Chapter 7

Factorization Theorems

This chapter highlights a few of the many factorization theorems for matrices. While some factorization results are relatively direct, others are iterative. While some factorization results serve to simplify the solution to linear systems, others are concerned with revealing the matrix eigenvalues. We consider both types of results here.

7.1 The PLU Decomposition

The PLU decomposition (or factorization) To achieve LU factorization we require a modified notion of the row reduced echelon form.

Definition 7.1.1. The *modified row echelon form* of a matrix is that form which satisfies all the conditions of the modified row reduced echelon form except that we do not require zeros to be above leading ones, and moreover we do not require leading ones, just nonzero entries.

For example the matrices below are in row echelon form.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & -7 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Most of the factorizations $A \in M_n(\mathbb{C})$ studied so far require one essential ingredient, namely the eigenvectors of A. While it was not emphasized when we studied Gaussian elimination, there is a LU-type factorization there. Assume for the moment that the only operations needed to carry A to its modified row echelon form are those that add a multiple of one row to another. The *modified row echelon form* of a matrix is that form which satisfies all the conditions of the modified row reduced echelon form except that we do not require zeros to be above leading ones, and moreover we do not require leading ones, just nonzero entries. Naturally it is easy to make the leading nonzero entries into leading ones by the multiplication by an appropriate identity matrix. That is not the point here. What we want to observe is that in this case the reduction is accomplished by the left multiplication of A by a sequence of lower triangular matrices of the form.

$$L = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & 0 & \\ \vdots & 0 & 1 & & \\ & c & & \ddots & \\ 0 & \cdots & & & 1 \end{bmatrix}$$

Since we pivot at the $(1,1)$-entry first, we eliminate all the entries in the first column below the first row. The product of all the matrices L to accomplish this has the form

$$L_1 = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & 0 & \\ c_{31} & 0 & 1 & & \\ \vdots & & & \ddots & \\ c_{n1} & 0 & \cdots & & 1 \end{bmatrix}$$

where $c_{k1} = -\frac{a_{k1}}{a_{11}}$. Thus, with the notation that $A = A_1$ has entries $a_{ij}^{(1)}$ this first phase of the reduction renders the matrix A_2 with entries $a_{ij}^{(2)}$

$$A_2 = L_1 A_1 = \begin{bmatrix} a_{11}^{(2)} & & \cdots & & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & & \vdots \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & & \\ \vdots & & & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & & a_{nn}^{(2)} \end{bmatrix}$$

Since we have assumed that no row interchanges are necessary to carry out the reduction we know that $a_{22}^{(2)} \neq 0$. The next part of the reduction process is the elimination of the elements in the second column below the second row, i.e. $a_{32}^{(2)} \to 0, \ \ldots a_{n2}^{(2)} \to 0$. Correspondingly, this can be achieved by a matrix of the form

$$L_2 = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & 0 & \\ 0 & c_{22} & 1 & & \\ \vdots & \vdots & & \ddots & \\ 0 & c_{n2} & \cdots & & 1 \end{bmatrix}$$

(What are the values c_{k2} ?) The result is the matrix A_3 given by

$$A_3 = L_2 A_2 = L_2 L_1 A_1 = \begin{bmatrix} a_{11}^{(3)} & & \cdots & & a_{1n}^{(3)} \\ 0 & a_{22}^{(3)} & \cdots & & \vdots \\ 0 & 0 & a_{33}^{(3)} & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{3n}^{(3)} & & a_{nn}^{(3)} \end{bmatrix}$$

Proceeding in this way through all the rows (columns) there results

$$A_n = L_{n-1} A_{n-1} = L_{n-1} \cdots L_2 L_1 A_1 = \begin{bmatrix} a_{11}^{(3)} & & \cdots & & a_{1n}^{(3)} \\ 0 & a_{22}^{(3)} & \cdots & & \vdots \\ 0 & 0 & a_{33}^{(3)} & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & a_{nn}^{(3)} \end{bmatrix}$$

The right side of the equation above is an upper triangular matrix. Denote it by U. Since each of the matrices $L_i, \ i = 1, \ldots n-1$ is invertible we can write

$$A = L_1^{-1} \cdots L_{n-1}^{-1} U$$

The lemma below is useful in this.

