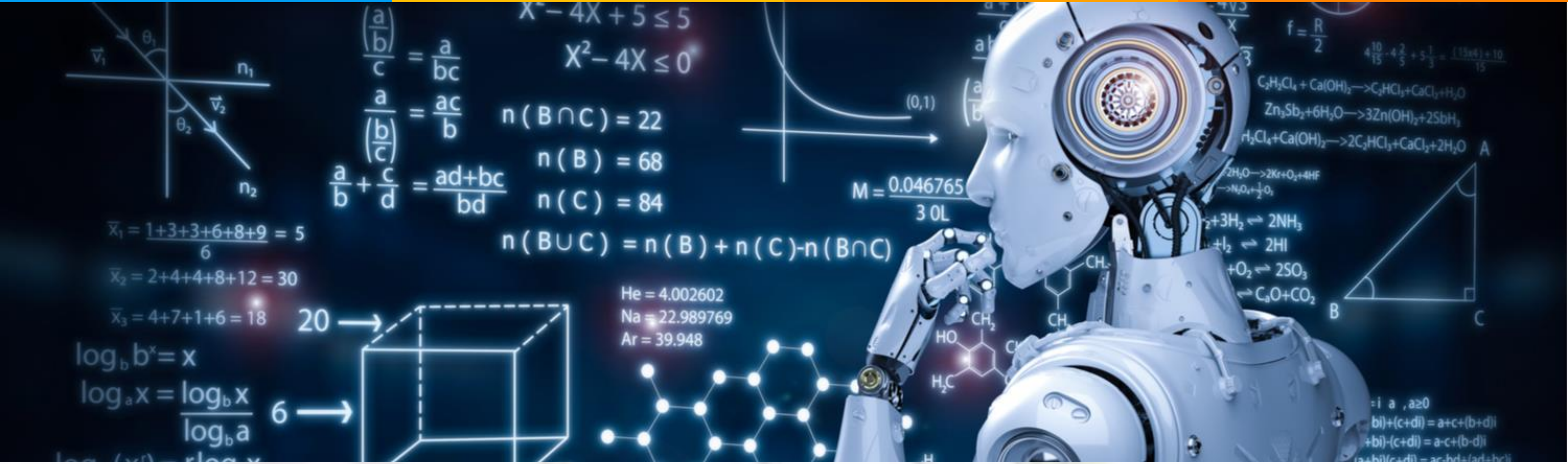
Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Lecture 09 Gram Schmidt

Recap

- Linear Combination
- Linear Independence, Dependence
  - Algebraic & Geometric meaning
- Basis
- Change of Basis – PCA
- We did not take traditional route to PCA with Eigen decomposition
- We interpreted the results of PCA intuitively as a linear combination

Problem with finding linear independence in general

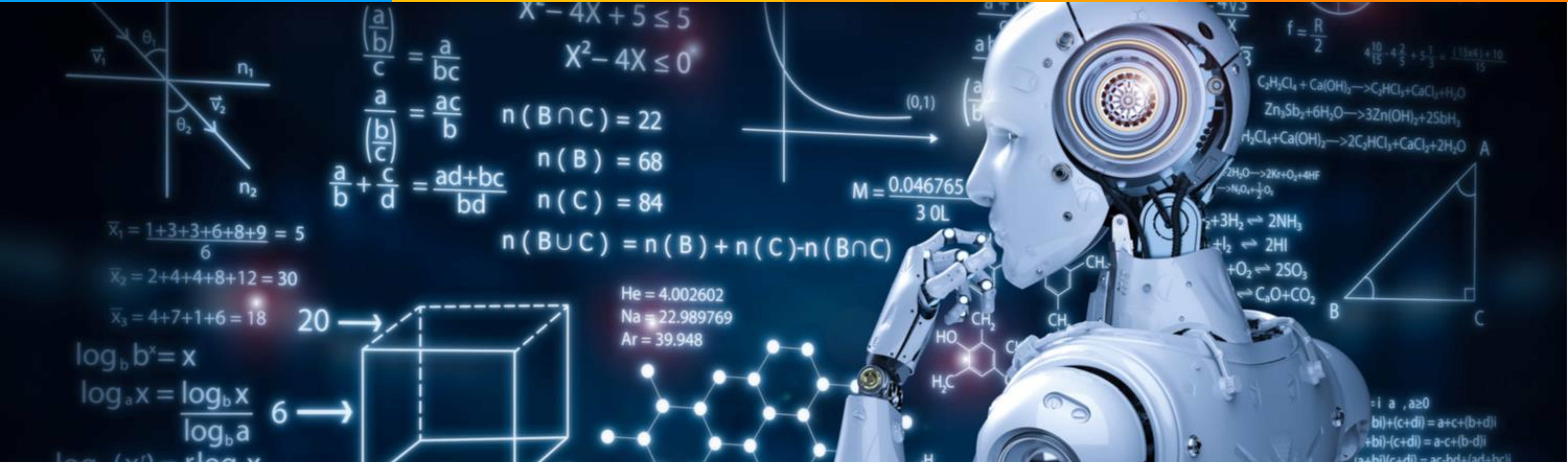
$$\{\mathbf{a_1}, \mathbf{a_2}, \ldots \mathbf{a_k}\} \qquad \beta_1, \beta_2, \ldots, \beta_n \neq 0$$

