MSBD 5007 HW2

RONG Shuo

March 27, 2025

Question1

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 6 \\ 5 \\ 7 \\ 4 \end{bmatrix}$. Use the QR factorization to solve the least square problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^4} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2.$$

Answer

We can calculate the QR factorization of $\boldsymbol{A}$ by Householder transformation.

$$\boldsymbol{Q} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} \sqrt{3} & \sqrt{3} & \frac{4}{\sqrt{3}} & \sqrt{3} \\ 0 & \sqrt{3} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{2}{\sqrt{3}} & \frac{2}{\sqrt{3}} \\ 0 & 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}$$

$$\boldsymbol{Q}^T \boldsymbol{b} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \\ 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 6\sqrt{3} \\ \sqrt{3} \\ \sqrt{3} \\ 2\sqrt{3} \end{bmatrix}$$

We need to solve $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^T\boldsymbol{b}$.

$$\begin{bmatrix} \sqrt{3} & \sqrt{3} & \frac{4}{\sqrt{3}} & \sqrt{3} \\ 0 & \sqrt{3} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{2}{\sqrt{3}} & \frac{2}{\sqrt{3}} \\ 0 & 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6\sqrt{3} \\ \sqrt{3} \\ \sqrt{3} \\ 2\sqrt{3} \end{bmatrix}$$

The solution is :

$$\begin{bmatrix} \frac{5}{2} \\ \frac{5}{2} \\ -\frac{3}{2} \\ 3 \end{bmatrix}$$

Question2

Find the best rank-1 approximation with respect to the Frobenius norm to the matrix.

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Answer

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 5 & 5 \\ 5 & 6 & 5 \\ 5 & 5 & 6 \end{bmatrix}$$

We need solve the equation,

$$-(\lambda - 16)(\lambda - 1)^2 = 0$$

The roots are $\lambda = 16, 1, 1$. We can calculate the normalized eigenvectors of $\lambda = 16$ is $\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, So the

rank-1 approximation is

$$4 * \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \frac{4}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Question3

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two matrices of dimensions $m \times n$. Consider the optimization problem

$$\min_{\boldsymbol{Q} \in \mathbb{R}^{m \times n}} \|\boldsymbol{X} - \boldsymbol{Q}\boldsymbol{Y}\|_F \text{ subject to } \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_m.$$

, where $\| \cdot \|_F$ denotes the Frobenius norm and I_m. is the $m \times n$ identity matrix. This problem aries when we want to align two datasets $\boldsymbol{X}$ and $\boldsymbol{Y}$ up to an orthogonal linear transformation. Prove that the solution is given by $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^T$, where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the singular value decomposition of $\boldsymbol{X}\boldsymbol{Y}^T$, i.e. $\boldsymbol{X}\boldsymbol{Y}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

Answer

$$\begin{aligned} \|\boldsymbol{X} - \boldsymbol{Q}\boldsymbol{Y}\|_F^2 &= \text{Tr}((\boldsymbol{X} - \boldsymbol{Q}\boldsymbol{Y})^T(\boldsymbol{X} - \boldsymbol{Q}\boldsymbol{Y})) \\ &= \text{Tr}(\boldsymbol{X}^T\boldsymbol{X}) - 2\text{Tr}(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}) + \text{Tr}(\boldsymbol{Y}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{Y}) \\ &= \text{Tr}(\boldsymbol{X}^T\boldsymbol{X}) - 2\text{Tr}(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}) + \text{Tr}(\boldsymbol{Y}^T\boldsymbol{Y}) \end{aligned}$$

So the minimization problem is equivalent to

$$\max_{\boldsymbol{Q} \in \mathbb{R}^{m \times m}} \text{Tr}(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{Y})$$

$$\text{trace}(\boldsymbol{Q}\boldsymbol{Y}\boldsymbol{X}^T) = \text{trace}(\boldsymbol{Q}(\boldsymbol{Y}\boldsymbol{X}^T))$$

