SI231b: Matrix Computations

Lecture 11: QR Factorization

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Oct. 17, 2022

Recap: Orthogonal Projection

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

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

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

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

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

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

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

Computing Orthonormal Basis

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**

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

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

How to compute the orthonormal basis?

Orthogonal projection of vector $\mathbf{a}$ onto vector $\mathbf{b}$

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

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

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

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

Algorithm: Gram-Schmidt Orthogonalization (numerically unstable)

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

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

for $i = 2, \ldots, n$

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

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

end

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

Modified Gram-Schmidt Orthogonalization

The (classic) Gram-Schmidt (CGS)

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

Modified Gram-Schmidt (MGS)

Instead of computing $\tilde{\mathbf{q}}_k = \mathbf{a}_k - (\mathbf{q}_1^T\mathbf{a}_k)\mathbf{q}_1 - (\mathbf{q}_2^T\mathbf{a}_k)\mathbf{q}_2 - \cdots - (\mathbf{q}_{k-1}^T\mathbf{a}_k)\mathbf{q}_{k-1}$, but

$$\begin{aligned}
\tilde{\mathbf{q}}_k^{(1)} &= \mathbf{a}_k - (\mathbf{q}_1^T\mathbf{a}_k)\mathbf{q}_1 \\
\tilde{\mathbf{q}}_k^{(2)} &= \tilde{\mathbf{q}}_k^{(1)} - (\mathbf{q}_2^T\tilde{\mathbf{q}}_k^{(1)})\mathbf{q}_2 \\
&\vdots \\
\tilde{\mathbf{q}}_k^{(j)} &= \tilde{\mathbf{q}}_k^{(j-1)} - (\mathbf{q}_j^T\tilde{\mathbf{q}}_k^{(j-1)})\mathbf{q}_j \\
&\vdots
\end{aligned}$$

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

Given $\mathbf{a}_1 = \begin{bmatrix}1 & \epsilon & 0 & 0\end{bmatrix}^T$, $\mathbf{a}_2 = \begin{bmatrix}1 & 0 & \epsilon & 0\end{bmatrix}^T$, $\mathbf{a}_3 = \begin{bmatrix}1 & 0 & 0 & \epsilon\end{bmatrix}^T$, compare classical and modified Gram-Schmidt for

$$\mathcal{V} = \text{span}\left\{\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3\right\}$$

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

Classical Gram-Schmidt

- $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2} = \begin{bmatrix}1 & \epsilon & 0 & 0\end{bmatrix}^T$
- $\mathbf{q}_2 = \mathbf{a}_2 - \mathbf{q}_1^T\mathbf{a}_2\mathbf{q}_1 = \begin{bmatrix}0 & -\epsilon & \epsilon & 0\end{bmatrix}^T$

  $\mathbf{q}_2 = \frac{\mathbf{q}_2}{\|\mathbf{q}_2\|_2} = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & -1 & 1 & 0\end{bmatrix}^T$
- $\mathbf{q}_3 = \mathbf{a}_3 - \mathbf{q}_1^T\mathbf{a}_3\mathbf{q}_1 - \mathbf{q}_2^T\mathbf{a}_3\mathbf{q}_2 = \begin{bmatrix}0 & -\epsilon & 0 & \epsilon\end{bmatrix}^T$

  $\mathbf{q}_3 = \frac{\mathbf{q}_3}{\|\mathbf{q}_3\|_2} = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & -1 & 0 & 1\end{bmatrix}^T$

Orthogonality is lost

Modified Gram-Schmidt

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

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

- $\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \mathbf{q}_1^T \mathbf{a}_2 \mathbf{q}_1 = \begin{bmatrix} 0 & -\epsilon & \epsilon & 0 \end{bmatrix}^T$

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

- $\tilde{\mathbf{q}}_3 = \mathbf{a}_3 - \mathbf{q}_1^T \mathbf{a}_3 \mathbf{q}_1 = \begin{bmatrix} 0 & -\epsilon & 0 & \epsilon \end{bmatrix}^T$

  $\tilde{\mathbf{q}}_3 = \tilde{\mathbf{q}}_3 - \mathbf{q}_2^T \tilde{\mathbf{q}}_3 \mathbf{q}_2 = \begin{bmatrix} 0 & -\frac{\epsilon}{2} & -\frac{\epsilon}{2} & \epsilon \end{bmatrix}^T$

  $\mathbf{q}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 & -1 & -1 & 2 \end{bmatrix}^T$

Orthogonality is preserved

For a full rank matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ $(m > n)$, the Gram-Schmidt procedure gives

$$\underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}}_{\mathbf{R}}$$

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

Full QR Factorization

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

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

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

- orthogonal matrix: a square matrix with orthonormal columns, i.e., $\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}} = \mathbf{I}_m$

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

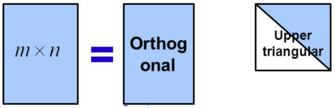

Figure 1: Reduced QR Factorization

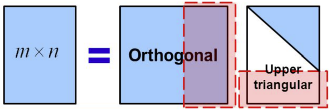

Figure 2: Full QR Factorization

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

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

$$\mathbf{A} = \mathbf{QR}$$

where

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

Reduced QR Factorization

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

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

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

QR Factorization Computations: 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 $\mathbf{x} \in \mathbb{R}^m$, find an orthogonal $\mathbf{H} \in \mathbb{R}^{m \times m}$ such that

$$\mathbf{Hx} = \begin{bmatrix} \beta \\ \mathbf{0} \end{bmatrix} = \beta \mathbf{e}_1, \qquad \text{for some } \beta \in \mathbb{R}.$$

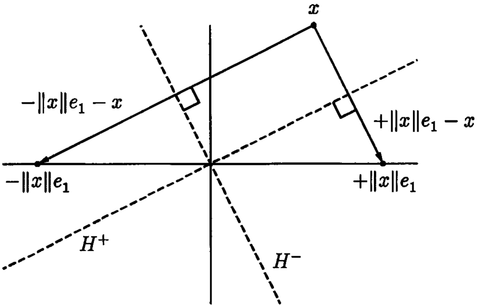

Figure 3: Householder reflection

- 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)

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

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

Householder QR

- let $\mathbf{H}_1 \in \mathbb{R}^{m\times m}$ be the Householder reflection w.r.t. $\mathbf{a}_1$. Transform $\mathbf{A}$ as

$$\mathbf{A}^{(1)} = \mathbf{H}_1\mathbf{A} = \begin{bmatrix} \times & \times & \dots & \times \\ 0 & {\color{red}\times} & \dots & \times \\ \vdots & \vdots & & \vdots \\ 0 & {\color{red}\times} & \dots & \times \end{bmatrix}$$

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

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

- by repeatedly applying the trick above, we can transform $\mathbf{A}$ as the desired $\mathbf{R}$

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 $\mathbf{A}^{(k-1)}_{k:m,k}$

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