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

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

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 are the Fourier coeffts. of f.

Then, $\| f_N - f \|_{L^2} \xrightarrow[N \to \infty]{} 0$

Subtle Issues: Points of discontinuities in f & possible extensions to over come these.

Theorem: If a sequence f_n converges uniformly to f as $n \to \infty$ 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

$$\|f_n(t) - f(t)\|_{L^2}^2 = \int_a^b |f_n(t) - f(t)|^2 \, dt \leq \int_a^b \varepsilon^2 \, dt = \varepsilon^2 (b-a)$$

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

Since ε can be chosen as small as desired,

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

$\blacksquare$

Ponder: Examine if Conv. in $\mathcal{L}^2 \implies$ 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 2π 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 $\{\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_N\}$, the orthogonal projection of $\bar{v} \in V$ onto V_0 is

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

Using Lemma above, f_N is the orthogonal projection of f onto space V_N.

$\Longrightarrow$ f_N is the element in V_N closest to f in L^2 sense

$$\therefore \quad \| f - f_N \|_{L^2} = \min_{g \in V_N} \| f - g \|_{L^2}$$

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_ks & b_ks are the Fourier coeffts. of f.

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

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

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.

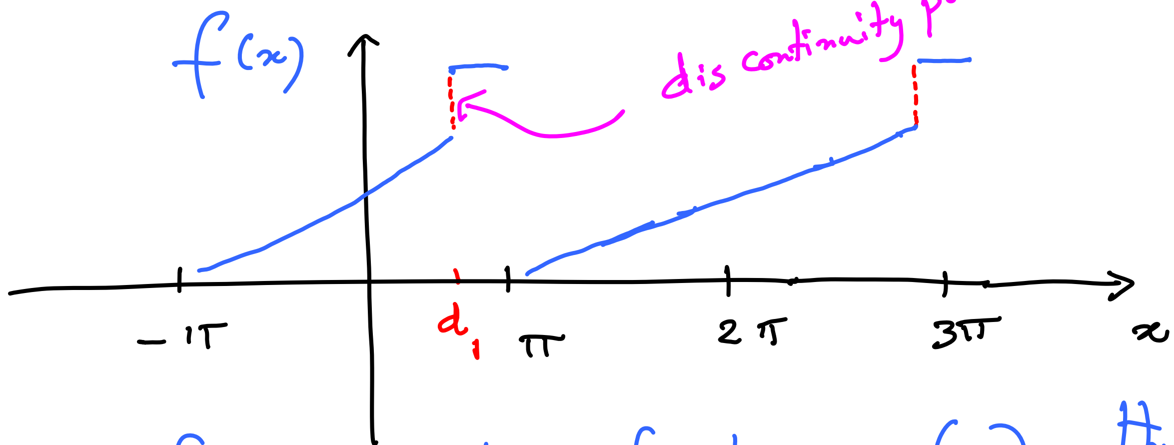
2) Approximate g uniformly (& therefore in L^2) by its Fourier series expansion.

We need to get an idea towards (1).

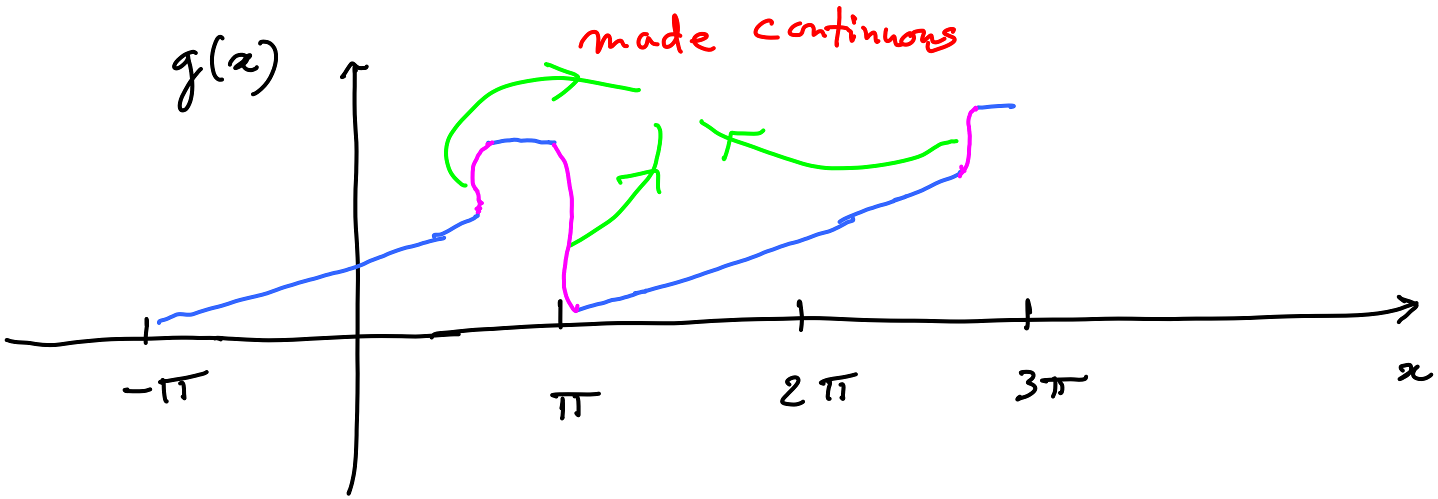
We have already established (2).