Lemma 7.1.1. *Suppose the lower triangular matrix $L \in M_n(\mathbb{C})$ has the* *form*

$$L = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & 0 & \\ & & 1 & & & \\ & & 0 & 1 & & \\ \vdots & & \vdots & c_{k+1,k} & \ddots & \\ & & & \vdots & \ddots & \\ 0 & \cdots & 0 & c_{nk} & & 1 \end{bmatrix} \longleftarrow k^{th} \text{ row}$$

Then L is invertible with inverse given by

$$L^{-1} = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & 0 & \\ & & 1 & & & \\ & & 0 & 1 & & \\ \vdots & & \vdots & -c_{k+1,k} & \ddots & \\ & & & \vdots & \ddots & \\ 0 & \cdots & 0 & -c_{nk} & & 1 \end{bmatrix} \longleftarrow k^{th} \text{ row}$$

Proof. Trivial

Lemma 7.1.2. *Suppose $L_1, L_2, \cdots, L_{n-1}$ are the matrices given above. Then the matrix $L = L_1^{-1} \cdots L_{n-1}^{-1}$ has the form*

$$L = \begin{bmatrix} 1 & & & & & & 0 \\ -c_{21} & 1 & & & & & \\ -c_{31} & -c_{32} & 1 & & & & \\ & & & 1 & & & \\ \vdots & \vdots & \vdots & -c_{k+1,k} & \ddots & & \\ & & & \vdots & & \ddots & \\ -c_{n1} & -c_{n2} & \cdots & -c_{nk} & \cdots & & 1 \end{bmatrix}$$

Proof. Trivial.

Applying these lemmas to the present situation we can say that when no row interchanges are needed we can factor and matrix $A \in M_n(\mathbb{C})$ as $A = LU$, where L is lower triangular and U is upper triangular. When row interchanges are needed and we let P be the permutation matrix that creates these row interchanges then the LU-factorization above can be carried out for the matrix PA. Thus $PA = LU$, where L is lower triangular and U is upper triangular. We call this the **PLU** factorization. Let us summarize this in the following theorem.

Theorem 7.1.1. *Let $A \in M_n(\mathbb{C})$. Then there is a permutation matrix $P \in M_n(\mathbb{C})$ and lower L and upper U triangular matrices ($\in M_n(\mathbb{C})$), such that $PA = LU$. Moreover, L can be taken to have ones on its diagonal. That is, $\ell_{ii} = 1, \quad i = 1, \ldots n$.*

By applying the result above to A^T it is easy to see that the matrix U can be taken to have the ones in its diagonal. The result is stated as a corollary.

Corollary 7.1.1. *Let $A \in M_n(\mathbb{C})$. Then there is a permutation matrix $P \in M_n(\mathbb{C})$ and lower and upper triangular matrices ($\in M_n(\mathbb{C})$) respectively, such that $PA = LU$. Moreover, U can be taken to have ones on its diagonal ($u_{ii} = 1, \quad i = 1, \ldots n$).*

The PLU decomposition can be put in service to solving the system $Ax = b$ as follows. Assume that $A \in M_n(\mathbb{C})$ is invertible. Determine the permutation matrix P in order that $PA = LU$, where L is lower triangular and U is upper triangular. Thus, we have

$$\begin{aligned} Ax &= b \\ PAx &= Pb \\ LUx &= Pb \end{aligned}$$

Solve the systems

$$\begin{aligned} Ly &= Pb \\ Ux &= y \end{aligned}$$

Then $LUx = Ly = Pb$ .Hence x is a solution to the system. The advantages of this formulation over the direct Gaussian elimination is that the systems $Ly = Pb$ and $Ux = y$ are triangular and hence are easy to solve. For example for the first of the systems, $Ly = Pb$, let the vector $Pb = \left[\hat{b}_1, \ldots, \hat{b}_n\right]^T$. Then it is easy to see that "back substitution" (aka "forward substitution") can be used to determine y. That is, we have the recursive relations

$$\begin{aligned} y_1 &= \frac{\hat{b}_1}{l_{11}} \\ y_2 &= \frac{\hat{b}_2 - l_{21} y_1}{l_{22}} \\ &\vdots \\ y_n &= \left(\hat{b}_n - \sum_{m=1}^{n-1} l_{nm} y_m \right) l_{nn}^{-1} \end{aligned}$$

