

## MOTIVATION :

Can we connect these issues of convergence over  $L^2[a, b]$  i.e., all square integrable functions in  $[a, b]$ ?

Suppose  $f \in L^2[a, b]$ . Let

$$f_N(x) = a_0 + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx) \quad k = 1, \dots, N$$

Here,  $a_k$  and  $b_k$  are obtained by projecting  $f$  onto  $\cos(kx)$  and  $\sin(kx)$ .

$f_N$  is the orthogonal projection of  $f$  onto a space  $V_N$ .  
i.e.,  $f_N$  is the element in  $V_N$  closest to  $f$  in the  $L^2$  sense

Theorem: Suppose  $f \in L^2[-\pi, \pi]$ . Let

$$f_N(x) = a_0 + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx)$$

where  $a_k$ 's &  $b_k$ 's  $^{k=1}$  are the Fourier coeffs. of  $f$ .

$$\text{Then, } \|f_N - f\|_{L^2} \xrightarrow{N \rightarrow \infty} 0$$

Subtle Issues:

Points of discontinuities in  $f$   
 $\exists$  possible extensions to overcome these.

Theorem: If a sequence  $f_n$  converges uniformly to  $f$  as  $n \rightarrow \infty$  over  $[a, b]$ , it also converges to  $f$  in  $L^2[a, b]$ .

Proof: From our definition of uniform convergence, for a tolerance  $\varepsilon > 0$  & integer  $N$ ,

$$|f_n(t) - f(t)| < \varepsilon \quad \forall n \geq N \text{ and } t \in [a, b]$$

Consider  $\|f_n^{(t)} - f^{(t)}\|_{L^2}^2 = \int_a^b |f_n(t) - f(t)|^2 dt$

$$\stackrel{1}{\leq} \int_a^b \varepsilon^2 dt = \varepsilon^2 (b-a)$$

$$\therefore \|f_n - f\| \leq \varepsilon \sqrt{b-a} \quad (n \geq N)$$

Since  $\varepsilon$  can be chosen as small as desired,

$$f_n \xrightarrow[n \rightarrow \infty]{} f \text{ in } L^2$$



Ponder : Examine if Conv. in  $L^2 \Rightarrow$  Conv. uniformly  
True / False

## Convergence in the mean

There may be cases where F.S. does not converge uniformly or pointwise. It may be useful to study if it converges in a weaker sense i.e., in  $L^2$  (in the mean)

Let us consider a  $2\pi$  periodic function, and  
let  $V = L^2[-\pi, \pi] / f(x) \in V \Rightarrow \int_{-\pi}^{\pi} |f(x)|^2 dx < \infty$

From our inner product spaces,

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx .$$

Let  $V_N$  be the space spanned by  $\{1, \cos(kx), \sin(kx)\}_{k=1}^N$ .

An element in  $V_N$  is  $c_0 + \sum_{k=1}^N c_k \cos(kx) + d_k \sin(kx)$

$c_k$ 's &  $d_k$ 's are possibly complex nos.

Let  $f_N(x) = a_0 + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx) \in V_N$   
 $a_k$  &  $b_k$  are obtained by projecting  $f(x)$  onto  $\cos(kx)$  &  $\sin(kx)$

Lemma : If  $V$  is an I.P. space and  $V_0$  is an  $N$ -dim. sub space with orthonormal basis  $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_N\}$ , the orthogonal projection of  $\underline{v} \in V$  onto  $V_0$  is

$$\underline{v}_0 = \sum_{j=1}^N \alpha_j \underline{e}_j, \quad \alpha_j = \langle \underline{v}, \underline{e}_j \rangle$$

Using Lemma above,  $f_N$  is the orthogonal projection of  $f$  onto space  $V_N$ .  
 $\Rightarrow f_N$  is the element in  $V_N$  closest to  $f$  in  $L^2$  sense  
 $\therefore \|f - f_N\|_{L^2} = \min_{g \in V_N} \|f - g\|_{L^2}$

Theorem: Suppose  $f$  is an element of  $L^2[-\pi, \pi]$ . Let  
 $f_N(x) = a_0 + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx)$ , where  
 $a_k$ 's &  $b_k$ 's are the Fourier coeffs. of  $f$ .

