QR factorisation and application to least squares

Matrix factorisations

Matrix factorisations are popular because they allow to perform some computations more easily

There are several different types of factorisations. Here, we study just the QR factorisation, which is useful for many least squares problems

Before we start, though, we need to learn a little about *orthogonal sets* and *orthogonal matrices*

Definition 1 (Orthogonal set of vectors)

The set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ is an **orthogonal set** if

$$\forall i, j = 1, \ldots, k, \quad i \neq j \implies \mathbf{v}_i \bullet \mathbf{v}_j = 0$$

Theorem 2

$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ with $\forall i$, $\mathbf{v}_i \neq \mathbf{0}$, orthogonal set $\implies$ $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ linearly independent

Definition 3 (Orthogonal basis)

Let S be a basis of the subspace $W \subset \mathbb{R}^n$ composed of an orthogonal set of vectors. We say S is an **orthogonal basis** of W

Proof of Theorem 2

Assume $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ orthogonal set with $\mathbf{v}_i \neq \mathbf{0}$ for all $i = 1, \dots, k$. Recall $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is LI if

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0} \iff c_1 = \cdots = c_k = 0$$

So assume $c_1, \dots, c_k \in \mathbb{R}$ are s.t. $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$. Recall that $\forall \mathbf{x} \in \mathbb{R}^k$, $\mathbf{0}_k \bullet \mathbf{x} = 0$. So for some $\mathbf{v}_i \in \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$

$$\begin{aligned} 0 &= \mathbf{0} \bullet \mathbf{v}_i \\ &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \bullet \mathbf{v}_i \\ &= c_1\mathbf{v}_1 \bullet \mathbf{v}_i + \cdots + c_k\mathbf{v}_k \bullet \mathbf{v}_i \end{aligned} \tag{1}$$

As $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ orthogonal, $\mathbf{v}_j \bullet \mathbf{v}_i = 0$ when $i \neq j$, (1) reduces to

$$c_i\mathbf{v}_i \bullet \mathbf{v}_i = 0 \iff c_i\|\mathbf{v}_i\|^2 = 0$$

As $\mathbf{v}_i \neq 0$ for all i, $\|\mathbf{v}_i\| \neq 0$ and so $c_i = 0$. This is true for all i, hence the result

Example – Vectors of the standard basis of $\mathbb{R}^3$

For $\mathbb{R}^3$, we denote

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

($\mathbb{R}^k$ for $k > 3$, we denote them $\mathbf{e}_i$)

Clearly, $\{\mathbf{i}, \mathbf{j}\}$, $\{\mathbf{i}, \mathbf{k}\}$, $\{\mathbf{j}, \mathbf{k}\}$ and $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ orthogonal sets. The standard basis vectors are also $\neq \mathbf{0}$, so the sets are LI. And

$$\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$$

is an orthogonal basis of $\mathbb{R}^3$ since it spans $\mathbb{R}^3$ and is LI

$$c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k} = c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Orthonormal version of things

Definition 4 (Orthonormal set)

The set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set and furthermore

$$\forall i = 1, \ldots, k, \quad \|\mathbf{v}_i\| = 1$$

Definition 5 (Orthonormal basis)

A basis of the subspace $W \subset \mathbb{R}^n$ is an **orthonormal basis** if the vectors composing it are an orthonormal set

$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ is orthonormal if

$$\mathbf{v}_i \bullet \mathbf{v}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Theorem 6

Let $Q \in \mathcal{M}_{mn}$. The columns of Q form an orthonormal set if and only if

$$Q^T Q = \mathbb{I}_n$$

Definition 7 (Orthogonal matrix)

$Q \in \mathcal{M}_n$ is an **orthogonal matrix** if its columns form an orthonormal set

So $Q \in \mathcal{M}_n$ orthogonal if $Q^T Q = \mathbb{I}$, i.e., $Q^T = Q^{-1}$

Theorem 8 (NSC for orthogonality)

$Q \in \mathcal{M}_n$ *orthogonal* $\iff Q^{-1} = Q^T$

Theorem 9 (Orthogonal matrices "encode" isometries)

Let $Q \in \mathcal{M}_n$. TFAE

1. *Q orthogonal*
2. $\forall \mathbf{x} \in \mathbb{R}^n$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$
3. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $Q\mathbf{x} \bullet Q\mathbf{y} = \mathbf{x} \bullet \mathbf{y}$

Theorem 10

Let $Q \in \mathcal{M}_n$ be orthogonal. Then

1. *The rows of Q form an orthonormal set*
2. *Q^{-1} orthogonal*
3. $\det Q = \pm 1$
4. $\forall \lambda \in \sigma(Q)$, $|\lambda| = 1$
5. *If $Q_2 \in \mathcal{M}_n$ also orthogonal, then QQ_2 orthogonal*

