Lecture 3: QR-Factorization

This lecture introduces the Gram–Schmidt orthonormalization process and the associated QR-factorization of matrices. It also outlines some applications of this factorization. This corresponds to section 2.6 of the textbook. In addition, supplementary information on other algorithms used to produce QR-factorizations is given.

1 Gram–Schmidt orthonormalization process

Consider n linearly independent vectors $u_1, \ldots, u_n$ in $\mathbb{C}^m$. Observe that we necessarily have $m \geq n$. We wish to 'orthonormalize' them, i.e., to create vectors $v_1, \ldots, v_n$ such that

$$(v_1, \ldots, v_k) \text{ is an orthonormal basis for } V_k := \operatorname{span}[u_1, \ldots, u_k] \quad \text{for all } 1k \in [1:n].$$

It is always possible to find such vectors, and in fact they are uniquely determined if the additional condition $\langle v_k, u_k \rangle > 0$ is imposed. The step-by-step construction is based on the following scheme.

Suppose that $v_1, \ldots, v_{k-1}$ have been obtained; search in V_k for a vector

$$\widetilde{v}_k = u_k + \sum_{i=1}^{k-1} c_{k,i}\, v_i \quad \text{such that} \quad \widetilde{v}_k \perp V_{k-1};$$

the conditions $0 = \langle \widetilde{v}_k, v_i \rangle = \langle u_k, v_i \rangle + c_{k,i}$, $i \in [1 : k-1]$, impose the choice $c_{k,i} = -\langle u_k, v_i \rangle$; now that $\widetilde{v}_k$ is completely determined, normalize it to obtain the vector $v_k = \dfrac{1}{\|\widetilde{v}_k\|}\widetilde{v}_k$.

For instance, let us write down explicitly all the steps in the orthonormalization process for the vectors

$$\underline{u_1 = [6,3,2]^\top}, \quad \underline{u_2 = [6,6,1]^\top}, \quad \underline{u_3 = [1,1,1]^\top}.$$

- $\widetilde{v}_1 = u_1, \quad \|\widetilde{v}_1\| = \sqrt{36+3+4} = 7, \quad \underline{v_1 = 1/7\,[6,3,2]^\top};$
- $\widetilde{v}_2 = u_2 + \alpha v_1, \quad 0 = \langle \widetilde{v}_2, v_1 \rangle \Rightarrow \alpha = -\langle u_2, v_1 \rangle = -(16+18+2)/7, \quad \underline{\alpha = -8},$
  $\widetilde{v}_2 = 1/7\,[7\cdot 6 - 8\cdot 6, 7\cdot 6 - 8\cdot 3, 7\cdot 1 - 8\cdot 2]^\top = 1/7\,[-6,18,-9]^\top = 3/7\,[-2,6,-3]^\top,$
  $\|\widetilde{v}_2\| = 3/7\sqrt{4+36+9} = 3, \quad \underline{v_2 = 1/7\,[-2,6,-3]^\top};$
- $\widetilde{v}_3 = u_3 + \beta v_2 + \gamma v_1, \quad \begin{array}{l} 0 = \langle \widetilde{v}_3, v_2 \rangle, \\ 0 = \langle \widetilde{v}_3, v_1 \rangle, \end{array} \Rightarrow \begin{array}{ll} \beta = -\langle u_3, v_2 \rangle = -(-2+6-3)/7, & \underline{\beta = -1/7}, \\ \gamma = -\langle u_3, v_1 \rangle = -(6+3+2)/7, & \underline{\gamma = -11/7}, \end{array}$
  $\widetilde{v}_3 = 1/49\,[49+2-66, 49-6-33, 49+3-22]^\top = 1/49\,[-15,10,30]^\top = 5/49\,[-3,2,6]^\top,$
  $\|\widetilde{v}_3\| = 5/49\sqrt{9+4+36} = 5/7, \quad \underline{v_3 = 1/7\,[-3,2,6]^\top}.$