Then  $f_N \xrightarrow{N \rightarrow \infty} f$  in  $L^2[-\pi, \pi]$

i.e.,  $\|f_N - f\|_{L^2} \xrightarrow{N \rightarrow \infty} 0$



The proof sketch involves 2 steps

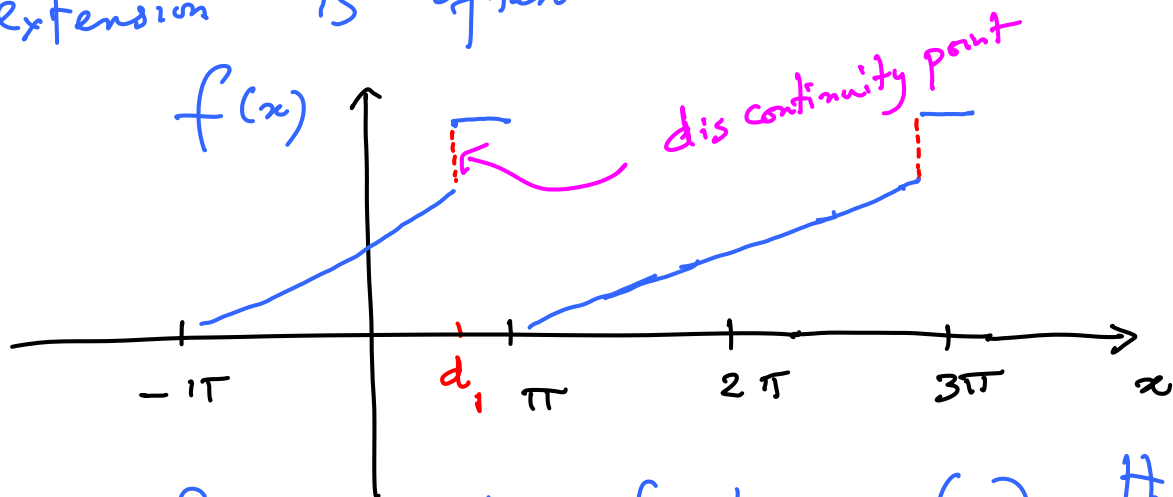
- 1) Show that any function in  $L^2[-\pi, \pi]$  can be approximated by a piecewise smooth periodic function  $g$ .
- 2) Approximate  $g$  uniformly ( & therefore in  $L^2$  ) by its Fourier series expansion.

We need to get an idea towards (1).

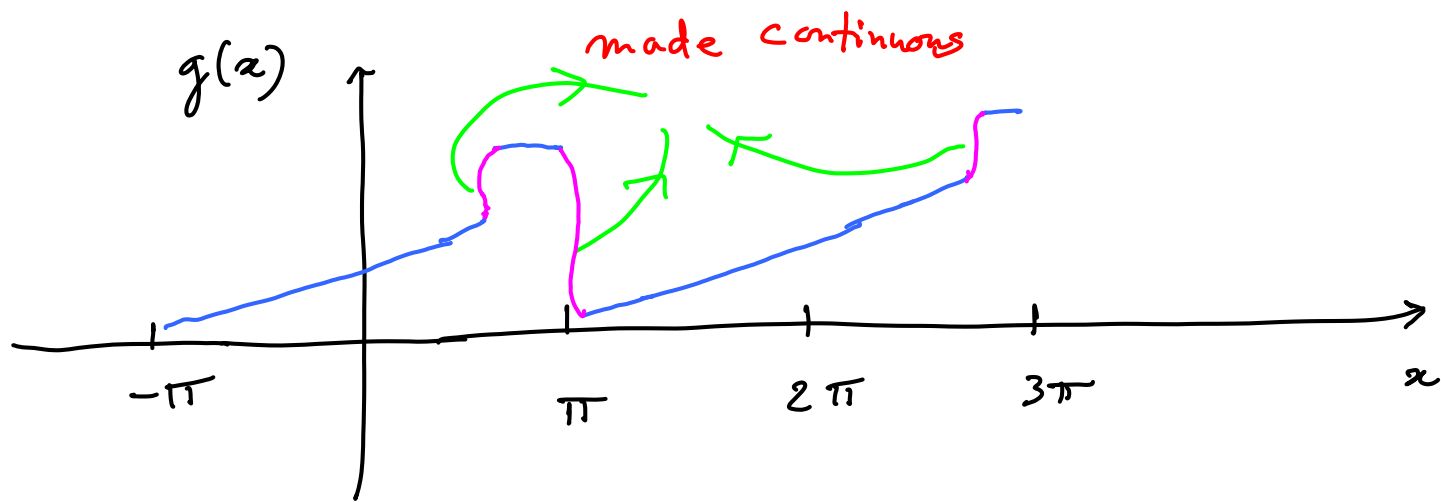
We have already established (2).

