Using the QR factorization to solve a matrix equation $A\mathbf{x} = \mathbf{b}$

First suppose A is square and its columns are linearly independent.
Then A is invertible.
It follows that there is a solution (because we can write $\mathbf{x} = A^{-1}\mathbf{b}$)
QR Solver Algorithm to find the solution in this case:

Find Q, R such that $A = QR$ and Q is column-orthogonal and R is triangular Compute vector $\mathbf{c} = Q^T\mathbf{b}$ Solve $R\mathbf{x} = \mathbf{c}$ using backward substitution, and return the solution.

Why is this correct?

- Let $\hat{\mathbf{x}}$ be the solution returned by the algorithm.
- We have $R\hat{\mathbf{x}} = Q^T\mathbf{b}$
- Multiply both sides by Q: $Q(R\hat{\mathbf{x}}) = Q(Q^T\mathbf{b})$
- Use associativity: $(QR)\hat{\mathbf{x}} = (QQ^T)\mathbf{b}$
- Substitute A for QR: $A\hat{\mathbf{x}} = (QQ^T)\mathbf{b}$
- Since Q and Q^T are inverses, we know QQ^T is identity matrix: $A\hat{\mathbf{x}} = \mathbb{1}\mathbf{b}$

Thus $A\hat{\mathbf{x}} = \mathbf{b}$.

Solving $A\mathbf{x} = \mathbf{b}$

What if columns of A are not independent?

Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ be columns of A.

Suppose $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are linearly dependent.

Then there is a basis consisting of a subset, say $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$

$$\left\{ \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \; : \; x_1, x_2, x_3, x_4 \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix} \; : \; x_1, x_2, x_4 \in \mathbb{R} \right\}$$

The least squares problem

Suppose A is an $m \times n$ matrix and its columns are linearly independent.

Since each column is an m-vector, dimension of column space is at most m, so $n \leq m$.

What if $n < m$? How can we solve the matrix equation $A\mathbf{x} = \mathbf{b}$?

Remark: There might not be a solution:

- Define $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $f(\mathbf{x}) = A\mathbf{x}$
- Dimension of $\operatorname{Im} f$ is n
- Dimension of co-domain is m.
- Thus f is not onto.

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{b}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{b}$$

Goal: An algorithm that, given equation $A\mathbf{x} = \mathbf{b}$, where columns are linearly independent, finds the vector $\hat{\mathbf{x}}$ minimizing $\|\mathbf{b} - A\hat{\mathbf{x}}\|$.

Solution: Same algorithm as we used for square A

The least squares problem

Recall...

High-Dimensional Fire Engine Lemma: The point in a vector space $\mathcal{V}$ closest to $\mathbf{b}$ is $\mathbf{b}^{||\mathcal{V}}$ and the distance is $\|\mathbf{b}^{\perp\mathcal{V}}\|$.

Given equation $A\mathbf{x} = \mathbf{b}$, let $\mathcal{V}$ be the column space of A.

We need to show that the QR Solver Algorithm returns the vector $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \mathbf{b}^{||\mathcal{V}}$.

The least squares problem

Suppose A is an $m \times n$ matrix and its columns are linearly independent.

Since each column is an m-vector, dimension of column space is at most m, so $n \leq m$.

What if $n < m$? How can we solve the matrix equation $A\mathbf{x} = \mathbf{b}$?

Remark: There might not be a solution:

- Define $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $f(\mathbf{x}) = A\mathbf{x}$
- Dimension of Im f is n
- Dimension of co-domain is m.
- Thus f is not onto.

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{b}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{b}$$

Goal: An algorithm that, given a matrix A whose columns are linearly independent and given $\mathbf{b}$, finds the vector $\hat{\mathbf{x}}$ minimizing $\|\mathbf{b} - A\hat{\mathbf{x}}\|$.

Solution: Same algorithm as we used for square A

The least squares problem

Recall...

