QR Algorithm

Contents

1. Householder Transformation (reflection operator)

$\downarrow$

2. QR Factorization Algorithm

3. Appendix: Gram-Schmidt Orthogonalization

(Review + perspective towards thin QR)

$m \geq n$

Recap

Matrix w/ orthonormal columns

$$Q = [\vec{q}_1 | \vec{q}_2 | \cdots | \vec{q}_n] \in \mathbb{R}^{m \times n}$$

(tall rectangle)

- $\vec{q}_1, \cdots, \vec{q}_n$ are orthonormal

$$\iff Q^T Q = I$$

- 2-norm preserving prop:

$$\|Q\vec{x}\|_2 = \|\vec{x}\|_2$$

- $\|Q\|_2 = 1$

$QQ^T \neq I$.

Orthogonal matrix

$$Q = [\vec{q}_1 | \cdots | \vec{q}_m] \in \mathbb{R}^{m \times m}$$

(square matrix)

- $\vec{q}_1, \cdots, \vec{q}_m$ are orthonormal

$$\iff Q^T Q = I$$

- $\|Q\vec{x}\|_2 = \|\vec{x}\|_2$
- $\|Q\|_2 = 1$

- $Q^{-1} = Q^T$
- $QQ^T = I$

Projection and Reflection Operators (cont')

From Monday (Lec. 20)

Summary: for given $\mathbf{v} \in \mathbb{R}^m$, a nonzero vector, let

$$P = \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} \in \mathbb{R}^{m \times m}.$$

Then the following matrices carry out geometric transformations

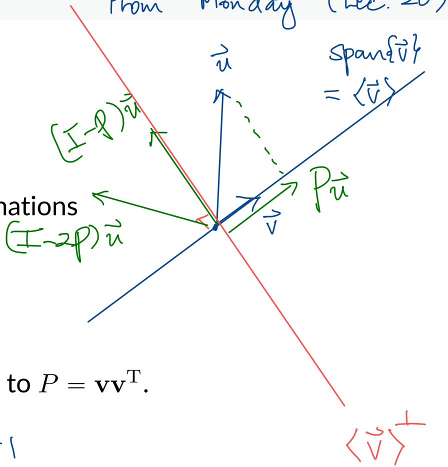
- Projection onto $\langle \mathbf{v} \rangle$: P
- Projection onto $\langle \mathbf{v} \rangle^{\perp}$: $I - P$
- Reflection across $\langle \mathbf{v} \rangle^{\perp}$: $I - 2P$

Note. If $\mathbf{v}$ were a unit vector, the definition of P simplifies to $P = \mathbf{v}\mathbf{v}^{\mathrm{T}}$.

$\Downarrow$

$\|\vec{v}\|_2 = 1 \Rightarrow \|\vec{v}\|_2^2 = 1 \Rightarrow \vec{v}^T\vec{v} = 1$

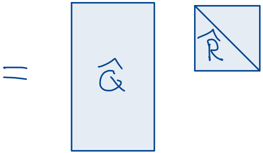
Householder transformation.



QR Factorization and Least Squares

$A \in \mathbb{R}^{m\times n}$, $m \geq n$

Thick QR $\begin{cases} Q \in \mathbb{R}^{m\times m} & \text{orthogonal} \\ R \in \mathbb{R}^{m\times n} & \text{upper-}\Delta \end{cases}$



Thin QR $\begin{cases} \hat{Q} \in \mathbb{R}^{m\times n} & \text{not orthogonal but has ONC.} \\ \hat{R} \in \mathbb{R}^{n\times n} & \text{upper-}\Delta \end{cases}$

<u>Normal Eqn</u> $\xrightarrow{\text{Thin QR}}$ <u>Upper-Δ system</u>

$$A^T A \tilde{x} = A^T \vec{b} \quad\xrightarrow{\text{Thin QR}}\quad \hat{R}\vec{x} = \hat{Q}^T \vec{b}$$

Let $Q \in \mathbb{R}^{m \times m}$ be orthogonal.

Recall that $Q^{-1} = Q^T$.

Practical advantage

- Finding inverse is super easy.

Theoretical implication

- Orthogonal matrices are invertible

- A is invertible.