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 2π periodic function $f(x)$ converges uniformly to $f(x)$ on $[-\pi, \pi]$

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 coeffts of $g(x)$

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

By choosing a large N_0, we can set

$$|g(x) - g_N(x)| < \varepsilon \quad \text{for all} \quad x \in [-\pi, \pi]$$

Now, $$\|g - g_N\|_{L^2}^2 = \int_{-\pi}^{\pi} |g(x) - g_N(x)|^2 \, dx$$

$$\leq \int_{-\pi}^{\pi} \varepsilon^2 \, dx = 2\pi \varepsilon^2$$

$$(\because N > N_0)$$

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

Consider $\| f - g_N \|$

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

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

$\therefore \; g_N(x) \in V_N$

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

$$\therefore \; \| f - f_N \| \leq \| f - g_N \| < \varepsilon\left(1 + \sqrt{2\pi}\right)$$

for $N > N_0$

$\Rightarrow$ Given a tolerance ε, we can get as closely 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 $n \times 1$ colm vector

We may need $\dfrac{d\alpha}{dx}$

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 2×2 matrix & $\underline{x}$ is a column vector 2×1.

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

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

(Denominator layout)

Consider $\underline{x}^T A \underline{x}$; Let $\underline{x}$ be a 2×1 column vector

$\underline{x}^T$ is a 1×2 row vector

A is a 2×2 matrix

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

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

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

$$\frac{d\,(\underbrace{\mathbf{x}^T A \mathbf{x}}_{\eta})}{d\underline{x}} = \begin{bmatrix} \dfrac{\partial \eta}{\partial x_1} \\[2ex] \dfrac{\partial \eta}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + a_{12}x_2 + a_{21}x_2 \\[2ex] a_{12}x_1 + a_{21}x_1 + 2a_{22}x_2 \end{bmatrix}$$

Rearranging slightly

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

$$= A\underline{x} + A^T\underline{x}$$

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

(If A is symmetric i.e., $A = A^T$, then $\frac{d}{d\underline{x}}(\underline{x}^T A \underline{x}) = 2A\underline{x}$)

Consider another example

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

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

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

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

Let us rearrange the terms

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

$$= \frac{\partial \underline{u}}{\partial \underline{x}} \cdot v + \frac{\partial V}{\partial x} \cdot u$$

Summarize some rules of matrix differentiation

Quantity	Result
1) a is not a function of $\underline{x}$: $\dfrac{\partial a}{\partial \underline{x}}$	0
2) a) $\dfrac{\partial}{\partial \underline{x}}(A\underline{x})$	A^T
b) $\dfrac{\partial}{\partial \underline{x}}(\underline{x}^T A)$	A
3) $\dfrac{\partial}{\partial \underline{x}}(A\underline{u})$	$\dfrac{\partial \underline{u}}{\partial \underline{x}} \cdot A^T$
4) $\dfrac{\partial}{\partial \underline{x}}(\underline{x}^T A \underline{x})$	$(A + A^T)\underline{x} = 2A\underline{x}$ 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? (Statistics of the data must play a role)

Kari Karhunen (1947) & Michel Loeve (1948)

Named after inventors Karhunen & Loeve

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

The covariance Σ_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., $\underline{y} = A\,\underline{x}$ $\Rightarrow$ $E(\underline{y}) = E(A\,\underline{x})$

$$\Rightarrow \underline{\mu}_y = A\,\underline{\mu}_x$$

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

$$\Sigma_y = E\left(\left(A\underline{x} - A\underline{\mu}_x\right)\left(A\underline{x} - A\underline{\mu}_x\right)^T\right)$$

$$= A\,E\left(\left(\underline{x} - \underline{\mu}_x\right)\left(\underline{x} - \underline{\mu}_x\right)^T\right)A^T \qquad \left(\because (AB)^T = B^T A^T\right)$$

$$= A\,\Sigma_x\,A^T$$

$$=$$

GOALS : What we may need

1) We may want $\underline{y}$ to be decorrelated i.e., Σ_y is a diagonal matrix, say Λ.

