Numerical Analysis Hilary Term 2025
Lecture 3: QR Factorization

So far the linear systems we treated had the same number of equations as unknowns (variables), so the problem was $Ax = b$ for a square matrix A. Very often in practice, we have more equations that we would like to satisfy than variables to fit them. It is then usually impossible to obtain a solution to $Ax = b$; a common approach is then to try minimise the difference between Ax and b. If we choose to minimise the Euclidean length of the vector, this leads to a *least-squares problem*:

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|, \qquad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ m \geq n, \tag{1}$$

where $\|y\| := \sqrt{y_1^2 + y_2^2 + \cdots + y_m^2} = \sqrt{y^T y}$. Least-squares problems (also known as *overdetermined* systems) are ubiquitous in applied mathematics and data science; linear regression is a basic example.

Our approach to solving the least-squares problem (1) is to find another factorisation that is suitable for rectangular matrices. In particular, we will use the QR-factorisation, which is the focus of this lecture.

Orthogonal matrices and sets

Definition: a square real matrix Q is **orthogonal** if $Q^\top = Q^{-1}$. This is true if, and only if, $Q^\top Q = I = QQ^\top$.

Example: The permutation matrices P in LU factorisation with partial pivoting are orthogonal.

Proposition. The product of orthogonal matrices is an orthogonal matrix.

Proof. If S and T are orthogonal, $(ST)^\top = T^\top S^\top$ so

$$(ST)^\top (ST) = T^\top S^\top ST = T^\top (S^\top S)T = T^\top T = I.$$

Definition: The **scalar (dot)(inner) product** of two vectors $x, y \in \mathbb{R}^n$ is

$$x^\top y = y^\top x = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}, \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Definition: Two vectors x, $y \in \mathbb{R}^n$ are **orthogonal** if $x^\top y = 0$. A set of vectors $\{u_1, u_2, \ldots, u_r\}$ is an **orthogonal set** if $u_i^\top u_j = 0$ for all $i, j \in \{1, 2, \ldots, r\}$ such that $i \neq j$. If an orthogonal set $\{u_1, u_2, \ldots, u_r\}$ additionally satisfies $\|u_i\| = 1$ for $i \in \{1, 2, \ldots, r\}$, then we say that the set is **orthonormal**.

Lemma. The columns of an orthogonal matrix Q form an orthonormal set, which is moreover an orthonormal basis for $\mathbb{R}^n$.

Proof. Suppose that $Q = [q_1 \;\; q_2 \;\; \cdots \;\; q_n]$, i.e., q_j is the jth column of Q. Then

$$I = Q^\top Q = \begin{bmatrix} q_1^\top \\ q_2^\top \\ \vdots \\ q_n^\top \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} q_1^\top q_1 & q_1^\top q_2 & \cdots & q_1^\top q_n \\ q_2^\top q_1 & q_2^\top q_2 & \cdots & q_2^\top q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^\top q_1 & q_2^\top q_n & \cdots & q_n^\top q_n \end{bmatrix}$$

Comparing the (i, j)th entries yields

$$q_i^\top q_j = (Q^\top Q)_{ij} = I_{ij} = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

Note that the columns of an orthogonal matrix are of length 1 as $q_i^\top q_i = 1$, so they form an orthonormal set.

To see that it forms a basis, let $x \in \mathbb{R}^n$ be any vector. One has $x = QQ^T x = Qc$ where $c = Q^T x$, so $x = \sum_{i=1}^n c_i q_i$. $\square$

Lemma. If $P \in \mathbb{R}^{n \times n}$ is orthogonal, then $\|Pu\| = \|u\|$ for all $u \in \mathbb{R}^n$.

Proof. $\|Pu\|^2 = (Pu)^\top Pu = u^\top (P^\top P)u = u^\top u = \|u\|^2$. $\square$

Definition: The **outer product** of two vectors x and $y \in \mathbb{R}^n$ is the n-by-n matrix

$$xy^\top = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{bmatrix}.$$

More usefully, if $z \in \mathbb{R}^n$, then

$$(xy^\top)z = xy^\top z = x(y^\top z) = \left(\sum_{i=1}^n y_i z_i \right) x.$$

Householder reflector

Definition: For $w \in \mathbb{R}^n$, $w \neq 0$, the **Householder** reflector $H(w) \in \mathbb{R}^{n \times n}$ is the matrix

$$H(w) = I - \frac{2}{w^\top w} ww^\top.$$

In particular, $H(w)x$ reflects $x \in \mathbb{R}^n$ over the plane $\mathcal{P}$ containing the origin that is normal to w. To see this, let $\{w, z_1, \ldots, z_{n-1}\}$ be an orthogonal set (note that such a set exists by e.g. Gram-Schmidt). Then, we have

