Column Spaces and QR

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Doug James (and Justin Solomon)
Announcements

- **HW1**: Updated $N$ and $h$ definitions. Posted online
- **Monday office hours:**
  - Alex Jin moved slot (starting next week)
  - Mon 3-5 PM in Lathrop Hall
Problem

\[ \text{cond} \ A^\top A \approx (\text{cond} \ A)^2 \]
Geometric Intuition

Least-squares fit is ambiguous!
When is \( \text{cond } A^\top A \approx 1 \)?
When Is $\text{cond } A^\top A \approx 1$?

$$\text{cond } I_{n \times n} = 1$$
\[\text{w.r.t. } \| \cdot \|_2\]

Desirable: $A^\top A \approx I_{n \times n}$
\[\text{then, } \text{cond } A^\top A \approx 1!\]
When Is $\text{cond } A^\top A \approx 1$?

\[
\text{cond } I_{n \times n} = 1
\]
(w.r.t. $\| \cdot \|_2$)

Desirable: $A^\top A \approx I_{n \times n}$
(then, $\text{cond } A^\top A \approx 1$!)

Doesn’t mean $A = I_{n \times n}$. 
Recall: Definition of Gram matrix

\[ Q^\top Q = \begin{pmatrix}
- & \vec{q}_1^\top & - \\
- & \vec{q}_2^\top & - \\
& \vdots & \\
- & \vec{q}_n^\top & \\
\end{pmatrix}
\begin{pmatrix}
\vec{q}_1 \\
\vec{q}_2 \\
\vdots \\
\vec{q}_n \\
\end{pmatrix}
\]

\[ = \begin{pmatrix}
\vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\
\vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\
& \vdots & \ddots & \vdots \\
\vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \\
\end{pmatrix} \]
When \( Q^\top Q = I_{n \times n} \)

\[
\vec{q}_i \cdot \vec{q}_j = \begin{cases} 
1 & \text{when } i = j \\
0 & \text{when } i \neq j 
\end{cases}
\]
When $Q^\top Q = I_{n \times n}$

$$\vec{q}_i \cdot \vec{q}_j = \begin{cases} 
1 & \text{when } i = j \\
0 & \text{when } i \neq j
\end{cases}$$

**Orthonormal; orthogonal matrix**

A set of vectors $\{\vec{v}_1, \cdots, \vec{v}_k\}$ is *orthonormal* if $\|\vec{v}_i\| = 1$ for all $i$ and $\vec{v}_i \cdot \vec{v}_j = 0$ for all $i \neq j$. A square matrix whose columns are orthonormal is called an *orthogonal* matrix.
Isometry Properties

\[ \| Q \vec{x} \|^2 = ? \]

\[ (Q \vec{x}) \cdot (Q \vec{y}) = ? \]
Geometric Interpretation

(a) Isometric

(b) Not isometric
Alternative Intuition for Least-Squares

\[ A^\top A\vec{x} = A^\top \vec{b} \leftrightarrow \min_{\vec{x}} \| A\vec{x} - \vec{b} \|_2 \]

Project \( \vec{b} \) onto the column space of \( A \).
Observation

Lemma: Column space invariance

For any $A \in \mathbb{R}^{m \times n}$ and invertible $B \in \mathbb{R}^{n \times n}$,

$$\text{col } A = \text{col } AB.$$
Observation

**Lemma: Column space invariance**

For any $A \in \mathbb{R}^{m \times n}$ and invertible $B \in \mathbb{R}^{n \times n}$,

$$\text{col } A = \text{col } AB.$$

Invertible *column* operations do not affect column space.
New Strategy

Apply column operations to $A$ until it is orthogonal; then, solve least-squares on the resulting orthogonal $Q$. 
New Factorization

\[ A = QR \]

- \( Q \) orthogonal
- \( R \) upper triangular
Using QR

\[ A^\top A \vec{x} = A^\top \vec{b}, \quad A = QR \]

\[ \rightarrow \vec{x} = R^{-1}Q^\top \vec{b} \]
Using QR

\[ A^\top A \vec{x} = A^\top \vec{b}, \quad A = QR \]

\[ \rightarrow \vec{x} = R^{-1}Q^\top \vec{b} \]

Didn’t need to compute \( A^\top A \) or \( (A^\top A)^{-1} \)!
"Which multiple of $\vec{a}$ is closest to $\vec{b}$?"

