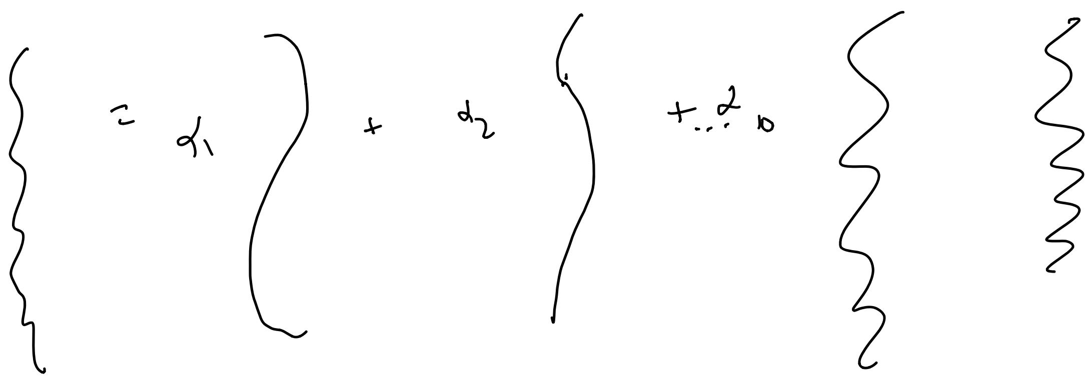
$$\underline{x} = \alpha_1 \underline{u}^{(1)} + \alpha_2 u^{(2)} + \cdots + \alpha_N u^{(n)}$$

$$[x]_{\beta} = \underline{\alpha}$$

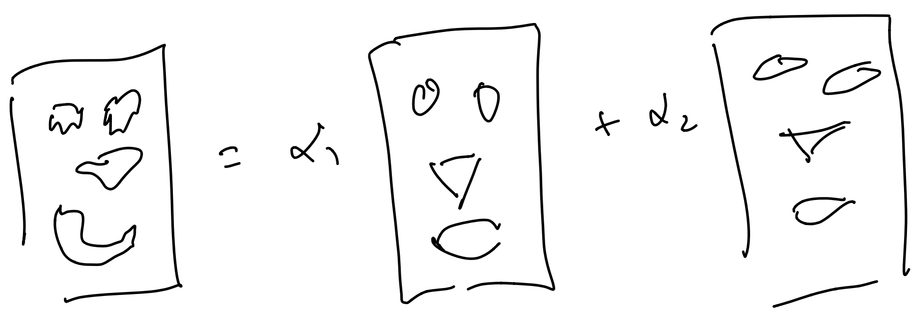
we write $\underline{x} \leftrightarrow \alpha$ are isomorphic

$\underline{x} = U\alpha$, U serves an invertible mapping

e.g. 1) Fourier modes decomp into waves



e.g. 2)



data driven.

notation:

inner product: $x, y \in \mathbb{R}^N$

$$\langle x, y \rangle = y^T x \in \mathbb{R}$$

$$= \sum_{i=1}^{N} y_i \cdot x_i$$

$$= x^T y$$

$x, y \in \mathbb{C}$

$$\langle x, y \rangle = y^* x \in \mathbb{C}$$

$$= (\bar{y}_1 \quad \bar{y}_N) \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

$$= \sum x_i \bar{y}_i.$$

Also, $\|x\|^2 = \langle x, x \rangle$.

