Fundamentals of linear algebra

- Inner product and distance
- Orthogonality; orthogonal sets and bases
- Orthogonal projections and decompositions
- Gram-Schmidt process
- QR-factorisation

Friday, 18 May 2018

Approximate solutions of linear systems

- It often happens that a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ which arises from a particular application, does not have an exact solution.

- In many cases, an approximate solution $\hat{\mathbf{x}}$ that makes the smallest difference between between $\mathbf{A}\hat{\mathbf{x}}$ and $\mathbf{b}$ is suitable.

- Such a solution is called a *least-squares solution* of $\mathbf{A}\mathbf{x} = \mathbf{b}$.

  To describe such solutions, we use the geometric concepts of length, distance and orthogonality for vectors in $\mathbb{R}^n$.

Inner product

Vectors in $\mathbb{R}^n$ can be regarded as $n \times 1$ matrices, e.g.:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

Their transposes are $1 \times n$ matrices:

$$\mathbf{v}^\top = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \qquad \mathbf{w}^\top = \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix}$$

Definition: The product $\quad \mathbf{v} \cdot \mathbf{w} \equiv \mathbf{v}^\top \mathbf{w}$

is called the *inner product* or *scalar product* (or *dot product*).

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$$

The result is a 1×1 matrix, that is, a scalar.

Inner product and orthogonality

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$$

Theorem: $\forall \{\mathbf{v}, \mathbf{w}, \mathbf{x}\} \in \mathbb{R}^n$:

- $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$
- $(\mathbf{v} + \mathbf{w}) \cdot \mathbf{x} = \mathbf{v} \cdot \mathbf{x} + \mathbf{w} \cdot \mathbf{x}$
- $(c\,\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (c\,\mathbf{w}) = c\,(\mathbf{v} \cdot \mathbf{w})$
- $\mathbf{v} \cdot \mathbf{v} \geqslant 0$, and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$

Definition: Two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ are called *orthogonal* to each other if $\mathbf{v} \cdot \mathbf{w} = 0$. The notation is $\mathbf{v} \perp \mathbf{w}$.

Note: $\forall\, \mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \perp \mathbf{0}$ because $\mathbf{0} \cdot \mathbf{v} = 0\ \forall\, \mathbf{v}$.

Norm and distance

Definitions:

- The length, or *norm*, of $\mathbf{v}$ is: $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\mathbf{v}^\top \mathbf{v}}$.

  $\forall\, c, \quad \|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$.

- A vector with a unit norm (length) is called a *unit vector*.

  If we divide a non-zero vector $\mathbf{v}$ by its length $\|\mathbf{v}\|$ we will obtain a *unit* (or *normalised*) vector $\mathbf{u}$:

$$\|\mathbf{u}\| = \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1.$$

Definition: The distance between two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ is

$$\mathrm{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

Pythagorean theorem: If and only if $\mathbf{v} \perp \mathbf{w}$,

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

Orthogonal complements

Definition: If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$ then $\mathbf{z}$ is said to be orthogonal to W.

The set of all vectors that are orthogonal to W is called the *orthogonal complement* of W and is denoted by $W^{\perp}$.

Example:

Let W be a plane through the origin in $\mathbb{R}^3$ and L be a line through the origin and perpendicular to W.

If $\mathbf{z} \in L$ and $\mathbf{w} \in W$

then $\mathbf{z} \cdot \mathbf{w} = 0$.

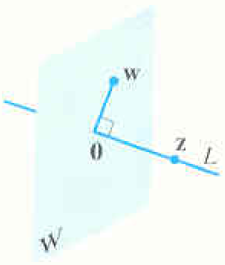

FIGURE 7
A plane and line through **0** as orthogonal complements.

Orthogonal complements

Notes:

L consists of all vectors orthogonal to $\mathbf{w} \in W$ and W consists of all vectors orthogonal to $\mathbf{z} \in L$, so

$$L = W^{\perp} \quad \text{and} \quad W = L^{\perp}$$

A vector $\mathbf{x} \in W^{\perp}$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.

$L^{\perp}$ and $W^{\perp}$ are subspaces of $\mathbb{R}^n$.

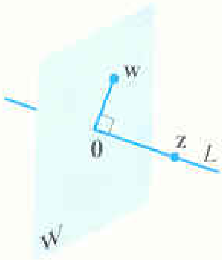

FIGURE 7
A plane and line through **0** as orthogonal complements.

