

SI231b: Matrix Computations

Lecture 9: QR Factorization

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology
ShanghaiTech University

Oct. 13, 2020

Recap: Orthogonal Projection

Suppose $A \in \mathbb{R}^{m \times n}$ ($m > n$) has full rank, to perform the orthogonal projection onto the column space of A , i.e., $\mathcal{R}(A)$, the orthogonal projector P

- ▶ when $\{q_1, q_2, \dots, q_n\}$ form an orthonormal basis of $\mathcal{R}(A)$,

$$P = QQ^T,$$

where $Q = [q_1, q_2, \dots, q_n]$ (already introduced in the last lecture)

- ▶ for arbitrary basis $\{a_1, a_2, \dots, a_n\}$ of $\mathcal{R}(A)$,

$$P = A(A^T A)^{-1}A^T,$$

where $A = [a_1, a_2, \dots, a_n]$ (you are supposed to work it out independently)

- ▶ QR Factorization
- ▶ QR Factorization by Gram-Schmidt Orthogonalization
- ▶ QR Factorization through Householder Reflection

Reduced QR Factorization

Given a basis $\{a_1, a_2, \dots, a_n\}$ of a subspace \mathcal{S} , we want to compute its orthonormal basis $\{q_1, q_2, \dots, q_n\}$

Key: through iterative process and using the fact that

- ▶ $\text{span}\{a_1, a_2, \dots, a_k\} = \text{span}\{q_1, q_2, \dots, q_k\}, k = 1, 2, \dots, n$
- ▶ $\text{span}\{a_1, a_2, \dots, a_k\} \subset \text{span}\{a_1, a_2, \dots, a_k, a_{k+1}\}$

This gives

$$\underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}}_A = \underbrace{\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}}_Q \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}}_R$$

with $r_{kk} \neq 0$. This is called the *reduced QR factorization* of A .

Full QR Factorization

Extending the reduced QR factorization by adding $m - n$ columns to Q so that

$$\tilde{Q} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n & q_{n+1} & \cdots & q_m \end{bmatrix}$$

is an orthogonal matrix ($\tilde{Q} \in \mathbb{R}^{m \times m}$)

- **orthogonal matrix**: a square matrix with orthonormal columns, i.e.,

$$\tilde{Q}^T \tilde{Q} = I_m$$

Then $A = \tilde{Q}\tilde{R}$ with $\tilde{R} = \begin{bmatrix} R \\ 0 \end{bmatrix}$



Figure 1: Reduced QR Factorization

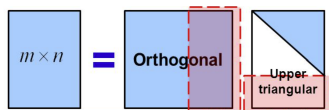


Figure 2: Full QR Factorization

How to compute the orthonormal basis?

Orthogonal projection of vector a onto vector b

$$\text{proj}_b(a) = \frac{\langle a, b \rangle}{\langle b, b \rangle} b,$$

where $\langle \rangle$ represents the inner product of two vectors.

$$q_1 = \frac{a_1}{\|a_1\|}$$

$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1$$

$$q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$$

$$\vdots$$

$$\tilde{q}_k = a_k - (q_1^T a_k) q_1 - (q_2^T a_k) q_2 - \cdots - (q_{k-1}^T a_k) q_{k-1}$$

$$q_k = \frac{\tilde{q}_k}{\|\tilde{q}_k\|}$$

Can you also explain in the context of projection onto subspaces?

Algorithm: Gram-Schmidt Orthogonalization (**numerically unstable**)

input: a collection of linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$

$\tilde{\mathbf{q}}_1 = \mathbf{a}_1$, $\mathbf{q}_1 = \tilde{\mathbf{q}}_1 / \|\tilde{\mathbf{q}}_1\|_2$

for $i = 2, \dots, n$

$\tilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^T \mathbf{a}_i) \mathbf{q}_j$

$\mathbf{q}_i = \tilde{\mathbf{q}}_i / \|\tilde{\mathbf{q}}_i\|_2$

end

output: $\mathbf{q}_1, \dots, \mathbf{q}_n$

► let $r_{ii} = \|\tilde{\mathbf{q}}_i\|_2$, $r_{ij} = \mathbf{q}_j^T \mathbf{a}_i$ for $j = 1, \dots, i-1$

► we see that $\mathbf{a}_i = \sum_{j=1}^i r_{ij} \mathbf{q}_j$ for all i , or, equivalently,

$$\mathbf{A} = \mathbf{Q}\mathbf{R},$$

where $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_n]$; \mathbf{R} is upper triangular with $[\mathbf{R}]_{ij} = r_{ij}$ for $j \leq i$

The (classic) Gram-Schmidt (CGS)

- ▶ gives orthogonal \tilde{q}_i in exact arithmetic
- ▶ is numerical unstable due to round off error accumulation on modern computers (returns non-orthogonal \tilde{q}_i)

Modified Gram-Schmidt (MGS)