Thm Let V be an inner product space and $\beta = \{u^{(1)}, \dots, u^{(n)}\}$ be a o.n. basis.

$$x = \sum_{i=1}^{n} \langle x, u^{(i)} \rangle u^{(i)}$$

pf: $x = \sum_{1}^{n} \alpha_i u^{(i)}$

$$\langle x, u^{(j)} \rangle = \left\langle \sum \alpha_i u^{(i)}, u^{(j)} \right\rangle$$

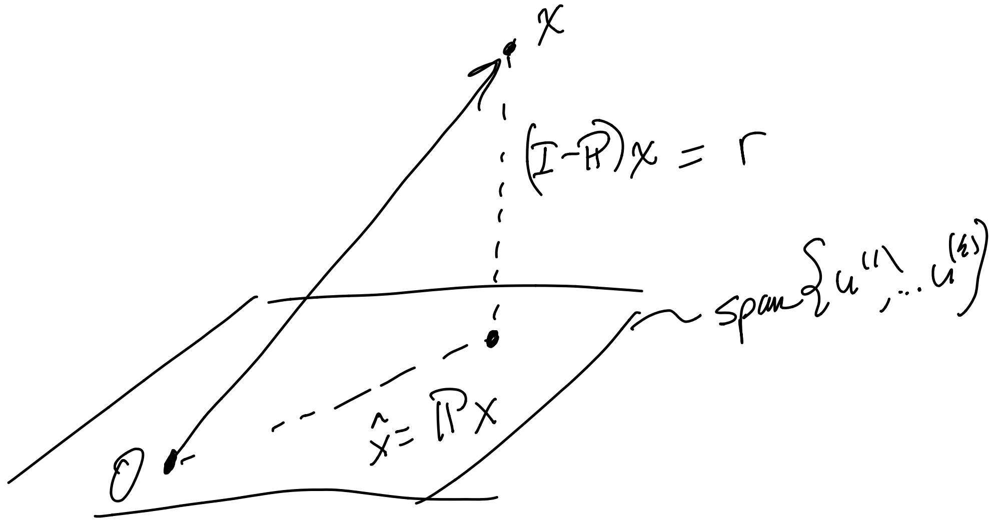
$$= \sum \alpha_i \langle u^{(i)}, u^{(j)} \rangle$$

$$= \alpha_j$$

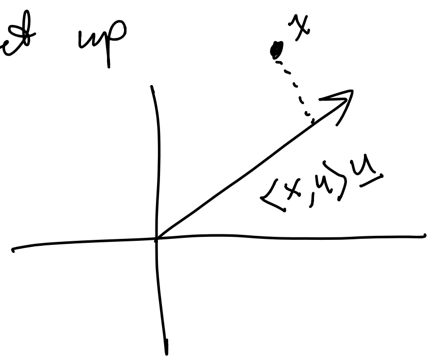
recall $x = \mathbb{P}x + (\mathbb{I} - \mathbb{P})x$.

$$x = \sum_{i=1}^{k} \langle x, u^{(i)} \rangle u^{(i)} + \sum_{i=k+1}^{n} \langle x, u^{(i)} \rangle u^{(i)}$$

$$= \underbrace{\sum u^{(i)} u^{(i)^T}}_{\mathbb{P}} x \quad +$$



1D set up



$$\hat{x} = \langle x, u \rangle u$$

maximize the length of the projected vectors ($\|\hat{x}\| \leq \|x\|$).

Data set $\{x^{(\mu)}\}$.

$$\underset{\substack{u \\ \|u\|=1}}{\text{maximize}} \sum_{\mu=1}^{P} \|\hat{x}^{(\mu)}\|^2$$

$$\|\hat{x}\|^2 = \langle \hat{x}, \hat{x} \rangle$$

$$= \langle \langle x,u\rangle u, \langle x,u\rangle u\rangle$$

$$= \langle x,u\rangle \overline{\langle x,u\rangle} \langle u,u\rangle$$

$$= |\langle x,u\rangle|^2 \in \mathbb{R}.$$ (so it makes sense to maximize.

