Karhunen–Loève theorem

In the theory of stochastic processes, the **Karhunen–Loève theorem** (named after Kari Karhunen and Michel Loève), also known as the **Kosambi–Karhunen–Loève theorem**[1][2] is a representation of a stochastic process as an infinite linear combination of orthogonal functions, analogous to a Fourier series representation of a function on a bounded interval. The transformation is also known as Hotelling transform and eigenvector transform, and is closely related to principal component analysis (PCA) technique widely used in image processing and in data analysis in many fields.[3]

Stochastic processes given by infinite series of this form were first considered by Damodar Dharmananda Kosambi.[4][5] There exist many such expansions of a stochastic process: if the process is indexed over $[a, b]$, any orthonormal basis of $L^2([a, b])$ yields an expansion thereof in that form. The importance of the Karhunen–Loève theorem is that it yields the best such basis in the sense that it minimizes the total mean squared error.

In contrast to a Fourier series where the coefficients are fixed numbers and the expansion basis consists of sinusoidal functions (that is, sine and cosine functions), the coefficients in the Karhunen–Loève theorem are random variables and the expansion basis depends on the process. In fact, the orthogonal basis functions used in this representation are determined by the covariance function of the process. One can think that the Karhunen–Loève transform adapts to the process in order to produce the best possible basis for its expansion.

In the case of a *centered* stochastic process $\{X_t\}_{t \in [a, b]}$ (*centered* means $\mathbf{E}[X_t] = 0$ for all $t \in [a, b]$) satisfying a technical continuity condition, X_t admits a decomposition

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

where Z_k are pairwise uncorrelated random variables and the functions e_k are continuous real-valued functions on $[a, b]$ that are pairwise orthogonal in $L^2([a, b])$. It is therefore sometimes said that the expansion is *bi-orthogonal* since the random coefficients Z_k are orthogonal in the probability space while the deterministic functions e_k are orthogonal in the time domain. The general case of a process X_t that is not centered can be brought back to the case of a centered process by considering $X_t - \mathbf{E}[X_t]$ which is a centered process.

Moreover, if the process is Gaussian, then the random variables Z_k are Gaussian and stochastically independent. This result generalizes the *Karhunen–Loève transform*. An important example of a centered real stochastic process on $[0, 1]$ is the Wiener process; the Karhunen–Loève theorem can be used to provide a canonical orthogonal representation for it. In this case the expansion consists of sinusoidal functions.

The above expansion into uncorrelated random variables is also known as the *Karhunen–Loève expansion* or *Karhunen–Loève decomposition*. The empirical version (i.e., with the coefficients computed from a sample) is known as the *Karhunen–Loève transform* (KLT), *principal component analysis, proper orthogonal decomposition (POD), empirical orthogonal functions* (a term used in meteorology and geophysics), or the *Hotelling transform*.

Contents

Formulation

- Throughout this article, we will consider a square-integrable zero-mean random process X_t defined over a probability space $(\Omega, F, \mathbf{P})$ and indexed over a closed interval $[a, b]$, with covariance function $K_X(s, t)$. We thus have:

$$\forall t \in [a,b] \qquad X_t \in L^2(\Omega, F, \mathbf{P}),$$
$$\forall t \in [a,b] \qquad \mathbf{E}[X_t] = 0,$$
$$\forall t, s \in [a,b] \qquad K_X(s,t) = \mathbf{E}[X_s X_t].$$

- We associate to K_X a linear operator T_{K_X} defined in the following way:

$$T_{K_X} \qquad : L^2([a,b]) \to L^2([a,b])$$
$$: f \mapsto T_{K_X} f = \int_a^b K_X(s, \cdot) f(s)\, ds$$

Since T_{K_X} is a linear operator, it makes sense to talk about its eigenvalues λ_k and eigenfunctions e_k, which are found solving the homogeneous Fredholm integral equation of the second kind

$$\int_a^b K_X(s,t) e_k(s)\, ds = \lambda_k e_k(t)$$

Statement of the theorem

Theorem. Let X_t be a zero-mean square-integrable stochastic process defined over a probability space $(\Omega, F, \mathbf{P})$ and indexed over a closed and bounded interval $[a, b]$, with continuous covariance function $K_X(s, t)$.

Then $K_X(s,t)$ is a Mercer kernel and letting e_k be an orthonormal basis on $L^2([a, b])$ formed by the eigenfunctions of T_{K_X} with respective eigenvalues λ_k, X_t admits the following representation

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

where the convergence is in L^2, uniform in t and

$$Z_k = \int_a^b X_t e_k(t)\, dt$$

Furthermore, the random variables Z_k have zero-mean, are uncorrelated and have variance λ_k

$$\mathbf{E}[Z_k] = 0, \ \forall k \in \mathbb{N} \qquad \text{and} \qquad \mathbf{E}[Z_i Z_j] = \delta_{ij} \lambda_j, \ \forall i, j \in \mathbb{N}$$

Note that by generalizations of Mercer's theorem we can replace the interval $[a, b]$ with other compact spaces C and the Lebesgue measure on $[a, b]$ with a Borel measure whose support is C.

Proof

- The covariance function K_X satisfies the definition of a Mercer kernel. By Mercer's theorem, there consequently exists a set $\{\lambda_k, e_k(t)\}$ of eigenvalues and eigenfunctions of T_{K_X} forming an orthonormal basis of $L^2([a,b])$, and K_X can be expressed as

$$K_X(s, t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t)$$

- The process X_t can be expanded in terms of the eigenfunctions e_k as:

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

where the coefficients (random variables) Z_k are given by the projection of X_t on the respective eigenfunctions

$$Z_k = \int_a^b X_t e_k(t)\, dt$$

- We may then derive

$$\mathbf{E}[Z_k] = \mathbf{E}\left[\int_a^b X_t e_k(t)\, dt\right] = \int_a^b \mathbf{E}[X_t] e_k(t) dt = 0$$

$$\begin{aligned}
\mathbf{E}[Z_i Z_j] &= \mathbf{E}\left[\int_a^b \int_a^b X_t X_s e_j(t) e_i(s)\, dt\, ds\right] \\
&= \int_a^b \int_a^b \mathbf{E}[X_t X_s]\, e_j(t) e_i(s)\, dt\, ds \\
&= \int_a^b \int_a^b K_X(s, t) e_j(t) e_i(s)\, dt\, ds \\
&= \int_a^b e_i(s) \left(\int_a^b K_X(s, t) e_j(t)\, dt\right)\, ds \\
&= \lambda_j \int_a^b e_i(s) e_j(s)\, ds \\
&= \delta_{ij} \lambda_j
\end{aligned}$$

where we have used the fact that the e_k are eigenfunctions of T_{K_X} and are orthonormal.