$$\beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + \ldots \beta_i \mathbf{a_i} + \ldots + \beta_n \mathbf{a_k} = 0$$

$$\mathbf{a_i} = \left(\frac{-\beta_1}{\beta_i}\right)\mathbf{a_1} + \left(\frac{-\beta_2}{\beta_i}\right)\mathbf{a_2} + \ldots + \left(\frac{-\beta_k}{\beta_i}\right)\mathbf{a_k}$$

- Infinite possibilities of beta 1 through k
- Solved by Gram Schmidt algorithm
- Exploits some elegant properties of orthogonal vectors

Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Orthogonality

Orthogonal & Orthonormal Vectors

- Two vectors a, b are orthogonal if $a^T b = 0$

$$a^T b = ||a|| ||b|| cos\theta = 0 \implies \theta = 90$$

- Two vectors a, b are orthonormal if they are
  - Orthogonal & $||a|| = ||b|| = 1$
- Extend the concept to set of vectors $\{\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_n}\}$

$$a_i^T a_j = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$$

Orthogonal vectors are linearly independent

- Geogebra demo
  - https://www.geogebra.org/calculator/eedbpvb9
  - https://www.geogebra.org/calculator/a644wefh
- Algebraically $\{\mathbf{a_1}, \mathbf{a_2}, \ldots \mathbf{a_k}\}$ $\beta_1, \beta_2, \ldots, \beta_k \neq 0$

$$\beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + \ldots \beta_i \mathbf{a_i} + \ldots + \beta_k \mathbf{a_k}$$

$$a_i^T(\beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + \ldots \beta_i \mathbf{a_i} + \ldots + \beta_k \mathbf{a_k})$$

$$= \beta_1 \mathbf{a_i}^T \mathbf{a_1} + \beta_2 \mathbf{a_i}^T \mathbf{a_2} + \ldots \beta_i \mathbf{a_i}^T \mathbf{a_i} + \ldots + \beta_k \mathbf{a_i}^T \mathbf{a_k})$$

$$= \beta_i \neq 0$$

Linear combination of orthonormal vectors

$$\{\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_k}\} \qquad \beta_1, \beta_2, ..., \beta_k \neq 0$$

$$x = \beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + ...\beta_i \mathbf{a_i} + ... + \beta_k \mathbf{a_k}$$

$$a_i^T x = a_i^T (\beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + ...\beta_i \mathbf{a_i} + ... + \beta_k \mathbf{a_k})$$

$$a_i^T x = \beta_1 \mathbf{a_i}^T \mathbf{a_1} + ...\beta_i \mathbf{a_i}^T \mathbf{a_i} + ... + \beta_k \mathbf{a_i}^T \mathbf{a_k})$$

$$a_i^T x = \beta_i \neq 0$$

Linear combination of orthonormal vectors (contd.)

$$\{\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_k}\} \qquad \beta_1, \beta_2, ..., \beta_k \neq 0$$