A similar formula applies to solve $Ux = y$. In this case we solve first for $x_n = y_n / u_{nn}$. The general formula is recursive with x_k being determined after $x_{k+1}, \ldots, x_n$. are determined using the formula

$$x_k = \left(y_k - \sum_{m=k+1}^{n} u_{km} y_m \right) u_{kk}^{-1}$$

In practice the step of determining and then multiplying by the permutation matrix is not actually carried out. Rather, an index array is generated, while the elimination step is accomplished that effectively interchanges a "pointer" to the row interchanges. This saves considerable time in solving potentially very large systems.

More general and instructive methods are available for accomplishing this LU factorization. Also, conditions are available for when no (nontrivial) permutation is required. We need the following lemma.

Lemma 7.1.3. *Let $A \in M_n(\mathbb{C})$ have the LU factorization $A = LU$, where L is lower triangular and U is upper triangular. For any partition of the matrix of the form*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

there are corresponding decompositions of the matrices L and U

$$L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

where the L_{ii} *and the* U_{ii}. *are lower and upper triangular respectively. Moreover, we have*

$$\begin{aligned} A_{11} &= L_{11}U_{11} \\ A_{21} &= L_{21}U_{11} \\ A_{12} &= L_{12}U_{22} \\ A_{22} &= L_{21}U_{12} + L_{22}U_{22} \end{aligned}$$

Thus $L_{11}U_{11}$ *is a LU factorization of* A_{11}.

With this lemma we can establish that almost every matrix can have a LU factorization.

Definition 7.1.2. Let $A \in M_n(\mathbb{C})$ and suppose that $1 \leq j \leq n$. The expression $\det(A\{1,\ldots,j\})$ means the determininant of the upper left $j \times j$ submatrix of A. These quaditities for $j = 1,\ldots,n$ are called the **principal determinants of** A**.**

Theorem 7.1.2. *Let* $A \in M_n(\mathbb{C})$ *and suppose that* A *has rank* k. *If*

$$\det(A\{1,\ldots,j\}) \neq 0 \quad \text{for} \quad j = 1,\ldots,k \tag{1}$$

then A *has a LU factorization* $A = LU$, *where* L *is lower triangular and* U *is upper triangular. Moreover, the factorization may be taken so that either* L *or* U *is nonsingular. In the case* $k = n$ *both* L *and* U *will be nonsingular.*

Proof. We carry out this LU factorization as a direct calculation in comparison to the Gaussian elimination method above. Let us propose to solve the equation $LU = A$ expressed as

$$\begin{bmatrix} l_{11} & & & & & \\ l_{21} & l_{22} & & & 0 & \\ l_{31} & l_{32} & l_{33} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ & & & & \ddots & \\ l_{n1} & l_{n2} & \cdots & \cdots & & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ & u_{22} & u_{23} & \cdots & u_{2n} \\ & & u_{33} & & \\ & 0 & & \ddots & \vdots \\ & & & \ddots & \\ & & & & u_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & \ddots & \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}$$

It is easy to see that $l_{11}u_{11} = a_{11}$. We can take, for example $l_{11} = 1$ and solve for u_{11}. The detminant condition assures us that $u_{11} \neq 0$. Next solve for the $(2,1)$-entry. We have $l_{21}u_{11} = a_{21}$. Since $u_{11} \neq 0$, solve for l_{21}. For the $(1,2)$-entry we have $l_{11}u_{12} = a_{12}$, which can be solved for u_{12} since $l_{11} \neq 0$. Finally, for the $(2,2)$-entry, $l_{12}u_{12} + l_{22}u_{22} = a_{22}$ is an equation with two unknowns. Assign $l_{22} = 1$ and solve for u_{22}. What is important to note is that the process carried out this way gives the factorization of the upper left 2×2 submatrix of A. Thus

$$\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Since $\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) \neq 0$, it follows that $\det\left(\begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}\right) \neq 0$ and we know that $\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$ is nonsingular as the diagonal elements are ones. Continue the factorization process through the $k \times k$ upper left submatrix of A.

