IB Paper 7: Linear Algebra Handout 6

7. Least Squares Solution of Ax = b and QR factorisation

Suppose we have carried out an experiment, in which the parameter b has been measured at different times t,

$b = 0.25$ at $t = -1$

$b = 1.0$ at $t = 0$

$b = 1.25$ at $t = 1$

$b = 3.50$ at $t = 2$

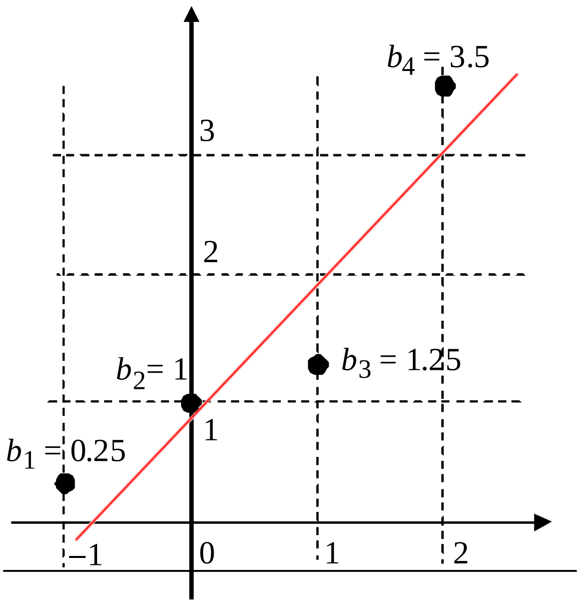
and that we are seeking to fit a relationship to the data:-

$$b = C + Dt$$

or for a quadratic fit

$$b = C + Dt + Et^2$$

where the constants C, D and E are to be found.



In matrix form the linear case is

general linear case

general quadratic case

$$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 0.25 \\ 1 \\ 1.25 \\ 3.5 \end{bmatrix} \qquad \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \cdots & \cdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix} \qquad \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \cdots & \cdots & \cdots \\ 1 & t_m & t_m^2 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix}$$

$$A \quad x = b$$

These equations are obviously *inconsistent* and there is no way C and D (and E) can be found to solve all of them. We need, instead, to find $\bar{\mathbf{x}}$ which represents in some sense "the best fit".

$$\mathbf{A}\bar{\mathbf{x}} \text{ as close as possible to } \mathbf{b}$$

Now the number of columns in **A** is the number of arbitrary constants in the function used for the fit, and the number of rows is the number of data points. For least squares problems, then, the $m \times n$ matrix **A** usually has the following properties:-

(i) $m > n$ (often $m >> n$)

(ii) the columns of **A** are independent. (rank of **A** is n.)

We will assume that (i) and (ii) hold.

The least squares solution for $\mathbf{x}$ $(= \bar{\mathbf{x}})$ minimises

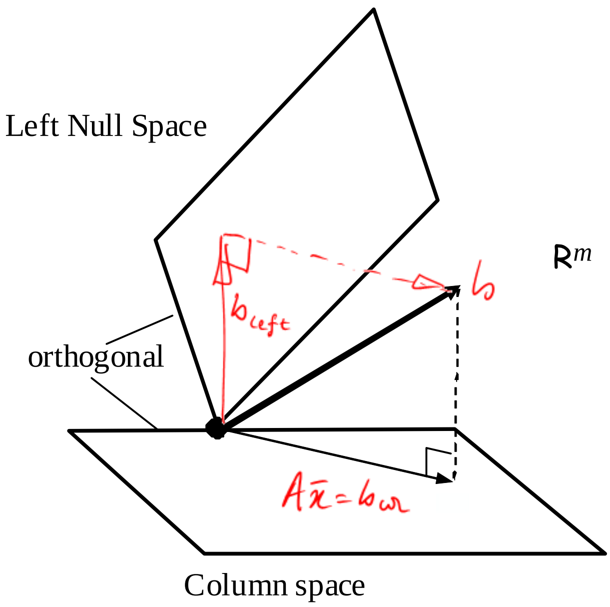
$$|\mathbf{Ax}-\mathbf{b}|^2 = (\mathbf{Ax}-\mathbf{b}) \,.\, (\mathbf{Ax}-\mathbf{b})$$

and this can be multiplied out and then partial differentiation used to find the minimum. A, perhaps more intuitive way, is based on geometrical reasoning.

