0.1 dot product

Derivation. Geometrically speaking the dot product shows how much two vectors are covarying in the same direction. Given two vectors v and w,

$$v \cdot w = ||w||\big(||v|| \cos\theta\big). \tag{1}$$

When written as

$$v \cdot \frac{w}{||w||} = ||v|| |\cos\theta, \tag{2}$$

this shows how much v is projected in the direction of w. This demonstrates if the magitude of both vectors are fixed, then the closer they are oriented, the more v is projected onto w. The greatest magnitude occurs when they overlap in the same direction $v \cdot w = ||v||||w||$. The product is subject to the size of v and w, so to compare how much one varies in the direction of the other will need normalization as (2).

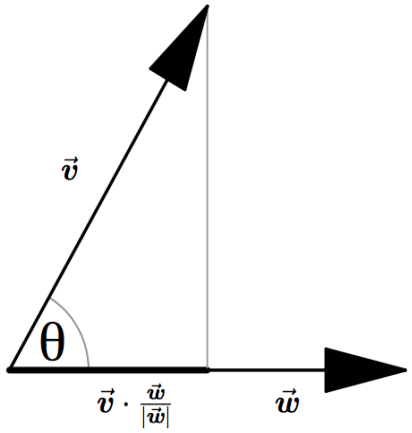

Figure 1: Dot product is a projection without normalization.

0.2 Visualize linear transformation of a function as a matrix

Suppose $T : V \in \mathcal{R}^n \to W \in \mathcal{R}^m$, hence $T(x_1, \ldots, x_n) = (b_1, \ldots, b_m)$, where x_j and b_i are the coordinate associated to the respective basis, and $M(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m))$.
What does a matrix do?

1. Each domain basis vector is tranformed to the codomain as a vector

$$Te_j = a_{11}f_1 + \cdots + a_{m1}f_m = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \tag{3}$$

   i.e. each basis vector $e_j \in V$ is mapped to the space spanned by f_i's, recorded by the unique values of the jth column vector of $M(T)$.

2. $Tv = T(x_1e_1 + \cdots + x_ne_n) = x_1Te_1 + \cdots + x_nTe_n = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \ldots x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \ldots \\ b_m \end{pmatrix}$, where $b_i = a_{ij}x_j$ or $b_if_i = \sum_j a_{ij}x_jf_i$. This helps visualize that each basis vector e_j is first mapped to the space of W recorded by a column vector and individually stretched by x_j's and combined to form a vector in W.

0.3 What is linear algebra?

- Solving system of equations $Ax = b$? If treated the equations as hyperplanes $a_i^T x = b_i$, then we are finding the intersection points.
- Finding the unique vector x that gets mapped to b by the linear function A?
  - If treated A as a linear operator on the same basis, then we are finding x that gets mapped by A to b in the same domain.
  - If treated A transforming between different basis for $x = x_1 e_1 + \cdots + x_n e_n$ and $b = b_1 f_1 + \cdots + b_m f_m$, then we are finding x in the domain that gets mapped by A to b in the codomain.

0.4 Invertible function

Definition: If $AB = I$, then B is the inverse function of A.

Example 1. Prove that if $ABw_j = w_j$, then $a_{j\cdot}^T b_{\cdot j} = 1$ and AB is an identity matrix.
Proof:
$ABw_j = A(\sum_i b_{ij} v_i) = \sum_i b_{ij}(Av_i) = \sum_i b_{ij}(\sum_k a_{ki} w_k) = \sum_i \sum_k b_{ij} a_{ki} w_k$. If $ABw_j = w_j$, then $\sum_i \sum_k b_{ij} a_{ki} w_k = w_j$ if and only if $\sum_i a_{ki} b_{ij} \delta_{kj} = a_{j\cdot}^T b_{\cdot j} = 1$. AB is an identity matrix.

0.5 How to solve eigenvalues of a matrix without determinant

Use Gauss elimination to the last pivot and find the eigenvalues.

0.6 Karhunen Luoeve Expansion

Q1: What is the expansion done on? The random variables at each random variable index? Or the entire random process/field expanded?
A1: Expanded as a random process, not individual random variables expanded (as the PCE does).
Q2: Where do you get the eigenvalues and eigenfunctions?
A2: In terms of continuous random process/fields, the integral (linear operator) of the continuous covariance function over a invariant function gives the eigenfunction and the corresponding eigenvalue. In a discrate random process, the integral on the covariance matrix (PSD) C is just the inner product of the convariance vector of a random variable and the eigenvector.
Q3: What "variance" is the eigenvalues associate with?
A3: The projection of a single realization of a random process on a eigenfunction gives a random coefficient. The variance of the collection of all the realizations of the random coefficient shows how much the random process varies in the direction of the eigenfunction. This variance equals to the eigenvalue!