rewriting:

$$= \langle x,u\rangle \langle u,x\rangle$$

$$= u^* x x^* u$$

$$\underset{\substack{u \\ \langle u,u\rangle=1}}{\text{maximize}} \frac{1}{P}\sum_{\mu=1}^{P} u^* x^{(\mu)} x^{(\mu)*} u$$

$$\max_{\substack{u \\ \|u\|=1}} u^* \underbrace{\left(\frac{1}{M}\sum_m x^{(m)} x^{(m)^*}\right)}_{C} u.$$

symmetric

Observation: i) $C \in \mathbb{R}^{N \times N} \Rightarrow C = C^T$

ii) $C \in \mathbb{C}^{N \times N} \Rightarrow C = C^*$

self adjoint / Hermetian

now define Lagrangian.

$$L(\underline{u}, \lambda) = u^* C u - \lambda (u^* u - 1)$$

$$\frac{\partial L}{\partial \underline{u}} = 2Cu - 2\lambda u = 0$$

$$\frac{\partial L}{\partial \lambda} = u^* u - 1 = 0.$$

$$Cu = \lambda u \quad , \quad u^* u = 1.$$

Properties of Eigenvectors, Eigenvalues of Hermitian Matrices.

1) Eigenvectors associated with distinct eigenvalues are orthogonal. (true for normal operators)

2) Eigenvalues of C are real.

3) If V is a real inner product space then the char poly splits. Hence Symmetric matrices are not defective geometric multiplicity = algebraic

4) In general eigenvalues of Hermitian matrices may be negative. Our C is special, i.e., non-neg. definite.

Alternative view:

$$\|x\|^2 = \|\hat{x}^{(\mu)}\|^2 + \|r^{(\mu)}\|^2$$

If we max $\|\hat{x}\|^2$ we min $\|r\|^2$.

$$r^{(\mu)} = \underline{x}^{(\mu)} - \underline{\hat{x}}^{(\mu)}$$

$$= \sum_{i=k+1}^{n} \alpha_i^{(\mu)} u^{(i)}$$

$$\text{mean square error} = \frac{1}{P}\sum_{\mu=1}^{P} \|r^{(\mu)}\|^2$$

$$= \frac{1}{P}\sum_{\mu} \langle r^{(\mu)}, r^{(\mu)} \rangle$$

$$= \frac{1}{P}\sum_{\mu} \langle \sum_{i=k+1}^{n} \alpha_i^{(\mu)} u^{(i)}, \sum_{j} \alpha_j^{(\mu)} u^{(j)} \rangle$$

$$= \frac{1}{P}\sum_{\mu} \sum_{ij} \alpha_i \overline{\alpha_j} \langle u^{(i)}, u^{(j)} \rangle$$

$$= \frac{1}{P} \sum_{\mu} \sum_{i=k+1}^{n} |\alpha_i|^2$$

Now the Lagrangian looks like:

$$L(u^{(k+1)}, \dots, u^{(n)}, \lambda_{k+1}, \dots \lambda_n) = \frac{1}{P} \sum_{\mu} \sum_{i=k+1}^{n} \left(\langle x^{(\mu)}, u^{(i)} \rangle \right)^2 - \sum \lambda_i \left(\langle u^{(i)}, u^{(i)} \rangle - 1 \right)$$

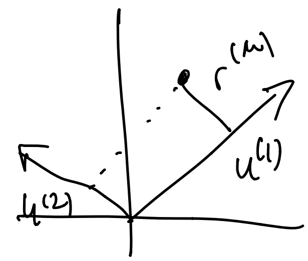
$$= \frac{1}{P} \sum_{i} u^{(i)*} C u^{(i)}$$

$$- \sum \lambda_i \cdot \left(\langle u^{(i)}, u^{(i)} \rangle - 1 \right)$$

$$\frac{\partial L}{\partial u^{(m)}} = 2 C u^{(m)} - 2 \lambda_m u^{(m)} = 0.$$

$$C u^{(m)} = \lambda_m u^{(m)}!$$

Review:



$$\|r^{(\mu)}\|^2 = \langle r^{(\mu)}, r^{(\mu)} \rangle$$

$$= \left(\langle u^{(2)}, x^{(\mu)} \rangle u^{(2)}, \langle u^{(2)}, x^{(\mu)} \rangle u^{(2)} \right)$$