$$x = \alpha_0 w + \sum_{i=1}^{n-1} \alpha_i z_i$$

for some coefficients $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{R}$. The reflection of x over $\mathcal{P}$ is then the vector $y \in \mathbb{R}^n$ defined by

$$y = -\alpha_0 w + \sum_{i=1}^{n-1} \alpha_i z_i.$$

(To convince yourself that this is correct notion of reflection, consider the case $n = 2$, $w = [1\ 0]^\top$ and $z = [0\ 1]^\top$.) Moreover, there holds

$$\begin{aligned} H(w)x &= \alpha_0 \left(I - \frac{2}{w^\top w} w w^\top \right) w + \sum_{i=1}^{n-1} \alpha_i \left(I - \frac{2}{w^\top w} w w^\top \right) z_i \\ &= \alpha_0 \left(w - \frac{2}{w^\top w} w (w^\top w) \right) + \sum_{i=1}^{n-1} \alpha_i \left(z_i - \frac{2}{w^\top w} w \underbrace{(w^\top z_i)}_{=0} \right) \qquad (2) \\ &= -\alpha_0 w + \sum_{i=1}^{n-1} \alpha_i z_i \\ &= y, \end{aligned}$$

where we used that w is orthogonal to z_i, $i \in \{1, \dots, n-1\}$.

The next two propositions concern some basic properties of $H(w)$.

Proposition. $H(w)$ is a symmetric orthogonal matrix.

Proof. Symmetry is straightforward to verify. For orthogonality,

$$\begin{aligned} H(w)H(w)^\top &= \left(I - \frac{2}{w^\top w} w w^\top \right) \left(I - \frac{2}{w^\top w} w w^\top \right) \\ &= I - \frac{4}{w^\top w} w w^\top + \frac{4}{(w^\top w)^2} w (w^\top w) w^\top \\ &= I. \end{aligned}$$

□

Proposition. The eigenpairs (eigenvalues, eigenvectors) of $H(w)$ are $(-1, w)$ and $(1, z_i)$ for $i \in \{1, \dots, n-1\}$. Consequently, $\det(H(w)) = -1$.

Proof. In line (2), we computed that $H(w)w = -w$ and $H(w)z_i = z_i$. □

The next result shows that, given a vector $u \in \mathbb{R}^n$, there exists a plane with normal $w \in \mathbb{R}^n$ so that the reflection of u over this plane coincides with the x_1 axis.

Lemma. Given $u \in \mathbb{R}^n$, there exists a $w \in \mathbb{R}^n$ such that

$$H(w)u = \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \qquad (3)$$

If $u = [\beta\ 0\ \cdots\ 0]^\top$, then $\alpha = -\sqrt{u^\top u}$. Otherwise, $\alpha = \pm\sqrt{u^\top u}$ depending on the choice of w.

Remark: Let $v \in \mathbb{R}^n$ denote the RHS of (3). Since $H(w)$ is an orthogonal matrix for any $w \in \mathbb{R}$, $w \neq 0$, it is necessary for the validity of the equality $H(w)u = v$ that $v^\top v = u^\top u$, i.e., $\alpha^2 = u^\top u$; hence we the only possible choice is $\alpha = \pm\sqrt{u^\top u}$.

Proof. Let $u \in \mathbb{R}^n$ and let $v \in \mathbb{R}^n$ denote the RHS of (3). For any fixed $\gamma \neq 0$, we define $w := \gamma(u - v)$. Note that our choice of α and γ ensures that $w \neq 0$ (particularly in the case that $u = [\beta \; 0 \; \cdots \; 0]$). Recall that $u^\top u = v^\top v$ because $H(w)$ is orthogonal. Thus,

$$\begin{aligned} w^\top w &= \gamma^2(u-v)^\top(u-v) = \gamma^2(u^\top u - 2u^\top v + v^\top v) \\ &= \gamma^2(u^\top u - 2u^\top v + u^\top u) = 2\gamma u^\top(\gamma(u-v)) \\ &= 2\gamma w^\top u, \end{aligned}$$

and so

$$H(w)u = \left(I - \frac{2}{w^\top w} w w^\top\right) u = u - \frac{2w^\top u}{w^\top w} w = u - \frac{1}{\gamma} w = u - (u - v) = v,$$

which completes the proof. □

Constructing the QR factorisation for square matrices

Let $A \in \mathbb{R}^{n\times n}$. Our goal is now to apply a sequence of Householder reflections to reduce A to an upper triangular matrix.