$\Leftrightarrow$ A is nonsingular.

$\Leftrightarrow$ $\det(A) \neq 0$

$\Leftrightarrow$ $A\vec{x} = \vec{b}$ has a unique.

$\Leftrightarrow$ Columns of A are linearly independent.

Householder Transformation

Motivation

$$\vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} \xrightarrow[\text{orthogonal}]{H} H\vec{z} = \begin{bmatrix} * \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Problem

Given $\mathbf{z} \in \mathbb{R}^m$, find an orthogonal matrix $H \in \mathbb{R}^{m \times m}$ such that $H\mathbf{z}$ is nonzero only in the first element.

- Since orthogonal matrices preserve the 2-norm, H must satisfy

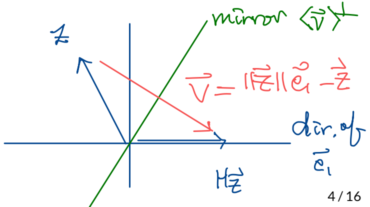
$$H\mathbf{z} = \begin{bmatrix} \pm \|\mathbf{z}\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \pm \|\mathbf{z}\|_2 \, \mathbf{e}_1 .$$

$$\Rightarrow \|H\vec{z}\|_2 = \left|\pm\|\vec{z}\|_2\right| \, \|\vec{e}_1\|_2 = \|\vec{z}\|_2 \cdot 1 = \|\vec{z}\|_2 .$$

- The **Householder transformation matrix** H defined by

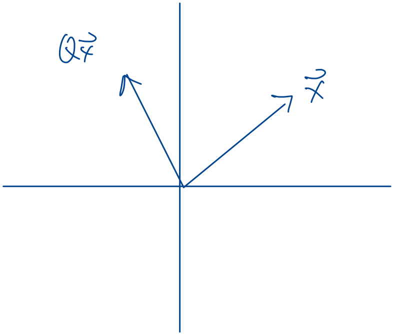
$$H = I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}, \quad \text{where } \mathbf{v} = \pm \|\mathbf{z}\|_2 \, \mathbf{e}_1 - \mathbf{z},$$

solves the problem. See Theorem 1 on the next slide.



Orthogonal matrix as transformation

$$\|Q\vec{x}\|_2 = \|\vec{x}\|_2$$



- reflection $\longrightarrow$ Householder trans.
- rotation $\longrightarrow$ Givens rotation

Properties of Householder Transformation

Theorem 1

Let $\mathbf{v} = \|\mathbf{z}\|_2 \, \mathbf{e}_1 - \mathbf{z}$ *and let* H *be the Householder transformation defined by*

$$H = I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}.$$

Then

1. H *is symmetric;* $(H^T = H)$
2. H *is orthogonal;*
3. $H\mathbf{z} = \|\mathbf{z}\|_2 \, \mathbf{e}_1$.

$\vec{v} \to \alpha\vec{v}$: $I - 2\frac{(\alpha\vec{v})(\alpha\vec{v})^T}{(\alpha\vec{v})^T(\alpha\vec{v})} = I - 2\frac{\cancel{\alpha^2}\vec{v}\vec{v}^T}{\cancel{\alpha^2}\vec{v}^T\vec{v}} = H$

- H is invariant under scaling of $\mathbf{v}$.
- If $\|\mathbf{v}\|_2 = 1$, then $H = I - 2\mathbf{v}\mathbf{v}^{\mathrm{T}}$.

Proof of (2) Since H is invariant under scaling of $\vec{v}$,

We assume that $\vec{v}$ is a unit vector so that

$$H = I - 2\vec{v}\vec{v}^T.$$

Now to prove that H is orthogonal, we need to show

$$H^T H = I.$$

$$H^T H = (I - 2\vec{v}\vec{v}^T)^T (I - 2\vec{v}\vec{v}^T)$$

$$= (\underbrace{I^T}_{= I} - 2\underbrace{(\vec{v}\vec{v}^T)^T}_{= \vec{v}\vec{v}^T})(I - 2\vec{v}\vec{v}^T)$$

