Optional study materials

- Approximate solutions: least-squares method
- Orthogonal factorisation of matrices (QR)
- Application of QR to least-squares method

- Scalar product $\mathbf{u} \cdot \mathbf{v} \equiv \mathbf{u}^\intercal \mathbf{v} = \sum_i \mathbf{u}_i \mathbf{v}_i$
- Norm: $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ and distance: $\mathrm{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
- Orthogonal / orthonormal vectors, sets, bases
- Orthogonal complements $\dim W^\perp + \dim W = n$
- Orthogonal projections and decompositions
- Gram-Schmidt process: $\mathbf{v}_1 = \mathbf{x}_1$, then

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

Theorem: *(the best approximation theorem)*

Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$, and $\check{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$.
Then $\check{\mathbf{y}}$ is the point in W closest to $\mathbf{y}$ in the sense that

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

The distance from $\mathbf{y}$ to W is given by $\|\check{\mathbf{y}} - \mathbf{y}\|$.

Notes:

- $\check{\mathbf{y}}$ is called the *best approximation* to $\mathbf{y}$ by elements of W.
- In a sense, we try to approximate $\mathbf{y}$ by some 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 $\mathbf{v} = \check{\mathbf{y}}$.

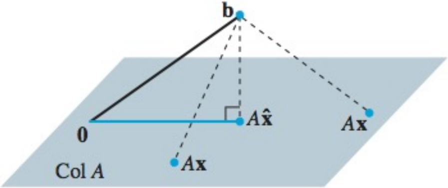
Definition: For an $m \times n$ matrix $\mathbf{A}$ and given $\mathbf{b} \in \mathbb{R}^m$, a *least-squares solution* of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\check{\mathbf{x}} \in \mathbb{R}^n$ such that $\|\mathbf{b} - \mathbf{A}\check{\mathbf{x}}\| \leqslant \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \quad \forall \mathbf{x} \in \mathbb{R}^n$.

Take $\check{\mathbf{b}} = \mathrm{proj}_{(\mathrm{Col}\,\mathbf{A})}\,\mathbf{b}$, then $\check{\mathbf{b}} \in \mathrm{Col}\,\mathbf{A}$ and $\exists\,\check{\mathbf{x}} : \ \mathbf{A}\check{\mathbf{x}} = \check{\mathbf{b}}$.

In $\mathrm{Col}\,\mathbf{A}$, $\check{\mathbf{b}}$ is the closest to $\mathbf{b}$, so $\check{\mathbf{x}}$ is a least-square solution.

Via orthogonal decomposition, $\mathbf{b} - \check{\mathbf{b}}$ is orthogonal to $\mathrm{Col}\,\mathbf{A}$, so

$$\begin{aligned} \mathbf{A}^{\mathsf{T}}(\mathbf{b} - \check{\mathbf{b}}) &= \mathbf{0} \\ \mathbf{A}^{\mathsf{T}}(\mathbf{b} - \mathbf{A}\check{\mathbf{x}}) &= \mathbf{0} \\ \mathbf{A}^{\mathsf{T}}\mathbf{b} - \mathbf{A}^{\mathsf{T}}\mathbf{A}\check{\mathbf{x}} &= \mathbf{0} \\ \mathbf{A}^{\mathsf{T}}\mathbf{A}\check{\mathbf{x}} &= \mathbf{A}^{\mathsf{T}}\mathbf{b} \end{aligned}$$



System $\mathbf{A}^{\mathsf{T}}\mathbf{A}\check{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$ is the *normal system* for $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Its non-empty solution (set) is the least-squares solution (set).

Example: Find a least-squares solution of the inconsistent system

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Solution:

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$\mathbf{A}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then

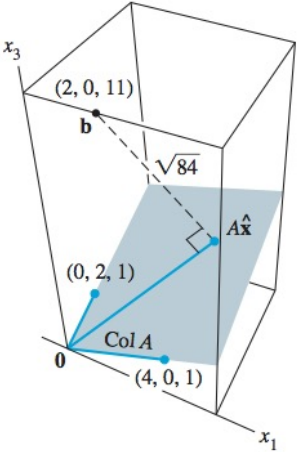
$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \check{\mathbf{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix} \qquad \Rightarrow \qquad \check{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

With $\check{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ being the least-squares solution, we can also calculate the *least-squares error* $\|\mathbf{b} - \mathbf{A}\check{\mathbf{x}}\|$:

$$\mathbf{A}\check{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

$$\mathbf{b} - \mathbf{A}\check{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

so $\|\mathbf{b} - \mathbf{A}\check{\mathbf{x}}\| = \sqrt{84}$.



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 $\operatorname{Col} \mathbf{A}$, 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, $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$.

Thus $\mathbf{Q}^\mathsf{T}\mathbf{A} = \mathbf{Q}^\mathsf{T}(\mathbf{QR}) = (\mathbf{Q}^\mathsf{T}\mathbf{Q})\mathbf{R} = \mathbf{IR} = \mathbf{R}$,

therefore $\mathbf{R} = \mathbf{Q}^\mathsf{T}\mathbf{A}$ (which makes it easy to calculate).

Proof: by construction

Proof: *(QR factorisation)*

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

An orthonormal basis $\{\mathbf{u}_1, \ldots \mathbf{u}_n\}$ for $\operatorname{Col}\mathbf{A}$ can be constructed by using the Gram-Schmidt process; denote $\mathbf{Q} = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$.

Then $\forall\, k = 1 \ldots n$: $\quad \mathbf{a}_k \in \operatorname{Span}\{\mathbf{a}_1, \ldots \mathbf{a}_k\} = \operatorname{Span}\{\mathbf{u}_1, \ldots \mathbf{u}_k\}$,
(as provided by the Gram-Schmidt process).

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

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

(note that $r_{kk} \neq 0$ due to the Gram-Schmidt algorithm).

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

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

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

Using $\mathbf{r}_k$ vectors, matrix $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 \ldots \mathbf{r}_n \end{bmatrix}$ is formed. 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 $\mathbf{A}\mathbf{x} = \mathbf{0}$).

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

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

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

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

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

So the QR factorisation is:

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

Sometimes, $\mathbf{A}^\mathsf{T}\mathbf{A}$ may be sensitive to round-off errors.
There is an alternative way to obtain least-squares solutions.

If $\mathbf{A}$ is an $m \times n$ matrix with linearly independent columns, it can be QR-factorised as $\mathbf{A} = \mathbf{Q}\mathbf{R}$.

Then $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique least-squares solution $\forall\, \mathbf{b} \in \mathbb{R}^m$:

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

In practice, $\check{\mathbf{x}}$ is obtained by solving

$$\mathbf{R}\check{\mathbf{x}} = \mathbf{Q}^\mathsf{T}\mathbf{b}$$

which is straightforward since $\mathbf{R}$ is upper-triangular.

Example (same matrix as for the QR example we had):

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

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

$$\mathbf{Q}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 1 \\ 2/\sqrt{12} \\ 2/\sqrt{6} \end{bmatrix} \quad \text{so} \quad \left[\begin{array}{ccc|c} 2 & 3/2 & 1 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} & 2/\sqrt{6} \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

thus $\check{\mathbf{x}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and, given that $\mathbf{A}\check{\mathbf{x}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, the error is $\|\mathbf{b} - \mathbf{A}\check{\mathbf{x}}\| = \sqrt{2}$.

Good luck with your further studies!