This starts by noting that

$$\mathbf{A}\bar{\mathbf{x}} = \bar{x}_1\mathbf{a}_1 + \bar{x}_2\mathbf{a}_2 + \bar{x}_3\mathbf{a}_3 + \ldots$$

lies in column space, so the nearest point will be at the end of the "perpendicular" dropped from $\mathbf{b}$ onto column space.



We saw in sections 5.3 that column space and the left nullspace of $\mathbf{A}$ were orthogonal complements. i.e. that for *any* vector

$$\mathbf{b} = \mathbf{b}_{col} + \mathbf{b}_{left}$$

where $\quad \mathbf{b}_{col} \,.\, \mathbf{b}_{left} = 0$

So we need to get rid of $\mathbf{b}_{left}$ and just concentrate on $\mathbf{b}_{col}$. We can do this by multiplying the original problem by $\mathbf{A}^t$.

$$\mathbf{A}^t\mathbf{A}\mathbf{x} = \mathbf{A}^t\mathbf{b}$$

$$= A^t b_{col} + \underbrace{\cancel{A^t b_{left}}}_{\to\, 0}$$

The solution of this is $\bar{\mathbf{x}}$.

For the specific example described above

$$\mathbf{A}^t\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^t\mathbf{b} \;\Rightarrow\; \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 0.25 \\ 1 \\ 1.25 \\ 3.5 \end{bmatrix}$$

i.e. $\quad \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \end{bmatrix} \quad \Rightarrow \quad C = 1 \quad$ and $\quad D = 1$

Best fit is $\; b = 1 + t$

Were we lucky that $\mathbf{A}^t\mathbf{A}$ turned out to be invertible/non-singular ?

7.1 Useful properties of the matrix $\mathbf{A}^t\mathbf{A}$

$\mathbf{A}^t\mathbf{A}$ is a benign matrix as the following properties show. We can, therefore, always tackle least squares problems using $\mathbf{A}^t\mathbf{A}\,\overline{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$.

1) Since $\mathbf{A}$ is $m \times n$, $\mathbf{A}^t$ is $n \times m$ and $\mathbf{A}^t\mathbf{A}$ is $n \times n$. i.e. $\mathbf{A}^t\mathbf{A}$ is a *square* matrix. The equation

$$\mathbf{A}^t\mathbf{A}\,\overline{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$$

thus represents n equations for n unknowns. Good start, but does $(A^tA)^{-1}$ exist?

2) $\mathbf{A}$ has independent columns, i.e. the rank of $\mathbf{A} = r = n$ and we saw earlier that this is also the number of independent rows of $\mathbf{A}$. The dimension of row space is thus also n. This means that the dimension of the null-space of $\mathbf{A}$ is $n - r = 0$, so that $\mathbf{A}\,\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = \mathbf{0}$

3) $\mathbf{A}^t\mathbf{A}\,\mathbf{x} = \mathbf{0}$ $\Rightarrow x^tA^tAx = 0$

$\Rightarrow (Ax)^tAx = 0$ i.e. $|Ax|^2 = 0 \Rightarrow Ax = 0 \Rightarrow x = 0$

i.e. the dimension of the null-space of $\mathbf{A}^t\mathbf{A}$ is also 0.

If follows that the rank of $\mathbf{A}^t\mathbf{A}$ (= the dimension of column space of $\mathbf{A}^t\mathbf{A}$) is also n. i.e. The column space of $\mathbf{A}^t\mathbf{A}$ is the whole of R^n (which is another way of saying that the matrix has an inverse).

In summary, the least squares solution to an inconsistent system $\mathbf{Ax} = \mathbf{b}$ of m equations in n unknowns satisfies

$$\mathbf{A}^t\mathbf{A}\,\overline{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$$

Assuming that the columns of $\mathbf{A}$ are independent, $\mathbf{A}^t\mathbf{A}$ is invertible and

$$\overline{\mathbf{x}} = \left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{b}$$

We note in passing, that the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$ is therefore

$$\mathbf{b}_{col} = \mathbf{A}\,\overline{\mathbf{x}} = \mathbf{A}\left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{b}$$

The expressions for $\overline{\mathbf{x}}$ and $\mathbf{b}_{col}$ are a bit of a handful $\Rightarrow$we need another method.

