5.7 Alternatives to the QR factorization:

(a) Can a matrix $A \in \mathbb{R}^{m\times n}$ be factored into $A = RQ$ where R is upper triangular and Q is orthogonal? How?

(b) Can a matrix $A \in \mathbb{R}^{m\times n}$ be factored into $A = QL$ where L is lower triangular?

5.7(a) First consider the case where $m \leq n$. We show that the factorization $A = RQ$ always exists in this case.

Consider the permutation matrix $P_m \in \mathbb{R}^{m\times m}$ where $P_m A$ swaps the rows of A such that $r_1 \leftrightarrow r_m$, $r_2 \leftrightarrow r_{m-1}$ and so on. Note that

$$P_m \approx \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix} \quad \text{where } P_m^{-1} = P_m^T = P_m$$

is a weird sort of diagonal matrix with the diagonal going the wrong direction. Also note that AP_n swaps the columns of A so that $c_1 \leftrightarrow c_n$, $c_2 \leftrightarrow c_{n-1}$ and so on.

Suppose $L \in \mathbb{R}^{m\times n}$ is lower triangular. Then $L_{ij} = 0$ for $i < j$.

Since

$$(P_m L)_{ij} = L_{m-i,j} \quad \text{and} \quad (P_m L P_n)_{ij} = L_{m-i,n-j}$$

we obtain that $(P_m L P_n)_{ij} = 0$ for $m-i < n-j$ or when $i > j + m - n$

Thus $P_m L P_n$ is upper triangular provided $m \leq n$

Let $A \in \mathbb{R}^{m\times n}$. Then $(P_m A P_n)^T = QR$ where $Q \in \mathbb{R}^{n\times n}$ is orthogonal and $R \in \mathbb{R}^{n\times m}$ is upper triangular. It follows that

$$P_m A P_n = (QR)^T = R^T Q^T$$

and since $m \le n$ then

$$A = P_m R^T Q^T P_n = P_m R^T P_n P_n Q^T P_n = \tilde{R}\tilde{Q}$$

where $\tilde{R} = P_m R^T P_n$ is upper triangular and $\tilde{Q} = P_n Q^T P_n$ is orthogonal.

Next consider the case $n < m$. If $A = RQ$, then Q being orthogonal implies $Q \in \mathbb{R}^{n \times n}$ and so $R \in \mathbb{R}^{m \times n}$. Since R is upper triangular, it follows that R has the block structure

$$R = \begin{bmatrix} \tilde{R} \\ \hline 0_{n-m,n} \end{bmatrix}$$

where $\tilde{R} \in \mathbb{R}^{n \times n}$ and $0_{n-m,n} \in \mathbb{R}^{n-m,n}$ is the zero matrix. Thus the last $n-m$ rows of R are zero. This implies the last $n-m$ rows of A would have to be zero as well, so A has block structure

$$A = \begin{bmatrix} \tilde{A} \\ \hline 0_{m-n,n} \end{bmatrix}$$

where $\tilde{A} \in \mathbb{R}^{n \times n}$. Now, since $\tilde{A}$ is square the previous case applies to obtain $\tilde{A} = \tilde{R}\tilde{Q}$. This can be extended to a factorization of A as

$$A = \begin{bmatrix} \tilde{A} \\ \hline 0_{m-n,n} \end{bmatrix} = \begin{bmatrix} \tilde{R}\tilde{Q} \\ \hline 0_{m-n,n} \end{bmatrix} = \begin{bmatrix} \tilde{R} \\ \hline 0_{m-n,n} \end{bmatrix} \left[\begin{array}{c|c} \tilde{Q} & 0_{n,m-n} \\ \hline 0_{m-n,n} & I \end{array}\right]$$

where I is the identity matrix.

Thus, $A = RQ$ only holds when $m \le n$ or if $n < m$ and the last $m-n$ rows of A are exactly zero.

5.7(b) Given $A \in \mathbb{R}^{m\times n}$ suppose $A^T \in \mathbb{R}^{n\times m}$ satisfies the conditions stated in part (a). Thus, either $n \leq m$ or the last $m-n$ columns of A or all zero. In this case $A^T = RQ$ where $R \in \mathbb{R}^{n\times m}$ is an upper triangular matrix and $Q \in \mathbb{R}^{m\times m}$ is orthogonal.

It follows that

$$A = (RQ)^T = Q^T R^T = Q^T L$$

where $L = R^T$ is lower triangular and Q^T is orthogonal.

