Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

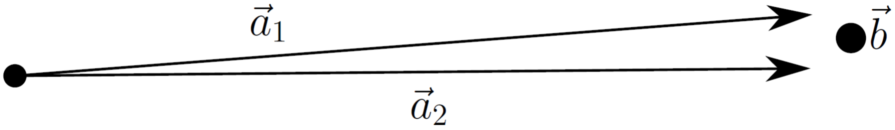
Lecture

Column Spaces and QR

Problem

$$cond\ A^T A \approx (cond\ A)^2$$

Geometric Intuition



Least-squares fit is ambiguous!

When Is $cond\ A^T A \approx 1$?

$$cond\ I_{n\times n} = 1$$

(w.r.t. $||\cdot||_2$)

When Is $cond\ A^T A \approx 1$?

$$cond\ I_{n\times n} = 1$$

(w.r.t. $||\cdot||_2$)

Desirable: $A^T A \approx I_{n\times n}$

(then, $cond\ A^T A \approx 1$!)

When Is $cond\ A^T A \approx 1$?

$$cond\ I_{n\times n} = 1$$

(w.r.t. $||\cdot||_2$)

Desirable: $A^T A \approx I_{n\times n}$

(then, $cond\ A^T A \approx 1$!)

Doesn't mean $A = I_{n\times n}$.

Recall: Definition of Gram matrix

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ \vdots & & \\ - & \vec{q}_n^\top & - \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

When $Q^T Q = I_{n \times n}$

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

When $Q^T Q = I_{n \times n}$

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

Orthonormal; orthogonal matrix

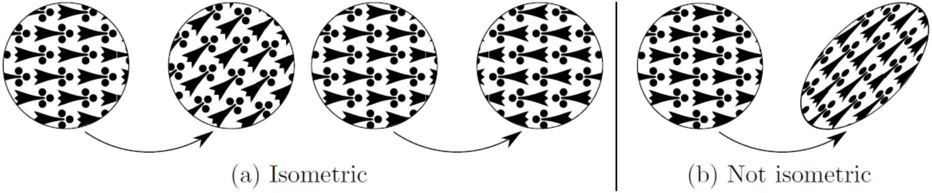
A set of vectors $\{\vec{v}_1, \cdots, \vec{v}_k\}$ is orthonormal if $||\vec{v}_i|| = 1$ for all i and $\vec{v}_i \cdot \vec{v}_j = 0$ for all $i \neq j$. A square matrix whose columns are orthonormal is called an orthogonal matrix.

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

$$A^T A\vec{x} = A^T b \leftrightarrow min_{\vec{x}}||A\vec{x} - \vec{b}||_2$$

Project onto the column space of A.

Observation

Lemma: Column space invariance

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

$$col\ A = col\ AB$$

.

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

$$A^T A\vec{x} = A^T\vec{b}, A = QR$$

$$\rightarrow \vec{x} = R^{-1}Q^T\vec{b}$$

Using QR

$$A^T A\vec{x} = A^T\vec{b}, A = QR$$

$$\to \vec{x} = R^{-1}Q^T\vec{b}$$

Didn't need to compute $A^T A$ or $(A^T A)^{-1}$

Vector Projection

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

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

Vector Projection

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

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

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

Vector Projection

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

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

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

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

Properties of Projection

$$proj_{\vec{a}}\vec{b} \| \vec{a}$$

$$\vec{a} \cdot (\vec{b} - proj_{\vec{a}}\vec{b}) = 0$$
$$\Rightarrow (\vec{b} - proj_{\vec{a}}\vec{b}) \perp \vec{a}$$

Orthonormal Projection

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

$$proj_{\hat{a}_i}\vec{b} = (\hat{a}_i \cdot \vec{b})\hat{a}_i$$

Orthonormal Projection

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

Orthonormal Projection

$$\left\| c_1\hat{a}_1 + c_2\hat{a}_2 + \cdots + c_k\hat{a}_k - \vec{b} \right\|_2^2 =$$

$$\sum_{i=1}^k \left(c_i^2 - 2c_i\vec{b}\cdot\hat{a}_i \right) + \|\vec{b}\|_2^2$$

$$\Rightarrow c_i = \vec{b}\cdot\hat{a}_i$$

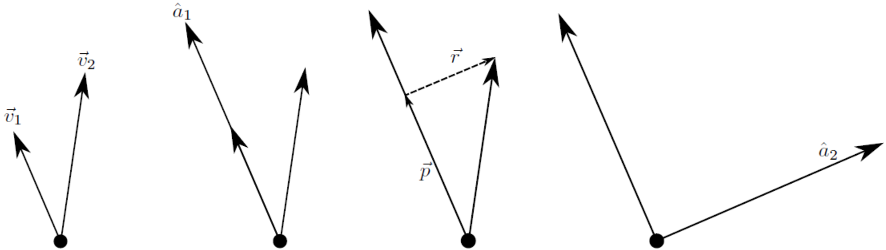
Orthonormal Projection

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

$$\Rightarrow c_i = \vec{b}\cdot\hat{a}_i$$

$$\Rightarrow \mathrm{proj}_{\mathrm{span}\{\hat{a}_1,\cdots,\hat{a}_k\}}\vec{b} = \left(\hat{a}_1\cdot\vec{b}\right)\hat{a}_1 + \cdots + \left(\hat{a}_k\cdot\vec{b}\right)\hat{a}_k$$

