Goal this week: SVD, PCA (Review CS132 Geometric Algorithms)

Review linear Algebra.

Vectors and Matrices.

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$$ → reals

is equivalent to a d-dimensional point $(\mathbb{R}^d)$.

⇝ column vector.

$v^T = [v_1, \dots, v_d]$ ⇝ row vector.

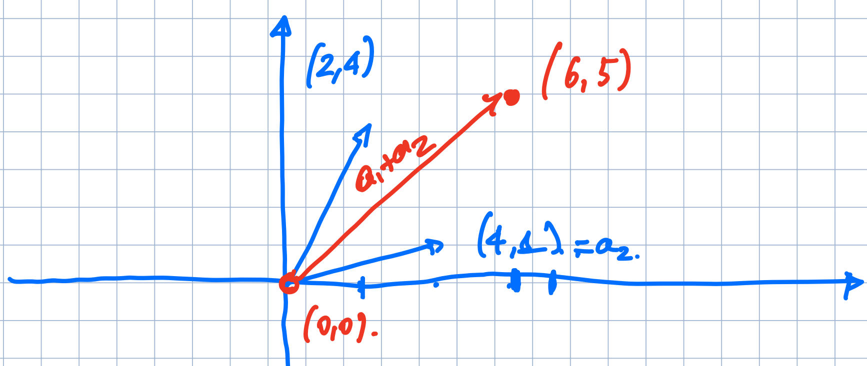
⇒ both will be stored as an array.

$$A^{n\times d} = \begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1d} \\ \vdots & & & \\ A_{21} & A_{22} & \dots & A_{2d} \\ & \vdots & & \\ A_{n1} & A_{n2} & \dots & A_{nd} \end{bmatrix}$$

a_1^T, b_1

$A_{i,j}$ — i: row index, j: col index.

$$A = \begin{bmatrix} | & & | \\ b_1 & \dots & b_d \\ | & & | \end{bmatrix}$$



$a_1 + a_2 = (2+4, 4+1) =$

$= (6, 5)$.

$$\begin{bmatrix} (1 & 2) \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (1,2)\cdot\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ (3,4)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix}$$

$$A \cdot x.$$

↳ dot product perspective of MATRIX VECTOR Multiplication.

"LINEAR COMBINATION of columns" perspective.

$$\begin{bmatrix} \boxed{\begin{matrix} 1 \\ 3 \end{matrix}} & \boxed{\begin{matrix} 2 \\ 4 \end{matrix}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} x_1 \\ 3x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ 4x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix}$$

$$2x_1 + 5x_2 - 7x_3 = 10.$$

$$-x_1 + 8x_2 - 2x_3 = -2.$$

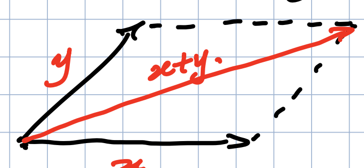
$$\underset{2\times3}{A} \cdot \overset{3\times1}{x} = \overset{2\times1}{b}.$$

$$\underbrace{\begin{bmatrix} (2 & 5 & -7) \\ (-1 & 8 & -2) \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \begin{bmatrix} 10 \\ -2 \end{bmatrix} = b.$$

We are trying to find a linear combination of the three 2dim columns of A such that this linear combination is equal to b.

$x, y \in \mathbb{R}^d$

$$x + y = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_d + y_d \end{bmatrix}.$$



$$\overset{n\times d}{A} + \overset{n\times d}{B} = \overset{n\times d}{C} \quad \text{where} \quad C_{ij} = A_{ij} + B_{ij} \quad \forall\ \begin{matrix} 1\le i\le n \\ 1\le j\le d. \end{matrix}$$

$$\overset{n\times m}{C} = \overset{n\times \bar d}{A} \cdot \overset{\bar d\times m}{B} \quad \text{where} \quad C_{ij} = \sum_{k=1}^{d} A_{ik} B_{kj} = a_i^T b_j$$

see here

$$\begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_i^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} \begin{bmatrix} | & & | & & | \\ b_1 & \cdots & b_j & \cdots & b_m \\ | & & | & & | \end{bmatrix}$$

$5 \cdot 3 = 3 \cdot 5 = 15.$

but $A \cdot B \neq B \cdot A$ (matrix multiplication is not commutative!).