$$= |\langle u^{(2)}, x^{(\mu)} \rangle|^2$$

$$\min. \sum_{i=k+1}^{N} \sum_{\mu} |\langle u^{(2)}, x^{(\mu)} \rangle|^2$$

$$= \max \sum_{i=1}^{k} \sum_{\mu} | \qquad |$$

For optimality we have:

$$\lambda_i = \sum_{\mu} | \langle u^{(i)}, x^{(\mu)} \rangle |^2$$

$$= \sum_{\mu} \langle u^{(i)}, x^{(\mu)} \rangle \langle x^{(\mu)}, u^{(i)} \rangle$$

$$= \sum_{\mu} u^{(i)*} x^{(\mu)} x^{(\mu)*} u^{(i)}$$

$$= u^{(i)*} C u^{(i)}$$

$$= \lambda_i$$

So, PCA maximizes $\sum_i u^{(i)*} C u^{(i)}$

or, $\leq \sum \lambda_i$

Ordered Basis $\{u^{(1)}, u^{(2)}, \ldots, u^{(n)}\}$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \underline{\underline{\lambda_n \geq 0}}$.

need to show this.

$$\lambda_m = \langle C u^{(m)}, u^{(m)} \rangle$$

Observe $C = \underline{X}\,\underline{X}^* / p$

$$\lambda_m = u^{(m)*}\, \underline{X}\,\underline{X}^*\, u^{(m)} / p.$$

property 1) C is non-negative def.

let $v = \underline{X}^* u^{(m)} / \sqrt{p}$

$$\begin{aligned} \lambda_m &= v^* v \\ &= \langle v, v \rangle \\ &= \|v\|^2 \geq 0 \quad \text{all } m. \end{aligned}$$

Mean Subtraction

$\text{mean}(x) = \frac{1}{P} \sum_{\mu=1}^{1} x^{(\mu)}$

$\tilde{x} = x - \text{mean}(x)$

property. $\text{mean}(\tilde{x}) = 0.$

$$\text{mean}\left(\tilde{x}^{(\mu)}\right) = \text{mean}\left(x - \text{mean}(x)\right)$$

$$= \text{mean}(x) - \text{mean}(x)$$

$$= 0.$$

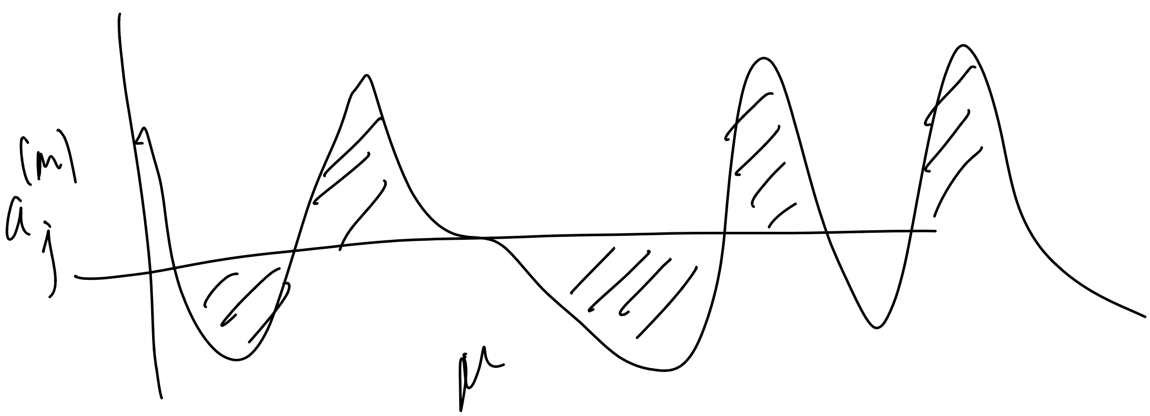
property If $\text{mean}(x^{(\mu)}) = 0$ then $\text{mean}\left(\alpha_j^{(\mu)}\right) = 0$