Now consider the blocked matrix form form A

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $k \times k$ and has rank k. Thus we know that the rows of the lower $(n-k) \times n$ matrix above, that is $\begin{bmatrix} A_{21} & A_{22} \end{bmatrix}$ can be written as a unique linear combination of the rows of the upper $k \times n$ matrix $\begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$. Thus

$$\begin{bmatrix} A_{21} & A_{22} \end{bmatrix} = C \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

for some $(n-k) \times k$ matrix C. Of course this means: $A_{21} = C\ A_{11}$ and $A_{22} = C\ A_{12}$. We consider the factorization

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

where the blocks L_{11} and U_{11} have just been determined. From the equations in the lemma above we solve to get $U_{12} = L_{11}^{-1}\ A_{12}$ and $L_{21} =$

$A_{12}U_{11}^{-1}$. Then

$$\begin{aligned} A_{22} &= L_{21}U_{12} + L_{22}U_{22} \\ &= A_{12}U_{11}^{-1}L_{11}^{-1}\ A_{12} + L_{22}U_{22} \\ &= A_{12}\ A_{11}^{-1}\ A_{12} + L_{22}U_{22} \\ &= C\ A_{11}\ A_{11}^{-1}\ A_{12} + L_{22}U_{22} \\ &= C\quad A_{12} + L_{22}U_{22} \\ &= A_{22} + L_{22}U_{22} \end{aligned}$$

Thus we solve $L_{22}U_{22} = 0$. Obviously, we can take for L_{22} any nonsingular matrix we wish and solve for U_{22} or conversely.

□

7.2 *LR* factorization

While the PLU factorization is useful for solving systems, the LR factorization can be used to determine eigenvalues. .

Let $A \in M_n$ be given. Then

$$A = A_1 = L_1R_1.$$

Then

$$\begin{gathered} L_1^{-1}A_1L_1 = R_1L_1 \equiv A_2 \\ A_2 = L_2R_2 \\ L_2^{-1}A_2L_2 = R_2L_2 \equiv A_3. \end{gathered}$$

Continue in this fashion to obtain

$$L_k^{-1}A_kL_k = R_kL_k \equiv A_{k+1} \tag{$\star$}$$

We define

$$\begin{gathered} P_k = L_1L_2\ldots L_k \\ Q_k = R_k\ldots R_2R_1. \end{gathered}$$

Then

$$P_kA_{k+1} = A_1P_k$$

for

$$\begin{aligned} A_{k+1} &= L_k^{-1} A_k L_k \\ &= L_k^{-1} L_{k-1}^{-1} A_{k-1} L_{k-1} L_k \\ &\vdots \\ &= P_k^{-1} A_1 P_k \end{aligned}$$

or

$$P_k\, A_{k+1} = A_1 P_k.$$

Hence

$$\begin{aligned} P_k Q_k &= P_{k-1}\, A_k Q_{k-1} \\ &= A_1 P_{k-1} Q_{k-1} \\ &= A_1 P_{k-2}\, A_{k-1} Q_{k-2} \\ &= A_1^2 P_{k-2} Q_{k-2} \\ &\;\vdots \\ &= A_1^k. \end{aligned}$$

Theorem 7.2.1 (Rutishauser). *Let $A \in M_n$ be given. Assume the eigenvalues of A satisfy*

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0.$$

Then $A \sim \Lambda = diag(\lambda_1 \ldots \lambda_n)$. Assume $A = S\Lambda S^{-1}$, and

$$Y \equiv S^{-1} = L_y R_y \qquad X = S = L_x R_x$$

where L_y and L_x are lower unit triangular matrices and R_y and R_x are upper triangular. Then A_k defined by $(\star)$ satisfy the result $\lim A_k$ is upper triangular.