Orthogonal complements

Theorem: Let $\mathbf{A}$ be an $m \times n$ matrix. Then

$$(\text{Row}\,\mathbf{A})^{\perp} = \text{Nul}\,\mathbf{A} \qquad \text{and} \qquad (\text{Col}\,\mathbf{A})^{\perp} = \text{Nul}\big(\mathbf{A}^{\top}\big)$$

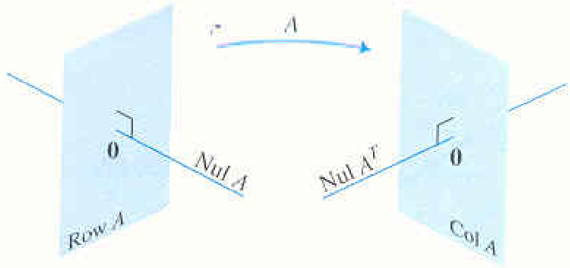

FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Orthogonal complements

Proof: If $\mathbf{x}$ is in $\mathrm{Nul}\,\mathbf{A}$, then $\mathbf{x}$ is orthogonal to each row of $\mathbf{A}$ given each row can be treated as vector in $\mathbb{R}^n$.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since the rows of $\mathbf{A}$ span $\mathrm{Row}\,\mathbf{A}$, $\mathbf{x}$ is orthogonal to $\mathrm{Row}\,\mathbf{A}$.

Conversely, if $\mathbf{x}$ is orthogonal to $\mathrm{Row}\,\mathbf{A}$, then $\mathbf{x}$ is certainly orthogonal to each row of $\mathbf{A}$, and hence $\mathbf{Ax} = \mathbf{0}$.

The statement $(\mathrm{Row}\,\mathbf{A})^{\perp} = \mathrm{Nul}\,\mathbf{A}$ is then true for any matrix including $\mathbf{A}^{\top}$. Therefore $(\mathrm{Col}\,\mathbf{A})^{\perp} = \big(\mathrm{Row}(\mathbf{A}^{\top})\big)^{\perp} = \mathrm{Nul}\,\mathbf{A}^{\top}$ because $\mathrm{Row}\,\mathbf{A}^{\top} = \mathrm{Col}\,\mathbf{A}$.

Orthogonal sets

Definition: A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is called an *orthogonal set* if each pair of vectors from the set is orthogonal:

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \qquad \forall \;\; i \neq j$$

Example: Show that this set is orthogonal:

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \;\; \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \;\; \mathbf{u}_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}.$$

Solution: Consider all the possible pairs:

$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0$
$\mathbf{u}_1 \cdot \mathbf{u}_3 = 3 \cdot (-1) + 1 \cdot (-4) + 1 \cdot 7 = 0$
$\mathbf{u}_2 \cdot \mathbf{u}_3 = -1 \cdot (-1) + 2 \cdot (-4) + 1 \cdot 7 = 0$

Each pair is orthogonal, thus $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.

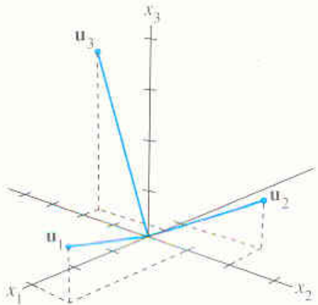

FIGURE 1

Orthogonal sets

Theorem: If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\} \in \mathbb{R}^n$ is an orthogonal set of non-zero vectors, then S is a linearly independent set and hence it is a basis for the subspace spanned by S.

Proof: If the set is linearly dependent, for some c_i not all zero,

$$\mathbf{0} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_p\mathbf{u}_p$$

Multiply this relation by $\mathbf{u}_1$ from both the sides:

$$\begin{aligned}
\mathbf{0} \cdot \mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\
0 &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \ldots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) \\
0 &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)
\end{aligned}$$

where only the first term remains since $\mathbf{u}_1 \perp \{\mathbf{u}_2, \ldots, \mathbf{u}_p\}$.

However $\mathbf{u}_1 \neq \mathbf{0}$ so $\mathbf{u}_1 \cdot \mathbf{u}_1 \neq 0$ and thus we must have $c_1 = 0$.

Similarly $c_2, \ldots c_p$ are also all zero. Thus S is linearly independent.