0.6.1 Mercer's Thm:

Given a continuous symmetric positive semidemfinite (PSD) kernel K on the random process/field X_s, the kernel can be expanded as

$$K(x, s) = \sum_i \lambda_i e_i(x) e_i(s). \tag{4}$$

Proof:
Given a covariance kernel (PSD) $K(x, s) = \mathbf{E}[X_x X_s]$, and $K : [a\ b] \times [a\ b] \to \mathcal{R}$, where x and s is

the continuous index of the random variables. Associated to K is a linear operator T_K on functions ϕ (covariance matrix operating on vectors is the discrete case) defined as

$$[T_K\phi](x) = \int_{[a\ b]} K(x,s)\phi(s)ds. \tag{5}$$

To visualize this function, if we fix x at a point, then the $K(x,s)$ and $\phi(s)$ are both functionals over the domain s. The integration (inner product in discrete s) acts as how much $K(x,s)$ is related to the mode $\phi(s)$. By testing over all x (all rows), one could possibly find a mode (eigenvector) that is perfectly stretched in the mode direction. Which is all rows of $K(x,s)$ relate to (by inner product) the mode with a single factor (eigenvalue). Therefore, if there is a basis $\phi(s)$ that best represents (captures the variance) the covarying feature of the random process in s. Then we could build a diagonal covariance matrix with this basis. **Since T_K is a linear operator, we can talk about the eigenfunctions and eigenvalues of T_K**. Therefore, the eigenfunctions associated to the linear operator has the representation

$$[T_K e_k](x) = \int_{[a\ b]} K(x,s)e_k(s)ds = \lambda_k e_k, \tag{6}$$

where e_k's are orthonormal wrt the linear operator $\int_{[a\ b]} e_k e_j ds = \delta_{kj}$. By plugging in $K(x,s) = \sum_i \lambda_i e_i(x)e_i(s)$ to the lhs equation

$$\int_{[a\ b]} \sum_i \lambda_i e_i(x)e_i(s)e_k(s)ds = \lambda_i e_k(x), \tag{7}$$

where orthogonality is used. This shows that in fact $K(x,s) = \sum_i \lambda_i e_i(x)e_i(s)$. ■
If X is a discrete random process/field, then $[T_K e_k](x_i) = \sum_j K(x_i,s_j)e_k(s_j) = \lambda_k e_k(x_i)$ is the inner product of the covariance vector of the ith random variable (ith row of the convariance matrix) and the eigenvector $e_k(\vec{s})$. From spectral theory of a PSD matrix, the covariance matrix (PSD) also yields an eigen-decomposition $K = V\Lambda V^T$, with $K_{ij} = \mathbf{E}[X_{x_i}X_{s_j}] = \sum_k \lambda_k e_k(x_i)e_k(s_j)$.
(Notice: the integral operator is not the expectation operator, so the expectation is only in the covariance kernel, but not shown explicitly)

0.6.2 Karhunen Luoeve's Thm:

Prove that a zero-mean square integrable stochastic process X_t with continuous covariance function can be represented as

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t) \tag{8}$$

where $Z_k = \int_a^b X_t e_k(t)dt$. Furthermore, prove $\mathbf{E}[Z_k] = 0$ and $\mathbf{E}[Z_k Z_j] = \delta_{kj}\lambda_k$ (the variance of Z_k, how much X_t varies in the direction of e_k in the mean, is just the eigenvalue of the covariance function).

Proof:
From the Mercer's Thm, $K(x,s) = \sum_i \lambda_i e_i(x)e_i(s)$. By projecting X_t onto the eigenfunctions obtained from K, $Z_k = \int_a^b X_t e_k(t)dt$, and we can form $X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$. Thus,

$$\mathbf{E}[Z_k] = \mathbf{E}[\int_a^b X_t e_k(t)dt] = \int_a^b \mathbf{E}[X_t]e_k(t)dt = 0, \tag{9}$$

and

$$\mathbf{E}[Z_i Z_j] = \delta_{ij}\lambda_i \text{ (This is a simple proof!)}. \tag{10}$$

Conclusion:
Eigenvalues represent the variance of the random process in the direction of the eigenvectors! This is why using the eigenvectors from decomposing the symmetric PSD covariance function makes sense!

0.7 Lyapunov Exponent

Given a linear dynamic system with known linear parameters a,b,c and d,

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{11}$$

The eigenvalues and eigenvectors of A can be obtained fairly easily, with eigenvalues being either complex or real. The orthonormal eigenvectors v_1 and v_2 are treated as the basis and the state x can be expanded as $x = c_1v_1 + c_2v_2$ (just as EOFs spanning the space of the multivariate space). Therefore the system becomes

$$\frac{d}{dt}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \tag{12}$$

This gives the evolution equation of the coefficients by using the invariant eigenvectors. The coefficients has the general solution form $c_i(t) = c_i(0)e^{\lambda_i t}$, where the stability is determined by the real or complex of the λ_i's. The coefficients determines how much the states are projected onto the eigenvectors at any time.

