Applied Linear Algebra in Data Analysis

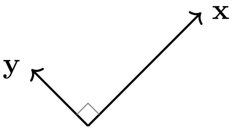
Orthogonality

Sivakumar Balasubramanian

Department of Bioengineering
Christian Medical College, Bagayam
Vellore 632002

Orthogonality

- Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal if $\mathbf{x}^\top \mathbf{y} = 0$.



- The set of non-zero vectors, $V = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_r\}$ is a set of mutually orthogonal vectors, if and only if,

$$\mathbf{v}_i^\top \mathbf{v}_j = 0, \ \ 1 \leq i, j \leq r \text{ and } i \neq j$$

- V is also a linearly independent set of vectors. Why?

Orthogonality

- If $\|\mathbf{v}_i\| = 1$, then V is an **orthonormal** set of vectors.

- A set of orthonormal vectors V also form an **orthonormal basis** of the subsapce $span\,(V)$.

Orthogonal Subspaces

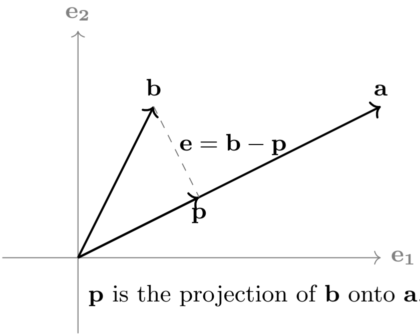
- Two subspaces $\mathcal{V}, \mathcal{W} \subset \mathbb{R}^n$ are orthogonal if every vector in one subspace is orthogonal to every vector in the other subspace.

$$\mathbf{v}^\top \mathbf{w} = 0, \;\; \forall \mathbf{v} \in \mathcal{V} \text{ and } \forall \mathbf{w} \in \mathcal{W} \implies \mathcal{V} \perp \mathcal{W}$$

- If $\mathcal{V} + \mathcal{W} = \mathbb{R}^n$, and $\mathcal{V} \perp \mathcal{W}$, then $\mathcal{V}$ and $\mathcal{W}$ are **orthogonal complements** of each other.

$$\mathcal{V}^\perp = \mathcal{W} \text{ or } \mathcal{W}^\perp = \mathcal{V}; \quad \left(\mathcal{V}^\perp\right)^\perp = \mathcal{V}$$

Orthogonal Projection onto Subspaces



$\mathbf{p}$ is the projection of $\mathbf{b}$ onto $\mathbf{a}$.

$\|\mathbf{e}\|$ is the distance of the point $\mathbf{b}$ from the line along $\mathbf{a}$. This distance is shortest when, $\mathbf{e} \perp \mathbf{a}$.

$$\mathbf{a}^\top(\mathbf{b} - \mathbf{p}) = \mathbf{a}^\top(\mathbf{b} - \alpha\mathbf{a}) = \mathbf{a}^\top\mathbf{b} - \alpha\mathbf{a}^\top\mathbf{a} =$$

$$\alpha = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}} \implies \mathbf{p} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a}$$

$$\mathbf{p} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a} = \mathbf{a}\frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}} = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}\mathbf{b} = \mathbf{P}\mathbf{b}$$

Orthogonal Projection onto Subspaces

- We can project vectors onto high dimensional subspaces.
- Consider the subspace $\mathcal{S} \subseteq \mathbb{R}^n$ spanned by the orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r\}$.
- We want to project a vector $\mathbf{b} \in \mathbb{R}^n$ onto $\mathcal{S}$
  $\mathbf{b}_{\mathcal{S}}$ – the orthogonal projection of $\mathbf{b}$ onto $\mathcal{S}$ is given by the following,

$$\mathbf{b}_{\mathcal{S}} = \mathbf{U}\mathbf{U}^\top \mathbf{b}; \;\; \mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_r \end{bmatrix}$$

$$\text{Projection matrix } \mathbf{P}_{\mathcal{S}} = \mathbf{U}\mathbf{U}^\top$$

- A projection matrix is **idempotent**, i.e. $\mathbf{P}^2 = \mathbf{P}$. What does this mean in terms of projecting a vector on to a subspace?

Orthogonal Projection onto Subspaces

- Consider two matrices $\mathbf{U}_1, \mathbf{U}_2$ whose columns form an orthonormal basis of the subspace $\mathcal{S} \subseteq \mathbb{R}^m$, $\mathcal{C}(\mathbf{U}_1) = \mathcal{C}(\mathbf{U}_2)$.

- The projection matrix onto the subspace $\mathcal{S}$, $\mathbf{U}_1\mathbf{U}_1^\top = \mathbf{U}_2\mathbf{U}_2^\top$. We get the same projection matrix irrespective of which orthonormal basis one uses.