Instead of computing $\tilde{q}_k = a_k - (q_1^T a_k)q_1 - (q_2^T a_k)q_2 - \cdots - (q_{k-1}^T a_k)q_{k-1}$,
but

$$\tilde{q}_k^{(1)} = a_k - (q_1^T a_k)q_1$$

$$\tilde{q}_k^{(2)} = \tilde{q}_k^{(1)} - (q_2^T \tilde{q}_k^{(1)})q_2$$

$$\vdots$$

$$\tilde{q}_k^{(j)} = \tilde{q}_k^{(j-1)} - (q_j^T \tilde{q}_k^{(j-1)})q_j$$

$$\vdots$$

Both CGS and MGS take $\mathcal{O}(2mn^2)$ flops

- ▶ a matrix $H \in \mathbb{R}^{m \times m}$ is called a **reflection matrix** if

$$H = I - 2P,$$

where P is an orthogonal projector.

- ▶ interpretation: denote $P^\perp = I - P$, and observe

$$x = Px + P^\perp x, \quad Hx = -Px + P^\perp x.$$

The vector Hx is a reflected version of x , with $\mathcal{R}(P)$ being the “mirror”

- ▶ a reflection matrix is orthogonal:

$$H^T H = (I - 2P)(I - 2P) = I - 4P + 4P^2 = I - 4P + 4P = I$$

► **Problem:** given $x \in \mathbb{R}^m$, find an orthogonal $H \in \mathbb{R}^{m \times m}$ such that

$$Hx = \begin{bmatrix} \beta \\ 0 \end{bmatrix} = \beta e_1, \quad \text{for some } \beta \in \mathbb{R}.$$

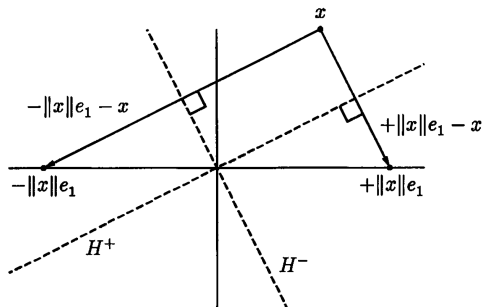


Figure 3: Householder reflection

- **Householder reflection:** let $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v} \neq \mathbf{0}$. Let

$$\mathbf{H} = \mathbf{I} - \frac{2}{\|\mathbf{v}\|_2^2} \mathbf{v} \mathbf{v}^T,$$

which is a reflection matrix with $\mathbf{P} = \mathbf{v} \mathbf{v}^T / \|\mathbf{v}\|_2^2$

- it can be verified that (try)

$$\mathbf{v} = \mathbf{x} \mp \|\mathbf{x}\|_2 \mathbf{e}_1 \implies \mathbf{H} \mathbf{x} = \pm \|\mathbf{x}\|_2 \mathbf{e}_1;$$

the sign above may be determined to be the one that maximizes $\|\mathbf{v}\|_2$, for the sake of numerical stability

- let $H_1 \in \mathbb{R}^{m \times m}$ be the Householder reflection w.r.t. a_1 . Transform A as

$$A^{(1)} = H_1 A = \begin{bmatrix} \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \dots & \times \end{bmatrix}$$

- let $\tilde{H}_2 \in \mathbb{R}^{(m-1) \times (m-1)}$ be the Householder reflection w.r.t. $A_{2:m,2}^{(1)}$ (marked red above). Transform $A^{(1)}$ as

$$A^{(2)} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{bmatrix}}_{=H_2} A^{(1)} = \begin{bmatrix} \times & \times & \dots & \times \\ 0 & \tilde{H}_2 A_{2:m,2}^{(1)} \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \dots & \times \\ 0 & \times & \times & \dots & \times \\ \vdots & 0 & \times & \dots & \times \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \times & \dots & \times \end{bmatrix}$$

- by repeatedly applying the trick above, we can transform A as the desired

R

$$A^{(0)} = A$$

for $k = 1, \dots, n-1$

$$A^{(k)} = H_k A^{(k-1)}, \text{ where}$$

$$H_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & \tilde{H}_k \end{bmatrix},$$

I_k is the $k \times k$ identity matrix; \tilde{H}_k is the Householder reflection of $A_{k:m,k}^{(k-1)}$

end

- ▶ H_k introduces zeros under the diagonal of the k -th column
- ▶ the above procedure results in

$$A^{(n-1)} = H_{n-1} \cdots H_2 H_1 A, \quad A^{(n-1)} \text{ taking an upper triangular form}$$

- ▶ by letting $R = A^{(n-1)}$, $Q = (H_{n-1} \cdots H_2 H_1)^T$, we obtain the full QR
- ▶ a popularly used method for QR decomposition

You are supposed to read

- ▶ Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, Johns Hopkins University Press, 2013.

Chapter 5.1 – 5.3