eigenvector Transform.*

Let $\phi_i$ and $\lambda_i$ be the ith eigenvector and eigenvalue of the correlation matrix $R_X$:

$$R_X \phi_i = \lambda_i \phi_i \qquad (i = 0, \cdots, N-1)$$

We can construct an $N \times N$ matrix $\Phi$

$$\Phi \stackrel{\triangle}{=} [\phi_0, \cdots, \phi_{N-1}]$$

Since the columns of $\Phi$ are the eigenvectors of a symmetric (Hermitian if $X$ is complex) matrix $R_X$, $\Phi$ is orthogonal (unitary):

$$\Phi^T \Phi = I$$

i.e.,

$$\Phi^{-1} = \Phi^T$$

and we have

$$R_X \Phi = \Phi \Lambda$$

where $\Lambda = diag(\lambda_0, \cdots, \lambda_{N-1})$. Or, we have

$$\Phi^{-1} R_X \Phi = \Phi^T R_X \Phi = \Lambda$$

We can now define the orthogonal (unitary if $X$ is complex) *Principal Component Transform* of $X$ by

$$\begin{cases} Y = \Phi^T X \\ X = \Phi Y \end{cases}$$

The ith component of the forward transform $Y = \Phi^T X$ is the projection of $X$ on $\phi_i$:

$$y_i = (\phi_i, X) = \phi_i^T X$$

and the inverse transform $X = \Phi Y$ represents $X$ in the N-dimensional space spanned by $\phi_i$ $(i = 0, 1, \cdots, N-1)$:

$$X = \sum_{i=0}^{N-1} y_i \phi_i$$

# KLT Completely Decorrelates the Signal

KLT is the optimal orthogonal transform in the following sense:

- KLT completely decorrelates the signal
- KLT optimally compacts the energy (information) contained in the signal.

The first property is simply due to the definition of KLT, and the second property is due to the fact that KLT redistributes the energy among the $N$ components in such a way that most of the energy is contained in a small number of components of $Y = \Phi^T X$.

To see the first property, consider the correlation matrix $R_Y$ of $Y$:

$$\begin{aligned} R_Y &= E(YY^T) = E[\Phi^T X(\Phi^T X)^T] \\ &= E[\Phi^T (XX^T)\Phi] = \Phi^T E(XX^T)\Phi \\ &= \Phi^T R_X \Phi = \Lambda \end{aligned}$$

We see that after KLT, the correlation matrix of the signal is diagonalized, i.e., the correlation $r_{ij} = 0$ between any two components $x_i$ and $x_j$ is always zero. In other words, the signal is completely decorrelated.

# KLT Optimally Compacts the Energy

Consider a general orthogonal transform pair defined as

$$\begin{cases} Y = A^T X \\ X = AY \end{cases}$$

where $X$ and $Y$ are N by 1 vectors and $A$ is an arbitrary N by N orthogonal matrix $A^{-1} = A^T$.

We represent $A$ by its column vectors $A_i, \quad (i = 0, \cdots, N-1)$ as

$$A = [A_0, \cdots, A_{N-1}]$$

or

$$A^T = \begin{bmatrix} A_0^T \\ \cdot \\ \cdot \\ A_{N-1}^T \end{bmatrix}$$

Now the ith component of $Y$ can be written as

$$y_i = A_i^T \, X$$

As we assume the mean vector of $X$ is zero $M_X = 0$ (and obviously we also have $M_Y = A^T M_x = 0$), we have $\Sigma_X = R_X$, and the variance of the ith element in both $X$ and $Y$ are

$$\sigma_{x_i}^2 = E(x_i^2) \stackrel{\triangle}{=} E(e_{x_i})$$

and

$$\sigma_{y_i}^2 = E(y_i^2) \stackrel{\triangle}{=} E(e_{y_i})$$

where $e_{x_i} \stackrel{\triangle}{=} x_i^2$ and $e_{y_i} \stackrel{\triangle}{=} y_i^2$ represent the energy contained in the ith component of $X$ and $Y$, respectively. In order words, the trace of $\Sigma_X$ (the sum of all the diagonal elements of the matrix) represents the expectation of the total amount of energy contained in the signal $X$

$$\text{Total energy contained in } X = tr\Sigma_X = \sum_{i=0}^{N-1} \sigma_{x_i}^2 = \sum_{i=0}^{N-1} E(x_i^2) = E(\sum_{i=0}^{N-1} e_{x_i})$$

