Mathematics Methods for Computer Science

Initial Observations

Orthogonality

Least-Square

Projections

. . .

Hausahaldar OB

Reduced QR

Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

Mathematics Methods for Computer Science

Initial Observations

Orthogonality

Least-Square

Duningtions

Gram-Schmidt

Householder QR

Reduced QR

Lecture

Column Spaces and QR

Initial Observations

Orthogonality

Least-Square

Projection

Cram Sahmid

Householder OR

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

Mathematics Methods for Computer Science

Geometric Intuition

Initial Observations

Orthogonality

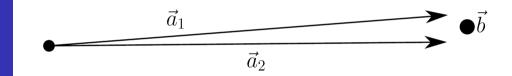
Least-Square

Draioction

Gram-Schmic

Householder QF

Reduced QR



Least-squares fit is ambiguous!

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmidt

Householder QR

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

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QF

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

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QF

Reduced QR

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

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

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

Mathematics Methods for Computer Science

Recall: Definition of Gram matrix

Initial Observations

Orthogonality

Least-Square

Proiectio

Gram-Schmid

Householder QI

$$Q^{\top}Q = \begin{pmatrix} - & \vec{q}_{1}^{\top} & - \\ - & \vec{q}_{2} & - \\ \vdots & & \\ - & \vec{q}_{n} \end{pmatrix} \begin{pmatrix} | & | & | & | \\ \vec{q}_{1} & \vec{q}_{2} & \cdots & \vec{q}_{n} \\ | & | & | & | \end{pmatrix}$$

$$= \begin{pmatrix} \vec{q}_{1} \cdot \vec{q}_{1} & \vec{q}_{1} \cdot \vec{q}_{2} & \cdots & \vec{q}_{1} \cdot \vec{q}_{n} \\ \vec{q}_{2} \cdot \vec{q}_{1} & \vec{q}_{2} \cdot \vec{q}_{2} & \cdots & \vec{q}_{2} \cdot \vec{q}_{n} \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_{n} \cdot \vec{q}_{1} & \vec{q}_{n} \cdot \vec{q}_{2} & \cdots & \vec{q}_{n} \cdot \vec{q}_{n} \end{pmatrix}$$

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmidt

Householder QR

$$\vec{q_i} \cdot \vec{q_j} = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}$$

Initial Observations

Orthogonality

Least-Square

Projectio

Gram-Schmid

Householder QI

Reduced QR

$$\vec{q_i} \cdot \vec{q_j} = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}$$

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR

$$||Q\vec{x}||^2 = ?$$

$$(Q\vec{x}) \cdot (Q\vec{y}) = ?$$

Mathematics Methods for Computer Science

Geometric Interpretation

Initial Observations

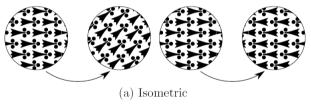
Orthogonality

Least-Square

Projectio

Grann-Schinic

Householder Q





(b) Not isometric

Alternative Intuition for Least-Squares

Initial Observations

Orthogonalit

Least-Squares

Projection

Gram-Schmid

Householder QR

Reduced QR

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

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

Initial Observation

Orthogonality

Least-Squares

Projections

Gram-Schmid

Householder QR

Reduced QR

Invertible column operations do not affect column space.

New Strategy

Initial Observations

Orthogonality

Least-Squares

Projections

_ _ _ .

Householder OF

Reduced QR

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

New Factorization

Initial Observations

Orthogonality

Least-Squares

Projection

Gram-Schmid

Householder QI

$$A = QR$$

- $oldsymbol{\cdot} Q$ orthogonal
- $\cdot R$ upper triangular

Initial Observations

Orthogonality

Least-Squares

Projection

Gram-Schmid

Householder QR

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

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

Initial Observations

Orthogonality

Least-Squares

_ . . .

Gram-Schmid

Householder QF

Reduced QR

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

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

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

Vector Projection

Initial Observations

Orthogonality

Least-Square

Projections

Gram-Schmidt

Householder OR

Reduced QR

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

Vector Projection

Initial Observations

Orthogonality

Least-Square

Projections

Gram-Schmidt

Householder QF

Reduced QR

"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

Initial Observations

Orthogonality

Least-Square

Projections

Gram-Schmidt

Householder QF

Reduced QR

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

$$mn_c||ca-b||$$

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

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

Properties of Projection

nitial Observations

Orthogonality

Least-Square

Projections

Gram-Schmid

Householder OR

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

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

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

Orthonormal Projection

Initial Observations

Orthogonalit

Least-Square

Projections

Gram-Schmidt

Householder Of

Reduced QR

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

Initial Observations

Orthogonalit

Least-Square

Projections

Gram-Schmidt

Householder QR

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

Orthonormal Projection

Initial Observations

Orthogonality

Least-Square

Projections

Gram-Schmidt

Householder QF

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

Mathematics Methods for Computer Science

Orthonormal Projection

Initial Observations

Orthogonality

Least-Square

Projections

Gram-Schmid

Householder QI

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

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

Mathematics Methods for Computer Science

Geometric Strategy for Orthogonalization

Initial Observation

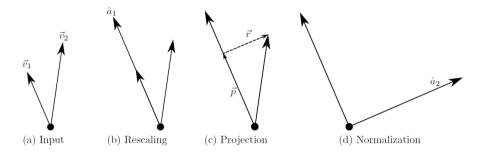
Orthogonalit

Least-Squar

Projection

Gram-Schmidt

Householder Q



Gram-Schmidt Orthogonalization

Initial Observations

Orthogonality

Least-Squar

Projectio

Gram-Schmidt

Householder QF

Reduced QR

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

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

• For i from 2 to k,

$$\vec{p_i} \equiv \operatorname{proj}_{\mathsf{span}} \{\hat{a}_1, \cdots, \hat{a}_{i-1}\} \vec{v_i}$$