Idea : An element  $f \in L^2[-\pi, \pi]$  may not be continuous. Even if it is continuous, its periodic extension is often not continuous.

Example :



Let us form another function  $g(x)$  that agrees with  $f(x)$  at all segments except on segments connecting the continuous components of  $f$ .



We are fine if we choose a differentiable periodic function  $g(x)$  /

$$\|f - g\|_{L^2} < \varepsilon \quad \text{for a tolerance } \varepsilon$$

From our last theorem, let us recall.

Theorem: The F.S. of a piecewise smooth  $2\pi$  periodic function  $f(x)$  converges uniformly to  $f(x)$  on  $[-\pi, \pi]$

$$\text{Let } g_N(x) = c_0 + \sum_{k=1}^N c_k \cos(kx) + d_k \sin(kx)$$

where  $c_k$ 's &  $d_k$ 's are Fourier coeffs of  $g(x)$

From the Theorem above, we can uniformly approximate  $g(x)$  by  $g_N(x)$ .

By choosing a large  $n > N_0$ , we can set  $|g(x) - g_n(x)| < \varepsilon$  for all  $x \in [-\pi, \pi]$

$$\begin{aligned} \text{Now, } \|g - g_n\|_{L^2}^2 &= \int_{-\pi}^{\pi} |g(x) - g_n(x)|^2 dx \\ &< \int_{-\pi}^{\pi} \varepsilon^2 dx = 2\pi \varepsilon^2 \\ &\left( \because n > N_0 \right) \end{aligned}$$

$$\therefore \|g - g_n\| = \varepsilon \sqrt{2\pi}$$

Consider  $\|f - g_N\|$

$$\begin{aligned} \|f - g_N\| &= \|f - g + g - g_N\| \\ &\leq \|f - g\| + \|g - g_N\| \quad (\because \text{Triangle inequality}) \\ &< \varepsilon + \varepsilon \sqrt{2\pi} \quad \text{for } N > N_0 \end{aligned}$$

Now,  $g_N(x)$  is a linear combination of  $\sin(kx)$  &  $\cos(kx)$  for values  $k = 1, \dots, N$   
 $\therefore g_N(x) \in V_N$

From our earlier lemma,  $f_N$  is the closest element from  $V_N$  to  $f$  in  $L^2$ .

$$\therefore \|f - f_N\| \leq \|f - g_N\| < \varepsilon (1 + \sqrt{2\pi}) \quad \text{for } N > N_0$$

$\Rightarrow$  Given a tolerance  $\varepsilon$ , we can get as close as we can to the original  $f \in L^2$  using the Fourier expansion.

## Matrix Calculus

For example:

$$\alpha = x^T x$$

(  $x^T$  is a  $1 \times n$  row vector  
 $x$  is a  $n \times 1$  column vector

We may  
need

$$\frac{d\alpha}{dx}$$

We can differentiate matrix quantities w.r.t. the elements  
in the reference matrices.  
Using the Jacobian matrix



Let us start with an example.

Consider  $\underline{y} = A \underline{x}$ ; Suppose  $A$  is a  $2 \times 2$  matrix  
and  $\underline{x}$  is a column vector  $2 \times 1$ .

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} x_1 + a_{12} x_2 \\ a_{21} x_1 + a_{22} x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\frac{\partial \underline{y}}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = A^T$$

(Denominator layout)

$$\frac{d(A\underline{x})}{d\underline{x}} = A^T$$

Consider  $\underline{x}^T A \underline{x}$  ; Let  $\underline{x}$  be a  $2 \times 1$  column vector  
 $\underline{x}^T$  is a  $1 \times 2$  row vector  
 $A$  is a  $2 \times 2$  matrix