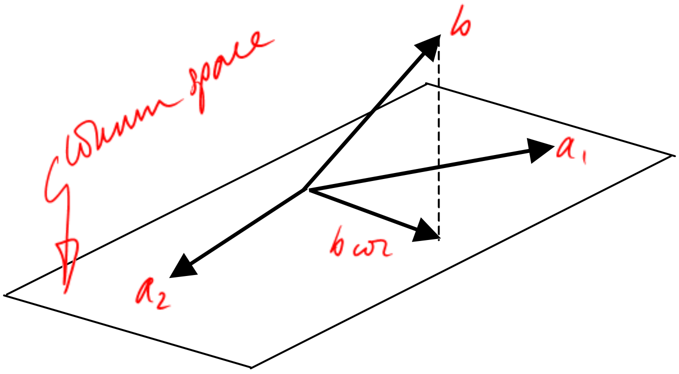
Note that $\left(\mathbf{A}^t\mathbf{A}\right)^{-1}$ is most certainly not $\mathbf{A}^{-1}\left(\mathbf{A}^t\right)^{-1}$

A is not square – it doesn't have an inverse

7.2 Orthogonal basis of Column Space - the Gram-Schmidt process.

The equation $\mathbf{A}^t\mathbf{A}\,\overline{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$ is fine, but we have to do quite a lot of work to follow through with this method when m and n are large (forming $\mathbf{A}^t\mathbf{A}$ alone takes $(2m-1)n^2$ operations, before we even set about solving the equation).

The reason for multiplying by $\mathbf{A}^t$ is so that we can remove the part of **b** that is not in the column space of **A**. Another way of doing this is to project **b** directly onto column space.



$$\mathbf{b}_{col} = \lambda_1\,\mathbf{a}_1 + \lambda_2\,\mathbf{a}_2 + \ldots$$

Finding the λ's, however, is a major exercise. Because the **a**'s are not orthogonal, dotting with $\mathbf{a}_1$, etc. doesn't help,

$$\mathbf{a}_1 \cdot \mathbf{b} = \lambda_1 a_1 . a_1 + \lambda_2 a_2 . a_2 + \ldots$$

If we do this with all of the **a**'s we will have a matrix to invert for the λ's.

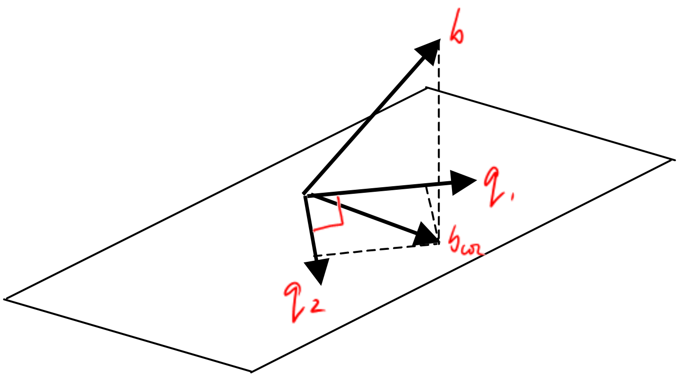
Think how much easier this would be if column space was aligned with our co-ordinate directions, so that, $\underline{i}$, $\underline{j}$, $\underline{k}$, $\underline{l}$ **...** lay in column space (and the other co-ordinate base vectors $\underline{m}$, $\underline{n}$, **...** lay in left-null space).

We would then write

$$\underline{b} = b_1\,\underline{i} + b_2\,\underline{j} + b_3\,\underline{k} + \ \ldots$$

and simply strip off the ones outside column space. Moreover, if we didn't have them already, we would generate the coefficients by

$$b_1 = \underline{i} \cdot \underline{b}, \quad b_2 = \underline{j} \cdot \underline{b} \text{ etc.}$$



The *Gram-Schmidt procedure* is a way of generating a set of mutually orthogonal unit vectors (orthogonal + unit = orthonormal) from an arbitrary set. Armed with these, taking projections is much easier.

$$b_{col} = \alpha_1 q_1 + \alpha_2 q_2 + \ldots$$

And to find the α's, we simply employ

$$q^1 . b = \alpha_1 \quad etc\ldots$$

We start with $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, the columns of **A** and derive the **q**'s as follows:-

1) Turn the first one into a unit vector $\mathbf{q}_1 = \dfrac{\mathbf{a}_1}{|\mathbf{a}_1|}$ q_1 is in column space

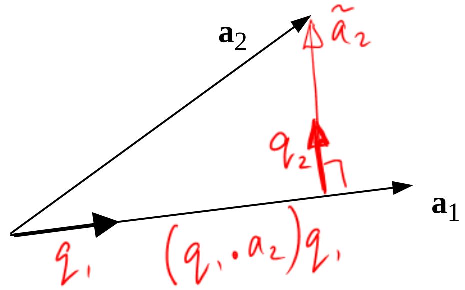
Remember, the notation $|\;|$ means the "length" of an n-dimensional vector

$|\mathbf{d}| = \sqrt{d_1^2 + d_2^2 + \ldots + d_n^2}$, generalised in the obvious fashion.