Orthogonal basis

Definition: An *orthogonal basis* for a subspace V of $\mathbb{R}^n$ is such a basis for V which is an orthogonal set.

Coordinates with respect to an orthogonal basis are easily found:

Theorem: Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace V of $\mathbb{R}^n$. For each $\mathbf{x} \in V$, the linear combination

$$\mathbf{x} = c_1\mathbf{u}_1 + \ldots + c_p\mathbf{u}_p \qquad \text{has the weights} \qquad c_j = \frac{\mathbf{x} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}$$

Proof: Compute all the inner products, as in the previous proof

$$\mathbf{x} \cdot \mathbf{u}_j = (c_1\mathbf{u}_1 + \ldots + c_p\mathbf{u}_p) \cdot \mathbf{u}_j = c_j(\mathbf{u}_j \cdot \mathbf{u}_j)$$

Since $\mathbf{u}_j \neq \mathbf{0}$, then $\mathbf{u}_j \cdot \mathbf{u}_j \neq 0$ and we can find $c_j = \dfrac{\mathbf{x} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}$

Orthogonal basis

Example: Find $[\mathbf{y}]_{\mathcal{B}}$ in the orthogonal basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$:

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$$

Solution: To find the coordinates $[\mathbf{y}]_{\mathcal{B}}$ we compute

$$\begin{aligned} \mathbf{y} \cdot \mathbf{u}_1 &= 6 \cdot 3 + 1 \cdot 1 - 8 \cdot 1 = 11, & \mathbf{u}_1 \cdot \mathbf{u}_1 &= 3^2 + 1^2 + 1^2 = 11 \\ \mathbf{y} \cdot \mathbf{u}_2 &= -6 \cdot 1 + 1 \cdot 2 - 8 \cdot 1 = -12, & \mathbf{u}_2 \cdot \mathbf{u}_2 &= (-1)^2 + 2^2 + 1^2 = 6 \\ \mathbf{y} \cdot \mathbf{u}_3 &= -6 \cdot 1 - 1 \cdot 4 - 8 \cdot 7 = -66, & \mathbf{u}_3 \cdot \mathbf{u}_3 &= (-1)^2 + (-4)^2 + 7^2 = 66 \end{aligned}$$

Thus

$$\mathbf{y} = \frac{11}{11}\,\mathbf{u}_1 - \frac{12}{6}\,\mathbf{u}_2 - \frac{66}{66}\,\mathbf{u}_3 \qquad \text{and} \qquad [\mathbf{y}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}$$

It is easy to find coordinates for an orthogonal basis.
For a non-orthogonal case we would need to solve a linear system.

Orthogonal projections

Given a non-zero vector $\mathbf{u} \in \mathbb{R}^n$, consider decomposing another vector $\mathbf{y} \in \mathbb{R}^n$ into the sum of two vectors, such that

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}, \qquad \hat{\mathbf{y}} = \alpha \mathbf{u}, \qquad \mathbf{z} \perp \mathbf{u}$$

Consider $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, which is orthogonal to $\mathbf{u}$ if and only if

$$0 = \mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u})$$

Hence

$$\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}, \qquad \hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\,\mathbf{u}$$

$\hat{\mathbf{y}}$ is the *orthogonal projection* of $\mathbf{y}$ onto $\mathbf{u}$, and $\mathbf{z}$ is called the component orthogonal to $\mathbf{u}$.

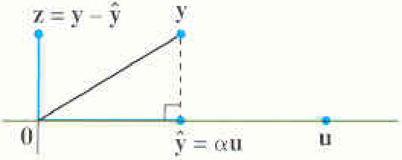

FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

Orthogonal projections

The orthogonal projection $\hat{\mathbf{y}}$ does not depend on the length of $\mathbf{u}$. Indeed, if we replace $\mathbf{u}$ by $\mathbf{u}' = \alpha\mathbf{u}$, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}'}{\mathbf{u}'\cdot\mathbf{u}'}\,\mathbf{u}' = \frac{\mathbf{y}\cdot(\alpha\mathbf{u})}{(\alpha\mathbf{u})\cdot(\alpha\mathbf{u})}\,(\alpha\mathbf{u}) = \frac{\alpha\,(\mathbf{y}\cdot\mathbf{u})}{\alpha^2\,(\mathbf{u}\cdot\mathbf{u})}\,\alpha\,\mathbf{u} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\,\mathbf{u}.$$

Definition:

$$\hat{\mathbf{y}} \equiv \mathrm{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\,\mathbf{u}$$

is the *orthogonal projection* of $\mathbf{y}$ onto $L = \mathrm{Span}\{\mathbf{u}\}$.