Suppose we are interested in  $\frac{d}{d\underline{x}} (\underline{x}^T A \underline{x})$

$$\underbrace{\begin{pmatrix} x_1 & x_2 \end{pmatrix}}_{\underline{x}^T} \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}}_A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = x_1 (a_{11}x_1 + a_{12}x_2) + x_2 (a_{21}x_1 + a_{22}x_2)$$

$$\frac{d}{dx} \underbrace{(x^T A x)}_{\eta} = \begin{bmatrix} \frac{\partial \eta}{\partial x_1} \\ \frac{\partial \eta}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + a_{12}x_2 + a_{21}x_2 \\ a_{12}x_1 + a_{21}x_1 + 2a_{22}x_2 \end{bmatrix}$$

← Rearranging slightly

$$\begin{aligned} & \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ & \equiv A \underline{x} + A^T \underline{x} \\ & \equiv (A + A^T) \underline{x} \end{aligned} \quad \left( \text{If } A \text{ is symmetric i.e., } A = A^T \right)$$

$$\left( \text{If } A \text{ is symmetric i.e., } A = A^T \right) \frac{d}{dx} (x^T A x) = 2 A x$$

Consider another example

$$\frac{\partial}{\partial \underline{x}} (\underline{u}^T \cdot \underline{v})$$

$$\eta = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\eta = u_1 v_1 + u_2 v_2$$

$$\frac{\partial \eta}{\partial \underline{x}} = \begin{bmatrix} u_1 \frac{\partial v_1}{\partial x_1} + v_1 \frac{\partial u_1}{\partial x_1} + u_2 \frac{\partial v_2}{\partial x_1} + v_2 \frac{\partial u_2}{\partial x_1} \\ u_1 \frac{\partial v_1}{\partial x_2} + v_1 \frac{\partial u_1}{\partial x_2} + u_2 \frac{\partial v_2}{\partial x_2} + v_2 \frac{\partial u_2}{\partial x_2} \end{bmatrix}$$

let us rearrange the terms

$$\begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_2}{\partial x_1} \\ \frac{\partial u_1}{\partial x_2} & \frac{\partial u_2}{\partial x_2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_2}{\partial x_1} \\ \frac{\partial v_1}{\partial x_2} & \frac{\partial v_2}{\partial x_2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$= \frac{\partial u_i}{\partial x_j} \cdot v_j + \frac{\partial v_j}{\partial x_i} \cdot u_i$$

Summarize some rules of matrix differentiation

<u>Quantity</u>	<u>Result</u>
1) $a$ is not a function of $\underline{x}$ $\frac{\partial a}{\partial \underline{x}}$	0
2) a) $\frac{\partial (A\underline{x})}{\partial \underline{x}}$	$A^T$
b) $\frac{\partial (\underline{x}^T A)}{\partial \underline{x}}$	$A$
3) $\frac{\partial (A\underline{u})}{\partial \underline{x}}$	$\frac{\partial \underline{u}}{\partial \underline{x}} \cdot A^T$
4) $\frac{\partial (\underline{x}^T A \underline{x})}{\partial \underline{x}}$	$(A + A^T) \underline{x} = 2 A \underline{x}$ if $A = A^T$ (i.e., for symmetric matrices)

## Karhunen Loeve Transform (KLT)

Many transforms we are familiar with have been signal independent  
i.e., we have a broad/generic framework to handle any signal  
with some properties associated with them e.g. Convergence etc.  
such that the transformation yields

- a) Energy compaction
- b) better analysis/signal properties in the transformed domain

- Qn: Is there a transform that can (Linear transform)
- a) yield energy compaction
  - b) decorrelate data
  - c) unitary  $(A A^T = I)$
  - d) data dependent ?  $(\text{Statistics of the data must play a role})$

Kari Karhunen & Michel Loeve (1948)  
(1947)

Named after inventors Karhunen & Loeve



Let us imagine vectors  $\underline{x} \in \mathbb{R}^N$  with a p.d.f. (column vectors)  
i.e.,  $N$  dimensional vectors

The covariance  $\Sigma_x$  of  $\underline{x}$  is  
$$\Sigma_x = E \left( (\underline{x} - \underline{\mu}_x) (\underline{x} - \underline{\mu}_x)^T \right)$$