$(AB)C = A(BC)$ (ASSOCIATIVE) but not commutative.

$A(B+C) = AB + AC$ (DISTRIBUTIVE)

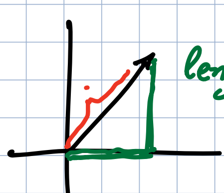
VECTOR-VECTOR PRODUCT.

$$x^T y = \overset{1\times d}{(x_1 \ \cdots \ x_d)} \overset{d\times 1}{\begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}} = x_1 y_1 + \cdots + x_d y_d = \sum_{i=1}^{d} x_i y_i = \langle x, y \rangle$$

\langle

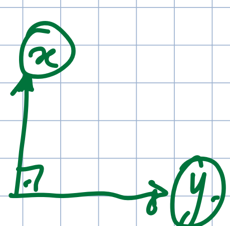
INNER PRODUCT

$$x^T \cdot y = \underbrace{\text{length}(x)\ \text{length}(y)}_{\text{A lot of geometry.}} . \cos(\theta) \in \mathbb{R}.$$



$\text{length}(x) = \|x\|_2 = \|x\| = \sqrt{x_1^2 + \cdots + x_d^2}$

if omitted



$\cos(90^\circ) = 0.$

generalizes Pythagoras theorem.

$$\left(\overset{d\times 1}{x}\ \overset{1\times k}{y^T}\right) = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_k \\ \vdots & x_i y_j & \vdots \\ x_d y_1 & \cdots & x_d y_k \end{pmatrix}^{d\times k} = \overset{d\times k}{C} \in \mathbb{R}^{d\times k}$$ OUTER Product

$x_i y_j \to (i,j)$-th entry.

$$\text{length}(x) = \|x\|_2 = \|x\| = \sqrt{x_1^2 + \dots + x_d^2} = \sqrt{\langle x, x\rangle} = \sqrt{x^T \cdot x}$$

($_2$: if omitted)

Euclidean NORM / ℓ_2.

$$\|x\|_2^2 = x^T \cdot x = \sum_{i=1}^{d} x_i^{2}$$

L_p norms (well-defined for our purposes for $p \in [1, +\infty)$.

$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p} \qquad \left[\ \|x\|_p^p = \sum_{i=1}^{d} |x_i|^p \ \right]$$

(^{p} → POWER; $_p$ denotes the norm)

$$\|x\|_1 = \sum_{i=1}^{d} |x_i| \qquad \|\underbrace{(1,-1)}\|_1 = |1| + |-1| = 2$$

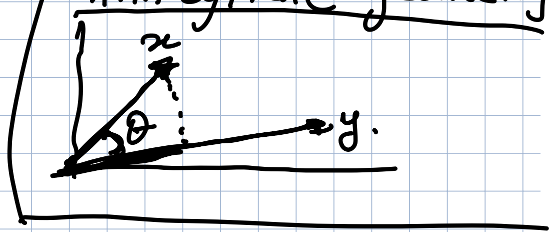
$$\|x\|_\infty = \max_{1 \le i \le d} |x_i|$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Q. $x^T y = x_1 y_1 + x_2 y_2$.

$$\|x\| = \sqrt{x_1^2 + x_2^2} \qquad \|y\| = \sqrt{y_1^2 + y_2^2}.$$

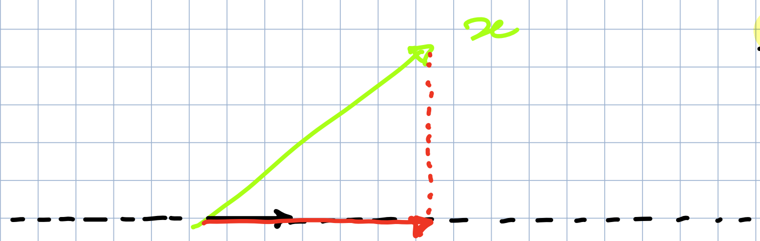
Analytical geometry.



$$\cos\theta = \frac{x_1 y_1 + x_2 y_2}{\sqrt{x_1^2 + x_2^2}\,\sqrt{y_1^2 + y_2^2}} \qquad \text{(projections)}$$

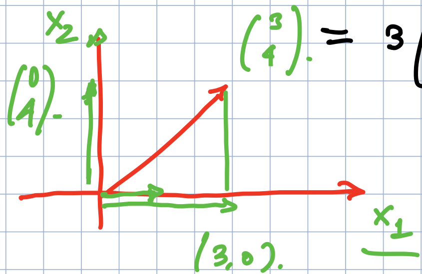
How do we project (why does trigonometry play a role?).