$$\min_c \| c\vec{a} - \vec{b} \|_2^2$$
"Which multiple of $\vec{a}$ is closest to $\vec{b}$?"

$$\min_c \| c\vec{a} - \vec{b} \|_2^2$$

$$c = \frac{\vec{a} \cdot \vec{b}}{\| \vec{a} \|_2^2}$$
Vector Projection

“Which multiple of $\vec{a}$ is closest to $\vec{b}$?”

$$\min_c \| c\vec{a} - \vec{b} \|_2^2$$

$$c = \frac{\vec{a} \cdot \vec{b}}{\| \vec{a} \|_2^2}$$

$$\text{proj}_{\vec{a}} \vec{b} = c\vec{a} = \frac{\vec{a} \cdot \vec{b}}{\| \vec{a} \|_2^2} \vec{a}$$
Properties of Projection

\[ \text{proj}_{\vec{a}} \vec{b} \parallel \vec{a} \]

\[ \vec{a} \cdot (\vec{b} - \text{proj}_{\vec{a}} \vec{b}) = 0 \]

\[ \iff (\vec{b} - \text{proj}_{\vec{a}} \vec{b}) \perp \vec{a} \]
Orthonormal Projection

Suppose $\hat{a}_1, \ldots, \hat{a}_k$ are orthonormal.

$$\text{proj}_{\hat{a}_i} \vec{b} = (\hat{a}_i \cdot \vec{b}) \hat{a}_i$$
Orthonormal Projection

\[\|c_1 \hat{a}_1 + c_2 \hat{a}_2 + \cdots + c_k \hat{a}_k - \vec{b}\|_2^2 = \sum_{i=1}^{k} \left( c_i^2 - 2c_i \vec{b} \cdot \hat{a}_i \right) + \|\vec{b}\|_2^2\]
Orthonormal Projection

\[ \| c_1 \hat{a}_1 + c_2 \hat{a}_2 + \cdots + c_k \hat{a}_k - \vec{b} \|_2^2 = \sum_{i=1}^{k} \left( c_i^2 - 2c_i \vec{b} \cdot \hat{a}_i \right) + \| \vec{b} \|_2^2 \]

\[ \implies c_i = \vec{b} \cdot \hat{a}_i \]
Orthonormal Projection

\[ \| c_1 \hat{a}_1 + c_2 \hat{a}_2 + \cdots + c_k \hat{a}_k - \vec{b} \|_2^2 = \sum_{i=1}^{k} \left( c_i^2 - 2c_i \vec{b} \cdot \hat{a}_i \right) + \| \vec{b} \|_2^2 \]

\[ \implies c_i = \vec{b} \cdot \hat{a}_i \]

\[ \implies \text{proj}_{\text{span} \{ \hat{a}_1, \ldots, \hat{a}_k \} \vec{b}} = (\hat{a}_1 \cdot \vec{b})\hat{a}_1 + \cdots + (\hat{a}_k \cdot \vec{b})\hat{a}_k \]
Geometric Strategy for Orthogonalization

(a) Input  (b) Rescaling  (c) Projection  (d) Normalization
Gram-Schmidt Orthogonalization

To orthogonalize $\mathbf{v}_1, \ldots, \mathbf{v}_k$:

1. $\hat{a}_1 \equiv \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$.

2. For $i$ from 2 to $k$,
   2.1 $\hat{p}_i \equiv \text{proj}_{\text{span}\{\hat{a}_1, \ldots, \hat{a}_{i-1}\}} \mathbf{v}_i$.
   2.2 $\hat{a}_i \equiv \frac{\mathbf{v}_i - \hat{p}_i}{\|\mathbf{v}_i - \hat{p}_i\|}$.
Gram-Schmidt Orthogonalization

To orthogonalize $\vec{v}_1, \ldots, \vec{v}_k$:

1. $\hat{a}_1 \equiv \frac{\vec{v}_1}{\| \vec{v}_1 \|}$.

2. For $i$ from 2 to $k$,
   2.1 $\vec{p}_i \equiv \text{proj}_{\text{span} \{ \hat{a}_1, \ldots, \hat{a}_{i-1} \}} \vec{v}_i$.
   2.2 $\hat{a}_i \equiv \frac{\vec{v}_i - \vec{p}_i}{\| \vec{v}_i - \vec{p}_i \|}$.