5.8 Relating QR and Cholesky factorizations:

(a) Take $A \in \mathbb{R}^{m\times n}$ and suppose we apply the Cholesky factorization to obtain $A^\top A = LL^\top$. Define $Q \equiv A(L^\top)^{-1}$. Show that the columns of Q are orthogonal. actually orthonormal

(b) Based on the previous part, suggest a relationship between the Cholesky factorization of $A^\top A$ and QR factorization of A.

5.8(a) To show the columns of Q are orthonormal it is enough to show $Q^TQ = I$. By definition

$$Q^TQ = \left(A(L^T)^{-1}\right)^T A(L^T)^{-1} = \left(\left(L^T\right)^{-1}\right)^T A^T A (L^T)^{-1}$$

$$= L^{-1} A^T A (L^T)^{-1} = L^{-1} L L^T (L^T)^{-1} = I$$

5.8(b) Since $Q = A(L^T)^{-1}$ then $A = QL^T = QR$ where

$R = L^T$ is upper triangular

and

$Q = A(L^T)^{-1}$ is orthogonal.

This obtains the QR factorization in terms of the Cholesky factorization of A^TA.

Alternatively, if $A = QR$, then defining $L = R^T$ yields that

$$A^TA = (QR)^T QR = R^T Q^T QR = R^T R = LL^T$$

which obtains the Cholesky factorization of A^TA from the QR factorization of A.

5.12 (Adapted from [50], §5.1) If $\vec{x}, \vec{y} \in \mathbb{R}^m$ with $\|\vec{x}\|_2 = \|\vec{y}\|_2$, write an algorithm for finding an orthogonal matrix Q such that $Q\vec{x} = \vec{y}$.

First consider the case when x and y are linearly dependent. Then $x = \alpha y$ for some $\alpha \in \mathbb{R}$ and $\|x\| = |\alpha| \|y\|$ implies that $\alpha = \pm 1$. It follows that either $x = y$ or $x = -y$ and one can take $Q = I$ or $Q = -I$ respectively.

If x and y are independent, we look for a Householder transform that maps x into y. Thus, find a unit vector v such that

$$Hx = (I - 2vv^T)x = y$$

Solving for v yields

$$2v(v \cdot x) = x - y$$

and since $x \neq y$ then $v \cdot x \neq 0$ and $v = \frac{x-y}{2x \cdot v}$.

Since v is a unit vector we immediately obtain that

$$v = \frac{x-y}{\|x-y\|}$$

So that $H = I - 2\frac{(x-y)(x-y)^T}{\|x-y\|^2}$.

6.3 Show that the eigenvalues of upper-triangular matrices $U \in \mathbb{R}^{n \times n}$ are exactly their diagonal elements.

6.3 If U is upper triangular than $U - \lambda I$ is upper triangular. It follows that $\det(U - \lambda I)$ is the product along the diagonal. Thus

$$\det(U - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda)$$

This implies $\det(U - \lambda I) = 0$ only when λ is one of the diagonal elements of U. Consequently, since the only times that

$$(U - \lambda I)x = 0$$

has a solution $x \neq 0$ is when $U - \lambda I$ is not invertible, it follows that the eigenvalues of U are exactly the diagonal elements.

6.6 (Suggested by J. Yeo) Suppose $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^n$ such that $\vec{u}^\top \vec{v} = 1$, and define $A \equiv \vec{u}\vec{v}^\top$.

(a) What are the eigenvalues of A?

(b) How many iterations does power iteration take to converge to the dominant eigenvalue of A?

6.6(a) First note that rank $A = 1$, therefore $\dim \mathrm{Nul}(A) = n-1$. This means 0 is an eigenvalue of A provided $n > 1$. Moreover, there can be at most one more eigenvalue since there are already $n-1$ linearly independent eigenvectors in $\mathrm{Nul}(A)$ corresponding to the zero eigenvalue.

Since

$$Au = uv^Tu = u(v \cdot u) = u(u \cdot v) = u(u^Tv) = u$$

it follows that u is another eigenvector with eigenvalue 1.

Thus, the eigenvalues are 0 and 1 provided $n > 1$ and only 1 when $n = 1$.

6.6(b) The dominant eigenvalue is 1. Let $\{\xi_1, \xi_2, \dots, \xi_n\}$ be an eigenbasis for A with corresponding eigenvalues $\lambda_1 = 1$ and $\lambda_2 = \dots = \lambda_n = 0$.
For $x \in \mathbb{R}^n$ chosen randomly let

$$x = c_1\xi_1 + c_2\xi_2 + \dots + c_n\xi_n$$

and assume $c_1 \neq 0$, which happens with probability one provided the vector x was chosen randomly.