Consider the SVD of $\boldsymbol{X}\boldsymbol{Y}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$. Let $\boldsymbol{M} = \boldsymbol{Y}\boldsymbol{X}^T$, and consider the SVD of $\boldsymbol{M} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T$.

$$\text{trace}(\boldsymbol{Q}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T) = \text{trace}(\boldsymbol{U}^T\boldsymbol{Q}\boldsymbol{V}\boldsymbol{\Sigma})$$

Let $\boldsymbol{W} = \boldsymbol{U}^T\boldsymbol{Q}\boldsymbol{V}$, and apparently $\boldsymbol{W}$ is an orthogonal matrix. We know the maximum trace is achieved when $\boldsymbol{W}$ is identity matrix, since for the diagonal entries $\sigma_{ii} \in \boldsymbol{\Sigma}, \sigma >= 0$, and for the element $w_{ij} \in \boldsymbol{W}, w_{ij} <= 1$.

$$\max \text{trace}(\boldsymbol{W}\boldsymbol{\Sigma}) = \text{trace}(\boldsymbol{I}\boldsymbol{\Sigma})$$
$$\boldsymbol{W} = \boldsymbol{U}^T\boldsymbol{Q}\boldsymbol{V} = \boldsymbol{I}$$
$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{I}\boldsymbol{V}^T = \boldsymbol{U}\boldsymbol{V}^T$$

In conclusion, the solution is given by $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^T$.

Question4

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a matrix with $m \geq n$ and full column rank, i.e., $\text{rank}(\boldsymbol{A}) = n$. A *Given rotation* matrix $\boldsymbol{G}_{i,j,\theta} \in \mathbb{R}^{n \times n}$ (for indices $1 \leq i \leq j \leq n$) is defined by

$$
\boldsymbol{G}_{i,j,\theta} = \begin{bmatrix} \boldsymbol{I}_{i-1} & & & & \\ & \cos(\theta) & & -\sin(\theta) & \\ & & \boldsymbol{I}_{j-i-1} & & \\ & \sin(\theta) & & \cos(\theta) & \\ & & & & \boldsymbol{I}_{n-j} \end{bmatrix}
$$

where $\boldsymbol{I}_k$ denotes the $k \times k$ identity matrix, and all unspecified entries are zero. Equivalently, $\boldsymbol{G}_{i,j,\theta}$ is identical to the identity matrix except for the four entries in its ith and jth rows and columns:

$$
g_{ii} = \cos\theta, g_{ij} = -\sin\theta, g_{ji} = \sin\theta, g_{jj} = \cos\theta.
$$

Answer the following questions:

    (a) Prove that $\boldsymbol{G}_{i,j,\theta}$ is an orthogonal matrix, i.e. show that

$$
G_{i,j,\theta}^T G_{i,j,\theta} = \boldsymbol{I}_n.
$$

    (b) Let $\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}$. Prove that for any indices i, j with $1 \leq i < j \leq n$, there exists an angle θ such that

$$
\boldsymbol{G}_{i,j,\theta}\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ \sqrt{v_i^2 + v_j^2} \\ v_{i+1} \\ \vdots \\ v_{j-1} \\ 0 \\ v_{j+1} \\ \vdots \\ v_n \end{bmatrix}
$$

In other words, show that the Given rotation can eliminate the jth entry of $\boldsymbol{v}$ by affecting only the ith and jth components.

    (c) An $m \times n$ $\boldsymbol{A}$ is called an *upper Hessenberg matrix* if

$$
a_{ij} = 0 \text{ whenever } i - j > 2.
$$

For $m \geq n$, the pattern of nonzero entries in $\boldsymbol{A}$ is depicted as

$$
\begin{bmatrix}
\times & \times & \cdots & \times \\
\times & \times & \cdots & \times \\
0 & \times & \ddots & \vdots \\
\vdots & \ddots & \ddots & \times \\
\vdots & \ddots & \ddots & \times \\
\vdots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0
\end{bmatrix}
$$

    Since $\boldsymbol{A}$ already contains many zeros in its lower triangular portion, the standard Householder QR algorithm can be optimzed. Develop an algorithm using Givens rotations to compute the QR decomposition of an upper Hessenberg matrix $\boldsymbol{A}$. Your algorithm should achieve a computational cost of $O(n^2)$ for the reduced (economy-size) QR factorization and $O(mn)$ for the full QR factorization.