- Let us now show that the convergence is in L^2. Let

$$S_N = \sum_{k=1}^{N} Z_k e_k(t).$$

Then:

$$\mathbf{E}\left[|X_t - S_N|^2\right] = \mathbf{E}\left[X_t^2\right] + \mathbf{E}\left[S_N^2\right] - 2\mathbf{E}\left[X_t S_N\right]$$

$$= K_X(t,t) + \mathbf{E}\left[\sum_{k=1}^{N}\sum_{l=1}^{N} Z_k Z_\ell e_k(t) e_\ell(t)\right] - 2\mathbf{E}\left[X_t \sum_{k=1}^{N} Z_k e_k(t)\right]$$

$$= K_X(t,t) + \sum_{k=1}^{N} \lambda_k e_k(t)^2 - 2\mathbf{E}\left[\sum_{k=1}^{N}\int_a^b X_t X_s e_k(s) e_k(t)\, ds\right]$$

$$= K_X(t,t) - \sum_{k=1}^{N} \lambda_k e_k(t)^2$$

which goes to 0 by Mercer's theorem.

Properties of the Karhunen–Loève transform

Special case: Gaussian distribution

Since the limit in the mean of jointly Gaussian random variables is jointly Gaussian, and jointly Gaussian random (centered) variables are independent if and only if they are orthogonal, we can also conclude:

Theorem. The variables Z_i have a joint Gaussian distribution and are stochastically independent if the original process $\{X_t\}_t$ is Gaussian.

In the Gaussian case, since the variables Z_i are independent, we can say more:

$$\lim_{N\to\infty} \sum_{i=1}^{N} e_i(t) Z_i(\omega) = X_t(\omega)$$

almost surely.

The Karhunen–Loève transform decorrelates the process

This is a consequence of the independence of the Z_k.

The Karhunen–Loève expansion minimizes the total mean square error

In the introduction, we mentioned that the truncated Karhunen–Loeve expansion was the best approximation of the original process in the sense that it reduces the total mean-square error resulting of its truncation. Because of this property, it is often said that the KL transform optimally compacts the energy.

More specifically, given any orthonormal basis $\{f_k\}$ of $L^2([a, b])$, we may decompose the process X_t as:

$$X_t(\omega) = \sum_{k=1}^{\infty} A_k(\omega) f_k(t)$$

where

$$A_k(\omega) = \int_a^b X_t(\omega) f_k(t)\, dt$$

and we may approximate X_t by the finite sum

$$\hat{X}_t(\omega) = \sum_{k=1}^{N} A_k(\omega) f_k(t)$$

for some integer N.

Claim. Of all such approximations, the KL approximation is the one that minimizes the total mean square error (provided we have arranged the eigenvalues in decreasing order).

[Proof]

Consider the error resulting from the truncation at the N-th term in the following orthonormal expansion:

$$\varepsilon_N(t) = \sum_{k=N+1}^{\infty} A_k(\omega) f_k(t)$$

The mean-square error $\varepsilon_N^2(t)$ can be written as:

$$\varepsilon_N^2(t) = \mathbf{E}\left[\sum_{i=N+1}^{\infty} \sum_{j=N+1}^{\infty} A_i(\omega) A_j(\omega) f_i(t) f_j(t) \right]$$

$$= \sum_{i=N+1}^{\infty} \sum_{j=N+1}^{\infty} \mathbf{E}\left[\int_a^b \int_a^b X_t X_s f_i(t) f_j(s)\, ds\, dt \right] f_i(t) f_j(t)$$

$$= \sum_{i=N+1}^{\infty} \sum_{j=N+1}^{\infty} f_i(t) f_j(t) \int_a^b \int_a^b K_X(s,t) f_i(t) f_j(s)\, ds\, dt$$

We then integrate this last equality over $[a, b]$. The orthonormality of the f_k yields:

$$\int_a^b \varepsilon_N^2(t)\, dt = \sum_{k=N+1}^{\infty} \int_a^b \int_a^b K_X(s,t) f_k(t) f_k(s)\, ds\, dt$$

The problem of minimizing the total mean-square error thus comes down to minimizing the right hand side of this equality subject to the constraint that the f_k be normalized. We hence introduce β_k, the Lagrangian multipliers associated with these constraints, and aim at minimizing the following function:

$$Er[f_k(t), k \in \{N+1, \ldots\}] = \sum_{k=N+1}^{\infty} \int_a^b \int_a^b K_X(s,t) f_k(t) f_k(s)\, ds\, dt - \beta_k \left(\int_a^b f_k(t) f_k(t)\, dt - 1 \right)$$

Differentiating with respect to $f_i(t)$ (this is a <u>functional derivative</u>) and setting the derivative to 0 yields:

$$\frac{\partial Er}{\partial f_i(t)} = \int_a^b \left(\int_a^b K_X(s,t) f_i(s)\, ds - \beta_i f_i(t) \right) dt = 0$$

which is satisfied in particular when

$$\int_a^b K_X(s,t) f_i(s)\, ds = \beta_i f_i(t).$$

In other words, when the f_k are chosen to be the eigenfunctions of T_{K_X}, hence resulting in the KL expansion.