By the power method

$$y = Ax = c_1\lambda_1\xi_1 + c_2\lambda_2\xi_2 + \dots + c_n\lambda_n\xi_n = c_1\xi_1$$

since $\lambda_1 = 1$ and $\lambda_2 = \dots = \lambda_n = 0$.

Therefore it takes one iteration for the power method to converge.

6.11 ("Epidemiology") Suppose $\vec{x}_0 \in \mathbb{R}^n$ contains sizes of different populations carrying a particular infection in year 0; for example, when tracking malaria we might take x_{01} to be the number of humans with malaria and x_{02} to be the number of mosquitoes carrying the disease. By writing relationships like "The average mosquito infects two humans," we can write a matrix M such that $\vec{x}_1 \equiv M\vec{x}_0$ predicts populations in year 1, $\vec{x}_2 \equiv M^2\vec{x}_0$ predicts populations in year 2, and so on.

(a) The spectral radius $\rho(M)$ is given by $\max_i |\lambda_i|$, where the eigenvalues of M are $\lambda_1, \ldots, \lambda_k$. Epidemiologists call this number the "reproduction number" $\mathcal{R}_0$ of M. Explain the difference between the cases $\mathcal{R}_0 < 1$ and $\mathcal{R}_0 > 1$ in terms of the spread of disease. Which case is more dangerous?

(b) Suppose we only care about proportions. For instance, we might use $M \in \mathbb{R}^{50\times 50}$ to model transmission of diseases between residents in each of the 50 states of the USA, and we only care about the fraction of the total people with a disease who live in each state. If $\vec{y}_0$ holds these proportions in year 0, give an iterative scheme to predict proportions in future years. Characterize behavior as time goes to infinity.

6.11(a) If $R_0 < 1$ then all eigenvalues of M satisfy $|\lambda_i| < 1$. In the case where M admits a basis of eigenvectors $\xi_1, \ldots, \xi_n$ it follows that

$$x_0 = c_1\xi_1 + \cdots + c_n\xi_n$$

and so $x_j = M^j x_0 = c_1\lambda_1^j\xi_1 + \cdots + c_n\lambda_n^j\xi_n$ implies

$$\|x_j\| \leq \|c_1\lambda_1^j\xi_1 + \cdots + c_n\lambda_n^j\xi_n\| \leq \left(\max|\lambda_i|^j\right)\left(|c_1|\|\xi_1\| + \cdots + |c_n|\|\xi_n\|\right)$$

$$= R_0^j\left(|c_1|\|\xi_1\| + \cdots + |c_n|\|\xi_n\|\right) \to 0 \quad \text{as} \quad j \to \infty$$

Therefore the number of infected people goes to zero over time and the epidemic dies out. The case when $k < n$ and there is not a basis of eigenvectors is more complicated but similar.

On the other hand if $R_0 > 1$ there exists an eigenvalue λ_i with $|\lambda_i| > 1$ with corresponding eigenvector ξ_i. In this if $x_0 = \xi_i$ then

$$x_j = M^j x_0 = c_i\lambda_i^j\xi_i$$

and so

$$\|x_j\| = |c_i||\lambda_i|^j\|\xi_i\| \to \infty \quad \text{as} \quad j \to \infty.$$

Therefore the number of infected people grows exponentially in time.

6.1.1b) If y_0 hold proportions of people who are infected at time $t=0$, then the entries of y_0 should be non-negative and sum to 1. Thus means that $y_0 = \frac{x_0}{\|x_0\|_1}$ where the 1-norm of a vector v is given

$$\|v\|_1 = \sum_{i=1}^{50} |v_i|$$

Since the populations at the next time are $x_1 = Mx_0$ it follows that

$$y_1 = \frac{x_1}{\|x_1\|_1} = \frac{Mx_0}{\|x_1\|_1} = \frac{My_0\|x_0\|_1}{\|x_1\|_1} = \frac{My_0}{\|My_0\|_1}$$

since y_1 is a unit vector with respect to the 1-norm. It follows that

$$y_j = \frac{M^j y_0}{\|M^j y_0\|}$$

at all future times.

As time goes to infinity $y_j \to \frac{\xi_i}{\|\xi_i\|_1}$ where ξ_i is the eigenvector corresponding to the largest eigenvalue of M further chosen so all entries of ξ_i are non-negative.