Proof. (Wilkinson) We have

$$\begin{aligned} A_1^k &= X\Lambda^k Y \\ &= X\Lambda^k L_y R_y \\ &= X\Lambda^k L_y \Lambda^{-k} \Lambda^k R_y. \end{aligned}$$

By the strict inequalities between the eigenvalues we have

$$(\Lambda^k L_y \Lambda^{-k})_{ij} = \begin{cases} 1 & i = j \\ \left(\dfrac{\lambda_i}{\lambda_j}\right)^k \ell_{ij} & i > j \\ 0 & i < j. \end{cases}$$

Hence $\Lambda^k L_y \Lambda^{-k} \to I$ (because $\frac{|\lambda_i|}{|\lambda_j|} < 1$ if $i > j$). Hence with

$$A_1^k = L_x R_x (\Lambda^k L_y \Lambda^{-k}) \Lambda^k R_y$$

and

$$A_1^k = P_k Q_k$$

we conclude that $\lim\limits_{k\to\infty} P_k = L_x$. Therefore

$$L_k = P_{k-1}^{-1} P_k \to I.$$

Finally we have that A_k must be upper triangular because

$$L_k^{-1} A_k = R_k$$

is upper triangular.

□

This exposes all the eigenvalues of A. Therefore the eigenvectors can be determined.

7.3 The *QR* algorithm

Certain numerical problems with the LU algorithm have led to the QR algorithm, which is based on the decomposition of the matrix A as

$$A = QR$$

where Q is unitary and R is upper triangular.

Theorem 7.3.1 (QR-factorization). *(i) Suppose A is in $M_{n,m}$ and $n \geq m$. Then there is a matrix $Q \in M_{n,m}$ with orthogonal columns and an upper triangular matrix*
$R \in M_m$ such that $A = QR$.

(ii) If $n = m$, then Q is unitary. If A is nonsingular the diagonal entries of R can be chosen to be positive.

(iii) If A is real; then Q and R may be chosen to be real.

Proof. (i) We proceed inductively. Let $a_1, \ldots, a_n$ denote the columns of A and $q_1, q_2, \ldots, q_m$ denote the columns of Q. The basic idea of the QR-factorization is to orthogonalize the columns of A from left to right. Then the columns can be expressed by the formulas $a_k = \sum_{i=1}^{k} c_k q_k, \quad k = 1, \ldots, n$. The coefficients of the expansion become, respectively, the entries of the k^{th} column of R, completed by $n - k$ zeros. (Of course, if the rank of A is less than m, we fill in arbitrary orthogonal vectors which we know exist as $m \leq n$.) For the details, first define $q_1 = a_1/\|a_1\|$. To compute q_2 we use the Gram–Schmidt procedure.

$$\begin{aligned}\hat{q}_2 &= a_2 - \langle q_1, a_1 \rangle q_1 \\ q_2 &= \hat{q}_2/\|\hat{q}_2\|.\end{aligned}$$

Tracing backwards note that

$$\begin{aligned}a_2 &= \hat{q}_2 + \langle q_1, a_1 \rangle q_1 \\ &= \|\hat{q}_2\| q_2 + \langle q_1, a_1 \rangle q_1.\end{aligned}$$

So we have

$$\begin{bmatrix} a_1 & a_2 & a_3 & \\ \downarrow & \downarrow & \downarrow & \ldots \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 & \\ \downarrow & \downarrow & \downarrow & \ldots \end{bmatrix} \begin{bmatrix} \|a_1\| & \langle q_1, a_1 \rangle & \ldots \\ 0 & \|\hat{q}_2\| & \\ \vdots & 0 & \\ 0 & 0 & \end{bmatrix}.$$

Instead of the full inductive step we compute q_3 and finish at that point

$$\begin{aligned}\hat{q}_3 &= a_3 - \langle q_1, a_3 \rangle q_1 - \langle q_2, a_3 \rangle q_2 \\ q_3 &= \hat{q}_3/\|\hat{q}_3\|.\end{aligned}$$

Hence

$$a_3 = \|\hat{q}_3\| q_3 + \langle q_1, a_3 \rangle q_1 + \langle q_2, a_3 \rangle q_2.$$

The third column of R is thus given by

$$r_3 = [\langle q_1, a_3\rangle, \langle q_2, a_3\rangle, \|\hat{q}_3\|, 0, 0, \dots, 0]^T.$$

