AFRICAN INSTITUTE FOR MATHEMATICAL SCIENCES, AIMS RWANDA

AIMS | African Institute for Mathematical Sciences RWANDA

THE LEAST SQUARES PROBLEM

Prof. Franck Kalala Mutombo
University of Lubumbashi

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using reflectors

Example: We will perform the QR decomposition of a 3×3 matrix A using Householder reflector.

Let

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

Step 1: Construct the First Householder Matrix Q_1

The goal is to zero out the elements below the first entry in the first column.

1. Define the first column that we want

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

2. Compute the norm $\|x_1\|$: $\|x_1\| = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using reflectors

3. Define e_1, the first unit vector: $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

4. Compute $v = x_1 + \|x_1\| e_1$:

$$v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \sqrt{3} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix}$$

5. Normalize v to obtain u:

The normalized vector u is:

$$u = \frac{v}{\|v\|} = \frac{1}{\sqrt{(1+\sqrt{3})^2 + 1^2 + 1^2}} \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix}$$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using reflectors

6. Form the Householder matrix $Q_1 = I - 2uu^T$:

Applying Q_1 to A will zero out all elements below the diagonal in the first column. Let's call the result $A' = Q_1 A$.

Step 2: Construct the Second Householder Matrix Q_2

Now, we need to zero out the element below the second pivot (second row, second column) in the matrix A').

1. Focus on the 2×1 subcolumn of A' from the second row onward:

Let x_2 be this subvector: $x_2 = \begin{bmatrix} a'_{22} \\ a'_{32} \end{bmatrix}$

2. Repeat steps similar to Step 1 to form Q_2, constructing a new Householder vector for this submatrix.

Step 3: Form Q and R

1. Construct Q : The matrix Q is formed by Q = Q1 Q2.
2. Construct R : The resulting matrix R is Q_2 Q_1 A, which will be an upper triangular matrix.

This yields matrices Q and R such that A = QR, where Q is orthogonal and R is upper triangular.

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using (Householder) reflectors

In this example, we perform the QR decomposition of a 3×3 matrix A using Householder transformations. We aim to find matrices Q and R such that $A = QR$, where Q is orthogonal and R is upper triangular.

Example

Let us perform the QR decomposition of the matrix

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{bmatrix}$$

Step 1: Find the First Householder Vector $\boldsymbol{u_1}$

1. Define the first column vector of A:

$$a_1 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using (Householder) reflectors

2. Calculate the norm of a_1:

$$\|a_1\| = \sqrt{1^2 + 3^2 + 2^2} = \sqrt{14} \approx 3.74$$

3. Define $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. 4. Compute the reflection vector u_1:

$$u_1 = a_1 - \|a_1\| e_1 = \begin{bmatrix} 1 - 3.74 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -2.74 \\ 3 \\ 2 \end{bmatrix}$$

5. Normalize u_1:

$$u_1 = \frac{1}{\|u_1\|} \begin{bmatrix} -2.74 \\ 3 \\ 2 \end{bmatrix} \approx \begin{bmatrix} -0.8018 \\ 0.8018 \\ 0.5345 \end{bmatrix}$$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using (Householder) reflectors

Step 2: Construct the First Householder Matrix H_1

The Householder matrix H_1 is given by:

$$H_1 = I - 2\frac{u_1 u_1^T}{u_1^T u_1}$$

Calculating this, we get:

$$H_1 = \begin{bmatrix} 0.267 & 0.802 & 0.535 \\ 0.802 & 0.123 & -0.585 \\ 0.535 & -0.585 & 0.610 \end{bmatrix}$$

Step 3: Apply H_1 to A

Calculate $A^{(1)} = H_1 A$:

$$A^{(1)} = \begin{bmatrix} 3.742 & 10.156 & 19.243 \\ 0 & -0.924 & -2.679 \\ 0 & 0.050 & 1.881 \end{bmatrix}$$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using (Householder) reflectors