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 Ψ 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 Ψ to comprise of eigen vectors of Σ_x.

$$\Sigma_x \Psi = \Psi \Lambda \qquad \text{A}$$

$$\left(\because \ \Sigma_x \psi_i = \lambda_i \psi_i \right)$$

$i = 1, 2, \cdots, N$

$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N)$

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

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

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

Since $y - \mu_y = \psi^{-1}(x - \mu_x)$

$= E\left(\psi^{-1}(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T \psi\right)$

$= \psi^{-1} E\left(\underline{x} - \underline{\mu}_x\right)(x - \mu_x)^T\psi$

Property 1 : For any real & symmetric matrix A, eigen vectors are 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, $A = A^T$

$$\langle A\underline{x}, \underline{y}\rangle = (A\underline{x})^T \underline{y} = \underline{x}^T A^T \underline{y} \quad \text{——— (I)}$$

Eqn (I) can be seen slightly differently as

$$\langle \underline{x}, A^T \underline{y}\rangle = \langle \underline{x}, A\underline{y}\rangle \quad \text{——— (II)}$$

Let $A\underline{x} = \lambda\underline{x}$ & $A\underline{y} = \mu\underline{y}$

$$\left.\begin{aligned} \langle A\underline{x}, \underline{y}\rangle &= \lambda\langle \underline{x}, \underline{y}\rangle \\ \langle \underline{x}, A\underline{y}\rangle &= \mu\langle \underline{x}, \underline{y}\rangle \end{aligned}\right\} \text{(iii)}$$

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

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

$\blacksquare$

Let us check if the energy conservation holds.

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 ψ^{-1} : $\psi^{-1} = \psi^T$

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

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

(Energy is conserved!)

MOTIVATION

Suppose we want to $\underline{\text{COMPACT}}$ 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 $\underline{y} = A\underline{x}$.

$$A := \left[\underline{a}_0 \ \ \underline{a}_1 \ \cdots \ \underline{a}_{n-1}\right]^T_{n \times n}$$ where

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

$$A^H := \left[\underline{a}_0^* \ \ \underline{a}^* \ \cdots \ \underline{a}^*_{n-1}\right]$$

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

$$A_Q := \left[\underline{a}_0 \ \underline{a}_1 \ \cdots \ \underline{a}_{Q-1} \ \underline{0} \ \underline{0} \ \underline{0} \ \cdots \ \underline{0}\right]^T$$

$$A_Q^H := \left[\underline{a}_0^* \ \underline{a}_1^* \ \cdots \ \underline{a}_{Q-1}^* \ \underline{0} \ \underline{0} \ \underline{0} \ \cdots \ \underline{0}\right]$$

We need to maximize

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

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

We can form a Lagrange multiplier within an optimization framework.

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

To simplify, let us compute

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

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

Now,

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

Let us rearrange (1) a bit

Consider $(\underline{x} - \underline{\mu}_x)^H A_Q^H A_Q (\underline{x} - \underline{\mu}_x)$

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

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

Using (2) in (1),

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

Let us simplify (3).

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

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

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

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

$$\frac{\partial J}{\partial a_k^*} = 0 \implies \Sigma_x \underline{a}_k - \lambda_k \underline{a}_k = 0$$

THIS IS OUR <u>EIGEN VALUE EQN.</u>

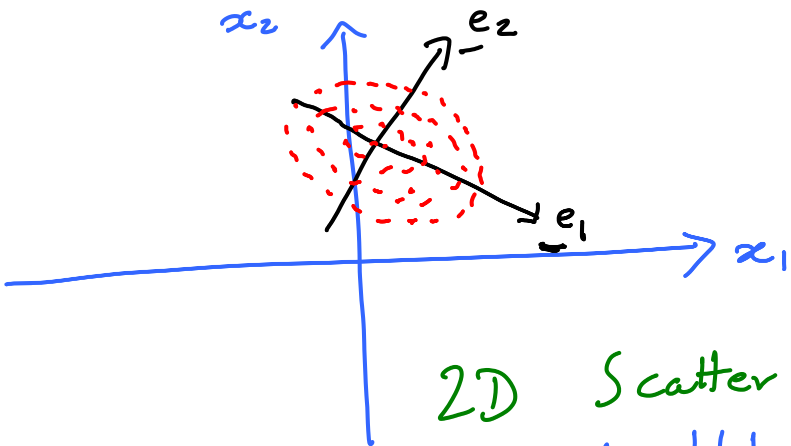
Applications of KL   Dimensionality Reduction

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.



2D Scatter Plot

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 'k':

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

<u>Step 3</u> : Compute $D - \mu_D$ and form a covariance matrix

$$C = (D - \underline{\mu}_D)^T (D - \underline{\mu}_D)$$

<u>Step 4</u> : Do an eigendecomposition on C

Let $[\{\underline{v}\}, \{\lambda\}] = \text{eig}(C)$

eigenvect. eigen values

Step 5 : Sort the eigen values in descending order.

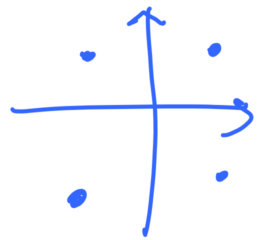
Step 6 : To retain $k < d$ dimensions,

Store $\{ \underline{v}_i \}_{i=1}^{k}$ 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

$$\left\{ \langle \underline{x}, \underline{v}_j \rangle \right\}_{j=1}^{k} \quad \text{components}$$



Example: Suppose i/p vectors are

$[1, 1]$, $[1, -1]$, $[-1, 1]$, $[-1, -1]$

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

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

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