Subspace $L = \mathrm{Span}\{\mathbf{u}\}$ is a line through $\mathbf{u}$ and $\mathbf{0}$.

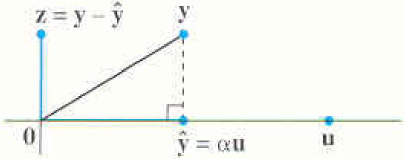

FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

Orthogonal projections

Example: Find orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ and write $\mathbf{y}$ as the sum of two vectors, one in $\mathrm{Span}\{\mathbf{u}\}$ and the other $\perp \mathbf{u}$,

$$\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

Solution: First we compute $\mathbf{y} \cdot \mathbf{u}$ and $\mathbf{u} \cdot \mathbf{u}$:

$$\mathbf{y} \cdot \mathbf{u} = \mathbf{y}^\top \mathbf{u} = \begin{bmatrix} 7 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40,$$

$$\mathbf{u} \cdot \mathbf{u} = \mathbf{u}^\top \mathbf{u} = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20.$$

Then the orthogonal projection and the orthogonal component are

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = \frac{40}{20} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Geometric interpretation

For two orthogonal basis vectors $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^2$:

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$

with

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}$$

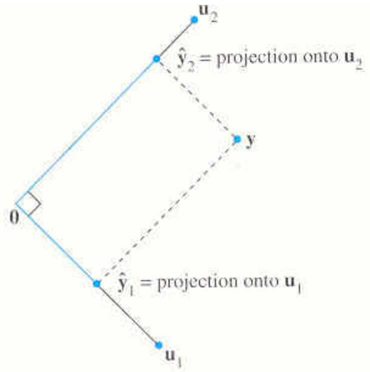

FIGURE 4 A vector decomposed into the sum of two projections.

The first term is the projection of $\mathbf{y}$ onto the line $\mathrm{Span}\{\mathbf{u}_1\}$, and the second term is the projection of $\mathbf{y}$ onto the line $\mathrm{Span}\{\mathbf{u}_2\}$.

Orthogonal decomposition

Theorem: (*the orthogonal decomposition theorem*)

Let W be a subspace of $\mathbb{R}^n$.

Then $\forall\, \mathbf{y} \in \mathbb{R}^n$ there is a unique decomposition

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z},$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis in W,

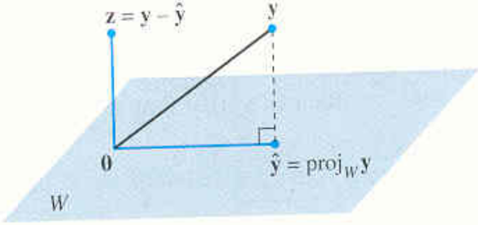

FIGURE 2 The orthogonal projection of $\mathbf{y}$ onto W.

$$\hat{\mathbf{y}} = \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}\, \mathbf{u}_i \qquad \text{and} \qquad \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}.$$

$\hat{\mathbf{y}} \equiv \mathrm{proj}_W\, \mathbf{y}$ is called the *orthogonal projection* of $\mathbf{y}$ onto W.

Orthogonal decomposition

Notes:

- The uniqueness of the decomposition indicates that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W but not on a particular basis used in W.
- If $\mathbf{y} \in W$ then $\mathrm{proj}_W \mathbf{y} = \mathbf{y}$.

Example: Let

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

(indeed, $\mathbf{u}_1 \cdot \mathbf{u}_2 = -4 + 5 - 1 = 0$ so the vectors are orthogonal)

Decompose $\mathbf{y}$ into a vector in W and a vector orthogonal to W.

