553.385 Numerical Linear Algebra, Spring 2022
Section 12

Hongyu Cheng

hongyucheng@jhu.edu

April 16, 2022

Definition 1 (Givens Rotation Matrix). *For $1 \leq i < j \leq n$, real matrix*

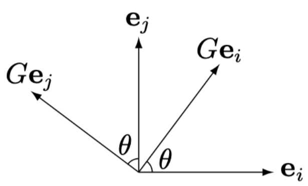
$$G_{ij}(\theta) = \begin{pmatrix} I_{i-1} & & & & \\ & \cos\theta & & -\sin\theta & \\ & & I_{j-i-1} & & \\ & \sin\theta & & \cos\theta & \\ & & & & I_{n-j} \end{pmatrix}$$
$$= I - (1-\cos\theta)(E_{ii} + E_{jj}) - \sin\theta\,(E_{ij} - E_{ji})$$

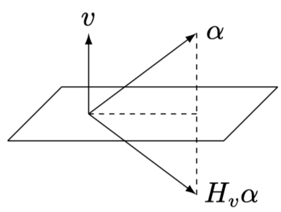
is called the Givens rotation matrix.

Definition 2 (Householder Matrix). $v \in \mathbb{C}^{n+1} \backslash \{0\}$*, the symmetric matrix*

$$H_v = I - \frac{2}{v^* v} v v^*$$

is called Householder matrix. In particular, for $w \in \mathbb{C}^n$ with $\|w\|_2 = 1$, $H_w = I - 2ww^*$.





Givens matrix $G_{ij}(\theta)$ can be viewed as a matrix representation of the rotation transformation in the (e_i, e_j) plane. Householder matrix H_v can be viewed as a matrix representation of the reflection transformation on the normal vector v. It's not hard to show that $G_{ij}(\theta)$ and H_v are both orthogonal (or unitary) matrices, and $\det(H_v) = -1$.

Theorem 1 (Sherman-Morrison). *Suppose $A \in \mathbb{R}^{n\times n}$ invertible, $u, v \in \mathbb{R}^n$ are column vectors. Then $A + uv^T$ is invertible iff $1 + v^T A^{-1} u \neq 0$. In this case,*

$$\left(A + uv^\top\right)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1} u}. \tag{1}$$

Proof sketch.

$$
\begin{aligned}
&\left(A+uv^\top\right)\left(A^{-1}-\frac{A^{-1}uv^\top A^{-1}}{1+v^\top A^{-1}u}\right)\\
=&AA^{-1}+uv^\top A^{-1}-\frac{AA^{-1}uv^\top A^{-1}+uv^\top A^{-1}uv^\top A^{-1}}{1+v^\top A^{-1}u}\\
=&I+uv^\top A^{-1}-\frac{uv^\top A^{-1}+uv^\top A^{-1}uv^\top A^{-1}}{1+v^\top A^{-1}u}\\
=&I+uv^\top A^{-1}-\frac{u\left(1+v^\top A^{-1}u\right)v^\top A^{-1}}{1+v^\top A^{-1}u}\\
=&I+uv^\top A^{-1}-uv^\top A^{-1}
\end{aligned}
$$

Another direction is similar. □

A general form of Sherman-Morrison is called Woodbury. In our Householder case, when $A=I$, we have,

$$\left(I+uv^\top\right)^{-1}=I-\frac{uv^\top}{1+v^\top u} \tag{2}$$

therefore,

$$H_w^{-1}=(I+(-2w)w^T)^{-1}\overset{(2)}{=}I-\frac{(-2w)w^T}{1+(-2w)^Tw}=I-2ww^T=H_w$$

Theorem 2. *For $A\in\mathbb{R}^{n\times n}$, there exists some Givens matrices $G_1,G_2,\dots,G_k$ such that $R=G_k\cdots G_2G_1A$ is upper triangular. In particular, when A is orthogonal matrix, $R=diag(1,\dots,1,\det(A))$.*

Proof sketch. $\forall a,b\in\mathbb{R},\ \exists\theta$ such that $\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}=\begin{pmatrix}\sqrt{a^2+b^2}\\ 0\end{pmatrix}$. □

Theorem 3. *For $A\in\mathbb{R}^{n\times n}$, there exists some Householder matrices $H_1,H_2,\dots,H_k$ such that $R=G_H\cdots H_2H_1A$ is upper triangular with nonnegative diagonal. In particular, when A is orthogonal matrix, $R=I$.*

Proof sketch. $\forall\alpha\in\mathbb{R}^n$, if $\beta=(\|\alpha\|_2,0,...,0)^T\neq\alpha$, then $v=\alpha-\beta$ satisfies $H_v\alpha=\beta$ (Note that here $\alpha-\beta$ may not be unit vector). □