Proof of 4 in Theorem 10

All statements in Theorem 10 are easy, but let's focus on 4

Let λ be an eigenvalue of $Q \in \mathcal{M}_n$ orthogonal, i.e., $\exists \mathbb{R}^n \ni \mathbf{x} \neq \mathbf{0}$ s.t.

$$Q\mathbf{x} = \lambda \mathbf{x}$$

Take the norm on both sides

$$\|Q\mathbf{x}\| = \|\lambda \mathbf{x}\|$$

From 2 in Theorem 9, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ and from the properties of norms, $\|\lambda\mathbf{x}\| = |\lambda| \; \|\mathbf{x}\|$, so we have

$$\|Q\mathbf{x}\| = \|\lambda \mathbf{x}\| \iff \|\mathbf{x}\| = |\lambda| \; \|\mathbf{x}\| \iff 1 = |\lambda|$$

(we can divide by $\|\mathbf{x}\|$ since $\mathbf{x} \neq \mathbf{0}$ as an eigenvector)

Projections

Definition 11 (Orthogonal projection onto a subspace)

$W \subset \mathbb{R}^n$ a subspace and $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ an orthogonal basis of W.
$\forall \mathbf{v} \in \mathbb{R}^n$, the **orthogonal projection** of **v onto** W is

$$\mathrm{proj}_W(\mathbf{v}) = \frac{\mathbf{u}_1 \bullet \mathbf{v}}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \cdots + \frac{\mathbf{u}_k \bullet \mathbf{v}}{\|\mathbf{u}_k\|^2}\mathbf{u}_k$$

Definition 12 (Component orthogonal to a subspace)

$W \subset \mathbb{R}^n$ a subspace and $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ an orthogonal basis of W.
$\forall \mathbf{v} \in \mathbb{R}^n$, the **component** of **v orthogonal to** W is

$$\mathrm{perp}_W(\mathbf{v}) = \mathbf{v} - \mathrm{proj}_W(\mathbf{v})$$

We could spend a long time on these interesting notions, but let us get back to the QR decomposition

What this aims to do is to construct an orthogonal basis for a subspace $W \subset \mathbb{R}^n$

To do this, we use the *Gram-Schmidt orthogonalisation process*, which turn s a basis of W into an orthogonal basis of W

Gram-Schmidt process

Theorem 13

$W \subset \mathbb{R}^n$ a subset and $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ a basis of W. Let

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{v}_1 \bullet \mathbf{x}_2}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{v}_1 \bullet \mathbf{x}_3}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{v}_2 \bullet \mathbf{x}_3}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{x}_k - \frac{\mathbf{v}_1 \bullet \mathbf{x}_k}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \cdots - \frac{\mathbf{v}_{k-1} \bullet \mathbf{x}_k}{\|\mathbf{v}_{k-1}\|^2}\mathbf{v}_{k-1}
\end{aligned}$$

and

$$W_1 = \mathrm{span}(\mathbf{x}_1), W_2 = \mathrm{span}(\mathbf{x}_1, \mathbf{x}_2), \ldots, W_k = \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

Then $\forall i = 1, \ldots, k$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ orthogonal basis for W_i

The QR factorisation

Theorem 14

Let $A \in \mathcal{M}_{mn}$ with LI columns. Then A can be factored as

$$A = QR$$

where $Q \in \mathcal{M}_{mn}$ has orthonormal columns and $R \in \mathcal{M}_n$ is nonsingular upper triangular

Back to least squares

So what was the point of all that..?

Theorem 15 (Least squares with QR factorisation)

$A \in \mathcal{M}_{mn}$ with LI columns, $\mathbf{b} \in \mathbb{R}^m$. If $A = QR$ is a QR factorisation of A, then the unique least squares solution $\tilde{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is

$$\tilde{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$$

Proof of Theorem 15

A has LI columns so

- least squares $A\mathbf{x} = \mathbf{b}$ has unique solution $\tilde{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$
- by Theorem 14, A can be written as $A = QR$ with $Q \in \mathcal{M}_{mn}$ with orthonormal columns and $R \in \mathcal{M}_n$ nonsingular and upper triangular

So

$$
\begin{aligned}
A^T A\tilde{\mathbf{x}} = A^T \mathbf{b} &\implies (QR)^T QR\tilde{\mathbf{x}} = (QR)^T \mathbf{b} \\
&\implies R^T Q^T QR\tilde{\mathbf{x}} = R^T Q^T \mathbf{b} \\
&\implies R^T \mathbb{I}_n R\tilde{\mathbf{x}} = R^T Q^T \mathbf{b} \\
&\implies R^T R\tilde{\mathbf{x}} = R^T Q^T \mathbf{b} \\
&\implies (R^T)^{-1} R\tilde{\mathbf{x}} = (R^T)^{-1} R^T Q^T \mathbf{b} \\
&\implies R\tilde{\mathbf{x}} = Q^T \mathbf{b} \\
&\implies \tilde{\mathbf{x}} = R^{-1} Q^T \mathbf{b}
\end{aligned}
$$