Orthogonal decomposition

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Solution:

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2.$$

$$\mathbf{y}\cdot\mathbf{u}_1 = 2 + 10 - 3 = 9 \qquad \mathbf{y}\cdot\mathbf{u}_2 = -2 + 2 + 3 = 3$$

$$\mathbf{u}_1\cdot\mathbf{u}_1 = 4 + 25 + 1 = 30 \qquad \mathbf{u}_2\cdot\mathbf{u}_2 = 4 + 1 + 1 = 6$$

$$\hat{\mathbf{y}} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

and

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

Orthogonal decomposition

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \hat{\mathbf{y}} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}, \quad \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

The theorem ensures that $(\mathbf{y} - \hat{\mathbf{y}}) \in W^{\perp}$.

We can verify that $(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 = 0$ and $(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_2 = 0$.

The final decomposition is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

Properties of projection

Theorem: *(the best approximation theorem)*

Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$, and $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$.

Then $\hat{\mathbf{y}}$ is the point in W closest to $\mathbf{y}$ in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \qquad \forall\, \mathbf{v} \neq \hat{\mathbf{y}}$$

Definition: The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W.

Notes:

- $\hat{\mathbf{y}}$ is called the *best approximation* to $\mathbf{y}$ by elements of W.
- In a sense, we approximate $\mathbf{y}$ by a variable vector $\mathbf{v} \in W$.

  The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathbf{y} - \mathbf{v}\|$, can be regarded as the 'error' incurred by using $\mathbf{v}$ in place of $\mathbf{y}$.

  This error is minimised when $\mathbf{v} = \hat{\mathbf{y}}$.

Properties of projection

Example: Find the distance from $\mathbf{y}$ to $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ where

$$\mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}$$

Solution: Distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$.

Vectors $\mathbf{u}_1, \mathbf{u}_2$ form an orthogonal basis for W, and

$$\mathbf{y} \cdot \mathbf{u}_1 = -5 + 10 + 10 = 15 \qquad \mathbf{y} \cdot \mathbf{u}_2 = -1 - 10 - 10 = -21$$

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = 25 + 4 + 1 = 30 \qquad \mathbf{u}_2 \cdot \mathbf{u}_2 = 1 + 4 + 1 = 6$$

Then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \frac{15}{30}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{21}{6}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

Properties of projection

$$\mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

Then

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

and the distance from $\mathbf{y}$ to W is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{0^2 + 3^2 + 6^2} = \sqrt{45} = 3\sqrt{5}$$

So, $\hat{\mathbf{y}}$ is the best approximation for $\mathbf{y}$ within $\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$; any other vector in $\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ will have a greater distance from $\mathbf{y}$.

Orthonormal set and basis

Each $\mathbf{y} \in \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ can be decomposed into a sum of p projections onto mutually orthogonal one-dimensional subspaces.

Definition: A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is called an *orthonormal set* if it is an orthogonal set of unit vectors.

If V is the subspace spanned by such a set, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an *orthonormal basis* for V, since this set is linearly independent.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$.

Any non-empty subset of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal too, and forms an orthonormal basis for the corresponding sub-space.

Orthonormal basis and projection

Theorem: If $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_p]$, where $\{\mathbf{u}_1,\ \ldots\ \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then:

$$\mathrm{proj}_W\, \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\, \mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p)\, \mathbf{u}_p$$

$$\mathrm{proj}_W\, \mathbf{y} = \mathbf{U}\mathbf{U}^\top \mathbf{y} \qquad \forall\, \mathbf{y} \in \mathbb{R}^n$$

Proof: By the definition of projection

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\, \mathbf{u}_1 + \ldots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\, \mathbf{u}_p$$

and taking into account that the basis is orthonormal

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = 1, \quad \mathbf{u}_2 \cdot \mathbf{u}_2 = 1, \quad \ldots \quad \mathbf{u}_p \cdot \mathbf{u}_p = 1$$

we immediately obtain

$$\mathrm{proj}_W\, \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\, \mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p)\, \mathbf{u}_p.$$

Orthonormal basis and projection

Proof (contnuing):

From $\operatorname{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\,\mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p)\,\mathbf{u}_p$ we see that $\operatorname{proj}_W \mathbf{y}$ is a linear combination of the columns of $\mathbf{U}$ with the coefficients $(\mathbf{y} \cdot \mathbf{u}_1),\ (\mathbf{y} \cdot \mathbf{u}_2),\ \ldots\ (\mathbf{y} \cdot \mathbf{u}_p)$.