$$\frac{1}{P} \sum_{\mu} \alpha_j^{(\mu)} = \frac{1}{P} \sum_{\mu} \langle x^{(\mu)}, u^{(j)} \rangle$$

$$= \left\langle \frac{1}{P} \sum_{\mu} x^{(\mu)}, u^{(j)} \right\rangle$$

$$= \left\langle 0, u^{(j)} \right\rangle$$

$$= 0.$$



property 2) The expansion coefs are uncorrelated on average (over data).

$$\alpha_j^{(\mu)} \bar{\alpha}_k^{(\mu)} = \langle x^{(\mu)}, u^{(j)} \rangle \overline{\langle x^{(\mu)}, u^{(k)} \rangle}$$

$$= \langle x^{(\mu)}, u^{(j)} \rangle \langle u^{(k)}, x^{(\mu)} \rangle$$

$$= u^{(j)*} x^{(\mu)} x^{(\mu)*} u^{(k)}$$

$$\frac{1}{P} \sum \alpha_j^{(\mu)} \bar{\alpha}_k^{(\mu)} = u^{(j)*} \left(\frac{1}{P} \sum x^{(\mu)} x^{(\mu)*} \right) u^{(k)}$$

$$= u^{(j)*} C u^{(k)}$$

$$= \lambda_k u^{(j)*} u^{(k)} = \lambda_k \delta_{jk}$$

$$= \begin{cases} 0 & j \neq k \\ \lambda_k & j = k \end{cases}$$

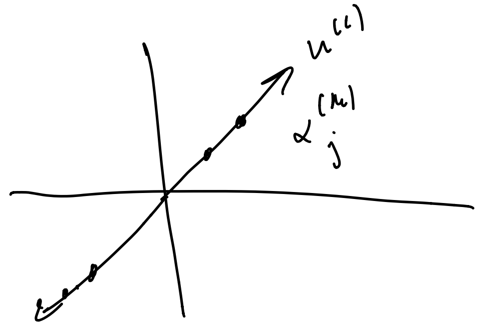
<u>Property</u> Assume mean$(x^{(\mu)}) = 0$. Then

$$\mathrm{Var}\left(\alpha_j^{(\mu)}\right) = \lambda_j .$$

$$\mathrm{Var}\left(\alpha_j^{(\mu)}\right) = \frac{1}{P}\sum_\mu \left(\alpha_j^{(\mu)} - \mathrm{mean}\left(\alpha_j^{(\mu)}\right)\right)^2$$

$$= \frac{1}{P}\sum_\mu \left(\alpha_j^{(\mu)}\right)^2$$ assume mean subtracted so mean$(\alpha_j^{(\mu)}) = 0$.

$$= \lambda_j$$



Shannon's Entropy

$$H = -\sum_{j=1}^{N} \tilde{\rho}_i \ln \tilde{\rho}_i \quad \text{where} \quad \sum \tilde{\rho}_i = 1.$$

Setting: measure quality of a basis in information theoretic terms.

NOT NEC. PCA

let $\beta = \{ v^{(1)}, \ldots, v^{(N)} \}$ be a basis for $\mathbb{R}^N$.

$$x^{(\mu)} = \sum_{j=1}^{N} \alpha_j^{(\mu)} v^{(j)}$$

define $\rho_j = \sum_\mu \left(\alpha_j^{(\mu)} \right)^2$, ie

the variance in the j'th coordinate direction.

$$\tilde{\rho}_j = \rho_j \Big/ \sum_j \rho_j \qquad \tilde{\rho}_1 \geq \tilde{\rho}_2 \geq \ldots$$

H measures the distribution of variance.

H is a minimum $(H=0)$ when

$\tilde{\rho}_1 = 1$, $\tilde{\rho}_i = 0$ $i>1$. (1D)

H is a maximum when $\rho_i = \frac{1}{N}$ all i.