Explained variance

An important observation is that since the random coefficients Z_k of the KL expansion are uncorrelated, the <u>Bienaymé formula</u> asserts that the variance of X_t is simply the sum of the variances of the individual components of the sum:

$$\mathrm{var}[X_t] = \sum_{k=0}^{\infty} e_k(t)^2\, \mathrm{var}[Z_k] = \sum_{k=1}^{\infty} \lambda_k e_k(t)^2$$

Integrating over $[a, b]$ and using the orthonormality of the e_k, we obtain that the total variance of the process is:

$$\int_a^b \mathrm{var}[X_t]\, dt = \sum_{k=1}^{\infty} \lambda_k$$

In particular, the total variance of the N-truncated approximation is

$$\sum_{k=1}^{N} \lambda_k.$$

As a result, the *N*-truncated expansion explains

$$\frac{\sum_{k=1}^{N} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}$$

of the variance; and if we are content with an approximation that explains, say, 95% of the variance, then we just have to determine an $N \in \mathbb{N}$ such that

$$\frac{\sum_{k=1}^{N} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k} \geq 0.95.$$

The Karhunen–Loève expansion has the minimum representation entropy property

Given a representation of $X_t = \sum_{k=1}^{\infty} W_k \varphi_k(t)$, for some orthonormal basis $\varphi_k(t)$ and random W_k, we let $p_k = \mathbb{E}[|W_k|^2]/\mathbb{E}[|X_t|_{L^2}^2]$, so that $\sum_{k=1}^{\infty} p_k = 1$. We may then define the representation <u>entropy</u> to be $H(\{\varphi_k\}) = -\sum_i p_k \log(p_k)$. Then we have $H(\{\varphi_k\}) \geq H(\{e_k\})$, for all choices of φ_k. That is, the KL-expansion has minimal representation entropy.

Proof:

Denote the coefficients obtained for the basis $e_k(t)$ as p_k, and for $\varphi_k(t)$ as q_k.

Choose $N \geq 1$. Note that since e_k minimizes the mean squared error, we have that

$$\mathbb{E}\left|\sum_{k=1}^{N} Z_k e_k(t) - X_t\right|_{L^2}^2 \leq \mathbb{E}\left|\sum_{k=1}^{N} W_k \varphi_k(t) - X_t\right|_{L^2}^2$$

Expanding the right hand size, we get:

$$\mathbb{E}\left|\sum_{k=1}^{N} W_k \varphi_k(t) - X_t\right|_{L^2}^2 = \mathbb{E}|X_t^2|_{L^2} + \sum_{k=1}^{N}\sum_{\ell=1}^{N} \mathbb{E}[W_\ell \varphi_\ell(t) W_k^* \varphi_k^*(t)]_{L^2} - \sum_{k=1}^{N} \mathbb{E}[W_k \varphi_k X_t^*]_{L^2} - \sum_{k=1}^{N} \mathbb{E}[X_t W_k^* \varphi_k^*(t)]_{L^2}$$

Using the orthonormality of $\varphi_k(t)$, and expanding X_t in the $\varphi_k(t)$ basis, we get that the right hand size is equal to:

$$\mathbb{E}[X_t]_{L^2}^2 - \sum_{k=1}^{N} \mathbb{E}[|W_k|^2]$$

We may perform indentitcal analysis for the $e_k(t)$, and so rewrite the above inequality as:

$$\mathbb{E}[X_t]_{L^2}^2 - \sum_{k=1}^{N} \mathbb{E}[|Z_k|^2] \leq \mathbb{E}[X_t]_{L^2}^2 - \sum_{k=1}^{N} \mathbb{E}[|W_k|^2]$$

Subtracting the common first term, and dividing by $\mathbb{E}[|X_t|_{L^2}^2]$, we obtain that:

$$\sum_{k=1}^{N} p_k \geq \sum_{k=1}^{N} q_k$$

This implies that:

$$-\sum_{k=1}^{\infty} p_k \log(p_k) \leq -\sum_{k=1}^{\infty} q_k \log(q_k)$$

Linear Karhunen–Loève approximations

Consider a whole class of signals we want to approximate over the first M vectors of a basis. These signals are modeled as realizations of a random vector $Y[n]$ of size N. To optimize the approximation we design a basis that minimizes the average approximation error. This section proves that optimal bases are Karhunen–Loeve bases that diagonalize the covariance matrix of Y. The random vector Y can be decomposed in an orthogonal basis

$$\{g_m\}_{0 \leq m \leq N}$$

as follows:

$$Y = \sum_{m=0}^{N-1} \langle Y, g_m \rangle \, g_m,$$

where each

$$\langle Y, g_m \rangle = \sum_{n=0}^{N-1} Y[n] g_m^*[n]$$

is a random variable. The approximation from the first $M \leq N$ vectors of the basis is

$$Y_M = \sum_{m=0}^{M-1} \langle Y, g_m \rangle \, g_m$$

The energy conservation in an orthogonal basis implies

$$\varepsilon[M] = \mathbf{E}\left\{\|Y - Y_M\|^2\right\} = \sum_{m=M}^{N-1} \mathbf{E}\left\{|\langle Y, g_m \rangle|^2\right\}$$

This error is related to the covariance of Y defined by

$$R[n, m] = \mathbf{E}\left\{Y[n] Y^*[m]\right\}$$

For any vector $x[n]$ we denote by K the covariance operator represented by this matrix,

$$\mathbf{E}\left\{|\langle Y, x \rangle|^2\right\} = \langle Kx, x \rangle = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} R[n, m] x[n] x^*[m]$$

The error $\varepsilon[M]$ is therefore a sum of the last $N - M$ coefficients of the covariance operator

$$\varepsilon[M] = \sum_{m=M}^{N-1} \langle Kg_m, g_m \rangle$$

The covariance operator K is Hermitian and Positive and is thus diagonalized in an orthogonal basis called a Karhunen–Loève basis. The following theorem states that a Karhunen–Loève basis is optimal for linear approximations.

Theorem (Optimality of Karhunen–Loève basis). Let K be a covariance operator. For all $M \geq 1$, the approximation error

$$\varepsilon[M] = \sum_{m=M}^{N-1} \langle Kg_m, g_m \rangle$$

is minimum if and only if

$$\{g_m\}_{0 \leq m < N}$$

is a Karhunen–Loeve basis ordered by decreasing eigenvalues.

$$\langle Kg_m, g_m \rangle \geq \langle Kg_{m+1}, g_{m+1} \rangle, \qquad 0 \leq m < N - 1.$$

Non-Linear approximation in bases

Linear approximations project the signal on M vectors a priori. The approximation can be made more precise by choosing the M orthogonal vectors depending on the signal properties. This section analyzes the general performance of these non-linear approximations. A signal $f \in \mathbf{H}$ is approximated with M vectors selected adaptively in an orthonormal basis for $\mathbf{H}$

$$\mathbf{B} = \{g_m\}_{m \in \mathbb{N}}$$

Let f_M be the projection of f over M vectors whose indices are in I_M:

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle \, g_m$$

The approximation error is the sum of the remaining coefficients

$$\varepsilon[M] = \left\{ \|f - f_M\|^2 \right\} = \sum_{m \notin I_M}^{N-1} \left\{ |\langle f, g_m \rangle|^2 \right\}$$

To minimize this error, the indices in I_M must correspond to the M vectors having the largest inner product amplitude

$$|\langle f, g_m \rangle| \, .$$

These are the vectors that best correlate f. They can thus be interpreted as the main features of f. The resulting error is necessarily smaller than the error of a linear approximation which selects the M approximation vectors independently of f. Let us sort

$$\{ |\langle f, g_m \rangle| \}_{m \in \mathbb{N}}$$

in decreasing order

$$|\langle f, g_{m_k} \rangle| \geq |\langle f, g_{m_{k+1}} \rangle| \, .$$

The best non-linear approximation is

$$f_M = \sum_{k=1}^{M} \langle f, g_{m_k} \rangle \, g_{m_k}$$

It can also be written as inner product thresholding:

$$f_M = \sum_{m=0}^{\infty} \theta_T \left(\langle f, g_m \rangle \right) g_m$$

with

$$T = |\langle f, g_{m_M} \rangle|, \qquad \theta_T(x) = \begin{cases} x & |x| \geq T \\ 0 & |x| < T \end{cases}$$

The non-linear error is

$$\varepsilon[M] = \left\{ \|f - f_M\|^2 \right\} = \sum_{k=M+1}^{\infty} \left\{ |\langle f, g_{m_k} \rangle|^2 \right\}$$

this error goes quickly to zero as M increases, if the sorted values of $|\langle f, g_{m_k} \rangle|$ have a fast decay as k increases. This decay is quantified by computing the $\mathbf{I^P}$ norm of the signal inner products in B:

$$\|f\|_{\mathbf{B},p} = \left(\sum_{m=0}^{\infty} |\langle f, g_m \rangle|^p \right)^{\frac{1}{p}}$$

The following theorem relates the decay of $\varepsilon[M]$ to $\|f\|_{\mathbf{B},p}$

Theorem (decay of error). If $\|f\|_{\mathbf{B},p} < \infty$ with $p < 2$ then

$$\varepsilon[M] \le \frac{\|f\|_{\mathrm{B},p}^2}{\frac{2}{p} - 1} M^{1 - \frac{2}{p}}$$

and

$$\varepsilon[M] = o\left(M^{1 - \frac{2}{p}}\right).$$

Conversely, if $\varepsilon[M] = o\left(M^{1 - \frac{2}{p}}\right)$ then

$\|f\|_{\mathrm{B},q} < \infty$ for any $q > p$.

Non-optimality of Karhunen–Loève bases

To further illustrate the differences between linear and non-linear approximations, we study the decomposition of a simple non-Gaussian random vector in a Karhunen–Loève basis. Processes whose realizations have a random translation are stationary. The Karhunen–Loève basis is then a Fourier basis and we study its performance. To simplify the analysis, consider a random vector $Y[n]$ of size N that is random shift modulo N of a deterministic signal $f[n]$ of zero mean

$$\sum_{n=0}^{N-1} f[n] = 0$$
$$Y[n] = f[(n - p) \bmod N]$$

The random shift P is uniformly distributed on $[0, N - 1]$:

$$\mathbf{Pr}(P = p) = \frac{1}{N}, \qquad 0 \le p < N$$

Clearly

$$\mathbf{E}\{Y[n]\} = \frac{1}{N} \sum_{p=0}^{N-1} f[(n - p) \bmod N] = 0$$

and

$$R[n, k] = \mathbf{E}\{Y[n]Y[k]\} = \frac{1}{N} \sum_{p=0}^{N-1} f[(n - p) \bmod N] f[(k - p) \bmod N] = \frac{1}{N} f \Theta \bar{f}[n - k], \quad \bar{f}[n] = f[-n]$$

Hence

$$R[n, k] = R_Y[n - k], \qquad R_Y[k] = \frac{1}{N} f \Theta \bar{f}[k]$$

Since R_Y is N periodic, Y is a circular stationary random vector. The covariance operator is a circular convolution with R_Y and is therefore diagonalized in the discrete Fourier Karhunen–Loève basis

$$\left\{ \frac{1}{\sqrt{N}} e^{i2\pi mn/N} \right\}_{0 \le m < N}.$$

The power spectrum is Fourier transform of R_Y:

$$P_Y[m] = \hat{R}_Y[m] = \frac{1}{N} \left| \hat{f}[m] \right|^2$$

Example: Consider an extreme case where $f[n] = \delta[n] - \delta[n - 1]$. A theorem stated above guarantees that the Fourier Karhunen–Loève basis produces a smaller expected approximation error than a canonical basis of Diracs $\{g_m[n] = \delta[n - m]\}_{0 \le m < N}$. Indeed, we do not know a priori the abscissa of the non-zero coefficients of Y, so there is no particular Dirac that is better adapted to perform the approximation. But the Fourier vectors cover the whole support of Y and thus absorb a part of the signal energy.

$$\mathbf{E}\left\{\left|\left\langle Y[n], \frac{1}{\sqrt{N}}e^{i2\pi mn/N}\right\rangle\right|^2\right\} = P_Y[m] = \frac{4}{N}\sin^2\left(\frac{\pi k}{N}\right)$$

Selecting higher frequency Fourier coefficients yields a better mean-square approximation than choosing a priori a few Dirac vectors to perform the approximation. The situation is totally different for non-linear approximations. If $f[n] = \delta[n] - \delta[n-1]$ then the discrete Fourier basis is extremely inefficient because f and hence Y have an energy that is almost uniformly spread among all Fourier vectors. In contrast, since f has only two non-zero coefficients in the Dirac basis, a non-linear approximation of Y with $M \geq 2$ gives zero error.[6]

Principal component analysis

We have established the Karhunen–Loève theorem and derived a few properties thereof. We also noted that one hurdle in its application was the numerical cost of determining the eigenvalues and eigenfunctions of its covariance operator through the Fredholm integral equation of the second kind

$$\int_a^b K_X(s,t)e_k(s)\,ds = \lambda_k e_k(t).$$

However, when applied to a discrete and finite process $(X_n)_{n\in\{1,\dots,N\}}$, the problem takes a much simpler form and standard algebra can be used to carry out the calculations.