$$= (I - 2\vec{v}\vec{v}^T)(I - 2\vec{v}\vec{v}^T)$$

$$= I^2 - 2\vec{v}\vec{v}^T - 2\vec{v}\vec{v}^T + 4\vec{v}\underbrace{\boxed{\vec{v}^T\vec{v}}}_{=1}\vec{v}^T$$

$$= I - \cancel{4\vec{v}\vec{v}^T} + \cancel{4\vec{v}\vec{v}^T}$$

$$= I$$

$\square$

Geometry Behind Householder Transformation

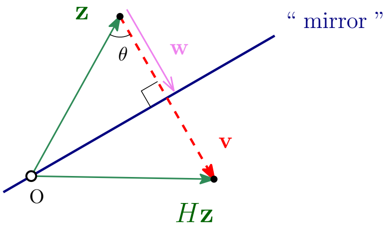
The Householder transformation matrix H is the reflector across $\langle \mathbf{v} \rangle^{\perp}$.

From any $\mathbf{z}$ to the "mirror":

$$\mathbf{w} = -\frac{\mathbf{z}^{\mathrm{T}}\mathbf{v}}{\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} \cdot \frac{\mathbf{v}}{\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} = -\mathbf{v}\frac{\mathbf{z}^{\mathrm{T}}\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} \,.$$

From any $\mathbf{z}$ to its reflection:

$$H\mathbf{z} - \mathbf{z} = -2\mathbf{v}\frac{\mathbf{z}^{\mathrm{T}}\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} \,.$$



Thus, for any $\mathbf{z}$,

$$H\mathbf{z} = \mathbf{z} - 2\mathbf{v}\frac{\mathbf{z}^{\mathrm{T}}\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} = \left(I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right)\mathbf{z} \quad \Longrightarrow \quad H = I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} \,.$$

QR Factorization Algorithm

Office Hours (unusual schedule)

- M 4:45 ~ 6:15
- T 9:00 ~ 10:30
- Th 9:00 ~ 10:30 (for 3607 only)

$A \in \mathbb{R}^{m \times n}$, $m \geq n$

$A = Q\,R$ where $Q \in \mathbb{R}^{m \times m}$ orthogonal

$R \in \mathbb{R}^{m \times n}$ upper-Δ.

QR Factorization Algorithm via Triangularization

- The Gram-Schmidt orthogonalization (thin QR factorization) is unstable in floating-point calculations.
- **Stable alternative:** Find orthogonal matrices $H_1, H_2, \ldots, H_n$ so that

$$\underbrace{H_n H_{n-1} \cdots H_2 H_1}_{=: Q^{\mathrm{T}}} A = R\,.$$

  introducing zeros one column at a time below diagonal terms. Householder matrices will do.
- As a product of orthogonal matrices, Q^{T} is also orthogonal and so $(Q^{\mathrm{T}})^{-1} = Q$. Therefore,

$$A = QR\,.$$

orthogonal

$$Q^T A = R$$

$$A = (Q^T)^{-1} R$$

$$= (Q^T)^T R$$

$$= QR$$

$\cdot \quad Q^T = H_n H_{n-1} \cdots H_2 H_1$. What is Q?

$$Q = (Q^T)^T = (H_n H_{n-1} \cdots H_2 H_1)^T$$

$$= H_1^T H_2^T \cdots H_{n-1}^T H_n^T$$

$\downarrow$ w/ orthogonality

$$= H_1 H_2 \cdots H_{n-1} H_n.$$

provided $H_1, \cdots, H_n$ are symmetric.

Illustration w/ 4x3 matrix

$$\underbrace{\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}}_{A} \xrightarrow{H_1} \underbrace{\begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}}_{H_1 A} \xrightarrow{H_2} \underbrace{\begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix}}_{H_2 H_1 A} \xrightarrow{H_3} \underbrace{\begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix}}_{H_3 H_2 H_1 A = R}$$