Denoting $\mathbf{x} = \begin{bmatrix} \mathbf{y} \cdot \mathbf{u}_1 \\ \mathbf{y} \cdot \mathbf{u}_2 \\ \vdots \\ \mathbf{y} \cdot \mathbf{u}_p \end{bmatrix}$ we can write $\operatorname{proj}_W \mathbf{y} = \mathbf{U}\mathbf{x}$.

In turn, the elements of $\mathbf{x}$ can be written as

$$\mathbf{u}_1^\top \mathbf{y}, \quad \mathbf{u}_2^\top \mathbf{y}, \quad \ldots \quad \mathbf{u}_p^\top \mathbf{y}$$

which are the entries of $\mathbf{U}^\top \mathbf{y}$.

Thus $\mathbf{x} = \mathbf{U}^\top \mathbf{y}$ and so $\operatorname{proj}_W \mathbf{y} = \mathbf{U}\mathbf{U}^\top \mathbf{y}$.

Gram-Schmidt process

Gram-Schmidt process is an algorithm for producing an orthogonal or orthonormal basis for any non-zero subspace of $\mathbb{R}^n$.

This is easy to understand with an example.

Example: Consider a linearly independent set

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

which is a basis for a subspace W of $\mathbb{R}^4$.

We aim to construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for W.

Solution:
Step 1: Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \mathrm{Span}\{\mathbf{x}_1\} = \mathrm{Span}\{\mathbf{v}_1\}$.

Gram-Schmidt process

Step 2: Vector $\mathbf{v}_2$ is then produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is,

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{proj}_{W_1}(\mathbf{x}_2) = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1.$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

For the ease of further calculations, renormalise $\mathbf{v}_2$ into $\mathbf{v}_2'$:

$$\mathbf{v}_2' = 4\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and then} \quad W_2 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2'\}$$

Gram-Schmidt process

Step 3: Produce $\mathbf{v}_3$ by subtracting from $\mathbf{x}_3$ its W_2-projection:

$$\mathrm{proj}_{W_2}(\mathbf{x}_3) = \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\, \mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\, \mathbf{v}_2$$

$$= \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}.$$

Then $\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{proj}_{W_2}(\mathbf{x}_3)$ is

$$\mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}; \qquad \mathbf{v}_3' = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

Gram-Schmidt process

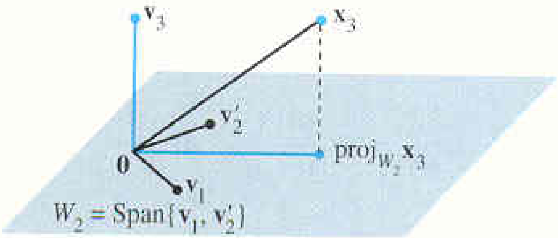

FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

Vector $\mathbf{v}_3 \in W$ given that $\mathbf{x}_3$ and $\text{proj}_W \mathbf{x}_3$ are both in W.

Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set in W and it is basis for W.

Gram-Schmidt process

Theorem: Given a basis $\mathbf{x}_1, \ldots, \mathbf{x}_p$ for a subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2$$

$$\ldots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2 - \ldots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\,\mathbf{v}_{p-1}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W.

In addition, $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for $1 \leqslant k \leqslant p$.

Orthonormal basis is then obtained by normalising $\mathbf{v}_i$ to unit vectors.

Gram-Schmidt process

Example: Construct an orthonormal basis for

$$\operatorname{Span}\left\{ \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\}$$

Solution:

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{\overbrace{3 \cdot 1 + 6 \cdot 2 + 0 \cdot 2}^{15}}{\underbrace{3^2 + 6^2 + 0^2}_{45}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

The corresponding orthonormal basis is

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Orthonormal matrices

Matrices with columns forming an orthonormal set are important for applications and computing algorithms for matrix computations.

The main properties of such (generally $m \times n$) matrices are:

Theorem: $\mathbf{U}$ has orthonormal columns if and only if $\mathbf{U}^\top \mathbf{U} = \mathbf{I}$.

Theorem: If $\mathbf{U}$ has orthonormal columns, then $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

(a) $\|\mathbf{Ux}\| = \|\mathbf{x}\|$

(b) $(\mathbf{Ux}) \cdot (\mathbf{Uy}) = \mathbf{x} \cdot \mathbf{y}$

(c) $(\mathbf{Ux}) \cdot (\mathbf{Uy}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

Properties (a) and (c) say that the linear mapping $\mathbf{x} \mapsto \mathbf{Ux}$ preserves the length and orthogonality.