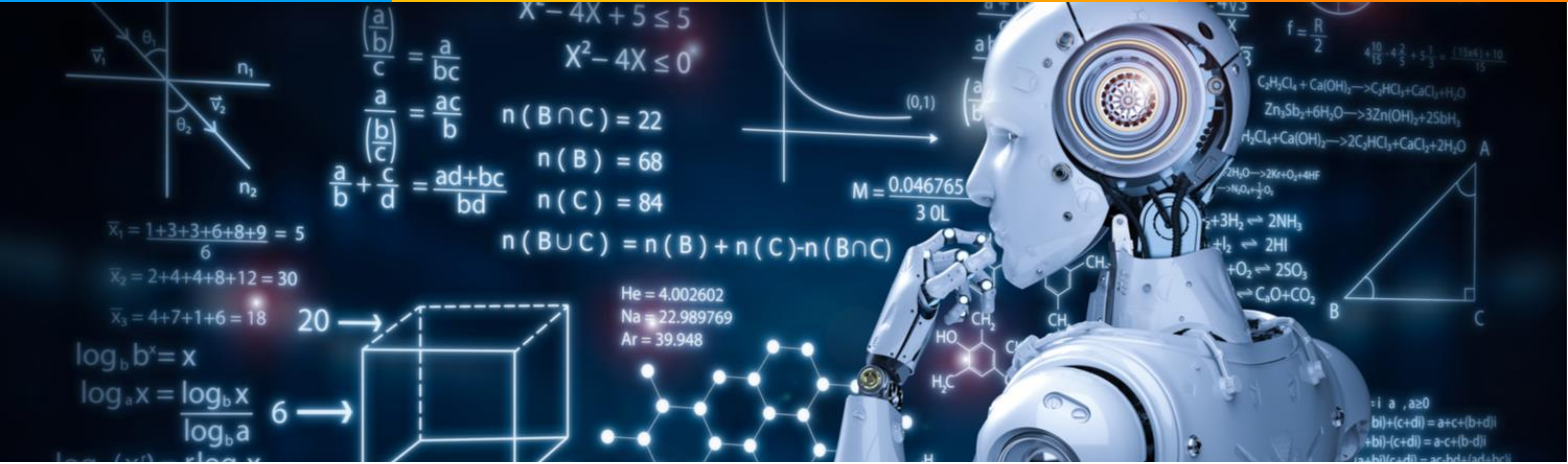
$$x = \beta_1\mathbf{a_1} + \beta_2\mathbf{a_2} + ...\beta_i\mathbf{a_i} + ... + \beta_k\mathbf{a_k}$$

$$a_i^T x = \beta_i \neq 0$$

Substituting for Beta1, Beta2 etc.

$$x = (a_1^T x)a_1 + (a_2^T x)a_2 + .. + (a_i^T x)a_i + ...(a_k^T x)a_k$$

Beta1, Beta2 are unique.