Q: How do we project x on b?



b $\quad \Pi_b(x) = $ projection of x on b?

Example: $\quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. then $\Pi_b(x) = ? = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$.



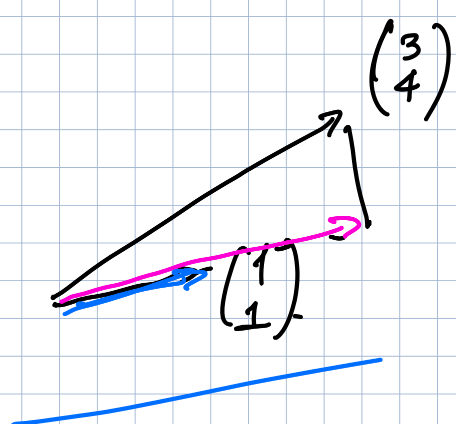
$$\begin{pmatrix} 3 \\ 4 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 4\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

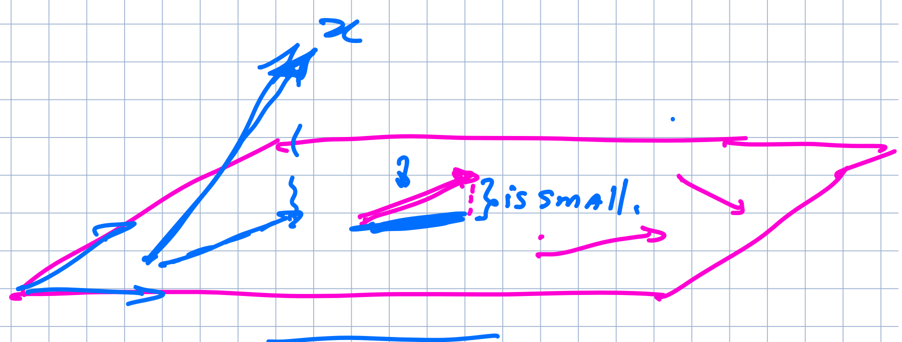
$$\Pi_{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}\left(\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$$

$$\Pi_{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}\left(\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

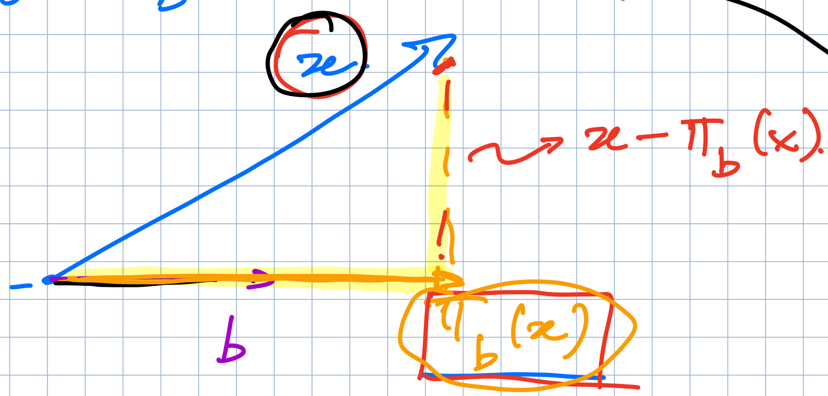
This example is easy, because $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ STANDARD ORTHONORMAL basis for $\mathbb{R}^2$.

($e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \dots, \; e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ is The STANDARD BASis for $\mathbb{R}^n$).





