SI231b: Matrix Computations

Lecture 11: QR Factorization

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Oct. 21, 2021

Recap: Orthogonal Projection

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

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

  $$\mathrm{P} = \mathrm{Q}\mathrm{Q}^T,$$

  where $\mathrm{Q} = [q_1,\ q_2,\ \cdots,\ q_n]$

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

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

  where $\mathrm{A} = [a_1,\ a_2,\ \cdots,\ a_n]$

Given a basis $\{a_1, \ a_2, \ \cdots, \ a_n\}$ of a subspace $\mathcal{S}$, how to compute its orthogonal/orthonormal basis $\{q_1, \ q_2, \ \cdots, \ q_n\}$?

Key: through iterative process and using the fact that

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

Gram-Schmidt orthogonalization.

Key: orthogonal projection of vector a onto vector b

$$\text{proj}_{\text{b}}(\text{a}) = \frac{< \text{a}, \text{b} >}{< \text{b}, \text{b} >}\text{b},$$

where $<>$ represents the inner product of two vectors.

How to compute the orthonormal basis?

Orthogonal projection of vector a onto vector b

$$\mathrm{proj_b(a)} = \frac{<\mathrm{a}, \mathrm{b}>}{<\mathrm{b}, \mathrm{b}>}\mathrm{b},$$

where $<>$ represents the inner product of two vectors.

$$\begin{aligned}
\mathrm{q}_1 &= \frac{\mathrm{a}_1}{\|\mathrm{a}_1\|} \\
\tilde{\mathrm{q}}_2 &= \mathrm{a}_2 - (\mathrm{q}_1^T \mathrm{a}_2)\mathrm{q}_1 \\
\mathrm{q}_2 &= \frac{\tilde{\mathrm{q}}_2}{\|\tilde{\mathrm{q}}_2\|} \\
&\vdots \\
\tilde{\mathrm{q}}_k &= \mathrm{a}_k - (\mathrm{q}_1^T \mathrm{a}_k)\mathrm{q}_1 - (\mathrm{q}_2^T \mathrm{a}_k)\mathrm{q}_2 - \cdots - (\mathrm{q}_{k-1}^T \mathrm{a}_k)\mathrm{q}_{k-1} \\
\mathrm{q}_k &= \frac{\tilde{\mathrm{q}}_k}{\|\tilde{\mathrm{q}}_k\|}
\end{aligned}$$

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

Algorithm: Gram-Schmidt Orthogonalization (numerically unstable)

input: a collection of linearly independent vectors $a_1, \ldots, a_n$

$\tilde{q}_1 = a_1$, $q_1 = \tilde{q}_1 / \|\tilde{q}_1\|_2$

for $i = 2, \ldots, n$

$\quad \tilde{q}_i = a_i - \sum_{j=1}^{i-1} (q_j^T a_i) q_j$

$\quad q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$

end

output: $q_1, \ldots, q_n$

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

$$
\begin{aligned}
\tilde{q}_k^{(1)} &= a_k - (q_1^T a_k)q_1 \\
\tilde{q}_k^{(2)} &= \tilde{q}_k^{(1)} - (q_2^T \tilde{q}_k^{(1)})q_2 \\
&\vdots \\
\tilde{q}_k^{(j)} &= \tilde{q}_k^{(j-1)} - (q_j^T \tilde{q}_k^{(j-1)})q_j \\
&\vdots
\end{aligned}
$$

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

Given $a_1 = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T$, $a_2 = \begin{bmatrix} 1 & 0 & \epsilon & 0 \end{bmatrix}^T$, $a_3 = \begin{bmatrix} 1 & 0 & 0 & \epsilon \end{bmatrix}^T$, compare classical and modified Gram-Schmidt for

$$\mathcal{V} = \operatorname{span}\{a_1,\ a_2,\ a_3\}$$

where the approximation $1 + \epsilon^2 = 1$ can be made.

Classical Gram-Schmidt

- $q_1 = \frac{a_1}{\|a_1\|_2} = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T$
- $q_2 = a_2 - q_1^T a_2 q_1 = \begin{bmatrix} 0 & -\epsilon & \epsilon & 0 \end{bmatrix}^T$

  $q_2 = \frac{q_2}{\|q_2\|_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 1 & 0 \end{bmatrix}^T$
- $q_3 = a_3 - q_1^T a_3 q_1 - q_2^T a_3 q_2 = \begin{bmatrix} 0 & -\epsilon & 0 & \epsilon \end{bmatrix}^T$

  $q_3 = \frac{q_3}{\|q_3\|_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix}^T$

Orthogonality is lost

Modified Gram-Schmidt

- $\tilde{q}_1 = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T$

  $q_1 = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T$

- $\tilde{q}_2 = a_2 - q_1^T a_2 q_1 = \begin{bmatrix} 0 & -\epsilon & \epsilon & 0 \end{bmatrix}^T$