0.8 Self-Adjoint and Normal Operator

linear functional: $\phi(v) =< v, u >$, u is a "unique vector" in V for this functional $\mathcal{R}^n \to \mathcal{R}$.
Define adjoint function from the linear functional: $\phi(v) =< Tv, w >$, one can find the unique vector of this linear functional to be $T^*w \in V$ satisfying $< Tv, w >=< v, T^*w >$, T^* is the adjoint of T.

Example:
Suppose $T : \mathcal{R}^3 \to \mathcal{R}^2$ and $T(x_1, x_2, x_3) = (x_1 + x_2, x_3)$. The adjoint T^* is $T^*(y_1, y_2)$, therefore $< T(x_1, x_2, x_3), (y_1, y_2) >= x_1y_1 + x_2y_1 + x_3y_2 =< (x_1, x_2, x_3), T^*(y_1, y_2) >\Rightarrow T^*(y_1, y_2) = (y_1, y_1, y_2)$.

0.9 Kernel PCA

Given a d-dim random variable $\boldsymbol{x}$ with N random samples, the general PCE operates on zero-centered data

$$\mathbf{E}(\boldsymbol{x}) = \frac{1}{N}\sum_i^N \boldsymbol{x}_i = \mathbf{0} \tag{13}$$

with covariance matrix

$$\mathbf{E}(\boldsymbol{x}^2) = \frac{1}{N}\sum_i^N \boldsymbol{x}_i\boldsymbol{x}_i^T \tag{14}$$

which is a matrix of the average of a set of outer product matrices, or can be seen as a matrix with entries of the averaged inner product of a single dimension r. v. with N samples.

0.10 Generalized Eigenvalue Decomposition (2nd kind)

0.10.1 Definition

Given a pair of Hermitian matrices $(A, B) \in \mathcal{S}^n$, the generalized eigenvalue is the roots of $p(\lambda) = \det(A - \lambda B) = 0$ satisfying the relationship

$$Av = \lambda Bv. \tag{15}$$

Notice the generalized eigenvectors need not be orthogonal! The standard eigenvalue decomposition is just ordering λ_i in nonincreasing order, and denote the maximum generalized eigenvalue by $\lambda_{\max}(A, B)$.

0.10.2 Special case

If $B \in \mathcal{S}^n_{++}$, we know that positive definite matrix has positive definite square root ($B^{\frac{1}{2}} = (B^{\frac{1}{2}})^*$), $B = B^{\frac{1}{2}}(B^{\frac{1}{2}})^* = B^{\frac{1}{2}}B^{\frac{1}{2}}$. Therefore

$$AV = B^{\frac{1}{2}}B^{\frac{1}{2}}V\Lambda \Rightarrow B^{-\frac{1}{2}}AB^{-\frac{1}{2}}B^{\frac{1}{2}}V = B^{\frac{1}{2}}V\Lambda \Rightarrow B^{-\frac{1}{2}}AB^{-\frac{1}{2}}V' = V'\Lambda, \tag{16}$$

implies the generalized eigenvalues of (A, B) are the same as the eigenvalues of $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ when $B \in \mathcal{S}^n_{++}$. Since $V'\Lambda V'^T$ is the eigenvalue decomposition of $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$, the *generalized eigenvalue decomposition*

$$B^{-\frac{1}{2}}AB^{-\frac{1}{2}} = V'\Lambda V'^T \Rightarrow A = B^{\frac{1}{2}}V'\Lambda V'^T B^{\frac{1}{2}} \Rightarrow A = Q\Lambda Q^T \qquad B = B^{\frac{1}{2}}B^{\frac{1}{2}} = B^{\frac{1}{2}}V'V'^{-1}B^{\frac{1}{2}} = QQ^T. \tag{17}$$

1 Chapter 8 Operators on Complex Vector Space

This chapter is aimed to find the space spanned by the generalized eigenvectors of an operator (not all operators have enough eigenvectors).

Definition of Generalized Eigenvectors:
v is the generalized eigenvector of T corresponding to λ if $v \in \mathbf{null}(T - \lambda I)^j$.

Properties of generalized eigenvectors

- Inclusion : $\{0\} = \mathbf{null}(T-\lambda)^0 \subset \mathbf{null}(T-\lambda)^1 \subset \cdots \subset \mathbf{null}(T-\lambda)^k \subset \cdots$
- Saturation : $\mathbf{null}(T-\lambda)^0 \subset \cdots \subset \mathbf{null}(T-\lambda)^m = \mathbf{null}(T-\lambda)^{m+1} = \cdots$
- Maximization : $\mathbf{null}(T-\lambda)^{\mathbf{dim}V} = \mathbf{null}(T-\lambda)^{\mathbf{dim}V+1} = \cdots$

Remarks : Space of the generalized eigenvectors is $\mathbf{null}(T-\lambda)^{\mathbf{dim}V}$!!

Defining a Nilpotent operator using generalized eigenvectors

- Operators going to zero at some power, and the entire vector space is the generalized eigenvector space with respect to the zero eigenvalue.

Remarks : The power for a nilpotent operator must be contained by $\mathbf{dim}V$, $N^{\mathbf{dim}V} = 0$.