Clearly. $\pi_b(x) = \lambda \cdot b$ for some $\lambda \in \mathbb{R}$.



$$x - \pi_b(x) \perp b$$

$$(x - \pi_b(x))^T b = 0 \Rightarrow x^T b = (\pi_b(x))^T b$$

$$(x - \pi_b(x))^T = (x^T - \pi_b^T(x))$$

$$\lambda\, b^T b = x^T b \Rightarrow \lambda = \frac{x^T \cdot b}{b^T b} = \frac{b^T x}{\|b\|_2^2}$$

$$\text{So } \underset{d\times 1}{\pi_b(x)} = \lambda \cdot b = \left(\frac{b^T x}{b^T b}\right) \cdot b = \underset{d\times d}{\left[\frac{(b\, b^T)}{(b^T b)}\right]} \underset{d\times 1}{x}$$

Projection MATRIX.

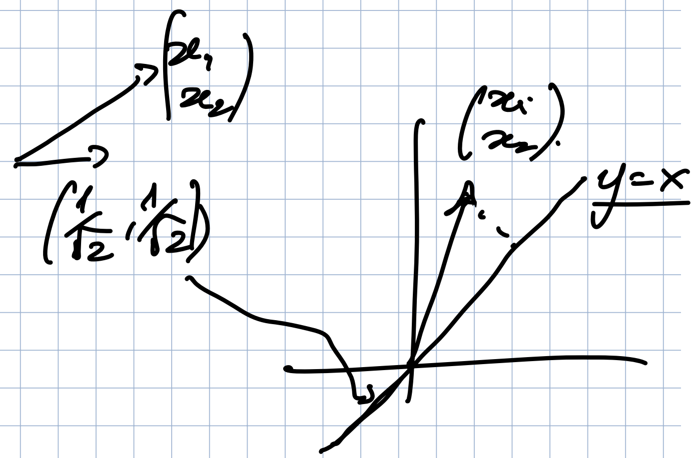
$$P = \frac{b\, b^T}{b^T b}$$

$$P^2 = \left(\frac{b\, b^T}{b^T b}\right)\left(\frac{b b^T}{b^T b}\right) = \frac{b\,\cancel{(b^T b)}\, b^T}{(b^T b)^{\cancel{2}}} = \frac{b\, b^T}{b^T b} = P.$$

$$\boxed{P^2 = P.}$$

$$x^T b = x_1 b_1 + \dots + x_d b_d = b^T x. \checkmark$$

$$b = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$



$$b^T b = \frac{1}{2} + \frac{1}{2} = 1.$$

$$b b^T = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

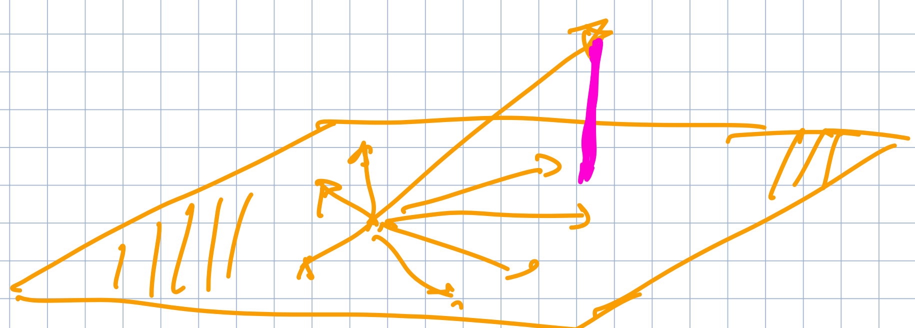
$$\underline{P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.} \qquad \boxed{Px} = \begin{pmatrix} \frac{1}{2} x_1 + \frac{1}{2} x_2 \\ \frac{1}{2} x_1 + \frac{1}{2} x_2 \end{pmatrix}$$

$$P^2 = P.$$

$$\frac{b b^T}{b^T b}.$$

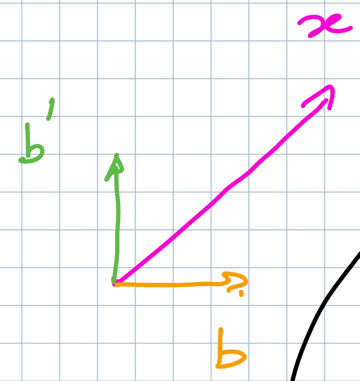
$$(Px) = \pi_b(x)$$

$$\boxed{\pi_b(\pi_b(x)) = \pi_b(x).}$$



$\{x_1, \dots, x_n\}$

$\mathbb{R}^3$



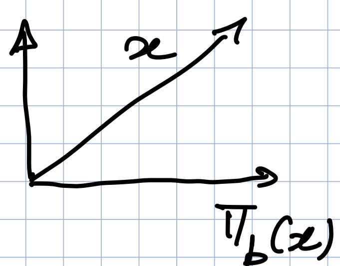
idea 1 find b'. (e.g. if $b = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, then $b' = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$). AND repeat.