Definition 3 (Gram-Schmidt orthogonalization)**.** *Given $A=[\alpha_1,\dots,\alpha_m]\in\mathbb{C}^{n\times m}$, $n\geq m$ with full column rank, define $r_{11}=\|\alpha_1\|_2$, $\mathbf{q}_1=\alpha_1/r_{11}$. Then, for $j=2,\dots,m$ set*

$$r_{ij}=\langle\alpha_j,\mathbf{q}_i\rangle,\ i=1,\dots,j-1,\quad \hat{\mathbf{q}}_\mathbf{j}=\alpha_j-\sum_{i=1}^{j-1}r_{ij}\mathbf{q}_i$$
$$r_{jj}=\|\hat{\mathbf{q}}_j\|_2,\quad \mathbf{q}_j=\hat{\mathbf{q}}_j/r_{jj}$$

Note that Gram-Schmidt orthogonalization gives orthonormal $[\mathbf{q_1},\dots,\mathbf{q}_m]$ such that

$$\alpha_j=\sum_{i=1}^{j}r_{ij}\mathbf{q}_i \quad\text{or}\quad A=QR$$

In particular, when $A \in \mathbb{R}^{n\times n}$ is invertible,

$$\begin{pmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_m\end{pmatrix} = \begin{pmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_m\end{pmatrix}\begin{pmatrix}\|\hat{\mathbf{q}}_1\|_2 & \mathbf{q}_1^T\alpha_2 & \cdots & \mathbf{q}_1^T\alpha_m \\ 0 & \|\hat{\mathbf{q}}_2\|_2 & \ddots & \vdots \\ 0 & 0 & \ddots & \mathbf{q}_{n-1}^T\alpha_m \\ 0 & 0 & 0 & \|\hat{\mathbf{q}}_m\|_2\end{pmatrix}.$$

Example 1 (QR decomposition). *Calculate the QR decomposition of* $A = \begin{pmatrix}2 & 1 & 1\\ 2 & 3 & 2\\ 1 & 1 & 3\end{pmatrix}$.

Solution 1.[Givens]

$$\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}2 & 1 & 1\\ 2 & 3 & 2\\ 1 & 1 & 3\end{pmatrix} = \begin{pmatrix}2\sqrt{2} & 2\sqrt{2} & \frac{3}{\sqrt{2}}\\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}}\\ 1 & 1 & 3\end{pmatrix}$$

$$\begin{pmatrix}\frac{2\sqrt{2}}{3} & 0 & \frac{1}{3}\\ 0 & 1 & 0\\ -\frac{1}{3} & 0 & \frac{2\sqrt{2}}{3}\end{pmatrix}\begin{pmatrix}2\sqrt{2} & 2\sqrt{2} & \frac{3}{\sqrt{2}}\\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}}\\ 1 & 1 & 3\end{pmatrix} = \begin{pmatrix}3 & 3 & 3\\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}}\\ 0 & 0 & \frac{3}{\sqrt{2}}\end{pmatrix} = R.$$

Thus, $A = QR$, where

$$Q = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\frac{2\sqrt{2}}{3} & 0 & -\frac{1}{3}\\ 0 & 1 & 0\\ \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3}\end{pmatrix} = \begin{pmatrix}\frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}}\\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}}\\ \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3}\end{pmatrix}$$

Solution 2.[Householder]

$$u = \begin{pmatrix}-1\\ 2\\ 1\end{pmatrix}, \quad H_u = \begin{pmatrix}\frac{2}{3} & \frac{2}{3} & \frac{1}{3}\\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3}\\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3}\end{pmatrix}, \quad H_u\begin{pmatrix}2 & 1 & 1\\ 2 & 3 & 2\\ 1 & 1 & 3\end{pmatrix} = \begin{pmatrix}3 & 3 & 3\\ 0 & -1 & -2\\ 0 & -1 & 1\end{pmatrix}$$

$$v = \begin{pmatrix}-1-\sqrt{2}\\ -1\end{pmatrix}, \quad H_v = \begin{pmatrix}-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}, \quad H_v\begin{pmatrix}-1 & -2\\ -1 & 1\end{pmatrix} = \begin{pmatrix}\sqrt{2} & \frac{1}{\sqrt{2}}\\ 0 & \frac{3}{\sqrt{2}}\end{pmatrix}$$

Thus, $A = QR$, where

$$Q = H_u\begin{pmatrix}1 & 0\\ 0 & H_v\end{pmatrix} = \begin{pmatrix}\frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}}\\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}}\\ \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3}\end{pmatrix}, \quad R = \begin{pmatrix}3 & 3 & 3\\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}}\\ 0 & 0 & \frac{3}{\sqrt{2}}\end{pmatrix}$$