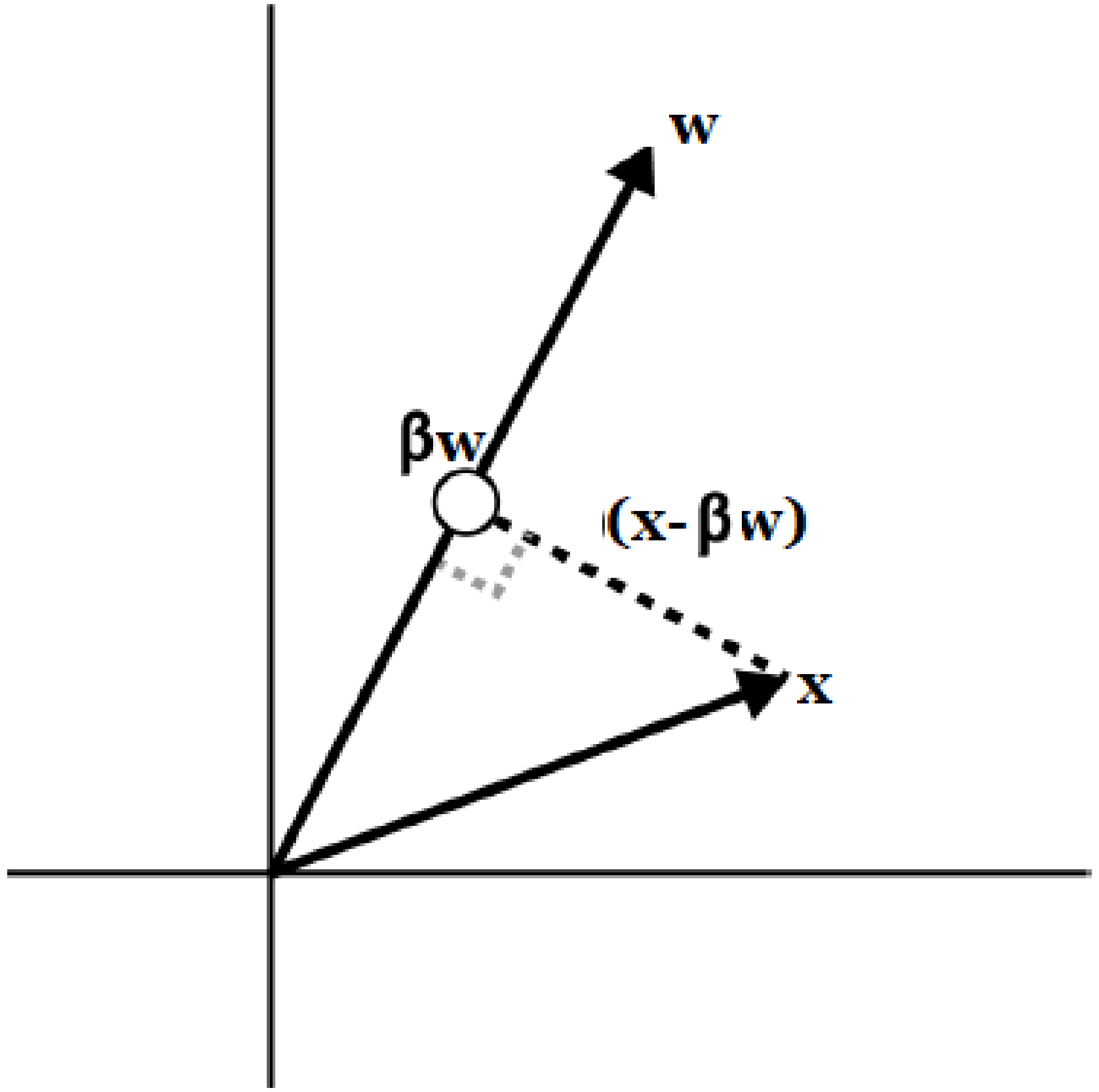


Gram Schmidt Algorithm

Extracting orthonormal set of vectors

- Do you recall how force resolution is done in perpendicular direction in Newtonian Mechanics?
- Intuition in 2D:
  - https://www.geogebra.org/calculator/wcwzbptu