Applying the lemma with u being the first column of A, we obtain

$$H(w)A = \left[\begin{array}{c|c} \alpha & \times \; \cdots \; \times \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right], \quad \text{where } \times = \text{general entry.}$$

Similarly for B, we can find $\hat{w} \in \mathbb{R}^{n-1}$ such that

$$H(\hat{w})B = \left[\begin{array}{c|c} \beta & \times \; \cdots \; \times \\ \hline 0 & \\ \vdots & C \\ 0 & \end{array}\right],$$

and then

$$\left[\begin{array}{c|c} 1 & 0 \; \cdots \; 0 \\ \hline 0 & \\ \vdots & H(\hat{w}) \\ 0 & \end{array}\right] H(w)A = \left[\begin{array}{cc|c} \alpha & \times & \times \; \cdots \; \times \\ 0 & \beta & \times \; \cdots \; \times \\ \hline 0 & 0 & \\ \vdots & \vdots & C \\ 0 & 0 & \end{array}\right].$$

Note that

$$\begin{bmatrix} 1 & 0 \\ 0 & H(\hat{w}) \end{bmatrix} = H(w_2), \quad \text{where} \quad w_2 = \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix} \in \mathbb{R}^n.$$

Thus, if we continue in this manner for the $n-1$ steps, we obtain

$$\underbrace{H(w_{n-1})\cdots H(w_3)H(w_2)H(w)}_{Q^\top} A = \begin{bmatrix} \alpha & \times & \cdots & \times \\ 0 & \beta & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma \end{bmatrix}.$$

Note that for readability, we are abusing notation: Each w_i depends on the columns of the matrix from the previous step. The matrix $Q^\top$ is orthogonal as it is the product of orthogonal (Householder) matrices, so we have constructively sketched the proof of the following result:

Theorem. Given any square matrix A, there exists an orthogonal matrix Q and an upper triangular matrix R such that $A = QR$.

Notes:

1. The existence of the QR factorisation can also be established using the Gram–Schmidt Process.

2. If u is already of the form $[\alpha\ 0\ \cdots\ 0]^\top$, we do not need to use a reflection and can use the identity matrix I in place of H.

3. Householder reflectors can be applied to a vector in $O(n)$ flops; $4n$ (if $w^\top w$ does not need to be computed each application) or $6n-1$ (if $w^\top w$ is computed each application) to be precise. To see this, note that $Hv = (I - 2ww^T)v = v - 2w(w^T v)/(w^\top w)$. Using this, the QR factorisation can be computed in $O(n^3)$ flops. You should verify that the constant for QR is greater than $\frac{2}{3}$, the constant on the n^3 term for the flops for LU. Thus, QR is asymptotically more expensive than LU.

QR factorisation of rectangular matrices

Now suppose that A is not square: $A \in \mathbb{R}^{m\times n}$ with $m \neq n$.

(a) If $m < n$, then we can apply the product of $m-1$ Householder matrices (of dimension $m \times m$) to obtain

$$\begin{bmatrix} \times & \cdots & \cdots & \cdots & \times \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \times & \cdots & \cdots & \cdots & \times \end{bmatrix} = A = QR = \begin{bmatrix} \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \times & \cdots & \times \end{bmatrix} \begin{bmatrix} \times & \cdots & \cdots & \cdots & \times \\ & \ddots & \ddots & \ddots & \vdots \\ & & \times & \cdots & \times \end{bmatrix},$$

where $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{m\times n}$ and the omitted entries in R are all zero (below the main diagonal).

(b) Let $m > n$. One option is the *full* QR factorisation, obtained by applying $n-1$

Householder reflections:

$$
\begin{bmatrix} \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \times & \cdots & \times \end{bmatrix} = A = Q_F R_F = Q_F \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} \times & \cdots & \cdots & \cdots & \times \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \times & \cdots & \cdots & \cdots & \times \end{bmatrix} \begin{bmatrix} \times & \cdots & \times \\ & \ddots & \vdots \\ & & \times \\ & & \\ & & \end{bmatrix},
$$

where $Q_F \in \mathbb{R}^{m\times m}$ and $R_F \in \mathbb{R}^{m\times n}$. The matrix $R \in \mathbb{R}^{n\times n}$ is the first n rows of R_F.

(c) Let $m > n$. The other option is the *thin QR* factorisation, obtain by selecting the first n columns of Q_F and the upper triangular matrix R from (b):

$$
\begin{bmatrix} \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \times & \cdots & \times \end{bmatrix} = A = QR = \begin{bmatrix} \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \times & \cdots & \times \end{bmatrix} \begin{bmatrix} \times & \cdots & \times \\ & \ddots & \vdots \\ & & \times \end{bmatrix},
$$

where $Q \in \mathbb{R}^{m\times n}$ and $R \in \mathbb{R}^{n\times n}$. Note that Q has orthonormal columns since the matrix Q_F in (b) is orthogonal. In particular, we can write $Q_F = [Q\ Q_\perp]$, where $Q_\perp \in \mathbb{R}^{m\times(m-n)}$ has orthonormal columns.