$$P' = \frac{(b' {b'}^T)}{\|(b')^T b'}$$

$$\Pi_{b'}(x) = P' \cdot x.$$

idea 2 $(I - P)x = x - Px = x - \Pi_b(x)$

$=$



Claim $P' = I - P$. (proof by example)

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \left(\text{on } b = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right).$$

$$I - P = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} \quad ✓$$

$$\frac{(b')(b')^T}{1} =$$

$1 = (b')^T b'$.

$$\left(\left\| \left(\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}} \right) \right\| = \sqrt{\left(\tfrac{1}{\sqrt{2}}\right)^2 + \left(-\tfrac{1}{\sqrt{2}}\right)^2} = 1 \right)$$

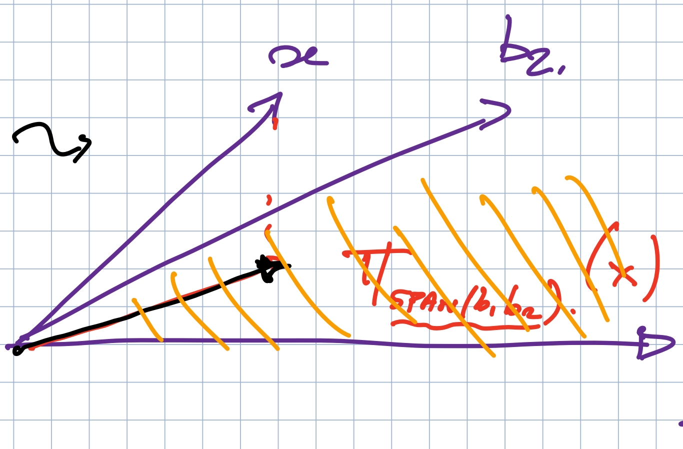
$$(I-P)^2 = (I-P)(I-P) = I - 2P + P^2$$

$$= I - 2P + P = I - P.$$

(Projection property, $X^2 = X$.).

Projection on general spaces.

$x \in \mathbb{R}^3$. $\underline{b_1, b_2} \in \mathbb{R}^3$.



$$\Pi_{\text{span}(b_1,b_2)}(x) = \Pi_U(x).$$

$U = \text{span}.\ (b_1, b_2)$

$\left(\text{span}(b_1, \dots, b_m).\right.$ more generally. $\left.\right).$

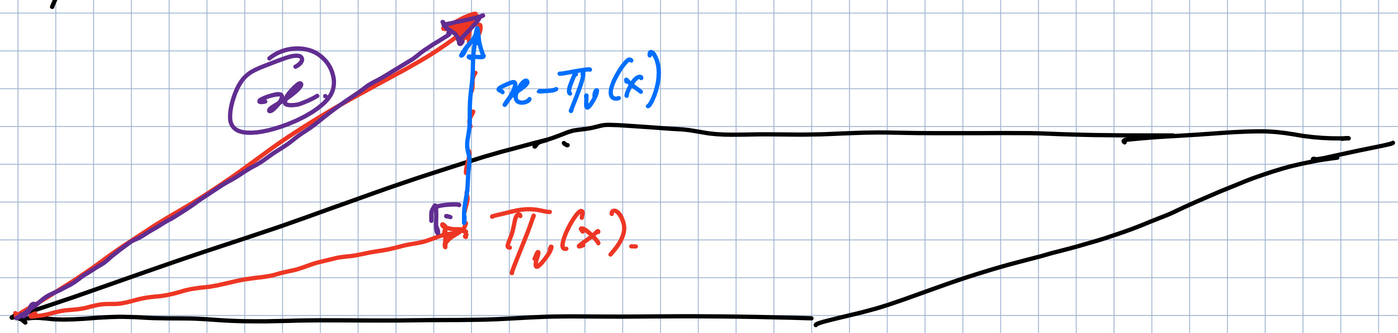
$$\Pi_U(x) = \lambda_1 b_1 + \dots + \lambda_m b_m \quad \text{for some } \lambda_1 \dots \lambda_m.$$

$$= \begin{array}{cc} d\times m. & m\times 1 \\ \begin{bmatrix} | & & | \\ b_1 & \cdots & b_m \\ | & & | \end{bmatrix} & \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \end{array} = B \cdot \lambda$$

LECTURE 19. (APRIL 6TH, 2023)

Cont. from last time.