Orthogonal component w.r.t. another vector

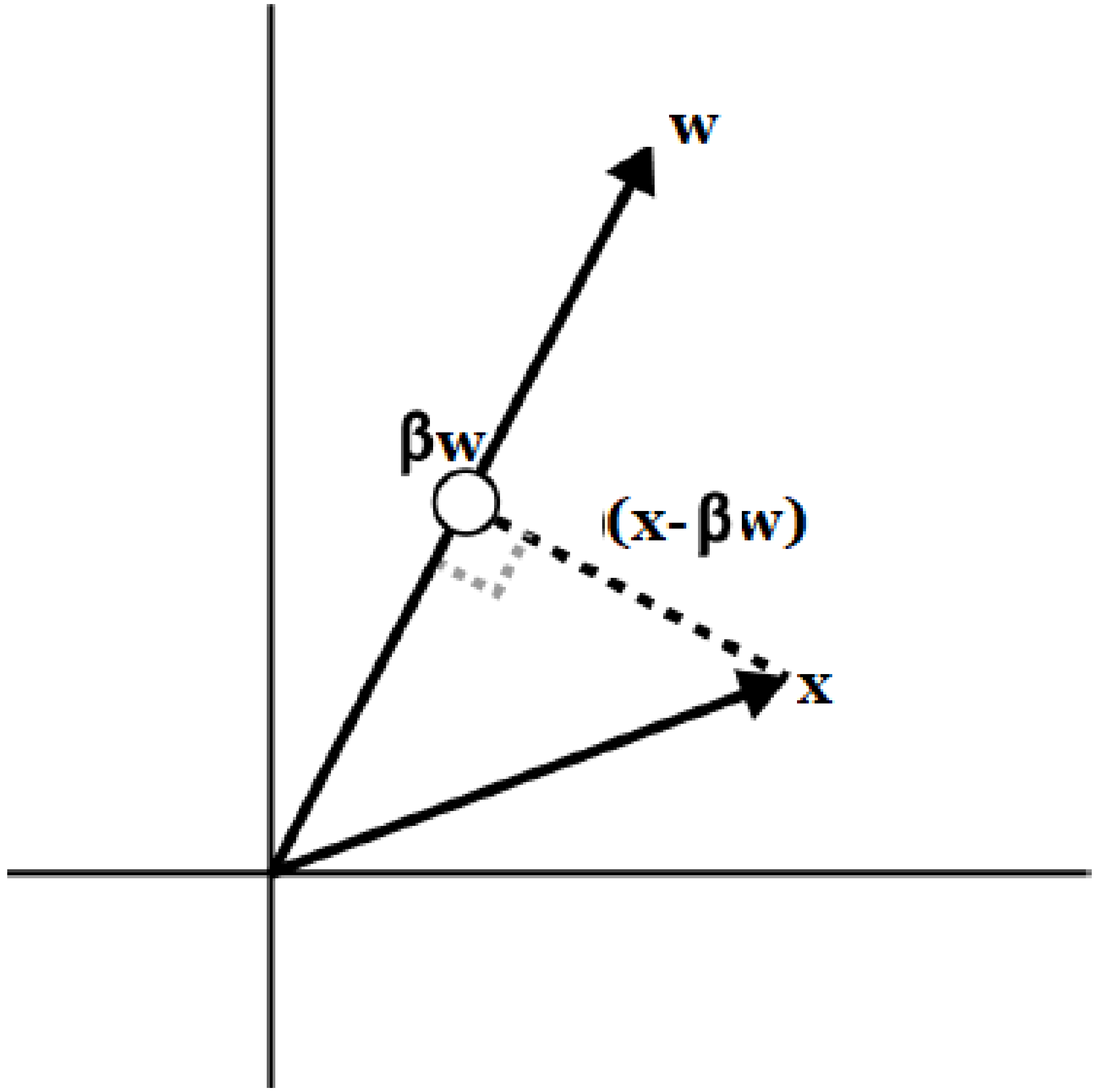


- Projection of x onto w βw
- Difference of projection vector βw and x is $x - \beta w$
- Projection vector βw is such as to minimize distance $x - \beta w$
- Then w and $x - \beta w$ are orthogonal

$$w^T(x - \beta w) = 0 \implies w^T x = \beta w^T w \implies \beta = \frac{w^T x}{w^T w}$$

$$\implies \beta w = \frac{w^T x}{\|w\|^2} w$$

Orthogonal component w.r.t. another vector



$$\beta = \frac{w^T x}{w^T w} \implies \beta w = \frac{w^T x}{\|w\|^2} w$$

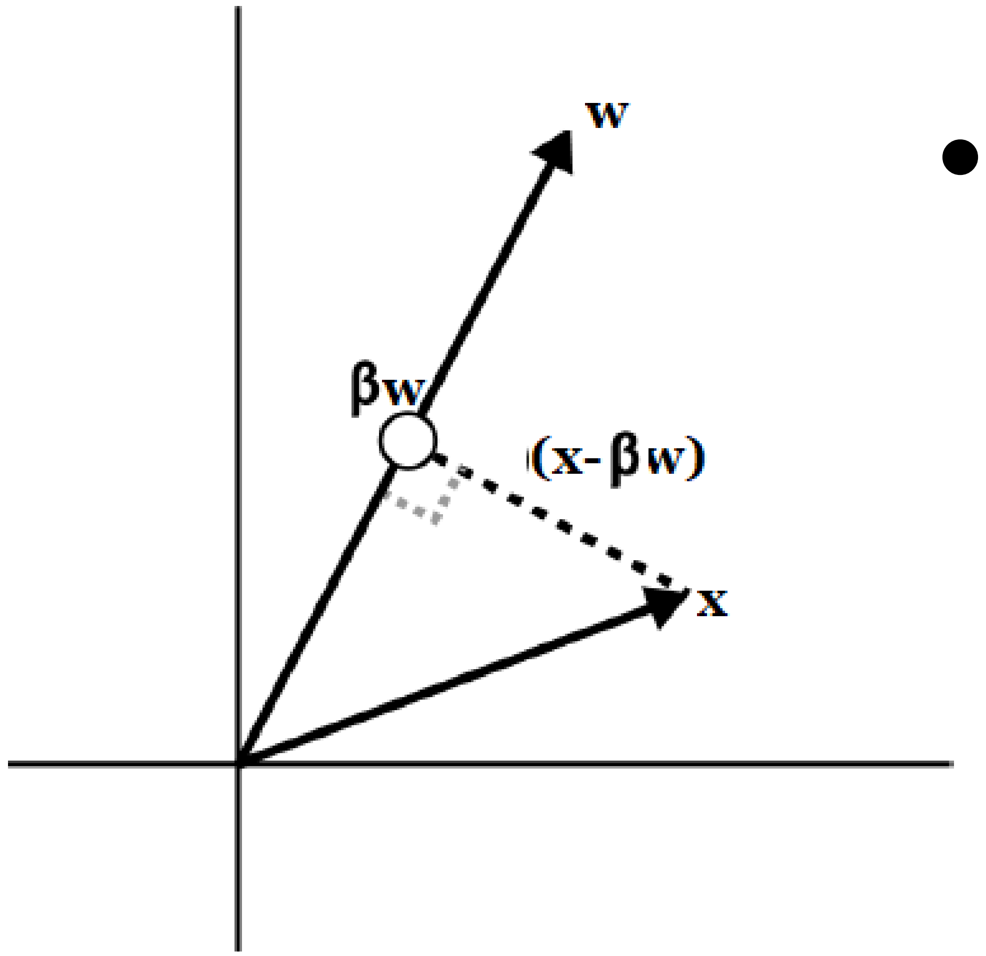
$$\beta w = \frac{w^T x}{\|w\|} \frac{w}{\|w\|}$$