H_1 : Householder transforming $\begin{bmatrix} \times \\ \times \\ \times \\ \times \end{bmatrix} \rightarrow \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$$H_2 = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & \tilde{H}_2 & \\ 0 & & & \end{array}\right]$$

where $\tilde{H}_2$ is H.T. $\begin{bmatrix} \times \\ \times \\ \times \end{bmatrix} \rightarrow \begin{bmatrix} \times \\ 0 \\ 0 \end{bmatrix}$

$$H_3 = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{$\tilde{H}_3$}} \\ 0 & 0 & & \end{array}\right]$$

where $\tilde{H}_3$ is H.T. $\begin{bmatrix} \times \\ \times \end{bmatrix} \rightarrow \begin{bmatrix} \times \\ 0 \end{bmatrix}$

Example (Calculation of Householder)

Find an orthogonal matrix H such that

$$H \underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}}_{\vec{z}} = \underbrace{\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\|\vec{z}\|_2 \vec{e}_1}$$

Soln Let $\vec{v} = \|\vec{z}\|_2 \vec{e}_1 - \vec{z} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$.

Then the Householder matrix obtained by

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

will do.

Scratch

- $\|\vec{z}\|_2^2 = 1^2 + (-1)^2 + 1^2 + (-1)^2$

  $= 1 + 1 + 1 + 1 = 4$

  $\|\vec{z}\|_2 = 2$

- $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^4$

So let's compute H:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & -1/2 & \cdots \\ \vdots & \ddots & \\ & & \ddots \end{bmatrix}$$

$$2 \frac{\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}} = \frac{2}{\vec{v}^T\vec{v}} \vec{v}\vec{v}^T$$

$$\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 1 \end{bmatrix}$$

$$\vec{v}^T\vec{v} = 1^2 + 1^2 + (-1)^2 + 1^2$$

$$= 4.$$

MATLAB Implementation: MYQR

(using Householder)

- Demo version
- not very efficient
- Suggestion for improvement found in practice problem

$m \geq n$

```
function [Q, R] = myqr(A)
  [m, n] = size(A);
  A0 = A;
  Q = eye(m);
  for j = 1:min(m,n)
      Aj = A(j:m, j:n);
      z = Aj(:, 1);
      v = z + sign0(z(1))*norm(z)*eye(length(z), 1);
      Hj = eye(length(v)) - 2/(v'*v) * v*v';
      Aj = Hj*Aj;
      H = eye(m);
      H(j:m, j:m) = Hj;
      Q = Q*H;
      A(j:m, j:n) = Aj;
  end
  R = A;
end
```

Q will go through successive update in loop.

(*) $\longrightarrow$

$\vec{v} = \pm\|\vec{z}\|\,\vec{e}_1 - \vec{z}$

Householder

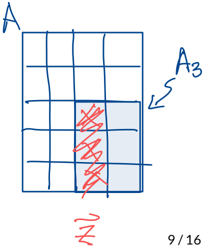
$\longrightarrow \quad H_j = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & \tilde{H}_j \end{array}\right]$

$Q = I$

$j=1: \quad Q = H_1$

$j=2: \quad Q = H_1 H_2$

$\vdots$



MATLAB Implementation: MYQR (cont')

"Signum" Sine, Sign

(continued from the previous page)

```
% local function
function y=sign0(x)
  y = ones(size(x));
  y(x < 0) = -1;
end
```

$$\mathrm{sign0}(x) = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

- The MATLAB command `qr` works similar to, but more efficiently than, this.
- The function finds the factorization in $\sim (2mn^2 - n^3/3)$ flops asymptotically.

Improvement

$$A_j = H_j A_j$$

$$H_j A_j = \underbrace{\left(I - \frac{2\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}}\right)}_{\text{mat}} \underbrace{A_j}_{\text{mat}}$$

$$= A_j - \frac{2}{\vec{v}^T\vec{v}} \underbrace{\vec{v}}_{\text{vec}} \underbrace{\left(\underbrace{\vec{v}^T}_{\text{vec}} \underbrace{A_j}_{\text{mat}}\right)}_{\text{vec}}$$

Which Reflector Is Better?