Note that a continuous process can also be sampled at N points in time in order to reduce the problem to a finite version.

We henceforth consider a random N-dimensional vector $X = (X_1\ X_2\ \dots\ X_N)^T$. As mentioned above, X could contain N samples of a signal but it can hold many more representations depending on the field of application. For instance it could be the answers to a survey or economic data in an econometrics analysis.

As in the continuous version, we assume that X is centered, otherwise we can let $X := X - \mu_X$ (where μ_X is the mean vector of X) which is centered.

Let us adapt the procedure to the discrete case.

Covariance matrix

Recall that the main implication and difficulty of the KL transformation is computing the eigenvectors of the linear operator associated to the covariance function, which are given by the solutions to the integral equation written above.

Define Σ, the covariance matrix of X, as an $N \times N$ matrix whose elements are given by:

$$\Sigma_{ij} = \mathbf{E}[X_i X_j], \qquad \forall i,j \in \{1,\dots,N\}$$

Rewriting the above integral equation to suit the discrete case, we observe that it turns into:

$$\sum_{j=1}^{N}\Sigma_{ij}e_j = \lambda e_i \quad \Leftrightarrow \quad \Sigma e = \lambda e$$

where $e = (e_1\ e_2\ \dots\ e_N)^T$ is an N-dimensional vector.

The integral equation thus reduces to a simple matrix eigenvalue problem, which explains why the PCA has such a broad domain of applications.

Since Σ is a positive definite symmetric matrix, it possesses a set of orthonormal eigenvectors forming a basis of $\mathbb{R}^N$, and we write $\{\lambda_i, \varphi_i\}_{i\in\{1,\dots,N\}}$ this set of eigenvalues and corresponding eigenvectors, listed in decreasing values of λ_i. Let also Φ be the orthonormal matrix consisting of these eigenvectors:

$$\Phi := (\varphi_1\ \varphi_2\ \dots\ \varphi_N)^T$$
$$\Phi^T\Phi = I$$

Principal component transform

It remains to perform the actual KL transformation, called the *principal component transform* in this case. Recall that the transform was found by expanding the process with respect to the basis spanned by the eigenvectors of the covariance function. In this case, we hence have:

$$X = \sum_{i=1}^{N} \langle \varphi_i, X \rangle \varphi_i = \sum_{i=1}^{N} \varphi_i^T X \varphi_i$$

In a more compact form, the principal component transform of X is defined by:

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

The *i*-th component of Y is $Y_i = \varphi_i^T X$, the projection of X on φ_i and the inverse transform $X = \Phi Y$ yields the expansion of X on the space spanned by the φ_i:

$$X = \sum_{i=1}^{N} Y_i \varphi_i = \sum_{i=1}^{N} \langle \varphi_i, X \rangle \varphi_i$$

As in the continuous case, we may reduce the dimensionality of the problem by truncating the sum at some $K \in \{1, \ldots, N\}$ such that

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \geq \alpha$$

where α is the explained variance threshold we wish to set.

We can also reduce the dimensionality through the use of multilevel dominant eigenvector estimation (MDEE).[7]

Examples

The Wiener process

There are numerous equivalent characterizations of the Wiener process which is a mathematical formalization of Brownian motion. Here we regard it as the centered standard Gaussian process $\mathbf{W}_t$ with covariance function

$$K_W(t, s) = \text{cov}(W_t, W_s) = \min(s, t).$$

We restrict the time domain to [*a*, *b*]=[0,1] without loss of generality.

The eigenvectors of the covariance kernel are easily determined. These are

$$e_k(t) = \sqrt{2} \sin\left(\left(k - \tfrac{1}{2}\right) \pi t\right)$$

and the corresponding eigenvalues are

$$\lambda_k = \frac{1}{(k - \tfrac{1}{2})^2 \pi^2}.$$

[Proof]

In order to find the eigenvalues and eigenvectors, we need to solve the integral equation:

$$\int_a^b K_W(s, t) e(s) \, ds = \lambda e(t) \qquad \forall t, 0 \leq t \leq 1$$

$$\int_0^1 \min(s, t) e(s) \, ds = \lambda e(t) \qquad \forall t, 0 \leq t \leq 1$$

$$\int_0^t s e(s) \, ds + t \int_t^1 e(s) \, ds = \lambda e(t) \qquad \forall t, 0 \leq t \leq 1$$

differentiating once with respect to t yields:

$$\int_t^1 e(s)\,ds = \lambda e'(t)$$

a second differentiation produces the following differential equation:

$$-e(t) = \lambda e''(t)$$

The general solution of which has the form:

$$e(t) = A\sin\left(\frac{t}{\sqrt{\lambda}}\right) + B\cos\left(\frac{t}{\sqrt{\lambda}}\right)$$

where A and B are two constants to be determined with the boundary conditions. Setting $t = 0$ in the initial integral equation gives $e(0) = 0$ which implies that $B = 0$ and similarly, setting $t = 1$ in the first differentiation yields $e'(1) = 0$, whence:

$$\cos\left(\frac{1}{\sqrt{\lambda}}\right) = 0$$

which in turn implies that eigenvalues of T_{K_X} are:

$$\lambda_k = \left(\frac{1}{(k - \frac{1}{2})\pi}\right)^2, \qquad k \ge 1$$

The corresponding eigenfunctions are thus of the form:

$$e_k(t) = A\sin\left((k - \frac{1}{2})\pi t\right), \qquad k \ge 1$$

A is then chosen so as to normalize e_k:

$$\int_0^1 e_k^2(t)\,dt = 1 \quad \Longrightarrow \quad A = \sqrt{2}$$

This gives the following representation of the Wiener process:

Theorem. There is a sequence $\{Z_i\}_i$ of independent Gaussian random variables with mean zero and variance 1 such that

$$W_t = \sqrt{2}\sum_{k=1}^\infty Z_k \frac{\sin\left(\left(k - \frac{1}{2}\right)\pi t\right)}{\left(k - \frac{1}{2}\right)\pi}.$$

Note that this representation is only valid for $t \in [0, 1]$. On larger intervals, the increments are not independent. As stated in the theorem, convergence is in the L^2 norm and uniform in t.