High-Dimensional Fire Engine Lemma: The point in a vector space $\mathcal{V}$ closest to $\mathbf{b}$ is $\mathbf{b}^{||\mathcal{V}}$ and the distance is $\|\mathbf{b}^{\perp\mathcal{V}}\|$.

Given equation $A\mathbf{x} = \mathbf{b}$, let $\mathcal{V}$ be the column space of A.

We need to show that the QR Solver Algorithm returns $\mathbf{b}^{||\mathcal{V}}$.

Representation of $\mathbf{b}^{||}$ in terms of columns of Q

Let Q be a column-orthogonal matrix. Let $\mathbf{b}$ be a vector, and write $\mathbf{b} = \mathbf{b}^{||} + \mathbf{b}^{\perp}$ where $\mathbf{b}^{||}$ is projection of $\mathbf{b}$ onto Col Q and $\mathbf{b}^{\perp}$ is projection orthogonal to Col Q.

Let $\mathbf{u}$ be the coordinate representation of $\mathbf{b}^{||}$ in terms of columns of Q.

By linear-combinations definition of matrix-vector multiplication,

$$\begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix} \begin{bmatrix} \mathbf{u} \end{bmatrix}$$

Multiply both sides on the left by Q^T:

$$\begin{bmatrix} & Q^T & \end{bmatrix} \begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix} = \begin{bmatrix} & Q^T & \end{bmatrix} \begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix} \begin{bmatrix} \mathbf{u} \end{bmatrix}$$

Representation of $\mathbf{b}^{||}$ in terms of columns of Q

Let Q be a column-orthogonal matrix. Let $\mathbf{b}$ be a vector, and write $\mathbf{b} = \mathbf{b}^{||} + \mathbf{b}^{\perp}$ where $\mathbf{b}^{||}$ is projection of $\mathbf{b}$ onto Col Q and $\mathbf{b}^{\perp}$ is projection orthogonal to Col Q.

Let $\mathbf{u}$ be the coordinate representation of $\mathbf{b}^{||}$ in terms of columns of Q.

Multiply both sides on the left by Q^T:

$$\begin{bmatrix} & & \\ & Q^T & \\ & & \end{bmatrix} \begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & Q^T & \\ & & \end{bmatrix} \begin{bmatrix} \\ Q \\ \\ \end{bmatrix} \begin{bmatrix} \mathbf{u} \end{bmatrix}$$

Substitute using $Q^T Q = \mathbb{1}$

$$\begin{bmatrix} & & \\ & Q^T & \\ & & \end{bmatrix} \begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix} = \mathbb{1} \begin{bmatrix} \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \end{bmatrix}$$

Representation of $\mathbf{b}^{||}$ in terms of columns of Q

Let Q be a column-orthogonal matrix. Let $\mathbf{b}$ be a vector, and write $\mathbf{b} = \mathbf{b}^{||} + \mathbf{b}^{\perp}$ where $\mathbf{b}^{||}$ is projection of $\mathbf{b}$ onto Col Q and $\mathbf{b}^{\perp}$ is projection orthogonal to Col Q.

Let $\mathbf{u}$ be the coordinate representation of $\mathbf{b}^{||}$ in terms of columns of Q.

- $Q^T \mathbf{b}^{||} = \mathbf{u}$

Since $\mathbf{b}^{\perp}$ is orthogonal to Col Q,

$$\mathbf{q}_i \cdot \mathbf{b}^{\perp} = 0 \text{ for every column } \mathbf{q}_i \text{ of } Q$$

Therefore, by dot-product definition of matrix-vector multiplication,

$$\left[\quad Q^T \quad \right] \left[\; \mathbf{b}^{\perp} \; \right] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Representation of $\mathbf{b}^{||}$ in terms of columns of Q

Let Q be a column-orthogonal matrix. Let $\mathbf{b}$ be a vector, and write $\mathbf{b} = \mathbf{b}^{||} + \mathbf{b}^{\perp}$ where $\mathbf{b}^{||}$ is projection of $\mathbf{b}$ onto Col Q and $\mathbf{b}^{\perp}$ is projection orthogonal to Col Q.