Step 4: Find the Second Householder Vector $\boldsymbol{u_2}$

1. Extract the 2x1 subvector from $A^{(1)}$:

$$a_2 = \begin{bmatrix} -0.924 \\ 0.050 \end{bmatrix}$$

2. Compute the norm of a_2:

$$\|a_2\| = \sqrt{(-0.924)^2 + (0.050)^2} \approx 0.926$$

3. Define $e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. 4. Compute the reflection vector u_2:

$$u_2 = a_2 - \|a_2\| e_2 = \begin{bmatrix} -1.850 \\ 0.050 \end{bmatrix}$$

5. Normalize u_2:

$$u_2 = \frac{1}{\|u_2\|} \begin{bmatrix} -1.850 \\ 0.050 \end{bmatrix} \approx \begin{bmatrix} -0.999 \\ 0.054 \end{bmatrix}$$

SOLUTION TO LEAST SQUARES PROBLEM

QR decomposition using (Householder) reflectors

Step 5: Construct the Second Householder Matrix $\boldsymbol{H_2}$

$$H_2 = \begin{bmatrix} -0.999 & 0.054 \\ 0.054 & 0.999 \end{bmatrix}$$

Embedding H_2 into a 3x3 matrix:

$$H_2^{(3x3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.999 & 0.054 \\ 0 & 0.054 & 0.999 \end{bmatrix}$$

Final Matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$

The final matrices for the QR decomposition are:

$$Q = H_1 H_2 = \begin{bmatrix} 0.267 & -0.772 & 0.577 \\ 0.802 & -0.154 & -0.577 \\ 0.535 & 0.617 & 0.577 \end{bmatrix}$$

$$R = \begin{bmatrix} 3.742 & 10.156 & 19.243 \\ 0 & 0.926 & 2.777 \\ 0 & 0 & 1.732 \end{bmatrix}$$

Thus, $A = QR$.

Applications of Qx with $Q^TQ = I$

1. solving least squares problems
2. solving linear equations
3. solving eigenvalue problems
4. finding the singular value decomposition

Stability of orthogonal transformations

We recall:

- Suppose Q is an orthogonal matrix, that is $Q^TQ = I$.
- If $x' = x + e$ is an approximation to x then $Qx' = Qx + Qe$ and $\|Qe\|_2 = \|e\|_2$.
- If $A \in \mathbb{R}^{m,n}$ and $A' = A + E$ is an approximation to A then $QA' = QA + QE$ and $\|QE\|_2 = \|E\|_2$.
- Conculsion: when an orthogonal transformation is applied to a vector or a matrix the error will not grow.

QR-Decomposition and Factorization

Definition 1. *Let* $A \in \mathbb{R}^{m,n}$ *with* $m \geq n \geq 1$. *We say that* $A = QR$ *is a QR-decomposition of* A *if* $Q \in \mathbb{R}^{m,m}$ *is square and orthogonal and*

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

with $R_1 \in \mathbb{R}^{n,n}$ *upper triangular and* $\mathbf{0} \in \mathbb{R}^{m-n,n}$ *the zero matrix. We say that* $A = QR$ *is a QR-factorization of* A *if* $Q \in \mathbb{R}^{m,n}$ *has orthonormal columns and* $R = R_1 \in \mathbb{R}^{n,n}$ *is upper triangular.*

QR decomp $\leftrightarrow$ QR fact

- It is easy to construct a QR-factorization from a QR-decomposition $A = QR$.
- We simply partition $Q = [Q_1, Q_2]$ and $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ where $Q_1 \in \mathbb{R}^{m,n}$ and $R_1 \in \mathbb{R}^{n,n}$.
- Then $A = Q_1 R_1$ is a QR-factorization of A.
- Conversely we construct the QR-decomposition from the QR-factorization $A = Q_1 R_1$ by extending the columns of Q_1 to an orthonormal basis $Q = [Q_1, Q_2]$ for $\mathbb{R}^m$ and defining $R = \left[\begin{smallmatrix} R_1 \\ 0 \end{smallmatrix}\right] \in \mathbb{R}^{m,n}$.
- If $m = n$ then the two factorizations are the same.

example

An example of a QR-decomposition is

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix} = QR,$$

while a QR-factorization $\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{R}_1$ is obtained by dropping the last column of $\boldsymbol{Q}$ and the last row of $\boldsymbol{R}$ so that

$$\boldsymbol{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} = \boldsymbol{Q}_1 \boldsymbol{R}_1$$