The Brownian bridge

Similarly the Brownian bridge $B_t = W_t - tW_1$ which is a stochastic process with covariance function

$$K_B(t, s) = \min(t, s) - ts$$

can be represented as the series

$$B_t = \sum_{k=1}^\infty Z_k \frac{\sqrt{2}\sin(k\pi t)}{k\pi}$$

Applications

Adaptive optics systems sometimes use K–L functions to reconstruct wave-front phase information (Dai 1996, JOSA A). Karhunen–Loève expansion is closely related to the Singular Value Decomposition. The latter has myriad applications in image processing, radar, seismology, and the like. If one has independent vector observations from a vector valued stochastic process then the left singular vectors are maximum likelihood estimates of the ensemble KL expansion.

Applications in signal estimation and detection

Detection of a known continuous signal $S(t)$

In communication, we usually have to decide whether a signal from a noisy channel contains valuable information. The following hypothesis testing is used for detecting continuous signal $s(t)$ from channel output $X(t)$, $N(t)$ is the channel noise, which is usually assumed zero mean Gaussian process with correlation function $R_N(t, s) = E[N(t)N(s)]$

$$H : X(t) = N(t),$$
$$K : X(t) = N(t) + s(t), \quad t \in (0, T)$$

Signal detection in white noise

When the channel noise is white, its correlation function is

$$R_N(t) = \tfrac{1}{2} N_0 \delta(t),$$

and it has constant power spectrum density. In physically practical channel, the noise power is finite, so:

$$S_N(f) = \begin{cases} \frac{N_0}{2} & |f| < w \\ 0 & |f| > w \end{cases}$$

Then the noise correlation function is sinc function with zeros at $\dfrac{n}{2\omega}, n \in \mathbf{Z}.$ Since are uncorrelated and gaussian, they are independent. Thus we can take samples from $X(t)$ with time spacing

$$\Delta t = \frac{n}{2\omega} \text{ within } (0,'' T'').$$

Let $X_i = X(i \, \Delta t)$. We have a total of $n = \dfrac{T}{\Delta t} = T(2\omega) = 2\omega T$ i.i.d observations $\{X_1, X_2, \dots, X_n\}$ to develop the likelihood-ratio test. Define signal $S_i = S(i \, \Delta t)$, the problem becomes,

$$H : X_i = N_i,$$

$$K : X_i = N_i + S_i, i = 1, 2, \dots, n.$$

The log-likelihood ratio

$$\mathcal{L}(\underline{x}) = \log \frac{\sum_{i=1}^n \left(2 S_i x_i - S_i^2 \right)}{2\sigma^2} \Leftrightarrow \Delta t \sum_{i=1}^n S_i x_i = \sum_{i=1}^n S(i \, \Delta t) x(i \, \Delta t) \, \Delta t \gtrless \lambda.2$$

As $t \rightarrow 0$, let:

$$G = \int_0^T S(t) x(t) \, dt.$$

Then G is the test statistics and the Neyman–Pearson optimum detector is

$$G(\underline{x}) > G_0 \Rightarrow K < G_0 \Rightarrow H.$$

As G is Gaussian, we can characterize it by finding its mean and variances. Then we get

$$H : G \sim N\left(0, \tfrac{1}{2} N_0 E\right)$$
$$K : G \sim N\left(E, \tfrac{1}{2} N_0 E\right)$$

where

$$\mathbf{E} = \int_0^T S^2(t)\,dt$$

is the signal energy.

The false alarm error

$$\alpha = \int_{G_0}^\infty N\left(0, \tfrac{1}{2}N_0 E\right)\,dG \Rightarrow G_0 = \sqrt{\tfrac{1}{2}N_0 E}\,\Phi^{-1}(1-\alpha)$$

And the probability of detection:

$$\beta = \int_{G_0}^\infty N\left(E, \tfrac{1}{2}N_0 E\right)\,dG = 1 - \Phi\left(\frac{G_0 - E}{\sqrt{\tfrac{1}{2}N_0 E}}\right) = \Phi\left(\sqrt{\frac{2E}{N_0}} - \Phi^{-1}(1-\alpha)\right),$$

where Φ is the cdf of standard normal, or Gaussian, variable.

Signal detection in colored noise

When N(t) is colored (correlated in time) Gaussian noise with zero mean and covariance function $R_N(t,s) = E[N(t)N(s)]$, we cannot sample independent discrete observations by evenly spacing the time. Instead, we can use K–L expansion to uncorrelate the noise process and get independent Gaussian observation 'samples'. The K–L expansion of $N(t)$:

$$N(t) = \sum_{i=1}^\infty N_i \Phi_i(t), \quad 0 < t < T,$$

where $N_i = \int N(t)\Phi_i(t)\,dt$ and the orthonormal bases $\{\Phi_i t\}$ are generated by kernel $R_N(t,s)$, i.e., solution to

$$\int_0^T R_N(t,s)\Phi_i(s)\,ds = \lambda_i \Phi_i(t), \quad \mathrm{var}[N_i] = \lambda_i.$$

Do the expansion:

$$S(t) = \sum_{i=1}^\infty S_i \Phi_i(t),$$

where $S_i = \int_0^T S(t)\Phi_i(t)\,dt$, then

$$X_i = \int_0^T X(t)\Phi_i(t)\,dt = N_i$$

under H and $N_i + S_i$ under K. Let $\overline{X} = \{X_1, X_2, \ldots\}$, we have

N_i are independent Gaussian r.v's with variance λ_i

under H: $\{X_i\}$ are independent Gaussian r.v's.

$$f_H[x(t)|0 < t < T] = f_H(\underline{x}) = \prod_{i=1}^\infty \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{x_i^2}{2\lambda_i}\right)$$

under K: $\{X_i - S_i\}$ are independent Gaussian r.v's.

$$f_K[x(t) \mid 0 < t < T] = f_K(\underline{x}) = \prod_{i=1}^\infty \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{(x_i - S_i)^2}{2\lambda_i}\right)$$