(a)

$$\boldsymbol{G}_{i,j,\theta}^{T}\boldsymbol{G}_{i,j,\theta} = \begin{bmatrix} \boldsymbol{I}_{i-1} & & & & \\ & \cos(\theta) & & \sin(\theta) & \\ & & \boldsymbol{I}_{j-i-1} & & \\ & -\sin(\theta) & & \cos(\theta) & \\ & & & & \boldsymbol{I}_{n-j} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{i-1} & & & & \\ & \cos(\theta) & & -\sin(\theta) & \\ & & \boldsymbol{I}_{j-i-1} & & \\ & \sin(\theta) & & \cos(\theta) & \\ & & & & \boldsymbol{I}_{n-j} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_{i-1} & & & & \\ & 1 & & 0 & \\ & & \boldsymbol{i}_{j-i-1} & & \\ & 0 & & 1 & \\ & & & & \boldsymbol{I}_{n-j} \end{bmatrix} = \boldsymbol{I}_{n}$$

(b)

We know,

$$(\boldsymbol{G}_{i,j,\theta}\boldsymbol{v})_{i} = \cos(\theta)v_{i} - \sin(\theta)v_{j} \quad (\boldsymbol{G}_{i,j,\theta}\boldsymbol{v})_{j} = \sin(\theta)v_{i} + \cos(\theta)v_{j}$$

From

$$(\boldsymbol{G}_{i,j,\theta}\boldsymbol{v})_{i} = \sqrt{v_{i}^{2} + v_{j}^{2}} \quad (\boldsymbol{G}_{i,j,\theta}\boldsymbol{v})_{j} = 0$$

We can derive that

$$\cos(\theta) = \frac{v_{i}}{\sqrt{v_{i}^{2} + v_{j}^{2}}}$$

$$\sin(\theta) = -\frac{v_{j}}{\sqrt{v_{i}^{2} + v_{j}^{2}}}$$

For any vector $\boldsymbol{v}$, apply $\boldsymbol{G}_{i,j,\theta}$ with θ such that:

$$\cos(\theta) = \frac{v_{i}}{\sqrt{v_{i}^{2} + v_{j}^{2}}}$$

$$\sin(\theta) = -\frac{v_{j}}{\sqrt{v_{i}^{2} + v_{j}^{2}}}$$

This yields:

$$\boldsymbol{G}_{i,j,\theta}\boldsymbol{v} = \begin{bmatrix} v_{1} \\ \vdots \\ v_{i-1} \\ \sqrt{v_{i}^{2} + v_{j}^{2}} \\ v_{i+1} \\ \vdots \\ v_{j-1} \\ 0 \\ v_{j+1} \\ \vdots \\ v_{n} \end{bmatrix}$$

(c)

We know that for a vector $\boldsymbol{v}_k$, which satisfies $v_i = 0$, if $i > k$ like:

$$\boldsymbol{v}_k = \begin{bmatrix} v_0 \\ \vdots \\ v_{k-1} \\ v_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We can design a matrix $\boldsymbol{G}_{k-1,k,\theta}$,(where θ is calculated by v_{k-1} and v_k) s.t.

$$\boldsymbol{G}_{k-1,k,\theta}\boldsymbol{v}_k = \begin{bmatrix} v_0 \\ \vdots \\ \sqrt{v_{k-1}^2 + v_k^2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{v}'_{k-1}$$

And $\|\boldsymbol{v}_k\| = \|\boldsymbol{v}'_{k-1}\|$

For vector($\boldsymbol{w}$) that follow the pattern of $\boldsymbol{v}_{k-1}$, we know that:

$$\boldsymbol{G}_{k-1,k,\theta}\boldsymbol{w} = \boldsymbol{w}$$

s.t.