We are given a set of vectors $b_1, \dots, b_m$ (linearly independent) that span a subspace $U = \text{span}(b_1, \dots, b_m) \subseteq \mathbb{R}^n$



How do we project x on U? $\exists \lambda_1, \dots, \lambda_m \in \mathbb{R}$. such that

$\Pi_U(x) = \lambda_1 b_1 + \dots + \lambda_m b_m$ or equivalently

$$\underbrace{\begin{bmatrix} | & & | \\ b_1 & \cdots & b_m \\ | & & | \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}}_{\lambda} = \Pi_U(x) \in \mathbb{R}^n.$$

Now we express the fact that $\overbrace{x - \Pi_U(x)}^{\shortparallel} \perp b_i$, $i = 1 \dots m$.

$$\left.\begin{array}{l} b_1^T \left(x - \Pi_U(x) \right) = 0 \\ \vdots \\ b_m^T \left(x - \Pi_U(x) \right) = 0 \end{array}\right\} \Rightarrow \begin{bmatrix} - b_1^T - \\ \vdots \\ - b_m^T - \end{bmatrix} (x - B\lambda) = 0.$$

Notice $\begin{bmatrix} - b_1^T - \\ \vdots \\ - b_m^T - \end{bmatrix} = \begin{bmatrix} | & & | \\ b_1 & \cdots & b_m \\ | & & | \end{bmatrix}^T$ So we can rewrite $\to$:

why is it invertible

$$B^T x = (B^T B)\lambda \implies \lambda = (B^T B)^{-1} B x.$$

So our projection is $\pi_V(x) = B\lambda = B(B^T B)^{-1} B^T x$.

The projection matrix. $\boxed{P = B(B^T B)^{-1} B^T}$

Example if $b_1, \dots, b_m$ is orthonormal. $(B^T B = I)$ then

$\boxed{P = B B^T}$

In this case where $b_i^T b_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$, this is equiv.

to $\pi_V(x) = B B^T x = \begin{bmatrix} | & & | \\ b_1 & \cdots & b_m \\ | & & | \end{bmatrix} \begin{bmatrix} - b_1^T - \\ \vdots \\ - b_m^T - \end{bmatrix} = \begin{bmatrix} | & & | \\ b_1 & \cdots & b_m \\ | & & | \end{bmatrix} \begin{bmatrix} b_1^T x \\ \vdots \\ b_m^T x \end{bmatrix}$

$$= (b_1^T x)\, b_1 + \cdots + (b_m^T x)\, b_m.$$

$\to$ Recall from previous lecture this is the projection of x on the 1dim subspace spanned by b_1. etc...

Reminder (important fact) Section 12.8 (online book. Blum - Hopcroft - Kannan).

Eigenvalue decomposition for symmetric ($A=A^T$) real matrices

THEOREM. Let A be a real symmetric matrix. Then,

1) The eigenvalues $\lambda_1, \dots, \lambda_n$ are real, as are the components of the corresponding eigenvectors $v_1, \dots, v_n$.

2) A is orthogonally diagonalizable,

$$A = \sum_{i=1}^{n} \lambda_i \, v_i v_i^T. \qquad \left(v_i^T v_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \right).$$

Equivalently, $A = V D V^T$ where $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

and $V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$, $(V^T = V^{-1})$. ∎

Theorem A real matrix A is orthogonally diagonizable if and only if A is symmetric.