In this way we see that the columns of Q are orthogonal and the matrix R is upper triangular, with an exception. That is the possibility that $\hat{q}_k = 0$ for some k. In this degenerate case we take q_k to be any vector orthogonal to the span of $a_1, a_2, \dots, a_m$, and we take $r_{kj} = 0$, $j = k, k+1 \dots m$. Also we note that if $\hat{q}_k = 0$, then a_k is linearly dependent on $a_1, a_2, \dots, a_{k-1}$, and hence on $q_1, q_2, \dots q_{k-1}$. Select the coefficients $r_{1k}, \dots, r_{k-1k}$ to reflect this dependence.

(ii) If $m = n$, the process above yields a unitary matrix. If A is nonsingular, the process above yields a matrix R with a positive diagonal.

(iii) If A is a real, all operators above can be carried out in real arithmetic.

□

Now what about the uniqueness of the decomposition? Essentially the uniqueness is true up to a multiplication by a diagonal matrix, except in the case when the matrix has rank is less than m, when there is no form of uniqueness. Suppose that the rank of A is m.

Then application of the Gram-Schmidt procedure yields a matrix R with positive diagonal. Suppose that A has two QR factorizations, QR and PS with upper triangular factors having positive diagonals. Then

$$P^*Q = SR^{-1}$$

We have that SR^{-1} is upper triangular and moreover has a positive diagonal. Also, P^*Q is unitary. We know that the only upper triangular unitary matrices are diagonal matrices, and finally the only unitary matrix with a positive diagonal is the identity matrix. Therefore $P^*Q = I$, which is to say that $P = Q$. We summarize as

Corollary 7.3.1. *Suppose A is in $M_{n,m}$ and $n \geq m$. If rank$(A) = m$ then the QR factorization of $A = QR$ with upper triangular matrix R having a positive diagonal is unique.*

The QR algorithm

The QR algorithm parallels the LR algorithm almost identically. Suppose A is in M_n Define

$$\begin{aligned} A_1 &= Q_1 R_1 \\ A_2 &\equiv R_1 Q_1. \end{aligned}$$

Also

$$Q_1^* A_1 Q = A_2.$$

Then decompose A_2 into a QR decomposition

$$A_2 = Q_2 R_2$$

and

$$Q_2^* A_2 Q_2 = R_2 Q_2 \equiv A_3.$$

Also

$$Q_2^* Q_1^* A_1 Q_1 Q_2 = R_2 Q_2 = A_3.$$

Proceed sequentially

$$\begin{aligned} A_k &= Q_k R_k \\ A_{k+1} &= R_k Q_k \\ Q_k^* A_k Q_k &= A_{k+1}. \end{aligned}$$

Let

$$\begin{aligned} P_k &= Q_1 Q_2 \dots Q_k \\ T_k &= R_k R_{k-1} \dots R_1. \end{aligned}$$

Then

$$P_k^* A_1 P_k = A_{k+1}.$$

whence

$$P_k A_{k+1} = A_1 P_k.$$

Also we have

$$\begin{aligned} P_k T_k &= P_{k-1} Q_k R_k T_{k-1} \\ &= P_{k-1} A_k T_{k-1} \\ &= A_1 P_{k-1} T_{k-1} \\ &= \ldots \\ &= A_1^k. \end{aligned}$$

Theorem 7.3.2. *Let $A \in M_n$ be given, and assume the eigenvalues of A satisfy*

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0.$$

Then the iterations A_k converge to a triangular matrix.