Unit vector in the direction of w

- Component of x perpendicular to w is

$$x - \beta w = x - \frac{w^T x}{\|w\|} \frac{w}{\|w\|}$$

Orthogonal component w.r.t. another vector



- Component of x orthogonal to w is

$$x - \beta w = x - \frac{w^T x}{\|w\|}\frac{w}{\|w\|}$$

- Component of a2 orthogonal to a1: $q_2 = a_2 - \frac{a_1^T a_2}{\|a_1\|}\frac{a_1}{\|a_1\|}$

$$q_1 = \frac{a_1}{\|a_1\|}$$

$$q_2 = a_2 - (q_1^T a_2) q_1$$

Linear combination of orthonormal vectors: Recap

$$\{\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_k}\} \qquad \beta_1, \beta_2, ..., \beta_n \neq 0$$

$$x = \beta_1\mathbf{a_1} + \beta_2\mathbf{a_2} + ...\beta_i\mathbf{a_i} + ... + \beta_n\mathbf{a_k}$$

$$a_i^T x = \beta_i \neq 0$$

Substituting for Beta1, Beta2 etc.

$$x = (a_1^T x)a_1 + (a_2^T x)a_2 + .. + (a_i^T x)a_i + ...(a_k^T x)a_k$$

Beta1, Beta2 are unique.

Gram Schmidt steps

- Used to find if $\mathbf{a_1, a_2, \ldots, a_k}$ are linearly independent
- Steps
  - Extract orthogonal vectors from a1, a2
  - Find component of a2 that is orthogonal to a1

$$q_2 = a_2 - (q_1^T a_2) q_1$$

  - Try to express normalized a2 orthogonal component (q2) as a linear combination of all vectors up to a1

$$x = (a_1^T x) a_1 + (a_2^T x) a_2 + .. + (a_i^T x) a_i + \ldots (a_k^T x) a_k$$

$$q_i = a_i - \left((q_1^T a_i) q_1 + (q_2^T a_i) q_2 + .. + (a_{i-1}^T a_i) q_{i-1} \right)$$

Gram Schmidt steps (contd.)

- This expression is the starting point in Gram-Schmidt algo
  - Express Extract q_i that is orthogonal to all previously extracted mutually orthogonal set from $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{k}$

$$q_i = a_i - \left((q_1^T a_i)q_1 + (q_2^T a_i)q_2 + .. + (a_{i-1}^T a_i)q_{i-1} \right)$$

- Loop over all k and check if q_i is 0
- If q_i is 0 (or very close to 0 considering floating point accuracy), then abandon.
  - Conclusion: $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{k}$ are linearly dependent
- If loop completes, then $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{k}$ are linearly independent


QUESTIONS

Literature Survey