Solution 3.[Gram-Schmidt]

$$\begin{aligned}
&\alpha_1 = (2,2,1), \quad \beta_1 = \alpha_1, && \gamma_1 = \tfrac{1}{3}\beta_1 = \left(\tfrac{2}{3}, \tfrac{2}{3}, \tfrac{1}{3}\right)\\
&\alpha_2 = (1,3,1), \quad \beta_2 = \alpha_2 - 3\gamma_1 = (-1,1,0), && \gamma_2 = \tfrac{1}{\sqrt{2}}\beta_2 = \left(-\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0\right)\\
&\alpha_3 = (1,2,3), \quad \beta_3 = \alpha_3 - 3\gamma_1 - \tfrac{1}{\sqrt{2}}\gamma_2 = \left(-\tfrac{1}{2}, -\tfrac{1}{2}, 2\right), && \gamma_3 = \tfrac{\sqrt{2}}{3}\beta_3 = \left(-\tfrac{1}{3\sqrt{2}}, -\tfrac{1}{3\sqrt{2}}, \tfrac{2\sqrt{2}}{3}\right).
\end{aligned}$$

Thus, $A = QR$, where

$$Q = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3} \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{3}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 3 \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{3}{\sqrt{2}} \end{pmatrix}.$$

Theorem 4 (Hadamard inequality)**.** *Given $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}^n \backslash \{0\}$, then*

$$|\det(\alpha_1, \alpha_2, \ldots, \alpha_n)| \leq \prod_{k=1}^{n} \|\alpha_k\|_2 .$$

The equality is achieved if and only if the vectors are orthogonal.

Proof. Let $A = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$. When $\det(A) = 0$, the equality holds. When $\det(A) \neq 0$, $A = QR$, where Q is unitary matrix, $R = (r_{ij}) = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_n \end{pmatrix}$ is upper triangular matrix, then

$$|\det(A)| = \prod_{k=1}^{n} |r_{kk}| \leqslant \prod_{k=1}^{n} \|\beta_k\|_2 = \prod_{k=1}^{n} \|Q\beta_k\|_2 = \prod_{k=1}^{n} \|\alpha_k\|_2$$

the equality is achieved $\Longleftrightarrow$ R is diagonal matrix, i.e. $\beta_1, \beta_2, \ldots, \beta_n$ are orthogonal $\Longleftrightarrow$ $\alpha_1, \alpha_2, \ldots, \alpha_n$ are orthogonal. □

Example 2 (Least Squares)**.** *Given dataset $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)\} \subseteq \mathbb{R}^{n+1}$, find $\theta = (\theta_1, \theta_2, \ldots, \theta_n)^T \in \mathbb{R}^n$ such that minimizes $f(\theta) = \sum_{i=1}^{m} \left\|\mathbf{x}_i^T \theta_i - y_i\right\|_2^2$.*

Solution. In section week 7 we have seen that we can consider this from a smooth convex optimization perspective, let $X = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \end{pmatrix}^T$, $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \end{pmatrix}$,

$$\nabla^2 f(\theta) = \nabla \left(\nabla \|X\theta - \mathbf{y}\|_2^2\right) = \nabla \left(2X^T(X\theta - \mathbf{y})\right) = 2A^T A \succeq 0$$

then any solution of $\nabla f(\theta) = 0$ attains the minimum, i.e.,

$$\nabla f(\theta^*) = 0 \Longleftrightarrow 2X^T(X\theta^* - \mathbf{y}) = 0 \Longleftrightarrow X^T X\theta^* = X^T \mathbf{y}.$$

Now we can consider this problem in another way. Consider the space generated by column vectors of A, let $U = \mathrm{Span}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, by homework 7 we have $\mathbb{R}^n = U \oplus U^\perp$, then $\mathbf{y} = \beta + \gamma$, where $\beta \in U$, $\gamma \in U^\perp$. Therefore, $f(\theta) = \|\beta - \sum_{i=1}^{m} \theta_i \alpha_i\|_2^2 + \|\gamma\|_2^2$. The minimum is attained if and only if $\beta = \sum_{i=1}^{n} \theta_i \alpha_i$, which is equivalent to $\mathbf{y} - \sum_{i=1}^{n} \theta_i \alpha_i \in U^\perp$, i.e., $\langle \alpha_i, \mathbf{y} - \sum_{i=1}^{n} \theta_i \alpha_i \rangle = 0$, $\forall i$, which in matrix form is $X^T X\theta = X^T \mathbf{y}$. The second method can be generalized to general inner product space with the linear system

$$\begin{pmatrix} \langle \alpha_1, \alpha_1 \rangle & \langle \alpha_1, \alpha_2 \rangle & \cdots & \langle \alpha_1, \alpha_n \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \langle \alpha_2, \alpha_2 \rangle & \cdots & \langle \alpha_2, \alpha_n \rangle \\ \vdots & \vdots & \cdots & \vdots \\ \langle \alpha_n, \alpha_1 \rangle & \langle \alpha_n, \alpha_2 \rangle & \cdots & \langle \alpha_n, \alpha_n \rangle \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix} = \begin{pmatrix} \langle \alpha_1, \mathbf{y} \rangle \\ \langle \alpha_2, \mathbf{y} \rangle \\ \vdots \\ \langle \alpha_n, \mathbf{y} \rangle \end{pmatrix}.$$