2 QR-factorization

Theorem 1. For a nonsingular $A \in \mathcal{M}_n$, there exists a unique pair of unitary matrix $Q \in \mathcal{M}_n$ and upper triangular matrix $R \in \mathcal{M}_n$ with positive diagonal entries such that

$$A = QR.$$

The QR-factorization can be used for the following tasks:

- solving linear systems according to

  $$[Ax = b] \iff [Qy = b, \quad y = Rx],$$

  since the system $y = Rx$ is easy to solve [backward substitution], and the system $Qy = b$ is even easier to solve [take $y = Q^* b$];
- calculate the (modulus of the) determinant and find the inverse [$|\det A| = \prod_{i=1}^n r_{i,i}$, $A^{-1} = R^{-1} Q^*$];
- find the Cholesky factorization of a positive definite matrix $B = A^* A \in \mathcal{M}_n$, $A \in \mathcal{M}_n$ being nonsingular (we will later see why every positive definite matrix can be factored in this way), i.e., find a factorization

  $$B = LL^*,$$

  where $L \in \mathcal{M}_n$ is lower triangular with positive diagonal entries [$L = R^*$];
- find a Schur's factorization of a matrix $A \in \mathcal{M}_n$ via the QR-algorithm defined by

  $$\begin{aligned}
  A_0 &:= A, & A_0 &=: Q_0 R_0, \\
  A_1 &:= R_0 Q_0, & A_1 &=: Q_1 R_1, \\
  &\vdots & &\vdots \\
  A_k &:= R_{k-1} Q_{k-1}, & A_1 &=: Q_k R_k, \\
  &\vdots & &\vdots
  \end{aligned}$$

  Note that A_k is always unitarily equivalent to A. If all the eigenvalues of A have distinct moduli, then A_k tends to an upper triangular matrix T (which is therefore unitarily equivalent to A, see Exercise 3). The eigenvalues of A are read on the diagonal of T.

In the general case of nonsingular or nonsquare matrices, the QR-factorization reads:

Theorem 2. For $A \in \mathcal{M}_{m \times n}$, $m \geq n$, there exists a matrix $Q \in \mathcal{M}_{m \times n}$ with orthonormal columns and an upper triangular matrix $R \in \mathcal{M}_n$ such that

$$A = QR.$$

Beware that the QR-factorization of a rectangular matrix A is not always understood with Q rectangular and R square, but sometimes with Q square and R rectangular, as with the MATLAB command `qr`.

3 Proof of Theorems 1 and 2

Uniqueness: Suppose that $A = Q_1 R_1 = Q_2 R_2$ where Q_1, Q_2 are unitary and R_1, R_2 are upper triangular with positive diagonal entries. Then

$$M := R_1 R_2^{-1} = Q_1^* Q_2.$$

Since M is a unitary (hence normal) matrix which is also upper triangular, it must be diagonal (see Lemma 4 of Lecture 2). Note also that the diagonal entries of M are positive (because the upper triangular matrices R_1 and R_2^{-1} have positive diagonal entries) and of modulus one (because M is a diagonal unitary matrix). We deduce that $M = I$, and consequently that

$$R_1 = R_2, \qquad Q_1 = Q_2.$$

Existence: Let us consider a matrix $A \in \mathcal{M}_{m \times n}$ with $m \geq n$, and let $u_1, \ldots, u_n \in \mathbb{C}^m$ denote its columns. We may assume that $u_1, \ldots, u_n$ are linearly independent (otherwise limiting arguments can be used, see Exercise 4). Then the result is just a matrix interpretation of the Gram–Schmidt orthonormalization process of m linearly independent vectors in $\mathbb{C}^n$. Indeed, the Gram–Schmidt algorithm produces orthonormal vectors $v_1, \ldots, v_n \in \mathbb{C}^m$ such that, for each $j \in [1 : n]$,

$$u_j = \sum_{k=1}^{j} r_{k,j}\, v_k = \sum_{k=1}^{n} r_{k,j}\, v_k, \tag{1}$$

with $r_{k,j} = 0$ for $k > j$, in other words, $R := [r_{i,j}]_{i,j=1}^n$ is an $n \times n$ upper triangular matrix. The n equations (1) reduce, in matrix form, to $A = QR$, where Q is the $m \times n$ matrix whose columns are the orthonormal vectors $v_1, \ldots, v_n$.