$$\hat{a}_i \equiv \frac{\vec{v}_i - \vec{p}_i}{\|\vec{v}_i - \vec{p}_i\|}$$

Gram-Schmidt Orthogonalization

Orthogonality

Least-Squar

Projection

Gram-Schmidt

Householder QF

Reduced QR

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

$$\bullet \quad \hat{a}_1 \equiv \frac{\vec{v}_1}{||\vec{v}_1||}$$

• For i from 2 to k,

$$\vec{p_i} \equiv \operatorname{proj}_{\mathsf{span}} \{\hat{a}_1, \cdots, \hat{a}_{i-1}\} \vec{v_i}$$

$$\hat{a}_i \equiv \frac{\vec{v}_i - \vec{p}_i}{\|\vec{v}_i - \vec{p}_i\|}$$

Claim

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

Implementation via Column Operations

itial Observations

Orthogonalit

Least-Square

Projection

Gram-Schmidt

Householder QF

Reduced QR

Post-multiplication!

- Rescaling to unit length: diagonal matrix
- Subtracting off projection: upper triangular substitution matrix

New Factorization

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmidt

Householder QR

$$A = QR$$

- ullet Q orthogonal
- ullet R upper-triangular

Bad Case

Initial Observations

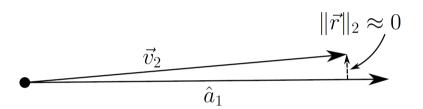
Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR



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

Two Strategies for QR

Initial Observations

Orthogonalit

Least-Square

Projections

Cram Sahmid

Householder QR

Reduced QR

Post-multiply by upper triangular matrices

Two Strategies for QR

Initial Observations

Orthogonality

Least-Square

rojectio

Gram-Schmid

Householder QR

- Post-multiply by upper triangular matrices
- Pre-multiply by orthogonal matrices
 New idea!

Mathematics Methods for Computer Science

"Easy" Class of Orthogonal Matrices

nitial Observations

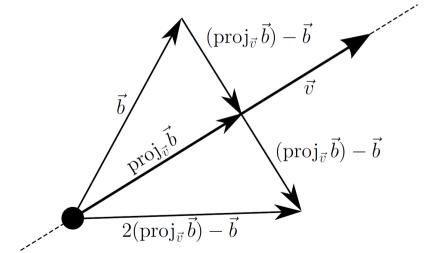
Orthogonality 8 1

Least-Square

Projection

Gram-Schmid

Householder QR



"Easy" Class of Orthogonal Matrices

Initial Observations

Orthogonalit

Least-Square

Proiection

Gram-Schmid

Householder QR

$$\begin{split} 2 \operatorname{proj}_{\vec{v}} \vec{b} - \vec{b} &= 2 \frac{\vec{v} \cdot \vec{b}}{\vec{v} \cdot \vec{v}} \vec{v} - \vec{b} \text{ by definition of projection} \\ &= 2 \vec{v} \cdot \frac{\vec{v}^\top \vec{b}}{\vec{v}^\top \vec{v}} - \vec{b} \text{ using matrix notation} \\ &= \left(\frac{2 \vec{v} \vec{v}^\top}{\vec{v}^\top \vec{v}} - I_{n \times n} \right) \vec{b} \\ &\equiv -H_{\vec{v}} \vec{b}, \text{ where } H_{\vec{v}} \equiv I_{n \times n} - \frac{2 \vec{v} \vec{v}^\top}{\vec{v}^\top \vec{v}} \end{split}$$

Analogy to Forward Substitution

Initial Observations

Orthogonalit

Least-Square

Projection

Cram Sahmida

Householder QR

Reduced QR

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR

Reduced QR

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

Initial Observations

Orthogonality 8 1

Least-Square

Projection

Gram-Schmidt

Householder QR

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

Later Steps

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR

Reduced QR

$$\vec{a} = \left(\begin{array}{c} \vec{a}_1 \\ \vec{a}_2 \end{array} \right) \mapsto H_{\vec{v}}\vec{a} = \left(\begin{array}{c} \vec{a}_1 \\ \overrightarrow{0} \end{array} \right)$$

Leave first k lines alone!

Householder QR

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

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

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR

Reduced 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

Initial Observations

Orthogonalit

Least-Square

Projection

Gram-Schmid

Householder Q

- 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

Initial Observations

Orthogonalit

Least-Square

Projectio

Gram-Schmic

Householder Qf

Reduced QR

- 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 $m \gg n$.

Initial Observations

Orthogonality

Least-Square

Projectio

Cram Sahmid

Householder OF

Reduced QR

Stability of Householder with shape of Gram-Schmidt.

Shape of R

Initial Observations

Orthogonality 8 1

Least-Square

Projection

Gram-Schmid

Householder QR

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

Reduced QR

Initial Observations

Orthogonality

Least-Square

Projection

Gram-Schmid

Householder QR

$$A = QR$$

$$= (Q_1Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$

$$= Q_1R_1$$