Let $\mathbf{u}$ be the coordinate representation of $\mathbf{b}^{||}$ in terms of columns of Q.

- $Q^T\mathbf{b}^{||} = \mathbf{u}$
- $Q^T\mathbf{b}^{\perp} = \mathbf{0}$

Therefore

$$Q^T\mathbf{b} = Q^T\left(\mathbf{b}^{||} + \mathbf{b}^{\perp}\right) = Q^T\mathbf{b}^{||} + Q^T\mathbf{b}^{\perp} = Q^T\mathbf{b}^{||} = \mathbf{u}$$

To go from representation $\mathbf{u}$ to $\mathbf{b}^{||}$, multiply by Q:

$$\begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix}\begin{bmatrix} \mathbf{u} \end{bmatrix} = \begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix}$$

Putting these together,

$$\begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix}\begin{bmatrix} & Q^T & \end{bmatrix}\begin{bmatrix} \\ \mathbf{b} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{b}^{||} \\ \\ \end{bmatrix}$$

Representation of $\mathbf{b}^{||}$ in terms of columns of Q

Let Q be a column-orthogonal matrix. Let $\mathbf{b}$ be a vector, and write $\mathbf{b} = \mathbf{b}^{||} + \mathbf{b}^{\perp}$ where $\mathbf{b}^{||}$ is projection of $\mathbf{b}$ onto Col Q and $\mathbf{b}^{\perp}$ is projection orthogonal to Col Q.

Summary:

- $QQ^T\mathbf{b} = \mathbf{b}^{||}$

QR Solver Algorithm for $A\mathbf{x} \approx \mathbf{b}$

Summary:

- $QQ^T\mathbf{b} = \mathbf{b}^{||}$

Proposed algorithm:

> Find Q, R such that $A = QR$ and Q is column-orthogonal and R is triangular
> Compute vector $\mathbf{c} = Q^T\mathbf{b}$
> Solve $R\mathbf{x} = \mathbf{c}$ using backward substitution, and return the solution $\hat{\mathbf{x}}$.

Goal: To show that the solution $\hat{\mathbf{x}}$ returned is the vector that minimizes $\|\mathbf{b} - A\hat{\mathbf{x}}\|$

Every vector of the form $A\mathbf{x}$ is in Col A ($=$ Col Q)

By the High-Dimensional Fire Engine Lemma, the vector in Col A closest to $\mathbf{b}$ is $\mathbf{b}^{||}$, the projection of $\mathbf{b}$ onto Col A.

Solution $\hat{\mathbf{x}}$ satisfies $R\hat{\mathbf{x}} = Q^T\mathbf{b}$

Multiply by Q: $QR\hat{\mathbf{x}} = QQ^T\mathbf{b}$

Therefore $A\hat{\mathbf{x}} = \mathbf{b}^{||}$

The Normal Equations

Let A be a matrix with linearly independent columns. Let QR be its QR factorization. We have given one algorithm for solving the least-squares problem $A\mathbf{x} \approx \mathbf{b}$:

> Find Q, R such that $A = QR$ and Q is column-orthogonal and R is triangular
> Compute vector $\mathbf{c} = Q^T\mathbf{b}$
> Solve $R\mathbf{x} = \mathbf{c}$ using backward substitution, and return the solution $\hat{\mathbf{x}}$.

However, there are other ways to find solution.

Not hard to show that

- $A^T A$ is an invertible matrix
- The solution to the matrix-vector equation $(A^T A)\mathbf{x} = A^T\mathbf{b}$ is the solution to the least-squares problem $A\mathbf{x} \approx \mathbf{b}$
- Can use another method (e.g. Gaussian elimination) to solve $(A^T A)\mathbf{x} = A^T\mathbf{b}$

The linear equations making up $A^T A\mathbf{x} = A^T\mathbf{b}$ are called the *normal equations*