2) Take $\mathbf{a}_2$ and form $\mathbf{q}_2$ by first subtracting off the bit that's parallel to $\mathbf{a}_1$ and then normalising

$$\tilde{\mathbf{a}}_2 = \mathbf{a}_2 - (\mathbf{q}_1.\mathbf{a}_2)\mathbf{q}_1$$

$$\mathbf{q}_2 = \frac{\tilde{\mathbf{a}}_2}{|\tilde{\mathbf{a}}_2|}$$



Check

$$q_1 \cdot \tilde{a}_2 = q_1 \cdot a_2 - (q_1 \cdot a_2)\, q_1 \cdot q_1 = 0$$

$$\Rightarrow q_1 \cdot q_2 = 0$$

$\mathbf{q}_2$ is in column space because it is a linear combination of $\mathbf{a}_2$ and $\mathbf{q}_1$

3) Repeat this process for the other $\mathbf{a}$'s.

$$\tilde{\mathbf{a}}_3 = \mathbf{a}_3 - (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 - (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2$$

$$\mathbf{q}_3 = \frac{\tilde{\mathbf{a}}_3}{|\tilde{\mathbf{a}}_3|} \qquad \text{etc.}$$

Check

$$\mathbf{q}_1 .\tilde{\mathbf{a}}_3 = \mathbf{q}_1 .\mathbf{a}_3 - (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 \cdot \mathbf{q}_1 - (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_1 \cdot \mathbf{q}_2 = \mathbf{0} \quad \Rightarrow \quad \mathbf{q}_1 .\mathbf{q}_3 = \mathbf{0}$$

$$q_2 \cdot \tilde{a}_3 = q_2 \cdot a_3 - (q_1 \cdot a_3)\, q_2 \cdot q_1 - (q_2 \cdot a_3)\, q_2 \cdot q_2 = 0 \quad \Rightarrow q_2 \cdot q_3 = 0$$

Note that, since the columns of **A** are independent, we never have $\tilde{\mathbf{a}}_k = \mathbf{0}$. So this *Gram-Schmidt orthogonalisation process*, will always furnish an orthonormal set of n vectors.

Any vector in the column space can, by definition, be written as a linear combination of the **a**'s and so as a linear combination of the **q**'s. i.e. $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ is an (orthonormal) basis for the column space of **A**.

Example

Perform Gram-Schmidt orthogonalisation on $\mathbf{a}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}$ $\mathbf{a}_2 = \begin{bmatrix}1\\0\\0\end{bmatrix}$ $\mathbf{a}_3 = \begin{bmatrix}2\\1\\0\end{bmatrix}$

1) $\mathbf{q}_1 = \dfrac{\mathbf{a}_1}{|\mathbf{a}_1|} = \begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}$ i.e. $\begin{bmatrix}1\\0\\1\end{bmatrix} = \sqrt{2}\begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}$

2) Subtract off the bit of $\mathbf{a}_2$ that is parallel to $\mathbf{q}_1$ and then create a unit vector

$$\tilde{\mathbf{a}}_2 = \mathbf{a}_2 - (\mathbf{q}_1.\mathbf{a}_2)\mathbf{q}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix} - \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}\frac{1}{2}\\0\\-\frac{1}{2}\end{bmatrix} = 1/\sqrt{2}\begin{bmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{bmatrix} \qquad \mathbf{q}_2 = \begin{bmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{bmatrix}$$

3) Subtract off the bits of $\mathbf{a}_3$ that are parallel to $\mathbf{q}_1$ and $\mathbf{q}_2$ and then create a unit vector

$$\tilde{\mathbf{a}}_3 = \mathbf{a}_3 - (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 - (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2 = \begin{bmatrix}2\\1\\0\end{bmatrix} - \sqrt{2}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} - \sqrt{2}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\-\frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}0\\1\\0\end{bmatrix} = q_3$$