Proof. Our hypothesis gives that A is diagonalizable, and we write $A \sim \Lambda = \operatorname{diag}(\lambda_1 \ldots \lambda_n)$. That is,

$$A_1 = S \Lambda S^{-1}$$

where $\Lambda = \operatorname{diag}(\lambda_1 \ldots \lambda_n)$. Let

$$\left. \begin{array}{l} X = S = Q_x R_x \\ Y = S^{-1} = L_y U_y \end{array} \right| \begin{array}{l} \text{here } QR \\ \text{here } LU. \end{array}$$

Then

$$\begin{aligned} A_1^k &= Q_x R_x \Lambda^k L_y U_y \\ &= Q_x R_x \Lambda^k L_y \Lambda^{-k} \Lambda^k U_y \\ &= Q_x (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_Y \end{aligned}$$

where

$$E_k = \Lambda^k L_y \Lambda^{-k} - I$$

$$(E_k)_{ij} = \begin{cases} 0 & i = j \\ (\lambda_i/\lambda_j)^k \ell_{ij} & i > j \\ 0 & i < j. \end{cases}$$

It follows that $I + R_x E_k R_x^{-1} \to I$, and $R_x \Lambda^{-k} U_y$ is upper triangular. Thus

$$Q_x (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_y = P_k T_k.$$

The matrix $I + R_x E_k R_x^{-1}$ can be QR factored as $\tilde{U}_k \tilde{R}_k$, and since $I + R_x E_k R_x^{-1} \to I$, it follows that we can assume both $\tilde{U}_k \to I$ and $\tilde{R}_k \to I$. Hence

$$A_1^k = Q_x \tilde{U}_k [\tilde{R}_k (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_y] = P_k T_k.$$

with the first factor unitary and the second factor upper triangular. Since we have assumed (by the eigenvalue condition) that A is nonsingular, this factorization is essentially unique, where possibly a multiplication by a diagonal matrix must be applied to give the upper triangular factor on the right a positive diagonal. Just what is the form of the diagonal matrix can be seen from the following. Let $\Lambda = |\Lambda|\, \Lambda_1$, where $|\Lambda|$ is the diagonal matrix of moduli of the elements of Λ and where Λ_1 is the unitary matrix of the signs of each eigenvalue respectively. We also take $U_y = \Lambda_2 \left(\Lambda_2^* U_y\right)$ where Λ_2 is a unitary matrix chosen so that $\Lambda_2^* U$ has a positive diagonal. Then

$$A_1^k = Q_x \tilde{U}_k \Lambda_2 \Lambda_1^k [\left(\Lambda_2 \Lambda_1^k\right)^{-1} \tilde{R}_k (I + R_x E_k R_x^{-1}) R_x \left(\Lambda_2 \Lambda_1^k\right) |\Lambda|^k \left(\Lambda_2^* U_y\right)] = P_k T_k.$$

From this we obtain P_k is essentially asymptotic to $Q_x \tilde{U}_k \Lambda_2 \Lambda_1^k$ and from this we obtain that

$$Q_k = P_{k-1}^{-1} P_k \to \Lambda_1$$

which is diagonal. Finally, it follows that A_k is upper triangular since

$$Q_k^{-1} A_k = R_k$$

In the limit therefore A is similar to an upper triangular matrix. □

Example 7.3.1. Apply the QR method to the matrix

$$A := \begin{bmatrix} 2.3 & 1 & 2 \\ 2 & 2 & 2.1 \\ 3 & 2 & 0 \end{bmatrix}$$

The matrix A has eigvenvalues 5.45, 0.723, -1.87. The successive iterations are

$$A_2 = \begin{bmatrix} 5.10 & -0.511 & 2.13 \\ 0.631 & 0.662 & 0.136 \\ 1.42 & -0.0202 & -1.44 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 5.51 & -1.02 & -0.36 \\ -0.0146 & 0.666 & 0.482 \\ 0.513 & 0.240 & -1.84 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 5.46 & -1.41 & 0.482 \\ -0.0372 & 0.495 & 0.672 \\ 0.169 & 0.815 & -1.62 \end{bmatrix} \qquad A_5 = \begin{bmatrix} 5.47 & -0.366 & -1.26 \\ -0.0404 & -0.462 & 1.39 \\ 0.0430 & 1.21 & -0.677 \end{bmatrix}$$

$$A_6 = \begin{bmatrix} 5.46 & -1.13 & -0.687 \\ -0.0184 & -1.52 & 0.813 \\ 0.00826 & 0.983 & 0.381 \end{bmatrix} \qquad A_7 = \begin{bmatrix} 5.45 & 0.529 & -1.18 \\ -0.00682 & -1.78 & 0.585 \\ 0.00115 & 0.414 & 0.638 \end{bmatrix}$$

$$A_8 = \begin{bmatrix} 5.43 & 0.684 & -1.09 \\ -0.000822 & -1.87 & 0.229 \\ 0.0000215 & 0.0659 & 0.729 \end{bmatrix}$$

Note the gradual appearance of the eigenvalues on the diagonal.
Remark. These iterations were carried out in precision 3 arithmetic, which affects the rate of convergence to triangular form.