Existence of QR Factorization

Lemma 2. *A matrix $A \in \mathbb{R}^{m,n}$ with linearly independent columns has a QR-decomposition and a QR-factorization. The QR-factorization is unique if R has positive diagonal entries.*

Proof. Existence

- Cholesky factorization $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{R}_1^T\boldsymbol{R}_1$,
- $\boldsymbol{R}_1 \in \mathbb{R}^{n,n}$ is upper triangular and nonsingular.
- $\boldsymbol{Q}_1 := \boldsymbol{A}\boldsymbol{R}_1^{-1}$
- $\boldsymbol{Q}_1^T\boldsymbol{Q}_1 = \boldsymbol{I}$.
- QR-factorization $\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{R}_1$

Uniqueness

- $\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{R}_1 \;\Rightarrow\; \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{R}_1^T\boldsymbol{R}_1$
- Cholesky factorization unique implies $\boldsymbol{R}_1$ unique and $\boldsymbol{Q}_1 = \boldsymbol{A}\boldsymbol{R}_1^{-1}$ unique.

$\square$

QR and Gram-Schmidt

If $\boldsymbol{A} \in \mathbb{R}^{m,n}$ has rank n, then the set of columns $\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}$ forms a basis for $\operatorname{span}(\boldsymbol{A})$ and the Gram-Schmidt orthogonalization process takes the form

$$\boldsymbol{v}_1 = \boldsymbol{a}_1, \quad \boldsymbol{v}_j = \boldsymbol{a}_j - \sum_{i=1}^{j-1} \frac{\boldsymbol{a}_j^T \boldsymbol{v}_i}{\boldsymbol{v}_i^T \boldsymbol{v}_i} \boldsymbol{v}_i, \text{ for } j = 2, \dots, n.$$

$\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is an orthogonal basis for $\operatorname{span}(\boldsymbol{A})$.

$$\boldsymbol{a}_1 = \boldsymbol{v}_1, \quad \boldsymbol{a}_j = \sum_{i=1}^{j-1} \rho_{ij} \boldsymbol{v}_i + \boldsymbol{v}_j, \text{ where } \rho_{ij} = \frac{\boldsymbol{a}_j^T \boldsymbol{v}_i}{\boldsymbol{v}_i^T \boldsymbol{v}_i}$$

GS=QR-factorization

$$\boldsymbol{a}_1 = \boldsymbol{v}_1, \quad \boldsymbol{a}_j = \sum_{i=1}^{j-1} \rho_{ij}\boldsymbol{v}_i + \boldsymbol{v}_j,$$

- $\boldsymbol{A} = \boldsymbol{V}\hat{\boldsymbol{R}}$, where $\boldsymbol{V} := [\boldsymbol{v}_1, \dots, \boldsymbol{v}_n] \in \mathbb{R}^{m,n}$ and $\hat{\boldsymbol{R}}$ is unit upper triangular.
- Since $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a basis for $\operatorname{span}(\boldsymbol{A})$ the matrix $\boldsymbol{D} := \operatorname{diag}(\|\boldsymbol{v}_1\|_2, \dots, \|\boldsymbol{v}_n\|_2)$ is nonsingular,
- the matrix $\boldsymbol{Q}_1 := \boldsymbol{V}\boldsymbol{D}^{-1} = [\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|_2}, \dots, \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|_2}]$ is orthogonal.
- Therefore, $\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{R}_1$, with $\boldsymbol{R}_1 := \boldsymbol{D}\hat{\boldsymbol{R}}$ is a QR-factorization of $\boldsymbol{A}$ with positive diagonal entries in $\boldsymbol{R}_1$.

QR and Least Squares

- Suppose $\boldsymbol{A} \in \mathbb{R}^{m,n}$ has rank n and let $\boldsymbol{b} \in \mathbb{R}^m$.
- Consider the least squares problem $\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$.
- Suppose $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ is a QR-decomposition of $\boldsymbol{A}$.
- We partition $\boldsymbol{Q}$ and $\boldsymbol{R}$ as $\boldsymbol{Q} = [\boldsymbol{Q}_1 \ \boldsymbol{Q}_2]$ and $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{0} \end{bmatrix}$, where $\boldsymbol{Q}_1 \in \mathbb{R}^{m,n}$ and $\boldsymbol{R}_1 \in \mathbb{R}^{n,n}$.