Hence, the log-LR is given by

$$\mathcal{L}(\underline{x}) = \sum_{i=1}^{\infty} \frac{2S_i x_i - S_i^2}{2\lambda_i}$$

and the optimum detector is

$$G = \sum_{i=1}^{\infty} S_i x_i \lambda_i > G_0 \Rightarrow K, < G_0 \Rightarrow H.$$

Define

$$k(t) = \sum_{i=1}^{\infty} \lambda_i S_i \Phi_i(t), 0 < t < T,$$

then $G = \int_0^T k(t)x(t)\, dt.$

How to find $k(t)$

Since

$$\int_0^T R_N(t, s)k(s)\, ds = \sum_{i=1}^{\infty} \lambda_i S_i \int_0^T R_N(t, s)\Phi_i(s)\, ds = \sum_{i=1}^{\infty} S_i \Phi_i(t) = S(t),$$

k(t) is the solution to

$$\int_0^T R_N(t, s)k(s)\, ds = S(t).$$

If $N(t)$ is wide-sense stationary,

$$\int_0^T R_N(t - s)k(s)\, ds = S(t),$$

which is known as the <u>Wiener–Hopf equation</u>. The equation can be solved by taking fourier transform, but not practically realizable since infinite spectrum needs spatial factorization. A special case which is easy to calculate $k(t)$ is white Gaussian noise.

$$\int_0^T \frac{N_0}{2} \delta(t - s)k(s)\, ds = S(t) \Rightarrow k(t) = CS(t), \quad 0 < t < T.$$

The corresponding impulse response is $h(t) = k(T - t) = CS(T - t)$. Let $C = 1$, this is just the result we arrived at in previous section for detecting of signal in white noise.

Test threshold for Neyman–Pearson detector

Since X(t) is a Gaussian process,

$$G = \int_0^T k(t)x(t)\, dt,$$

is a Gaussian random variable that can be characterized by its mean and variance.

$$\mathbf{E}[G \mid H] = \int_0^T k(t)\mathbf{E}[x(t) \mid H] \, dt = 0$$

$$\mathbf{E}[G \mid K] = \int_0^T k(t)\mathbf{E}[x(t) \mid K] \, dt = \int_0^T k(t)S(t) \, dt \equiv \rho$$

$$\mathbf{E}[G^2 \mid H] = \int_0^T \int_0^T k(t)k(s)R_N(t,s) \, dt \, ds = \int_0^T k(t) \left(\int_0^T k(s)R_N(t,s) \, ds \right) = \int_0^T k(t)S(t) \, dt = \rho$$

$$\mathbf{var}[G \mid H] = \mathbf{E}[G^2 \mid H] - (\mathbf{E}[G \mid H])^2 = \rho$$

$$\mathbf{E}[G^2 \mid K] = \int_0^T \int_0^T k(t)k(s)\mathbf{E}[x(t)x(s)] \, dt \, ds = \int_0^T \int_0^T k(t)k(s)(R_N(t,s) + S(t)S(s)) \, dt \, ds = \rho + \rho^2$$

$$\mathbf{var}[G \mid K] = \mathbf{E}[G^2|K] - (\mathbf{E}[G|K])^2 = \rho + \rho^2 - \rho^2 = \rho$$

Hence, we obtain the distributions of H and K:

$$H : G \sim N(0, \rho)$$
$$K : G \sim N(\rho, \rho)$$

The false alarm error is

$$\alpha = \int_{G_0}^\infty N(0, \rho) \, dG = 1 - \Phi\left(\frac{G_0}{\sqrt{\rho}} \right).$$

So the test threshold for the Neyman–Pearson optimum detector is

$$G_0 = \sqrt{\rho}\Phi^{-1}(1 - \alpha).$$

Its power of detection is

$$\beta = \int_{G_0}^\infty N(\rho, \rho) \, dG = \Phi\left(\sqrt{\rho} - \Phi^{-1}(1 - \alpha) \right)$$

When the noise is white Gaussian process, the signal power is

$$\rho = \int_0^T k(t)S(t) \, dt = \int_0^T S(t)^2 \, dt = E.$$

Prewhitening

For some type of colored noise, a typical practise is to add a prewhitening filter before the matched filter to transform the colored noise into white noise. For example, N(t) is a wide-sense stationary colored noise with correlation function

$$R_N(\tau) = \frac{BN_0}{4}e^{-B|\tau|}$$
$$S_N(f) = \frac{N_0}{2(1 + (\frac{w}{B})^2)}$$

The transfer function of prewhitening filter is

$$H(f) = 1 + j\frac{w}{B}.$$

Detection of a Gaussian random signal in Additive white Gaussian noise (AWGN)

When the signal we want to detect from the noisy channel is also random, for example, a white Gaussian process $X(t)$, we can still implement K–L expansion to get independent sequence of observation. In this case, the detection problem is described as follows:

$$H_0 : Y(t) = N(t)$$
$$H_1 : Y(t) = N(t) + X(t), \quad 0 < t < T.$$

$X(t)$ is a random process with correlation function $R_X(t,s) = E\{X(t)X(s)\}$

The K–L expansion of $X(t)$ is

$$X(t) = \sum_{i=1}^{\infty} X_i \Phi_i(t),$$

where

$$X_i = \int_0^T X(t) \Phi_i(t)\, dt$$

and $\Phi_i(t)$ are solutions to

$$\int_0^T R_X(t, s) \Phi_i(s)\, ds = \lambda_i \Phi_i(t).$$

So X_i's are independent sequence of r.v's with zero mean and variance λ_i. Expanding $Y(t)$ and $N(t)$ by $\Phi_i(t)$, we get

$$Y_i = \int_0^T Y(t) \Phi_i(t)\, dt = \int_0^T [N(t) + X(t)] \Phi_i(t) = N_i + X_i,$$

where

$$N_i = \int_0^T N(t) \Phi_i(t)\, dt.$$

As $N(t)$ is Gaussian white noise, N_i's are i.i.d sequence of r.v with zero mean and variance $\frac{1}{2} N_0$, then the problem is simplified as follows,

$$H_0 : Y_i = N_i$$
$$H_1 : Y_i = N_i + X_i$$

The Neyman–Pearson optimal test:

$$\Lambda = \frac{f_Y \mid H_1}{f_Y \mid H_0} = Ce^{-\sum_{i=1}^{\infty} \frac{y_i^2}{2} \frac{\lambda_i}{\frac{1}{2}N_0(\frac{1}{2}N_0+\lambda_i)}},$$

so the log-likelihood ratio is

$$\mathcal{L} = \ln(\Lambda) = K - \sum_{i=1}^{\infty} \frac{1}{2} y_i^2 \frac{\lambda_i}{\frac{N_0}{2}\left(\frac{N_0}{2} + \lambda_i\right)}.$$