These properties are important for many computer algorithms.

Orthogonal matrices

The above theorems are particularly useful for square matrices.

Definition: An *orthogonal matrix* $\mathbf{U}$ is a square invertible matrix such that $\mathbf{U}^{-1} = \mathbf{U}^{\top}$.

Notes:

- An orthogonal matrix has orthonormal columns.
- Any square matrix with orthonormal columns is orthogonal.
- An orthogonal matrix must have orthonormal rows too.

Example:

$$\mathbf{U} = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}.$$

is orthogonal because it is square and its columns are orthonormal.

QR factorisation of matrices

Theorem: *QR factorisation*
An $m \times n$ matrix $\mathbf{A}$ with linearly independent columns can be factorised as $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is an $m \times n$ matrix with columns forming an orthonormal basis for $\mathrm{Col}\,\mathbf{A}$, and $\mathbf{R}$ is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

Note: Since $\mathbf{Q}$ is an orthonormal matrix, $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$.

Thus $\mathbf{Q}^\top \mathbf{A} = \mathbf{Q}^\top (\mathbf{QR}) = (\mathbf{Q}^\top \mathbf{Q})\mathbf{R} = \mathbf{IR} = \mathbf{R}$, so $\mathbf{R} = \mathbf{Q}^\top \mathbf{A}$.

Proof: The columns of $\mathbf{A}$ form a basis $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ for $\mathrm{Col}\,\mathbf{A}$.

An orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\,\mathbf{A}$ can be constructed by using the Gram-Schmidt process.

Then $\forall\, k = 1 \ldots n \quad \mathbf{a}_k \in \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

Therefore, there are constants $r_{1k}, \ldots, r_{kk}$, such that

$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + \ldots + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + \ldots + 0 \cdot \mathbf{u}_n$

In case $r_{kk} < 0$, multiply r_{kk} and $\mathbf{u}_k$ by -1 so that all $r_{kk} > 0$.

QR factorisation of matrices

Proof (continuing): Rewriting the same equation in vector form,

$$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + \ldots + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + \ldots + 0 \cdot \mathbf{u}_n$$

$$= \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{Q}\mathbf{r}_k$$

From $\mathbf{r}_k$ vectors, we form matrix $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 \ldots \mathbf{r}_n \end{bmatrix}$. Then

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \ldots \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Q}\mathbf{r}_1 \ldots \mathbf{Q}\mathbf{r}_n \end{bmatrix} = \mathbf{Q}\mathbf{R}$$

By construction, $\mathbf{R}$ is triangular with positive diagonal entries.

It can be shown that $\mathbf{R}$ is invertible because the columns of $\mathbf{A}$ are linearly independent (consider $\mathbf{R}\mathbf{x} = \mathbf{0}$ given that $\mathbf{A}\mathbf{x} = \mathbf{0}$).

QR factorisation of matrices

Example:

Find a QR decomposition of: $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Solution: Earlier we have found an orthogonal basis for $\operatorname{Col}\mathbf{A}$ as

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

Upon normalisation we obtain

$$\mathbf{Q} = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}.$$

QR factorisation of matrices

Then

$$\mathbf{R} = \mathbf{Q}^\top \mathbf{A} = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1//\sqrt{6} \end{bmatrix}^\top \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{12} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}.$$

Summary

- Inner product $\mathbf{u} \cdot \mathbf{v} \equiv \mathbf{u}^\top \mathbf{v} = \sum_i \mathbf{u}_i \mathbf{v}_i$
- Norm $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ and distance $\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
- Orthogonal vectors, complements, sets, bases
- Orthogonal projections and decompositions
- The best approximation theorem
- Gram-Schmidt process: $\mathbf{v}_1 = \mathbf{x}_1$, then

$$\mathbf{v}_i = \mathbf{x}_i + \sum_{j=1}^{i-1} \left(-\frac{\mathbf{x}_i \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j} \mathbf{v}_j \right) \qquad i = 2, \ldots p$$

- $\mathbf{A} = \mathbf{QR}$ factorisation: $\mathbf{A} \xrightarrow{\text{Gram-Schmidt}} \mathbf{Q}$, then $\mathbf{R} = \mathbf{Q}^\top \mathbf{A}$.