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = \|\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{Q}^T\boldsymbol{b}\|_2^2 &= \|\begin{bmatrix} \boldsymbol{R}_1\boldsymbol{x} - \boldsymbol{Q}_1^T\boldsymbol{b} \\ -\boldsymbol{Q}_2^T\boldsymbol{b} \end{bmatrix}\|_2^2 \\ &= \|\boldsymbol{R}_1\boldsymbol{x} - \boldsymbol{Q}_1^T\boldsymbol{b}\|_2^2 + \|\boldsymbol{Q}_2^T\boldsymbol{b}\|_2^2.\end{aligned}$$

- Thus $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2 \geq \|\boldsymbol{Q}_2^T\boldsymbol{b}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$ with equality if $\boldsymbol{R}_1\boldsymbol{x} = \boldsymbol{Q}_1^T\boldsymbol{b}$.

Least Squares using QR

1. Find a QR-factorization $\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{R}_1$ of $\boldsymbol{A}$.
2. Solve $\boldsymbol{R}_1 \boldsymbol{x} = \boldsymbol{Q}_1^T \boldsymbol{b}$ for the least squares solution $\boldsymbol{x}$.

Consider the least squares problem with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The least squares solution $\boldsymbol{x}$ is found by solving the system

$$\begin{bmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

and we find $\boldsymbol{x} = [1, 0, 0]^T$.

Finding QR

- Gram Schmidt is numerically unstable
- Use instead orthogonal transformations
- they are stable
- study Householder transformations and Givens rotations

Householder Transformations

A matrix $\boldsymbol{H} \in \mathbb{R}^{n,n}$ of the form

$$\boldsymbol{H} := \boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^T, \text{ where } \boldsymbol{u} \in \mathbb{R}^n \text{ and } \boldsymbol{u}^T\boldsymbol{u} = 2$$

is called a Householder transformation.

- For $n = 2$ we find $\boldsymbol{H} = \begin{bmatrix} 1-u_1^2 & -u_1u_2 \\ -u_2u_1 & 1-u_2^2 \end{bmatrix}$.
- A Householder transformation is symmetric and orthogonal. In particular,

$$\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{H}^2 = (\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^T)(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^T) = \boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^T + \boldsymbol{u}(\boldsymbol{u}^T\boldsymbol{u})\boldsymbol{u}^T = \boldsymbol{I}.$$

Alternative representations

- Householder himself used $\boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^T$, where $\boldsymbol{u}^T\boldsymbol{u} = 1$.
- For any nonzero $\boldsymbol{v} \in \mathbb{R}^n$ the matrix

$$\boldsymbol{H} := \boldsymbol{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$$

  is a Householder transformation.
- In fact $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^T$, where $\boldsymbol{u} := \sqrt{2}\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}$.

Transformation

Lemma 3. *Suppose* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *with* $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$ *and* $\boldsymbol{v} := \boldsymbol{x} - \boldsymbol{y} \neq \boldsymbol{0}$. *Then* $\left(\boldsymbol{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}\right)\boldsymbol{x} = \boldsymbol{y}$.

Proof. Since $\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{y}$ we have

$$\boldsymbol{v}^T\boldsymbol{v} = (\boldsymbol{x} - \boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}) = 2\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{y}^T\boldsymbol{x} = 2\boldsymbol{v}^T\boldsymbol{x}. \tag{1}$$

But then $\left(\boldsymbol{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}\right)\boldsymbol{x} = \boldsymbol{x} - \frac{2\boldsymbol{v}^T\boldsymbol{x}}{\boldsymbol{v}^T\boldsymbol{v}}\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{v} = \boldsymbol{y}$. □

Geometric Interpretation $Hx = y$

Hx is the mirror image (reflection) of x

$$H = I - \frac{2vv^T}{v^Tv} = P - \frac{vv^T}{v^Tv}, \text{ where } P := I - \frac{vv^T}{v^Tv}$$

Px is the orthogonal projection of x into $\mathrm{span}\{x+y\}$

$$Px = x - \frac{v^Tx}{v^Tv}v \overset{(1)}{=} x - \frac{1}{2}v = \frac{1}{2}(x+y)$$