7.4 Least Squares

As we know, if $A \in M_{n,m}$ with $m < n$ it is generally not possible to solve the overdetermined system

$$Ax = b.$$

For example, suppose we have the data $\{(x_i, y_i)\}_{i=1}^n$, with the x-coordinates distinct. We may wish to "fit" a straight to this data. This means we want to find coefficients m and b so that

$$b + mx_i = y_i, \qquad i = 1, \dots, n. \tag{$\star$}$$

Taking the matrix and data vector

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \qquad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and $z = [b, m]^T$, the system $(\star)$ becomes $Az = b$. Usually $n \gg 2$. Hence there is virtually no hope to determine a unique solution to system.

However, there are numerous ways to determine constants m and b so that the resulting line represents the data. For example, owing to the distinctness of the x-coordinates, it is possible to solve any 2×2 subsystem of

$Az = b$. Other variations exist. A new 2×2 system could be created by creating two averages of the data, say left and right, and solving. Assume the sequence $\{x_j\}$ is ordered from least to greatest. Define $x_\ell = \frac{1}{k}\sum_{j=1}^{k} x_j$ and $x_r = \frac{1}{n-k}\sum_{j=k+1}^{n} x_j$. Let y_ℓ and y_r denote the corresponding averages for the ordinates. Then define the intercept b and slope m by solving the system

$$\begin{bmatrix} 1 & x_\ell \\ 1 & x_r \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} y_\ell \\ y_r \end{bmatrix}$$

While this will normally give a reasonable approximating line, its value has little utility beyond its naive simplicity and visual appearance. What is desired is to establish a criteria for choosing the line.

Define the **residual** of the approximation $r = b - Az$. It makes perfect sense to consider finding $z = [b, m]^T$ for which the residual is minimized in some norm. Any norm can be selected here, but on practical grounds the best norm to use is the Euclidean norm $\|\cdot\|_2$. The vector Az that yields the minimal norm residual is the one for which $(b - Az) \perp Aw$, for we are seeking the nearest value in the Aw to the vector b. It can be found by select the one for the solution, Az, for which

$$b - Ax \perp Aw \quad \text{all } w.$$

This means

$$\langle b - Ax, Ay \rangle = 0 \quad \text{all } y$$

or

$$\langle A^T(b - Ay), y \rangle = 0 \quad \text{all } y$$

or

$$A^T(b - Ay) = 0$$

$$\boxed{A^T Ay = A^T b.} \quad \text{Normal Equations}$$

The **least squares** solution to $Ax = b$ is given by the solution to the normal equation

$$A^T Ay = A^T b.$$

Suppose we have the QR decomposition for A. Then if A is real

$$A^TA = R^TQ^TQR = R^TR$$
$$A^Ty = R^TQy.$$

Hence the normal equations become

$$R^TRx = R^TQy.$$

Assuming that the rank of A is m, we must have that R and hence R^T is invertible. Therefore we have the least squares solution is given by the triangular system

$$Rx = Qy.$$

7.5 Exercises

1. If $A \in M(\mathbb{C})$ has rank k, show that there is a permutation matrix P such that PA has its first k principal determinants nonzero.
2. For the least squares fit of a straight line determine R and Q.
3. In the case of data

$$A^TA = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \qquad A^Tb = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}.$$

4. In attempting to solve a quadratic fit we have the model

$$c + bx_i + ax_i^2 = y_i \qquad i = 1, \dots, n.$$

The system is

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

The normal equations have the matrix and data given by

$$A^TA = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^2 & \Sigma x_i^4 \end{bmatrix} \qquad A^Tb = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}.$$

5. Find the normal equations for the least squares fit of data to a polynomial of degree k.