[To explain the other QR-factorization, let us complete $v_1, \ldots, v_n$ with $v_{m+1}, \ldots, v_n$ to form an orthonormal basis $(v_1, \ldots, v_m)$ of $\mathbb{C}^m$. The analogs of the equations (1), i.e., $u_j = \sum_{k=1}^m r_{k,j}\, v_k$ with $r_{k,j} = 0$ for $k > j$, read $A = QR$, where Q is the $m \times m$ orthogonal matrix with columns $v_1, \ldots, v_m$ and R is an $m \times n$ upper triangular matrix.]

To illustrate the matrix interpretation, observe that the orthonormalization carried out in Section 1 translates into the factorization [identify all the entries]

$$\begin{bmatrix} 6 & 6 & 1 \\ 3 & 6 & 1 \\ 2 & 1 & 1 \end{bmatrix} = \underbrace{\frac{1}{7}\begin{bmatrix} 6 & -2 & -3 \\ 3 & 6 & 2 \\ 2 & -3 & 6 \end{bmatrix}}_{\text{unitary}} \underbrace{\begin{bmatrix} 7 & 8 & 11/7 \\ 0 & 3 & 1/7 \\ 0 & 0 & 5/7 \end{bmatrix}}_{\text{upper triangular}}.$$

An aside: other QR-factorization algorithms

The Gram–Schmidt algorithm has the disadvantage that small imprecisions in the calculation of inner products accumulate quickly and lead to effective loss of orthogonality. Alternative ways to obtain a QR-factorization are presented below on some examples. They are based on the following idea and exploits the fact that the computed product of unitary matrices gives a unitary matrix with acceptable error.

Multiply the (real) matrix A on the left by some orthogonal matrices Q_i which 'eliminate' some entries below the main 'diagonal', until the result is an upper triangular matrix R, thus

$$Q_k \cdots Q_2 Q_1 A = R \quad \text{yields} \quad A = QR, \quad \text{with} \quad Q = Q_1^* Q_2^* \cdots Q_k^*.$$

We mainly know two types of unitary transformations, namely rotations and reflexions. Therefore, the two methods we describe are associated with Givens rotations [preferred when A is sparse] and Householder reflections [preferred when A is dense].

Givens rotations

The matrix

$$\Omega^{[i,j]} := \begin{array}{cc} & \begin{array}{ccccccc} & & i & & j & & \end{array} \\ \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \\ \end{array} & \begin{bmatrix} \ddots & & & & & & \\ & 1 & \vdots & 0 & \vdots & 0 & \\ & \cdots & \cos\theta & \cdots & \sin\theta & \cdots & \\ & 0 & \vdots & \ddots & \vdots & 0 & \\ & \cdots & -\sin\theta & \cdots & \cos\theta & \cdots & \\ & 0 & \vdots & 0 & \vdots & 1 & \\ & & & & & & \ddots \end{bmatrix} \end{array}$$