Now, suppose we consider a linear transformation of  $\underline{x}$   
by  $A$ . i.e.,  $\underline{y} = A \underline{x} \Rightarrow E(\underline{y}) = E(A \underline{x})$   
$$\underline{\mu}_y = A \underline{\mu}_x$$

$$\Sigma_y = E \left( (\underline{y} - \underline{\mu}_y) (\underline{y} - \underline{\mu}_y)^T \right)$$

$$\begin{aligned}
\Sigma_y &= E \left( (A \underline{x} - A \underline{\mu}_x) (A \underline{x} - A \underline{\mu}_x)^T \right) \\
&= A E \left( (\underline{x} - \underline{\mu}_x) (\underline{x} - \underline{\mu}_x)^T \right) A^T \quad ( \because (AB)^T = B^T A^T ) \\
&= A \Sigma_x A^T
\end{aligned}$$

GOALS :      What we may need

- 1) We may want  $\underline{y}$  to be de correlated i.e.,  $\Sigma_y$  is a diagonal matrix, say  $\Lambda$ .
- 2) We still need energy compaction i.e., place energy of the signal non-uniformly i.e., from high to low over the signal dimensions.

Let  $\psi$  be a unitary transformation matrix.

For goal 1: We can achieve diagonalization if

$\underline{y} - \underline{\mu}_y$  is linearly related to  $\underline{x} - \underline{\mu}_x$ .

Suppose  $\underline{y} - \underline{\mu}_y = \psi^{-1} (\underline{x} - \underline{\mu}_x)$

Let us choose  $\psi$  to comprise of eigen vectors of  $Z_x$ .

$$\Sigma_x \psi = \psi \Lambda$$

(A)

$$\left( \because \sum_{i=1}^N \lambda_i \psi_i = \lambda_i \psi_i \right)$$
$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$$

$$\Sigma_y = \Psi^{-1} \Sigma_x \Psi.$$

We are able to achieve our goal  
of forcing  $\Sigma_y = \Lambda$  by an  
appropriate transformation  $A = \Psi^{-1}$

Consider  $E((\underline{y} - \underline{\mu}_y)(\underline{y} - \underline{\mu}_y)^T)$   
Since  $\underline{y} - \underline{\mu}_y = \Psi^{-1}(\underline{x} - \underline{\mu}_x)$   
=  $E(\Psi^{-1}(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T \Psi)$   
=  $\Psi^{-1} E(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T \Psi$

Property 1:

For any real matrix  $A$ , eigen vectors are  
symmetric  
orthogonal if eigen values are distinct.

Proof:

Consider  $\langle A \underline{x}, \underline{y} \rangle$  for vectors  $\underline{x}$  and  $\underline{y}$   
that are eigen vectors.

Since  $A$  is symmetric,  $A = A^T$

$$\langle A \underline{x}, \underline{y} \rangle = \underbrace{(A \underline{x})^T \underline{y}}_{\underline{x}^T A^T \underline{y}} \quad \textcircled{I}$$

Eqn  $\textcircled{I}$  can be seen slightly differently as

$$\langle \underline{x}, A^T \underline{y} \rangle = \langle \underline{x}, A \underline{y} \rangle \quad \textcircled{II}$$

$$\text{Let } A_{\underline{x}} = \lambda \underline{x} \quad \& \quad A_{\underline{y}} = \mu \underline{y}$$

$$\langle A_{\underline{x}}, \underline{y} \rangle = \lambda \langle \underline{x}, \underline{y} \rangle$$

$$\langle \underline{x}, A_{\underline{y}} \rangle = \mu \langle \underline{x}, \underline{y} \rangle$$

$$\text{But, } \lambda \neq \mu \implies \langle \underline{x}, \underline{y} \rangle = 0.$$

$$\implies \underline{x} \perp \underline{y}.$$



Let us check if the energy conservation holds.

$$\text{Energy in } \underline{x} = E_x = E \left( (\underline{x} - \underline{\mu}_x)^T (\underline{x} - \underline{\mu}_x) \right)$$

Consider energy in  $\underline{y}$ .

$$E_y = E \left( (\underline{y} - \underline{\mu}_y)^T (\underline{y} - \underline{\mu}_y) \right)$$

With  $\underline{y} = \Psi^{-1} \underline{x}$  where  $\Psi^{-1} : \Psi^{-1} = \Psi^T$

$$E_y = E \left( (\underline{x} - \underline{\mu}_x)^T \underbrace{(\Psi^{-1})^T \Psi^{-1}}_{\mathbf{I}} (\underline{x} - \underline{\mu}_x) \right)$$

$$E_y = E \left( (\underline{x} - \underline{\mu}_x)^T (\underline{x} - \underline{\mu}_x) \right) = E_x \quad (\text{Energy is conserved!})$$

## MOTIVATION

Suppose we want to COMPACT energy within the first  $Q$  components of  $\underline{y}$ , can we construct a transformation that does this.