In preparation for what is coming next, let us rewrite this as a relationship between the **a**'s and the **q**'s in the form

$$\begin{bmatrix}1\\0\\1\end{bmatrix} = \sqrt{2}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} \qquad \begin{bmatrix}1\\0\\0\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} + \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\-\frac{1}{\sqrt{2}}\end{bmatrix}$$

$$\begin{bmatrix}2\\1\\0\end{bmatrix} = \sqrt{2}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} + \sqrt{2}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\-\frac{1}{\sqrt{2}}\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix}$$

7.3 QR factorisation of A

If we assemble the three vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ from the previous section as the columns of a matrix **A**, and vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ as those of a matrix **Q**, then we have

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

Then
$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$$

i.e. We have constructed another matrix factorisation

$$\boxed{\mathbf{A} = \mathbf{Q}\,\mathbf{R}}$$

The matrix **Q** has mutually orthogonal unit vectors and the matrix **R** is upper triangular.

Before writing down the general form of this factorisation (we have done it for a 3 × 3 one), we can tidy up the relationship between the **a**'s and the **q**'s. You can see that $\mathbf{a}_3$ for example satisfies

$$\mathbf{a}_3 = (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2 + \tilde{\mathbf{a}}_3$$

$$= (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2 + |\tilde{\mathbf{a}}_3|\mathbf{q}_3$$

Taking the dot product with $\mathbf{q}_3$ gives a neater formula for $|\tilde{\mathbf{a}}_3|$

$$\mathbf{q}_3 . \mathbf{a}_3 = |\tilde{\mathbf{a}}_3|$$

so that

$$\mathbf{a}_3 = (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2 + (\mathbf{q}_3.\mathbf{a}_3)\,\mathbf{q}_3$$

The general formula is clear

$$\mathbf{a}_1 = (\mathbf{q}_1.\mathbf{a}_1)\mathbf{q}_1$$

$$\mathbf{a}_2 = (\mathbf{q}_1.\mathbf{a}_2)\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{a}_2)\mathbf{q}_2$$

$$\mathbf{a}_3 = (\mathbf{q}_1.\mathbf{a}_3)\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{a}_3)\mathbf{q}_2 + (\mathbf{q}_3.\mathbf{a}_3)\,\mathbf{q}_3$$

etc.

Writing the Gram-Schmidt process as a relationship between matrices (see Section 2.6) :-

$$\mathbf{A} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & & \mathbf{a}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}$$

Then

$$\mathbf{A} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & & \mathbf{a}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} \mathbf{q}_1.\mathbf{a}_1 & \mathbf{q}_1.\mathbf{a}_2 & \mathbf{q}_1.\mathbf{a}_3 & \cdots & \mathbf{q}_1.\mathbf{a}_n \\ 0 & \mathbf{q}_2.\mathbf{a}_2 & \mathbf{q}_2.\mathbf{a}_3 & \cdots & \mathbf{q}_2.\mathbf{a}_n \\ 0 & 0 & \mathbf{q}_3.\mathbf{a}_3 & \cdots & \mathbf{q}_2.\mathbf{a}_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \mathbf{q}_n.\mathbf{a}_n \end{bmatrix}$$

i.e. $\qquad \mathbf{A} \qquad = \qquad \mathbf{Q} \qquad\qquad \mathbf{R}$

For the general case $\quad A = m \times n$, Q is same shape as A $(m \times n)$. R square $(n \times n)$

works for any A provided $\text{rank}(A) = n$

The columns of **Q** are mutually orthogonal vectors which span the column space of **A**.

The matrix **R** is square, upper triangular with non-zero elements down the diagonal. It therefore has rank n and is invertible. See section 4.2 where we discussed this issue for **L**.