Orthogonal Projection onto Subspaces

- Two subspaces $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ are said to be **complementary subspaces** of $\mathbb{R}^n$, when

$$\mathcal{X} + \mathcal{Y} = \mathbb{R}^n \quad \text{and} \quad \mathcal{X} \cap \mathcal{Y} = \{\mathbf{0}\}$$

- For complementary subspaces $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$, then any vector $\mathbf{v} \in \mathbb{R}^n$ can be uniquely represented as,

$$\mathbf{v} = \mathbf{v}_{\mathcal{X}} + \mathbf{v}_{\mathcal{Y}}, \quad \mathbf{v}_{\mathcal{X}} \in \mathcal{X}, \ \mathbf{v}_{\mathcal{Y}} \in \mathcal{Y}$$

  $\mathbf{v}_{\mathcal{X}}, \mathbf{v}_{\mathcal{Y}}$ are the components of $\mathbf{v}$ in $\mathcal{X}$ and $\mathcal{Y}$, respectively.

- When $\mathcal{X} \perp \mathcal{Y}$, then $\mathbf{v}_{\mathcal{X}}^\top \mathbf{v}_{\mathcal{Y}} = 0$; $\mathbf{v}_{\mathcal{X}}, \mathbf{v}_{\mathcal{Y}}$ are orthogonal components.

Relationship between the Four Fundamental Subspaces of A

- $\mathcal{C}(\mathbf{A}), \mathcal{N}\left(\mathbf{A}^\top\right) \subseteq \mathbb{R}^n$ are orthogonal complements.

$$\mathcal{C}(\mathbf{A}) \perp \mathcal{N}\left(\mathbf{A}^\top\right) \implies \mathcal{C}(\mathbf{A}) + \mathcal{N}\left(\mathbf{A}^\top\right) = \mathbb{R}^n$$

- $\mathcal{C}\left(\mathbf{A}^\top\right), \mathcal{N}(\mathbf{A}) \subseteq \mathbb{R}^m$ are orthogonal complements.

$$\mathcal{C}\left(\mathbf{A}^\top\right) \perp \mathcal{N}(\mathbf{A}) \implies \mathcal{C}\left(\mathbf{A}^\top\right) + \mathcal{N}(\mathbf{A}) = \mathbb{R}^m$$

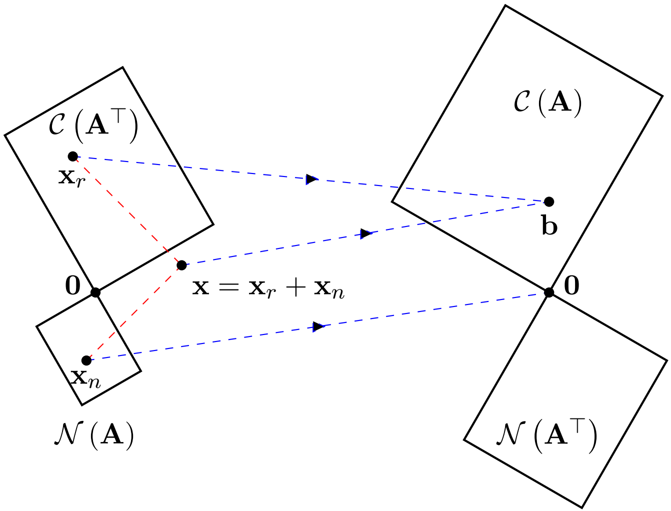
Orthogonal Projection onto Subspaces

- An orthogonal projection matrix $\mathbf{P}_{\mathcal{S}}$ onto a subspace $\mathcal{S}$ represents a linear mapping, $\mathbf{P}_{\mathcal{S}} : \mathbb{R}^n \to \mathbb{R}^n$. What are the four fundamental subspaces of $\mathbf{P}_{\mathcal{S}}$?

$$\mathcal{C}\left(\mathbf{P}_{\mathcal{S}}\right) = \mathcal{S}; \quad \mathcal{N}\left(\mathbf{P}_{\mathcal{S}}\right) = \mathcal{S}^{\perp}$$

$$\mathcal{N}\left(\mathbf{P}_{\mathcal{S}}^{\top}\right) = \mathcal{S}^{\perp}; \quad \mathcal{C}\left(\mathbf{P}_{\mathcal{S}}^{\top}\right) = \mathcal{S}$$

Relationship between the Four Fundamental Spaces



- $\mathbf{x}_r$ and $\mathbf{x}_n$ are the components of $\mathbf{x} \in \mathbb{R}^m$ in the row space and nullspace of $\mathbf{A}$.
- **Nullspace** $\mathcal{N}(\mathbf{A})$ is mapped to $\mathbf{0}$.

$$\mathbf{A}\mathbf{x}_n = \mathbf{0}$$

- **Row space** $\mathcal{C}(\mathbf{A}^\top)$ is mapped to the **column space** $\mathcal{C}(\mathbf{A})$.

$$\mathbf{A}\mathbf{x}_r = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n) = \mathbf{A}\mathbf{x} = \mathbf{b}$$

- The mapping from the **row space** to the **column space** is invertible, i.e. every $\mathbf{x}_r$ is mapped to a unique element in $\mathcal{C}(\mathbf{A})$
- What sort of mapping does $\mathbf{A}^\top$ do?