To be more precise, suppose  $A$  is a linear transform

such that  $\underline{y} = A \underline{x}$ .

$$A := \begin{bmatrix} \underline{a}_0 & \underline{a}_1 & \dots & \underline{a}_{n-1} \end{bmatrix}_{n \times n}^T \quad \text{where}$$

$\underline{a}_n$  is a  $n \times 1$  column vector

$$A^H := \begin{bmatrix} \underline{a}_0^* & \underline{a}_1^* & \dots & \underline{a}_{n-1}^* \end{bmatrix}$$



To consider energy in the first  $Q$  components of  $-y$ ,  
 let us null out  $k > Q$  components in  $A \in A^H$ .

$$A_Q := \begin{bmatrix} \underline{a}_0 & \underline{a}_1 & \dots & \underline{a}_{Q-1} & 0 & 0 & 0 & \dots & 0 \end{bmatrix}^T$$

$$A_Q^H := \begin{bmatrix} \underline{a}_0^* & \underline{a}_1^* & \dots & \underline{a}_{Q-1}^* & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

We need to maximize

$$E_y^{(Q)} = E \left( \left( \underline{x} - \underline{\mu}_x \right)^H A_Q^H A_Q \left( \underline{x} - \underline{\mu}_x \right) \right)$$

Subject to:  $\underline{a}_k^* \underline{a}_k = 1$ ;  $\underline{a}_k^* \underline{a}_l = 0$  ( $k \neq l$ )

We can form a Lagrange multiplier within an optimization framework.

$$J = \max_{\left\{ \underline{a}_k^* \right\}_{k=0}^{Q-1}} E_y^{(Q)} + \sum_{k=0}^{Q-1} \lambda_k \left( 1 - \underline{a}_k^{*T} \underline{a}_k = 0 \right)$$

To simplify, let us compute

$$A_Q^H A_Q$$

$$= \sum_{k=0}^{Q-1} \begin{bmatrix} \underline{a}_k^* & \underline{a}_k^T \end{bmatrix}$$

$$\begin{bmatrix} \underline{a}_{Q-1}^* & 1 & 0 & \dots & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}^T$$

Now,

$$J = \max_{\left\{ \underline{a}_k^* \right\}_{k=0}^{Q-1}} E \left( \left( \underline{x} - \underline{\mu}_x \right)^H \sum_{k=0}^{Q-1} \underline{a}_k^* \underline{a}_k^T \left( \underline{x} - \underline{\mu}_x \right) \right) + \sum_{k=0}^{Q-1} \lambda_k \left( 1 - \underline{a}_k^{*T} \underline{a}_k = 0 \right) \quad \textcircled{1}$$

Let us rearrange  $\textcircled{1}$  a bit

Consider

$$= \sum_{k=0}^{Q-1} \underbrace{\left( \underline{x} - \underline{\mu}_x \right)^H \underline{a}_k^*}_{\text{scalar}} \underbrace{\underline{a}_k^T \left( \underline{x} - \underline{\mu}_x \right)}_{\text{scalar}}$$

$$= \sum_{k=0}^{Q-1} \underline{a}_{-k}^T (\underline{x} - \underline{\mu}_x) (\underline{x} - \underline{\mu}_x)^H \underline{a}_{-k}^* \quad (2)$$

Using (2) in (1),

$$J = \max_{\left\{ \underline{a}_{-k}^* \right\}_{k=0}^{Q-1}} \left[ \sum_{k=0}^{Q-1} \underline{a}_{-k}^T (\underline{x} - \underline{\mu}_x) (\underline{x} - \underline{\mu}_x)^H \underline{a}_{-k}^* + \sum_{k=0}^{Q-1} \lambda_k (1 - \underline{a}_{-k}^{*T} \underline{a}_{-k} = 0) \right] \quad (3)$$

Let us simplify (3).

$$J = \max_{\left\{ \underline{a}_k^* \right\}_{k=0}^{Q-1}} \sum_{k=0}^{Q-1} \left[ \underline{a}_k^T \sum_x \underline{a}_k^* + \lambda_k \left( 1 - \underline{a}_k^{*T} \underline{a}_k \right) \right] \quad (4)$$

where  $\sum_x = E \left( (x - \underline{\mu}_x) (x - \underline{\mu}_x)^H \right)$

To solve for (4),  $\frac{\partial J}{\partial \underline{a}_k^*} = 0 \quad \forall k = 0, 1, \dots, Q-1$

$$\frac{d}{dx} (\underline{y}^T \underline{x}) = \underline{y}$$

$$\frac{\partial J}{\partial \underline{a}_k^*} = 0 \Rightarrow \sum_x \underline{a}_k - \lambda_k \underline{a}_k = 0$$

This is our EIGEN VALUE EQN.