7.4 The Matrix Q

We met *square* matrices like **Q** in Part IA Maths and studied all of their properties. We have to be careful here, because these matrices are in general *rectangular* with $m > n$. It is still true that

$$\underbrace{\mathbf{Q}^t}_{n \times m}\underbrace{\mathbf{Q}}_{m \times n} = \underbrace{\mathbf{I}}_{n \times n}$$

since $\mathbf{Q}^t\mathbf{Q} = \begin{bmatrix} \leftarrow & \mathbf{q}_1 & \rightarrow \\ \leftarrow & \mathbf{q}_2 & \rightarrow \\ \cdots & \cdots & \cdots \\ \leftarrow & \mathbf{q}_n & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 \cdot \mathbf{q}_1 & \mathbf{q}_1 \cdot \mathbf{q}_2 & \cdots & \mathbf{q}_1 \cdot \mathbf{q}_n \\ \mathbf{q}_2 \cdot \mathbf{q}_1 & \mathbf{q}_2 \cdot \mathbf{q}_2 & \cdots & \mathbf{q}_2 \cdot \mathbf{q}_n \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{q}_n \cdot \mathbf{q}_1 & \mathbf{q}_n \cdot \mathbf{q}_2 & \cdots & \mathbf{q}_n \cdot \mathbf{q}_n \end{bmatrix}$

and the **q**'s are orthogonal unit vectors.

$$q_i \cdot q_j = 0 \quad i \neq j$$
$$= 1 \quad i = j$$

But this does *not* imply that $\mathbf{Q}^{-1} = \mathbf{Q}^t$. *These matrices are not square* (*in general*)

Q *does not have an inverse* (*in general*)

Note also that $\quad \mathbf{QQ}^t \neq \mathbf{I} \quad$ unless Q is square

7.5 Simplification of the Least Squares solution to Ax = b

This is now much less effort using the QR decomposition. Given the set of equations **Ax** = **b**, where **A** is an $m \times n$ matrix whose columns are independent ($m \geq n$ and rank of **A** = n) then the least squares solution satisfies

$$\mathbf{A}^t\mathbf{A}\,\bar{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$$

$$\Rightarrow (QR)^t QR\bar{x} = (QR)^t b$$

$$\Rightarrow R^t Q^t QR\bar{x} = R^t Q^t b$$

But $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$, so that $\qquad \mathbf{R}^t\mathbf{R}\,\bar{\mathbf{x}} = \mathbf{R}^t\mathbf{Q}^t\mathbf{b}$

Further, the square matrix **R** is invertible, which means that so is $\mathbf{R}^t$. It follows that

$$\boxed{\mathbf{R}\,\bar{\mathbf{x}} = \mathbf{Q}^t\mathbf{b}}$$

The right hand side is simply a matrix multiplying a vector and the solution for $\bar{\mathbf{x}}$ is found by back-substitution (**R** is an upper triangular matrix).

Example

Find the least squares solution for the problem at the beginning of section 7

$$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 0.25 \\ 1 \\ 1.25 \\ 3.5 \end{bmatrix}$$

Step 1: QR decomposition of A

1) $\qquad \mathbf{a}_1 = 2\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = 2\mathbf{q}_1 \qquad \mathbf{q}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$

2) $\qquad \tilde{\mathbf{a}}_2 = \mathbf{a}_2 - (\mathbf{q}_1.\mathbf{a}_2)\mathbf{q}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} - 1\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -3/2 \\ -1/2 \\ 1/2 \\ 3/2 \end{bmatrix} = \sqrt{5}\,\mathbf{q}_2 \qquad \mathbf{q}_2 = \begin{bmatrix} -\frac{3}{2\sqrt{5}} \\ -\frac{1}{2\sqrt{5}} \\ \frac{1}{2\sqrt{5}} \\ \frac{3}{2\sqrt{5}} \end{bmatrix}$

A = Q R

3) $$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{3}{2\sqrt{5}} \\ \frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ \frac{1}{2} & \frac{1}{2\sqrt{5}} \\ \frac{1}{2} & \frac{3}{2\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & \sqrt{5} \end{bmatrix}$$

Step 2

Solve $\mathbf{R}\bar{\mathbf{x}} = \mathbf{Q}^{t}\mathbf{b}$ by back-substitution.

$$\begin{bmatrix} 2 & 1 \\ 0 & \sqrt{5} \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{3}{2\sqrt{5}} & -\frac{1}{2\sqrt{5}} & \frac{1}{2\sqrt{5}} & \frac{3}{2\sqrt{5}} \end{bmatrix} \begin{bmatrix} 0.25 \\ 1 \\ 1.25 \\ 3.5 \end{bmatrix} = \begin{bmatrix} 3 \\ \sqrt{5} \end{bmatrix} \qquad C = 1, \quad D = 1$$

7.6 Operation Count and Robustness of QR

QR factorisation is more costly than LU decomposition (the cost is primarily in the Gram-Schmidt process). LU is thus preferable for solving sets of *consistent* equations. For *inconsistent* equations (i.e. a genuine least-squares problem), QR is more cost effective than solving $\mathbf{A}^{t}\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^{t}\mathbf{b}$ by LU decomposition. In addition, the matrix $\mathbf{A}^{t}\mathbf{A}$ is often numerically poorly conditioned, so it is a not a good idea to go via $\mathbf{A}^{t}\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^{t}\mathbf{b}$.