Claim

$\text{span} \{ \vec{v}_1, \ldots, \vec{v}_i \} = \text{span} \{ \hat{a}_1, \ldots, \hat{a}_i \}$ for all $i$. 
Implementation via Column Operations

*Post-multiplication!*

1. Rescaling to unit length: diagonal matrix
2. Subtracting off projection: upper triangular substitution matrix
New Factorization

\[ A = QR \]

- \( Q \) orthogonal
- \( R \) upper-triangular
Bad Case

\[ \vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 1 + \varepsilon \end{pmatrix} \]

\[ \| \vec{r} \|_2 \approx 0 \]
Two Strategies for QR

1. Post-multiply by upper triangular matrices
   Done!
Two Strategies for QR

1. Post-multiply by upper triangular matrices
   Done!

2. Pre-multiply by orthogonal matrices
   New idea!
“Easy” Class of Orthogonal Matrices

\[ \text{proj}_v \vec{b} \]

\[ 2(\text{proj}_v \vec{b}) - \vec{b} \]

\[ (\text{proj}_v \vec{b}) - \vec{b} \]
Reflection Matrices

$$2\text{proj}_\vec{v} \vec{b} - \vec{b} = 2\vec{v} \cdot \frac{\vec{b}}{\vec{v} \cdot \vec{v}} \vec{v} - \vec{b} \text{ by definition of projection}$$

$$= 2\vec{v} \cdot \frac{\vec{v}^\top \vec{b}}{\vec{v}^\top \vec{v}} - \vec{b} \text{ using matrix notation}$$

$$= \left( 2\vec{v} \vec{v}^\top \vec{v}^\top \vec{v} - I_{n \times n} \right) \vec{b}$$

$$\equiv -H_\vec{v} \vec{b}, \text{ where } H_\vec{v} \equiv I_{n \times n} - \frac{2\vec{v} \vec{v}^\top}{\vec{v}^\top \vec{v}}.$$. 

CS 205A: Mathematical Methods
If \( \vec{a} \) is first column,

\[
    c \vec{e}_1 = H v \vec{a}
\]

\[\Rightarrow \quad \vec{v} = (\vec{a} - c \vec{e}_1) \cdot \frac{\vec{v}^\top \vec{v}}{2 \vec{v}^\top \vec{a}}\]
Analogy to Forward Substitution

If $\vec{a}$ is first column,

$$c\vec{e}_1 = H_\vec{v}\vec{a}$$

$$\implies \vec{v} = (\vec{a} - c\vec{e}_1) \cdot \frac{\vec{v}^\top \vec{v}}{2\vec{v}^\top \vec{a}}$$

Choose $\vec{v} = \vec{a} - c\vec{e}_1$

$$\implies c = \pm \|\vec{a}\|_2$$
After One Step

\[ H_v A = \begin{pmatrix} c & \times & \times & \times \\ 0 & \times & \times & \times \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \times & \times & \times \end{pmatrix} \]
Later Steps

\[ \vec{a} = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \end{pmatrix} \mapsto H_v \vec{a} = \begin{pmatrix} \vec{a}_1 \\ 0 \end{pmatrix} \]

Leave first \( k \) lines alone!
Householder QR

\[ R = H \vec{v}_n \cdots H \vec{v}_1 A \]

\[ Q = H^\top \vec{v}_1 \cdots H^\top \vec{v}_n \]
Householder QR

\[ R = H\vec{v}_n \cdots H\vec{v}_1 A \]
\[ Q = H\vec{v}_1 \cdots H\vec{v}_n \]

Can store \( Q \) implicitly by storing \( \vec{v}_i \)'s!
Slightly Different Output

- Gram-Schmidt: $Q \in \mathbb{R}^{m \times n}$, $R \in \mathbb{R}^{n \times n}$
- Householder: $Q \in \mathbb{R}^{m \times m}$, $R \in \mathbb{R}^{m \times n}$
Slightly Different Output

- **Gram-Schmidt:** \( Q \in \mathbb{R}^{m \times n}, \ R \in \mathbb{R}^{n \times n} \)
- **Householder:** \( Q \in \mathbb{R}^{m \times m}, \ R \in \mathbb{R}^{m \times n} \)

Typical least-squares case:
\[ A \in \mathbb{R}^{m \times n} \text{ has } m \gg n. \]
Desired

Stability of Householder with shape of Gram-Schmidt.
Shape of $R$

$$R = \begin{pmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
Reduced QR

\[ A = QR \]

\[ = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \]

\[ = Q_1 R_1 \]