$$\boldsymbol{G}_{k-1,k,\theta} \begin{bmatrix} \boldsymbol{w}_0 & \cdots & \boldsymbol{w}_{n-1} & \boldsymbol{v}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_0 & \cdots & \boldsymbol{w}_{n-1} & \boldsymbol{v}'_{k-1} \end{bmatrix}$$

We know $\boldsymbol{G}_{i,j,\theta}$ is orthogonal. According to this, we can do the following steps to get the QR decomposition of an upper Hessenberg matrix $\boldsymbol{A}$:

$$\boldsymbol{G}_{1,2,\theta_1}\boldsymbol{G}_{2,3,\theta_2}\cdots\boldsymbol{G}_{n-1,n,\theta_{n-1}}\boldsymbol{A} = \boldsymbol{R}$$
$$\boldsymbol{A} = \boldsymbol{G}_{n-1,n,\theta_{n-1}}^T \cdots \boldsymbol{G}_{2,3,\theta_2}^T \boldsymbol{G}_{1,2,\theta_1}^T \boldsymbol{R} = \boldsymbol{QR}$$

Also, it's easy to see:

$$G_{k-1,k,\theta}\boldsymbol{e}_i = \boldsymbol{e}_i, \forall i \neq k$$

And for the multiplication of 2 matrices $\boldsymbol{G}_{k-1,k,\theta}\boldsymbol{G}_{k,k+1,\theta}$, we have:

$$\boldsymbol{G}_{k-1,k,\theta}\boldsymbol{G}_{k,k+1,\theta} = \boldsymbol{G}_{k-1,k,\theta} \begin{bmatrix} \boldsymbol{e}_0 & \cdots & \boldsymbol{e}_{k-1} & \boldsymbol{v}_k & \boldsymbol{v}_{k+1} & \boldsymbol{e}_{k+2} & \cdots & \boldsymbol{e}_n \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{e}_0 & \cdots & \boldsymbol{e}_{k-1} & \boldsymbol{v}'_k & \boldsymbol{v}'_{k+1} & \boldsymbol{e}_{k+2} & \cdots & \boldsymbol{e}_n \end{bmatrix}$$

So we know, for any multiplication of any two rotation matrix, we cost $O(n)$ time to calculate the product.

Because of the time cost of multiplication like $\boldsymbol{G}_{i,j,\theta}\boldsymbol{R}$ and $\boldsymbol{G}_{i,j,\theta}\boldsymbol{G}_{m,n,\theta}$ is $O(n)$, and we need to do this for n times, so the total time cost is $O(n^2)$. And the time cost of putting zero if we want full QR factorization is $O(mn)$.

Algorithm 1 Optimized QR Decomposition of an Upper Hessenberg Matrix using Givens Rotations

---

Require: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (upper Hessenberg matrix)
Ensure: $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$ (orthogonal), $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ (upper triangular)
1: Initialize $\boldsymbol{R} \leftarrow \boldsymbol{A}$
2: Initialize $\boldsymbol{Q}^T \leftarrow \boldsymbol{I}_m$
3: **for** $j = n$ to 1 **do**
4:     **for** $i = j + 1$ to m **do**
5:         **if** $r_{i,j} \neq 0$ **then**
6:             Compute Givens rotation $\boldsymbol{G}_{i,j,\theta}$ to zero out $r_{i,j}$
7:             Apply $\boldsymbol{G}_{i,j,\theta}$ from the left: $\boldsymbol{R} \leftarrow \boldsymbol{G}_{i,j,\theta} \boldsymbol{R}$
8:             Accumulate $\boldsymbol{Q}^T \leftarrow \boldsymbol{Q}^T \boldsymbol{G}_{i,j,\theta}$
9:         **end if**
10:     **end for**
11: **end for**
12: Compute $\boldsymbol{Q} \leftarrow (\boldsymbol{Q}^T)^T$
13: **return** $\boldsymbol{Q}, \boldsymbol{R}$