H is a function of the data and the basis.

<u>proposition.</u> PCA is the basis that minimizes H over all o.n. bases

define $K_j = \sum_{k=1}^{j} \tilde{\rho}_k$ variance captured by $\{v^{(1)}, \dots, v^{(N)}\}$.

$K_0 \equiv 0$

$$\sum_{k=1}^{j} \tilde{\rho} \leq \underbrace{\sum_{k=1}^{j} \tilde{\lambda}_k}_{\text{total variance captured by PCA.}}$$

Observe $\quad K_{j+1} - K_j = \sum_{k=1}^{j+1} \tilde{\rho}_k - \sum_{k=1}^{j} \tilde{\rho}_k$

$$= \rho_{j+1} .$$

$$H = - \sum_{j=0}^{N-1} (K_{j+1} - K_j) \ln (K_{j+1} - K_j)$$

$$\frac{\partial H}{\partial K_m} = - \sum_{j=0}^{N-1} \left(\frac{\partial K_{j+1}}{\partial K_m} - \frac{\partial K_j}{\partial K_m} \right) \ln (K_{j+1} - K_j)$$

$$- \sum_{j=0}^{N-1} (K_{j+1} - K_j) \frac{1}{K_{j+1} - K_j} \left[\frac{\partial K_{j+1}}{\partial \alpha_m} - \frac{\partial K_j}{\partial \alpha_m} \right]$$

$$= -\sum_{j=0}^{N-1} \Big[\left(\delta_{j+1,m} - \delta_{j,m}\right) \ln \left(K_{j+1} - K_j\right) + \left(\delta_{j+1,m} - \delta_{j,m}\right)$$

$$= -\sum_{j=0}^{N-1} \left[\delta_{j+1,m} - \delta_{j,m} \right] \left[\ln \left(K_{j+1} - K_j\right) + 1 \right]$$

$$= -\sum_{j=0}^{N-1} \delta_{j+1,m} \left[\ln \left(K_{j+1} - K_j\right) + 1 \right]$$

$$+ \sum_{j=0}^{N-1} \delta_{j,m} \left[\ln \left(K_{j+1} - K_j\right) + 1 \right]$$

$$= \quad m = j+1 \quad - \left[\ln \left(K_m - K_{m-1}\right) + 1 \right]$$

$$+ \ln \left(K_{m+1} - K_m\right) + 1$$

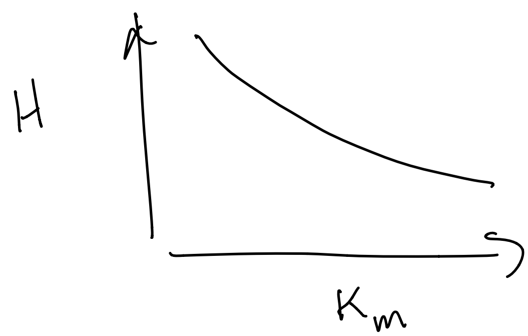
$$= \ln \frac{\left(K_{m+1} - K_m\right)}{K_m - K_{m-1})}$$

$$= \ln \frac{\rho_m}{\rho_{m-1}}$$

but $\rho_m < \rho_{m-1}$

$$\Rightarrow \quad \frac{\rho_m}{\rho_{m-1}} < 1.$$

$$\frac{\partial H}{\partial K_m} < 0. \quad H \text{ decreasing w/ } K_m.$$



$$K_m = \sum_{j=1}^{m} \tilde{\rho}_j$$

the largest K_m comes from PCA
(max variance.)
so H is min for PCA.

recall $\max \sum_{\mu, i} \left(\langle x^{(\mu)}, u^{(i)} \rangle \right)^2$

$$\sum_i u^{(i)*} C u^{(i)}$$

$$= \sum_i \lambda_i$$