Since

$$\widehat{X}_i = \frac{\lambda_i}{\frac{N_0}{2}\left(\frac{N_0}{2} + \lambda_i\right)}$$

is just the minimum-mean-square estimate of X_i given Y_i's,

K–L expansion has the following property: If

$$f(t) = \sum f_i \Phi_i(t), \quad g(t) = \sum g_i \Phi_i(t),$$

where

$$f_i = \int_0^T f(t) \Phi_i(t)\, dt, \quad g_i = \int_0^T g(t) \Phi_i(t)\, dt.$$

then

$$\sum_{i=1}^{\infty} f_i g_i = \int_0^T g(t) f(t)\, dt.$$

So let

$$\widehat{X}(t \mid T) = \sum_{i=1}^{\infty} \widehat{X}_i \Phi_i(t), \quad \mathcal{L} = K + \frac{1}{N_0} \int_0^T Y(t) \widehat{X}(t \mid T)\, dt.$$

Noncausal filter $Q(t,s)$ can be used to get the estimate through

$$\widehat{X}(t \mid T) = \int_0^T Q(t,s) Y(s)\, ds.$$

By orthogonality principle, $Q(t,s)$ satisfies

$$\int_0^T Q(t,s) R_X(s,t)\, ds + \frac{N_0}{2} Q(t,\lambda) = R_X(t,\lambda), 0 < \lambda < T, 0 < t < T.$$

However, for practical reasons, it's necessary to further derive the causal filter $h(t,s)$, where $h(t,s) = 0$ for $s > t$, to get estimate $\widehat{X}(t \mid t)$. Specifically,

$$Q(t,s) = h(t,s) + h(s,t) - \int_0^T h(\lambda,t) h(s,\lambda)\, d\lambda$$

See also

- Principal component analysis
- Proper orthogonal decomposition
- Polynomial chaos

Notes

1. Sapatnekar, Sachin (2011), "Overcoming variations in nanometer-scale technologies", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, **1** (1): 5–18, Bibcode:2011IJEST...1....5S (https://ui.adsabs.harvard.edu/abs/2011IJEST...1....5S), CiteSeerX 10.1.1.300.5659 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.300.5659), doi:10.1109/jetcas.2011.2138250 (https://doi.org/10.1109%2Fjetcas.2011.2138250)
2. Ghoman, Satyajit; Wang, Zhicun; Chen, PC; Kapania, Rakesh (2012), *A POD-based Reduced Order Design Scheme for Shape Optimization of Air Vehicles*
3. Karhunen–Loeve transform (KLT) (http://fourier.eng.hmc.edu/e161/lectures/klt/node3.html), Computer Image Processing and Analysis (E161) lectures, Harvey Mudd College
4. Raju, C.K. (2009), "Kosambi the Mathematician", *Economic and Political Weekly*, **44** (20): 33–45
5. Kosambi, D. D. (1943), "Statistics in Function Space", *Journal of the Indian Mathematical Society*, **7**: 76–88, MR 0009816 (https://www.ams.org/mathscinet-getitem?mr=0009816).
6. A wavelet tour of signal processing-Stéphane Mallat
7. X. Tang, "Texture information in run-length matrices," IEEE Transactions on Image Processing, vol. 7, No. 11, pp. 1602–1609, Nov. 1998

References

- Stark, Henry; Woods, John W. (1986). *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice-Hall, Inc. ISBN 978-0-13-711706-2. OL 21138080M (https://openlibrary.org/books/OL21138080M).
- Ghanem, Roger; Spanos, Pol (1991). *Stochastic finite elements: a spectral approach*. Springer-Verlag. ISBN 978-0-387-97456-9. OL 1865197M (https://openlibrary.org/books/OL1865197M).
- Guikhman, I.; Skorokhod, A. (1977). *Introduction a la Théorie des Processus Aléatoires*. Éditions MIR.
- Simon, B. (1979). *Functional Integration and Quantum Physics*. Academic Press.
- Karhunen, Kari (1947). "Über lineare Methoden in der Wahrscheinlichkeitsrechnung". *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys*. **37**: 1–79.

- Loève, M. (1978). *Probability theory.* Vol. II, 4th ed*. Graduate Texts in Mathematics. **46**. Springer-Verlag. ISBN 978-0-387-90262-3.*
- Dai, G. (1996). "Modal wave-front reconstruction with Zernike polynomials and Karhunen–Loeve functions". *JOSA A*. **13** (6): 1218. Bibcode:1996JOSAA..13.1218D (https://ui.adsabs.harvard.edu/abs/1996JOSAA..13.1218D). doi:10.1364/JOSAA.13.001218 (https://doi.org/10.1364%2FJOSAA.13.001218).
- Wu B., Zhu J., Najm F.(2005) "A Non-parametric Approach for Dynamic Range Estimation of Nonlinear Systems". In Proceedings of Design Automation Conference(841-844) 2005
- Wu B., Zhu J., Najm F.(2006) "Dynamic Range Estimation". IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 25 Issue:9 (1618–1636) 2006
- Jorgensen, Palle E. T.; Song, Myung-Sin (2007). "Entropy Encoding, Hilbert Space and Karhunen–Loeve Transforms". *Journal of Mathematical Physics*. **48** (10): 103503. arXiv:math-ph/0701056 (https://arxiv.org/abs/math-ph/0701056). Bibcode:2007JMP....48j3503J (https://ui.adsabs.harvard.edu/abs/2007JMP....48j3503J). doi:10.1063/1.2793569 (https://doi.org/10.1063%2F1.2793569).

External links

- *Mathematica* KarhunenLoeveDecomposition (http://reference.wolfram.com/mathematica/ref/KarhunenLoeveDecomposition.html) function.
- *E161: Computer Image Processing and Analysis* notes by Pr. Ruye Wang at Harvey Mudd College [1] (http://fourier.eng.hmc.edu/e161/lectures/klt/klt.html)