## Applications of KL

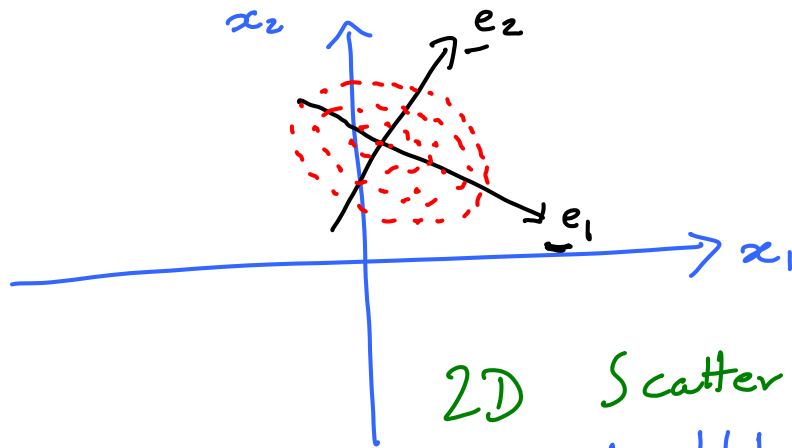
## Dimensionality Reduction

MOTIVATION : Often we find data in higher dimensions.

Each dimension can be attributed to an independent coordinate.

Qns:

- 1) Can we reduce the dimensionality of the data by trading off the reduction in dimensionality to reconstruction
- 2) We still need a linear transformation to accomplish this.



Suppose we need the 2D scatter plot to be collapsed to a line,  
 Intuition tells us to go in the direction of  
 $\underline{e}_1$  i.e., project all the points  $\langle \underline{x}, \underline{e}_1 \rangle$  & this is good enough

## PROCEDURE

---

Given :  $N$  vectors of dimension ' $d$ ' i.e.,  $\mathbb{R}^d$ .

Step 1 : Stack the vectors in rows to form a data matrix

$$D = \begin{bmatrix} \underline{x_1} \\ \underline{x_2} \\ \vdots \\ \underline{x_N} \end{bmatrix}_{N \times d}$$

Step 2 : Compute the mean  $\underline{\mu}_D$

$$\underline{\mu}_D = \frac{1}{N} \sum_{i=1}^N \underline{x_i}$$



There is a mean in each dimension  $c_k'$ :

$$\mu_k = \frac{1}{N} \sum_{i=1}^N d_{ik} \quad \text{where } D := [d_{ij}]$$

Step 3 : Compute  $D - \mu_D$  and form a  
Covariance matrix

$$C = (D - \mu_D)^T (D - \mu_D)$$

Step 4 : Do an eigen decomposition on  $C$   
Let  $[\underbrace{\{\underline{v}_i\}}_{\text{eigen vect.}}, \underbrace{\{\lambda_i\}}_{\text{eigen values}}] = \text{eig}(C)$

Step 5: Sort the eigen values in descending order.

---

Step 6: To retain  $k < d$  dimensions,  
Store  $\{v_i\}_{i=1}^k$  and  $\{\lambda_i\}_{i=1}^k$   
These are the 'dominant' eigen values  
and eigen vectors

Step 7: To get a representation of  $\underline{x} \in \mathbb{R}^d$   
 in  $\mathbb{R}^k$ , project & obtain  
 $\left\{ \langle \underline{x}, \underline{v}_j \rangle \right\}_{j=1}^k$  Components

Example: Suppose i/p vectors are  $[1, 1]$ ,  $[1, -1]$ ,  $[-1, 1]$ ,  $[-1, -1]$

$D = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}$ 
 $\mu_D = [0 \ 0]$ ;
 $C = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

$\lambda_1 = \lambda_2 = 4$ ;
 $\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 
 $\underline{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Both are dominant directions & energy is the same in both!