The Gram-Schmidt process can, for large *n*, become ill-conditioned (you are finding the **q**'s by a process of subtracting a large number of things and then normalising to unity). There are other ways of finding a **Q**, but these are beyond the scope of this course.

7.7 Projection onto Column Space

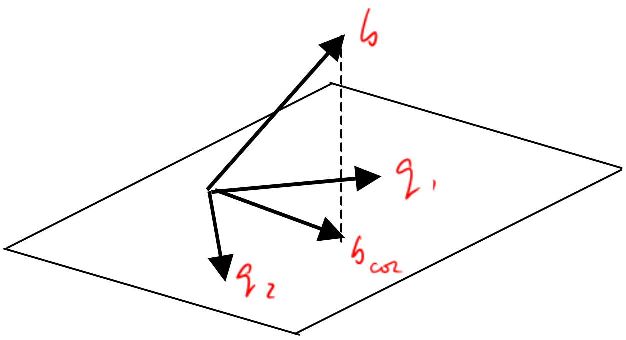
In section 7.1, we showed that $\mathbf{b}_{col}$, the projection of **b** onto the column space of **A**, satisfies

$$\mathbf{b}_{col} = \mathbf{A}\left(\mathbf{A}^{t}\mathbf{A}\right)^{-1}\mathbf{A}^{t}\mathbf{b}$$

i.e. $\mathbf{P}\mathbf{b} = \mathbf{b}_{col}$ where the *projection matrix* **P** is given by

$$\mathbf{P} = \mathbf{A}\left(\mathbf{A}^{t}\mathbf{A}\right)^{-1}\mathbf{A}^{t}$$

This rather complicated expression for **P**, was the reason that we developed the QR method.



There are a number of other applications were it is useful to be able to easily project onto column space, and the QR decomposition should give us a much simpler expression for this projection.

If we have performed the decomposition **A** = **QR**, then

$$\mathbf{P} = QR(R^tQ^tQR)^{-1}R^tQ^t$$
$$= QR(R^tR)^{-1}R^tQ^t$$
$$= QRR^{-1}(R^t)^{-1}R^tQ^t$$

R is square & upper triangular so $(R^t)^{-1}$ and R^{-1} exist

i.e. $\mathbf{P} = QQ^t$

N.B. $Q^tQ = I$ $\quad P = QQ^t \neq I$

This is as expected (!) since

$$\mathbf{b}_{col} = (\mathbf{q}_1.\mathbf{b})\mathbf{q}_1 + (\mathbf{q}_2.\mathbf{b})\mathbf{q}_2 + \ldots + (\mathbf{q}_n.\mathbf{b})\,\mathbf{q}_n$$

$$\mathbf{b}_{col} = \begin{bmatrix} \uparrow & \uparrow & \ldots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & & \mathbf{q}_n \\ \downarrow & \downarrow & \ldots & \downarrow \end{bmatrix} \begin{bmatrix} \mathbf{q}_1.\mathbf{b} \\ \mathbf{q}_2.\mathbf{b} \\ \ldots \\ \mathbf{q}_n.\mathbf{b} \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \ldots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & & \mathbf{q}_n \\ \downarrow & \downarrow & \ldots & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{q}_1 & \rightarrow \\ \leftarrow & \mathbf{q}_2 & \rightarrow \\ \ldots & \ldots & \ldots \\ \leftarrow & \mathbf{q}_n & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \mathbf{b} \\ \downarrow \end{bmatrix} = \mathbf{QQ}^t\mathbf{b}$$

You can now do Examples Paper 2 Q1-3

Key Points from Lecture

QR Decomposition

A = **QR**, where **Q** us the same shape as **A** and the columns of **Q** are orthonormal, and **R** is square, upper-triangular and invertible. When $m = n$ and so all matrices are square, **Q** is an orthogonal matrix.

Least squares solution of Ax = b using QR

Solve $\mathbf{R}\,\bar{\mathbf{x}} = \mathbf{Q}^t\mathbf{b}$ by back-substitution.