Since an orthogonal transform $A$ does not change the length of a vector X, i.e., $\| Y \| = \| AX \| = \| X \|$, where

$$\| X \| \triangleq \sqrt{\sum_{i=0}^{N-1} x_i^2} = \sqrt{\sum_{i=0}^{N-1} e_{x_i}}$$

the total energy contained in the signal vector $X$ is conserved after the orthogonal transform. (This conclusion can also be obtained from the fact that orthogonal transforms do not change the trace of a matrix.)

We next define

$$S_m(A) \triangleq \sum_{i=0}^{m-1} E(y_i^2) = \sum_{i=0}^{m-1} \sigma_{y_i}^2 = \sum_{i=0}^{m-1} E(e_{y_i})$$

where $m \leq N$. $S_m(A)$ is a function of the transform matrix $A$ and represents the amount of energy contained in the first $m$ components of $Y = A^T X$. Since the total energy is conserved, $S_m(A)$ also represents the percentage of energy contained in the first $m$ components. In the following we will show that $S_m(A)$ is maximized if and only if the transform $A$ is the KLT:

$$S_m(\Phi) \geq S_m(A)$$

i.e., KLT optimally compacts energy into a few components of the signal.

Consider

$$\begin{aligned} S_m(A) &\triangleq \sum_{i=0}^{m-1} E(y_i^2) = \sum_{i=0}^{m-1} E[A_i^T X (A_i^T X)^T] \\ &= \sum_{i=0}^{m-1} E[A_i^T X (X^T A_i)] = \sum_{i=0}^{m-1} A_i^T E(X X^T) A_i \\ &= \sum_{i=0}^{m-1} A_i^T R_X A_i \end{aligned} \tag{1}$$

Now we need to find a transform matrix $A$ so that

$$\begin{cases} S_m(A) \rightarrow max \\ \text{subject to } A_j^T A_j = 1 \quad (j = 0, \cdots, m-1) \end{cases}$$

The constraint $A_j^T A_j = 1$ is to guarantee that the column vectors in $A$ are normalized. This constrained optimization problem can be solved by Lagrange multiplier method as shown below.

We let

$$
\begin{aligned}
& \frac{\partial}{\partial A_i}[S_m(A) - \sum_{j=0}^{m-1} \lambda_j (A_j^T A_j - 1)] = 0 \\
= \quad & \frac{\partial}{\partial A_i}[\sum_{j=0}^{m-1} (A_j^T R_X A_j - \lambda_j A_j^T A_j + \lambda_j)] \\
= \quad & \frac{\partial}{\partial A_i}[A_i^T R_X A_i - \lambda_i A_i^T A_i] \\
\overset{*}{=} \quad & 2 R_x A_i - 2\lambda_i A_i = 0
\end{aligned}
$$

(* the last equal sign is due to explanation in the handout of review of linear algebra.) We see that the column vectors of $A$ must be the eigenvectors of $R_X$:

$$R_X A_i = \lambda_i A_i \qquad (i = 0, \cdots, m-1)$$

i.e., the transform matrix must be

$$A = [A_0, \cdots, A_{N-1}] = \Phi = [\phi_0, \cdots, \phi_{N-1}]$$

Thus we have proved that the optimal transform is indeed KLT, and

$$S_m(\Phi) = \sum_{i=0}^{m-1} \phi_i^T R_X \phi_i = \sum_{i=0}^{m-1} \lambda_i$$

where the ith eigenvalue $\lambda_i$ of $R_X$ is also the average (expectation) energy contained in the ith component of the signal. If we choose those $\phi_i's$ that correspond to the $m$ largest eigenvalues of $R_X$: $\lambda_0 \geq \lambda_1 \geq \cdots \lambda_m \cdots \geq \lambda_{N-1}$, then $S_m(\Phi)$ will achieve maximum.