corresponds to a rotation along the two-dimensional space $\mathrm{span}[e_i, e_j]$. The rows of the matrix $\Omega^{[i,j]}A$ are the same as the rows of A, except for the i-th and j-th rows, which are linear combinations of the i-th and j-th rows of A. By choosing θ appropriately, we may introduce a zero at a prescribed position on one of these rows. Consider for example the matrix $A = \begin{bmatrix} 6 & 6 & 1 \\ 3 & 6 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ of the end of Section 1. We pick $\Omega^{[1,2]}$ so that $\Omega^{[1,2]}A = \begin{bmatrix} 6 & 6 & 1 \\ 0 & \times & \times \\ \times & \times & \times \end{bmatrix}$. Then we pick $\Omega^{[1,3]}$ so that $\Omega^{[1,3]}\Omega^{[1,2]}A = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$. Finally, thanks to the leading zeros in the second and third rows, we can pick $\Omega^{[2,3]}$ so that $\Omega^{[2,3]}\Omega^{[1,3]}\Omega^{[1,2]}A = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$. The matrix $\left[\Omega^{[2,3]}\Omega^{[1,3]}\Omega^{[1,2]}\right]^*$ is the unitary matrix required in the factorization of A.

Householder reflections

The reflection in the direction of a vector v transforms v into $-v$ while leaving the space $v^\perp$ unchanged. It can therefore be expressed through the Hermitian unitary matrix

$$H_v := I - \frac{2}{\|v\|^2} vv^*.$$

Consider the matrix $A = \begin{bmatrix} 6 & 6 & 1 \\ 3 & 6 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ once again. We may transform $u_1 = [6,3,2]^\top$ into $7e_1 = [7,0,0]^\top$ by way of the reflection in the direction $v_1 = u_1 - 7e_1 = [-1,3,2]^\top$. The latter is represented by the matrix

$$H_{v_1} = I - \frac{2}{\|v_1\|^2} v_1 v_1^* = I - \frac{1}{7}\begin{bmatrix} 1 & -3 & -2 \\ -3 & 9 & 6 \\ -2 & 6 & 4 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 6 & 3 & 2 \\ 3 & -2 & -6 \\ 2 & -6 & 3 \end{bmatrix}.$$

Then the matrix $H_{v_1}A$ has the form $\begin{bmatrix} 7 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$, where the precise expression for the second column is

$$H_{v_1}u_2 = u_2 - \frac{\langle v_1, u_2\rangle}{7} v_1 = u_2 - 2v_1 = \begin{bmatrix} 8 \\ 0 \\ -3 \end{bmatrix}.$$

To cut the argument short, we may observe at this point that the multiplication of $H_{v_1}A$ on the left by the permutation matrix $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ [which can be interpreted as $H_{e_2 - e_3}$] exchanges the second and third rows, thus gives an upper triangular matrix. In conclusion, the orthogonal matrix Q has been obtained as

$$(PH_{v_1})^* = H_{v_1}^* P^* = H_{v_1}P = \frac{1}{7}\begin{bmatrix} 6 & 2 & 3 \\ 3 & -6 & -2 \\ 2 & 3 & -6 \end{bmatrix}.$$

4 Exercises

Ex.1: Prove that a matrix $A \in \mathcal{M}_{m \times n}$, $m \leq n$, can be factored as $A = LP$ where $L \in M_m$ is lower triangular and $P \in \mathcal{M}_{m \times n}$ has orthonormal rows.

Ex.2: Prove the uniqueness of the Cholesky factorization of a positive definite matrix.

Ex.3: Exercise 5 p. 117.

Ex.4: Fill in the details of the following argument: for $A \in \mathcal{M}_{m \times n}$ with $m \geq n$, there exists a sequence of matrices $A_k \in \mathcal{M}_{m \times n}$ with linearly independent columns such that $A_k \to A$ as $k \to \infty$; each A_k can be written as $A_k = Q_k R_k$ where $Q_k \in \mathcal{M}_{m \times n}$ has orthonormal columns and $R_k \in \mathcal{M}_n$ is upper triangular; there exists a subsequence (Q_{k_j}) converging to a matrix $Q \in \mathcal{M}_{m \times n}$ with orthonormal columns, and the sequence (R_{k_j}) converges to an upper triangular matrix $R \in \mathcal{M}_n$; taking the limit when $j \to \infty$ yields $A = QR$.

Ex.5: Fill in the numerical details in the section on Givens rotations.