Singular Value Decomposition (SVD)

FOR ANY MATRIX $A \in \mathbb{R}^{m \times n}$ there exist orthonormal matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ AND A DIAGONAL matrix

$$\Sigma = \left(\begin{array}{cccc|ccccc} \sigma_1 & & & & \cdots & 0 & \cdots & & 0 \\ & \ddots & & & & & & & \\ & & & \sigma_r & & & & & \\ \hline & & & & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 0 & 0 \cdots & 0 \end{array} \right)^{m \times n} \quad (\text{if } m < n)$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ & & & & 0 \end{bmatrix} \quad \text{if } m = n.$$

$$\Sigma = \left[\begin{array}{ccccc} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ \hline & & & 0 & \\ & & & & \ddots \quad 0 \\ 0 & & & & \vdots \\ \vdots & & & & \\ 0 & & & & 0 \end{array}\right] \quad \text{if } m > n.$$

With diagonal entries $\underbrace{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r}_{\text{singular values}} > \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$

$r = \text{rank}(A)$

such that $A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$.

U: left singular vectors

Σ: singular values.

V^T: right singular vectors

$\boxed{\text{LEMMA (Analog of eigenvalues and eigenvector)}}$

$$A v_i = \sigma_i u_i \quad (1) \quad \text{and} \quad A^T u_i = \sigma_i v_i. \quad (2)$$

PROOF

Since V is orthogonal, $A \cdot V = U \Sigma V^T V = U\Sigma$ which gives equation (1)

$$A^T u_i = \left(\sum_{k=1}^{r} \sigma_k u_k v_k^T\right)^T u_i = \left(\sum_{k=1}^{r} \sigma_k v_k u_k^T\right) u_i = \sigma_i v_i \overset{1}{(u_i u_i^T)} = \sigma_i v_i$$

BEST k-RANK APPROX.

Define $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \quad (k = 1, 2, \ldots, r)$.

LEMMA The rows of A_k are the projections of the rows of A

onto the subspace.

$$\overset{m\times n}{A} = \left[\begin{array}{ccc|ccc} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{array}\right] \cdot \Sigma \cdot \left[\begin{array}{c} - \; v_1^T \; - \\ \vdots \\ - \; v_r^T \; - \\ \hline - \; v_{r+1}^T \; - \\ \vdots \\ - \; v_n^T \; - \end{array}\right]$$

$\operatorname{rank}(A) = r. \leq \min(m,n).$ → when equality holds, A is full rank.

$R(A) = R([u_1, \ldots, u_r])$. $\quad N(A^T) = R([u_{r+1}, \ldots, u_m])$.

$R(A^T) = R([v_1, \ldots, v_r])$. $\quad N(A) = R([v_{r+1}, \ldots, v_n])$.

The 2-norm of a matrix. $\|A\|_2 \triangleq \sup_x \dfrac{\|Ax\|_2}{\|x\|_2} = \sigma_1.$ → give geom interpretation

The Frobenius norm $\|A\|_F^2 \triangleq \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sigma_1^2 + \cdots + \sigma_r^2.$

Theorem For any k-rank matrix B: $\|A-A_k\|_F \leq \|A-B\|_F$, $\|A-A_k\|_2 \leq \|A-B\|_2$. Furthermore, $\|A-A_k\|_2 = \sigma_{k+1}^2$.

$\boxed{\text{Important Remark}}$

$$A^TA = V\Sigma^T\Sigma V^T = V\Sigma^2 V^T.$$
$$AA^T = U\Sigma^2 U^T.$$

Thus σ_i^2 are the eigenvalues of A^TA, AA^T and $\{v_i\}$ $\{u_i\}$ are the eigenvectors respectively.

LEAST SQUARES VIA SVD

$$\min_x \|Ax-b\|_2^2$$

$$\|Ax-b\|_2^2 = \|(U\Sigma V^T)x - b\|_2^2 = \|U^T(AVV^Tx - b)\|_2^2 =$$

$$= \|\Sigma(V^Tx) - U^Tb\|_2^2 = \|\Sigma z - U^Tb\|_2^2$$

$$= \sum_{i=1}^{r} (\sigma_i z_i - u_i^T b)^2 + \sum_{i=r+1}^{m} (u_i^T b)^2$$

The opt solution is given by $z_i = \frac{u_i^T b}{\sigma_i}$, $i = 1 \dots r$.

AND the obj. becomes $\sum_{i=r+1}^{m} (u_i^T b)^2 = \min_x \|Ax-b\|_2^2$.

Let's find the actual x^*:

$$x^* = V z^* \Rightarrow \boxed{x^* = \sum_{i=1}^{r} \left(\frac{u_i^T b}{\sigma_i}\right) v_i}$$