Gram-Schmidt Orthogonalization

- Given a linearly independent set of vectors $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n\}$, where $\mathbf{x}_i \in \mathbb{R}^m, \; \forall i \in \{1, 2, \dots n\}$, how can we find a orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots \mathbf{u}_n\}$ for $span\left(\mathcal{B}\right)$? $\longrightarrow$ **Gram-Schmidt Algorithm**
- Its an iterative procedure that can also detect if a given set $\mathcal{B}$ is linearly dependent.

Data: $\{\mathbf{x}_i\}_{i=1}^n$
Result: Return an orthonormal basis $\{\mathbf{u}_i\}_{i=1}^n$ if the set $\mathcal{B}$ is linearly independent, else return nothing.
for $i = 1, 2, \dots n$ **do**
    1. $\tilde{\mathbf{q}}_i = \mathbf{x}_i - \sum_{j=1}^{i-1} \left(\mathbf{u}_j^\top \mathbf{x}_i\right) \mathbf{u}_i$ $\longrightarrow$(**Orthogonalization step**);
    2. **If** $\tilde{\mathbf{q}}_i = 0$ **then return**;
    3. $\mathbf{u}_i = \tilde{\mathbf{q}}_i / \left\|\tilde{\mathbf{q}}_i\right\|$ $\longrightarrow$(**Normalization step**);
end
return $\{\mathbf{u}_i\}_{i=1}^n$;

Gram-Schmidt Orthogonalization

- The algorithm can also be conveniently represented in a matrix form.

$$\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_n\}$$

$$\text{Let } \mathbf{U}_1 = 0_{m\times 1} \quad \text{and} \quad \mathbf{U}_i = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_{i-1} \end{bmatrix} \in \mathbb{R}^{m\times(i-1)}$$

$$\mathbf{U}_i^\top \mathbf{x}_i = \begin{bmatrix} \mathbf{u}_1^\top \mathbf{x}_i \\ \mathbf{u}_2^\top \mathbf{x}_i \\ \vdots \\ \mathbf{u}_{i-1}^\top \mathbf{x}_i \end{bmatrix} \quad \text{and} \quad \mathbf{U}_i \mathbf{U}_i^\top \mathbf{x}_i = \sum_{j=1}^{i-1} \left(\mathbf{u}_j^\top \mathbf{x}_i\right) \mathbf{u}_j$$

$$\mathbf{u}_i = \frac{\left(\mathbf{I} - \mathbf{U}_i \mathbf{U}_i^\top\right) \mathbf{x}_i}{\left\| \left(\mathbf{I} - \mathbf{U}_i \mathbf{U}_i^\top\right) \mathbf{x}_i \right\|}$$

QR Decomposition

- Gram-Schmidt procedure leads us to another form of matrix decomposition – **QR decomposition**.
- Given a matrix $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{n \times n}$, whose columns form a linearly independent set. Gramm-Schmidt algorithm produces a orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots \mathbf{q}_n\}$ for $\mathcal{C}(\mathbf{A})$.

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{r_1} \quad \text{and} \quad \mathbf{q}_i = \frac{\mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^\top \mathbf{a}_i) \mathbf{q}_j}{r_k}$$

where, $r_1 = \|\mathbf{a}_1\|$ and $r_k = \left\| \mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^\top \mathbf{a}_i) \mathbf{q}_j \right\|$.

$$\mathbf{a}_1 = r_1 \mathbf{q}_1 \quad \text{and} \quad \mathbf{a}_i = r_i \mathbf{q}_i + \sum_{j=1}^{i-1} \left(\mathbf{q}_j^\top \mathbf{a}_i \right) \mathbf{q}_j$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \dots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \dots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_1 & \mathbf{q}_1^\top \mathbf{a}_2 & \mathbf{q}_1^\top \mathbf{a}_3 & \dots & \mathbf{q}_1^\top \mathbf{a}_n \\ 0 & r_2 & \mathbf{q}_2^\top \mathbf{a}_3 & \dots & \mathbf{q}_2^\top \mathbf{a}_n \\ 0 & 0 & r_2 & \dots & \mathbf{q}_3^\top \mathbf{a}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_n \end{bmatrix} = \mathbf{QR}$$

QR Decomposition

$$\mathbf{A} = \mathbf{QR}; \;\; \mathbf{A}, \mathbf{Q} \in \mathbb{R}^{m \times n}, \;\; \mathbf{R} \in \mathbb{R}^{n \times n}$$

- The columns of $\mathbf{Q}$ form an orthonormal basis for $\mathcal{C}(\mathbf{A})$, and $\mathbf{R}$ is upper-triangular.

- $\mathbf{A} = \mathbf{QR}$ can be used for used to solve $\mathbf{Ax} = \mathbf{b}$.

$$\mathbf{Ax} = \mathbf{QRx} = \mathbf{b} \implies \mathbf{Rx} = \mathbf{Q}^{-1}\mathbf{b} = \mathbf{Q}^{\top}\mathbf{b}$$