Geometric Strategy for Orthogonalization



(a) Input (b) Rescaling (c) Projection (d) Normalization

Gram-Schmidt Orthogonalization

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

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

Gram-Schmidt Orthogonalization

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

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

Claim

$span \{\vec{v}_1, \cdots, \vec{v}_i\} = span\ \hat{a}_1, \cdots, \hat{a}_i$ for all i.

Implementation via Column Operations

Post-multiplication!

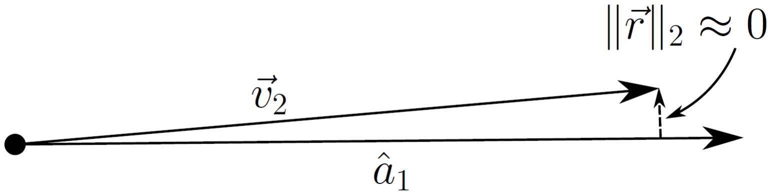
1. Rescaling to unit length: diagonal matrix
2. Subtracting off projection: upper triangular substitution matrix

New Factorization

$$A = QR$$

- Q orthogonal
- R upper-triangular

Bad Case



$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ 1+\varepsilon \end{pmatrix}$$

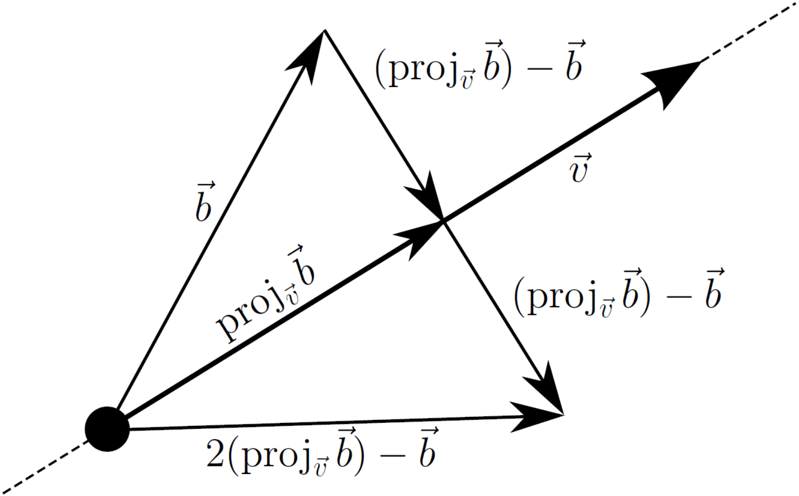
Two Strategies for QR

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



"Easy" Class of Orthogonal Matrices

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

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

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

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

Analogy to Forward Substitution

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

$$c\vec{e}_1 = H_{\vec{v}}\vec{a} \Rightarrow \vec{v} = (\vec{a} - c\vec{e}_1) \cdot \frac{\vec{v}^T\vec{v}}{2\vec{v}^T\vec{a}}$$

Analogy to Forward Substitution

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

$$c\vec{e}_1 = H_{\vec{v}}\vec{a}$$
$$\Rightarrow \vec{v} = (\vec{a} - c\vec{e}_1) \cdot \frac{\vec{v}^T\vec{v}}{2\vec{v}^T\vec{a}}$$

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

$$\Rightarrow c = \pm||\vec{a}||_2$$

After One Step

$$H_{\vec{v}}A = \begin{pmatrix} c & \times & \times & \times \\ 0 & \times & \times & \times \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \times & \times & \times \end{pmatrix}$$

Later Steps

$$\vec{a} = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \end{pmatrix} \mapsto H_{\vec{v}}\vec{a} = \begin{pmatrix} \vec{a}_1 \\ \vec{0} \end{pmatrix}$$

Leave first k lines alone!

Householder QR

$$R = H_{\vec{v}_n} \cdots H_{\vec{v}_1} A$$

$$Q = H_{\vec{v}_1}^T \cdots H_{\vec{v}_n}^T$$

Householder QR

$$R = H_{\vec{v}_n} \cdots H_{\vec{v}_1} A$$

$$Q = H_{\vec{v}_1}^T \cdots H_{\vec{v}_n}^T$$

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:
$A \in \mathbb{R}^{m \times n}$ has $m \gg n$.

Desired

Stability of Householder with shape of Gram-Schmidt.

Shape of R

$$R = \left(\begin{array}{ccc} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Reduced QR

$$
\begin{aligned}
A &= QR \\
&= (Q_1 Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \\
&= Q_1 R_1
\end{aligned}
$$