  $q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 1 & 0 \end{bmatrix}^T$

- $\tilde{q}_3 = a_3 - q_1^T a_3 q_1 = \begin{bmatrix} 0 & -\epsilon & 0 & \epsilon \end{bmatrix}^T$

  $\tilde{q}_3 = \tilde{q}_3 - q_2^T \tilde{q}_3 q_2 = \begin{bmatrix} 0 & -\frac{\epsilon}{2} & -\frac{\epsilon}{2} & \epsilon \end{bmatrix}^T$

  $q_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 & -1 & -1 & 2 \end{bmatrix}^T$

Orthogonality is preserved

Reduced QR Factorization

For a full rank matrix $A \in \mathbb{R}^{m \times n}$ ($m > n$), the Gram-Schmidt procedure 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 $\tilde{R} = \begin{bmatrix} R \\ 0 \end{bmatrix}$

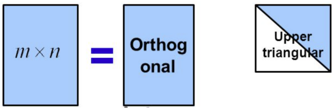

Figure 1: Reduced QR Factorization

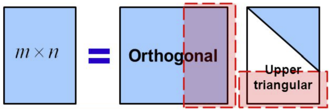

Figure 2: Full QR Factorization

One of the Top 10 Algorithms in the 20th Century[1]

Given a rectangular matrix $A \in \mathbb{R}^{m \times n}$, A can be factorized into the form

$$A = QR$$

where

- $Q \in \mathbb{R}^{m \times m}$ is an orthogonal matrix
- $R \in \mathbb{R}^{m \times n}$ is upper-triangular

Reduced QR Factorization

For $m > n$, the reduced QR factorization given by

- $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns
- $R \in \mathbb{R}^{n \times n}$ is upper-triangular
- also called 'economic' QR factorization in some cases

[1] https://doi.ieeecomputersociety.org/10.1109/MCISE.2000.814652

Reflection Matrices

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

$$\mathbf{H} = \mathbf{I} - 2\mathbf{P},$$

where $\mathbf{P}$ is an orthogonal projector.

- interpretation: denote $\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P}$, and observe

$$\mathbf{x} = \mathbf{P}\mathbf{x} + \mathbf{P}^{\perp}\mathbf{x}, \qquad \mathbf{H}\mathbf{x} = -\mathbf{P}\mathbf{x} + \mathbf{P}^{\perp}\mathbf{x}.$$

The vector $\mathbf{H}\mathbf{x}$ is a reflected version of $\mathbf{x}$, with $\mathcal{R}(\mathbf{P}^{\perp})$ being the "mirror"

- a reflection matrix is orthogonal:

$$\mathbf{H}^T\mathbf{H} = (\mathbf{I} - 2\mathbf{P})(\mathbf{I} - 2\mathbf{P}) = \mathbf{I} - 4\mathbf{P} + 4\mathbf{P}^2 = \mathbf{I} - 4\mathbf{P} + 4\mathbf{P} = \mathbf{I}$$

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

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

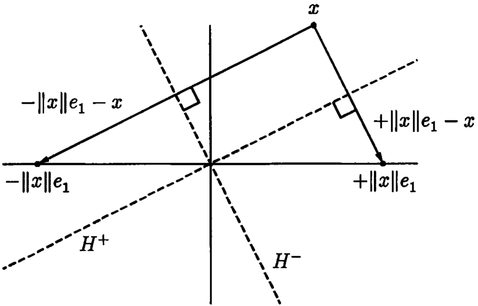

Figure 3: Householder reflection

- Householder reflection: let $v \in \mathbb{R}^m$, $v \neq 0$. Let
$$H = I - \frac{2}{\|v\|_2^2} vv^T,$$
which is a reflection matrix with $P = vv^T / \|v\|_2^2$

- it can be verified that (try)
$$v = x \mp \|x\|_2 e_1 \quad \Longrightarrow \quad Hx = \pm \|x\|_2 e_1;$$
the sign above may be determined to be the one that maximizes $\|v\|_2$, 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

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

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

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

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

Householder QR

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

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

$\mathbf{A}^{(k)} = \mathbf{H}_k \mathbf{A}^{(k-1)}$, where

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{I}_{k-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_k \end{bmatrix},$$

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

end

- $\mathbf{H}_k$ introduces zeros under the diagonal of the k-th column
- the above procedure results in

$$\mathbf{A}^{(n-1)} = \mathbf{H}_{n-1} \cdots \mathbf{H}_2 \mathbf{H}_1 \mathbf{A}, \quad \mathbf{A}^{(n-1)} \text{ taking an upper triangular form}$$

- by letting $\mathbf{R} = \mathbf{A}^{(n-1)}$, $\mathbf{Q} = (\mathbf{H}_{n-1} \cdots \mathbf{H}_2 \mathbf{H}_1)^T$, we obtain the full QR
- a popularly used method for QR decomposition

Readings

You are supposed to read

- Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra*, SIAM, 1997.

  Lecture 6, 8, 11

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

  Chapter 5.1 – 5.3