Matrix powers also come up in the analysis of a time-invariant linear dynamical system with an input. We have

$$x_{t+2} = Ax_{t+1} + Bu_{t+1} = A(Ax_t + Bu_t) = A^2 x_t + ABu_t + Bu_{t+1}.$$

Iterating this over ℓ periods we obtain

$$x_{t+\ell} = A^\ell x_t + A^{\ell-1}Bu_t + A^{\ell-2}Bu_{t+1} + \cdots + Bu_{t+\ell-1}. \tag{10.2}$$

(The first term agrees with the formula for $x_{t+\ell}$ with no input.) The other terms are readily interpreted. The term $A^j Bu_{t+\ell-j}$ is the contribution to the state $x_{t+\ell}$ due to the input at time $t+\ell-j$.

10.4 QR factorization

Matrices with orthonormal columns. As an application of Gram matrices, we can express the condition that the n-vectors $a_1, \ldots, a_k$ are orthonormal in a simple way using matrix notation:

$$A^T A = I,$$

where A is the $n \times k$ matrix with columns $a_1, \ldots, a_k$. There is no standard term for a matrix whose columns are orthonormal: We refer to a matrix whose columns are orthonormal as 'a matrix whose columns are orthonormal'. But a *square* matrix that satisfies $A^T A = I$ is called *orthogonal*; its columns are an orthonormal basis. Orthogonal matrices have many uses, and arise in many applications.

We have already encountered some orthogonal matrices, including identity matrices, 2-D reflections and rotations (page 129), and permutation matrices (page 132).

Norm, inner product, and angle properties. Suppose the columns of the $m \times n$ matrix A are orthonormal, and x and y are any n-vectors. We let $f : \mathbf{R}^n \to \mathbf{R}^m$ be the function that maps z to Az. Then we have the following:

- $\|Ax\| = \|x\|$. That is, f is *norm preserving*.
- $(Ax)^T(Ay) = x^T y$. f preserves the inner product between vectors.
- $\angle(Ax, Ay) = \angle(x, y)$. f also preserves angles between vectors.

Note that in each of the three equations above, the vectors appearing in the left- and right-hand sides have different dimensions, m on the left and n on the right.

We can verify these properties using simple matrix properties. We start with the second statement, that multiplication by A preserves the inner product. We have

$$\begin{aligned}
(Ax)^T(Ay) &= (x^T A^T)(Ay) \\
&= x^T(A^T A)y \\
&= x^T I y \\
&= x^T y.
\end{aligned}$$

In the first line, we use the transpose-of-product rule; in the second, we re-associate a product of 4 matrices (considering the row vector x^T and column vector x as matrices); in the third line, we use $A^TA = I$; and in the fourth line, we use $Iy = y$.

From the second property we can derive the first one: By taking $y = x$ we get $(Ax)^T(Ax) = x^Tx$; taking the squareroot of each side gives $\|Ax\| = \|x\|$. The third property, angle preservation, follows from the first two, since

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^Ty}{\|x\|\|y\|}\right) = \angle(x, y).$$

QR factorization. We can express the result of the Gram–Schmidt algorithm described in §5.4 in a compact form using matrices. Let A be an $n \times k$ matrix with linearly independent columns $a_1, \ldots, a_k$. By the independence-dimension inequality, A is tall or square. Let Q be the $n \times k$ matrix with columns $q_1, \ldots, q_k$, the orthonormal vectors produced by the Gram–Schmidt algorithm applied to the n-vectors $a_1, \ldots, a_k$. Orthonormality of $q_1, \ldots, q_k$ is expressed in matrix form as $Q^TQ = I$. We express the equation relating a_i and q_i,

$$a_i = (q_1^Ta_i)q_1 + \cdots + (q_{i-1}^Ta_i)q_{i-1} + \|\tilde{q}_i\|q_i,$$

where $\tilde{q}_i$ is the vector obtained in the first step of the Gram–Schmidt algorithm, as

$$a_i = R_{1i}q_1 + \cdots + R_{ii}q_i,$$

where $R_{ij} = q_i^Ta_j$ for $i < j$ and $R_{ii} = \|\tilde{q}_i\|$. Defining $R_{ij} = 0$ for $i > j$, we can write the equations above in compact matrix form as

$$A = QR.$$

This is called the *QR factorization* of A, since it expresses the matrix A as a product of two matrices, Q and R. The $n \times k$ matrix Q has orthonormal columns, and the $k \times k$ matrix R is upper triangular, with positive diagonal elements. If A is square, with linearly independent columns, then Q is orthogonal and the QR factorization expresses A as a product of two square matrices.

The attributes of the matrices Q and R in the QR factorization come directly from the Gram–Schmidt algorithm. The equation $Q^TQ = I$ follows from the orthonormality of the vectors $q_1, \ldots, q_k$. The matrix R is upper triangular because each vector a_i is a linear combination of $q_1, \ldots, q_i$.

The Gram–Schmidt algorithm is not the only algorithm for QR factorization. Several other QR factorization algorithms exist, that are more reliable in the presence of round-off errors. (These QR factorization methods may also change the *order* in which the columns of A are processed.)

Sparse QR factorization. There are algorithms for QR factorization that efficiently handle the case when the matrix A is sparse. In this case the matrix Q is stored in a special format that requires much less memory than if it were stored as a generic $n \times k$ matrix, *i.e.*, nk numbers. The flop count for these sparse QR factorizations is also much smaller than $2nk^2$.