Recall:

$$H = I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}, \quad \text{where } \mathbf{v} = \pm \|\mathbf{z}\|_2 \, \mathbf{e}_1 - \mathbf{z},$$

In `myqr.m`, the statement

```
v = z + sign0(z(1))*norm(z)*eye(length(z), 1);
```

defines $\mathbf{v}$ slightly differently, namely,

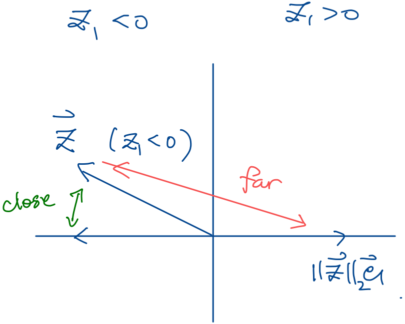
$$\mathbf{v} = \mathbf{z} \pm \|\mathbf{z}\|_2 \, \mathbf{e}_1.$$

This does not cause any difference since H is invariant under scaling of $\mathbf{v}$; see p. 5.

Which Reflector Is Better? (cont')

The sign of $\pm \|\mathbf{z}\|_2$ is chosen so as to avoid possible catastrophic cancellation in forming $\mathbf{v}$:

$$\mathbf{v} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} + \begin{bmatrix} \pm \|\mathbf{z}\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} z_1 \pm \|\mathbf{z}\|_2 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$



Subtractive cancellation may arise when $z_1 \approx \pm \|\mathbf{z}\|_2$.

- if $z_1 > 0$, use $z_1 + \|\mathbf{z}\|_2$;
- if $z_1 < 0$, use $z_1 - \|\mathbf{z}\|_2$;
- if $z_1 = 0$, either works.

For numerical stability, it is desirable to reflect $\mathbf{z}$ to the vector $s \|\mathbf{z}\|_2 \mathbf{e}_1$ that is not too close to $\mathbf{z}$ itself. (Trefethen & Bau)

Appendix: Gram-Schmidt Orthogonalization

The Gram–Schmidt Procedure

Problem: Orthogonalization

Given $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$, construct orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ such that

$$\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}, \quad \text{for any } k \in \mathbb{N}[1, n].$$

- **Strategy.** At the jth step, find a unit vector $\mathbf{q}_j \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_j\}$ that is orthogonal to $\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}$.
- **Key Observation.** The vector $\mathbf{v}_j$ defined by

$$\mathbf{v}_j = \mathbf{a}_j - \left(\mathbf{q}_1^{\mathrm{T}}\mathbf{a}_j\right)\mathbf{q}_1 - \left(\mathbf{q}_2^{\mathrm{T}}\mathbf{a}_j\right)\mathbf{q}_2 - \cdots - \left(\mathbf{q}_{j-1}^{\mathrm{T}}\mathbf{a}_j\right)\mathbf{q}_{j-1}$$

satisfies the required properties.

GS Algorithm

The Gram–Schmidt iteration is outlined below:

$$
\begin{aligned}
\mathbf{q}_1 &= \frac{\mathbf{a}_1}{r_{11}}, \\
\mathbf{q}_2 &= \frac{\mathbf{a}_2 - r_{12}\mathbf{q}_1}{r_{22}}, \\
\mathbf{q}_3 &= \frac{\mathbf{a}_3 - r_{13}\mathbf{q}_1 - r_{23}\mathbf{q}_2}{r_{33}}, \\
&\ \vdots \\
\mathbf{q}_n &= \frac{\mathbf{a}_n - \sum_{i=1}^{n-1} r_{in}\mathbf{q}_i}{r_{nn}},
\end{aligned}
$$

where

$$
r_{ij} = \begin{cases} \mathbf{q}_i^{\mathrm{T}}\mathbf{a}_j, & \text{if } i \neq j \\ \pm\left\|\mathbf{a}_j - \displaystyle\sum_{k=1}^{j-1} r_{kj}\mathbf{q}_k\right\|_2, & \text{if } i = j \end{cases}.
$$

GS Procedure and Thin QR Factorization

- The GS algorithm, written as a matrix equation, yields a **thin QR factorization**:

$$A = \underbrace{\begin{bmatrix} \\ \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \\ \\ \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \\ \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ \\ \\ \end{bmatrix}}_{\hat{Q}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}}_{\hat{R}} = \widehat{Q}\widehat{R}$$

- While it is an important tool for theoretical work, the GS algorithm is numerically unstable.