Math 642:550 — Summer 2003
MTTh 6:15–8:45 PM Hill 525
Prof. Bumby

Supplement 8: Schur's unitary triangularization theorem.

1. Introduction

Section 5.6 of the text attempts to explain the significance of the **similarity** between matrices A and B given by the equation

$$B = M^{-1}AM, \tag{1}$$

but appears to lose sight of this goal as it collects diverse theorems about finding special matrices similar to a given one. We saw early in the chapter that if the columns of M form a **basis of eigenvectors** for A, then B, given by (1), is diagonal, but not all matrices have a basis of eigenvectors.

There are two things to explain: first, the role of M as a change of basis; second, the benefits that come from changing the basis. We will describe the first in order to have a proper setting for the second, and then select **Schur's Unitary Triangularization Theorem** as an illustration of the second. Although many applications only need that every matrix is similar, over the complex numbers, to a triangular matrix, there is little extra work in obtaining the stronger result and it leads to a useful characterization of **normal matrices**..

A real orthogonal matrix represents a **rigid** change of coordinates. Unitary matrices are the appropriate analog when the vector spaces must be considered over the complex numbers.

Change of Basis. When m by n matrices are first introduced, they are used to describe linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e., as concrete things that will later be included in the general theory of vector spaces. When the generality is introduced, it is not always made clear that one works with it by producing very specific matrices that express things described abstractly. To see how this is done, let us look more closely at how matrices express linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The entries in a column used to write an element of $\mathbb{R}^n$ are the coefficients in the unique expression of that element in terms of the standard basis. If we have a linear transformation into $\mathbb{R}^m$, the image of this vector is the same combination of the images of the standard basis. The images of each of the n vectors in the standard basis of $\mathbb{R}^n$, like any vector in $\mathbb{R}^m$ is written as a column of m numbers. The matrix representing the linear transformation is the matrix with these columns, and **the definition of matrix multiplication models the way in which the image of a general vector is found from the image of the basis vectors**.

If you want to do the same thing for abstract vector spaces, the first thing to do is to choose bases for the spaces. This allows us to use the coefficients of the representation of a vector in terms of the chosen basis in exactly the same way as the standard basis is used in $\mathbb{R}^m$.

Row operations introduce **new bases on the codomain** while retaining **the same basis on the domain**. This means that features of the domain, like the nullspace, continue to be described in terms of the same coordinates. However, row operations cause features of the codomain, like the actual image of the function, to have descriptions in different bases. In terms of the standard basis, this image is the **column space** of the given matrix. The change of coordinates given by the reduction to echelon form identifies a set of vectors that can be used as a basis for this space, but one must go back to the original matrix to find the descriptions of these vectors in terms of the standard basis.

So far, so good. Unfortunately, at some point, you may want to think of $\mathbb{R}^m$ as an abstract vector space and choose a basis for it other than the standard basis. The intrinsic properties of a linear transformation should not depend on the basis that was chosen, so this leads to various equivalences between matrices, but there is some difficulty describing all this in words.

In an attempt to clarify matters, a new level of abstract structure will be introduced: a **Vector Space with Basis** (VSwB). Making the basis part of the structure allows us to distinguish two ways of describing the same space. If the basis is different, then we have different VSwB's even if we feel that we have the same vector spaces. In particular, a **change of basis** amounts to a composition with the matrix representing the **identity linear transformation** from the space with one basis to the **same space** with a **different basis**.

If you have one basis of a space, then to obtain the columns of a matrix representing the identity map to that space with a second basis, it is necessary to express the vectors in the **first** basis in terms of the **second**. If this information is not directly available, it is still possible to characterize it as the solution of a system of linear equations. In particular, if you have a matrix giving the **second** basis in terms of the **first**, the change of coordinates is expressed by the inverse of that matrix.

Since eigenvectors are characterized as vectors whose image is parallel to itself, it is necessary to be able to **compare** vectors with their images. Thus, this question only makes sense for mappings of a space to **itself** — not just to a space of the same dimension, but to **the same space**. In particular, any change of basis must be applied to domain and codomain at the same time.

If, for example, you discover a basis of eigenvectors for the matrix A and want a matrix B expressing the same linear transformation with respect to this basis, you need three steps: (1) perform the identity transformation from the space with the eigenvector basis to the space with the standard basis; (2) apply the given transformation, using its description in the standard basis; (3) perform the identity transformation from the space with the standard basis to the space with the eigenvector basis. These matrices are written **right-to-left**, since they represent functions applied to **column** vectors. Our description of the relation between linear transformations and matrices tells us that matrix (1) has columns that describe the eigenvectors in terms of the standard basis. This is M in equation (1). Matrix (2) is the given matrix A. Finally, the columns of matrix (3) must express the standard basis in terms of the eigenvector basis. We have seen that this described M^{-1}. This is equation (1).

2. Block Multiplication

If we have $C = AB$, the (i, j) entry of C is the product of row i of A with column j of B. If the rows of A and/or the columns of B are partitioned in some way, there is a **corresponding partition** of the rows and/or columns of C and a given (i, j) describes a location in a particular block of C based on the same interpretation of i with respect to the rows of A and j with respect to the columns of B.

It is also possible to interpret a **consistent** partition of columns of A **and** rows of B. This corresponds to partitioning a basis of the space that is simultaneously the domain of A and the codomain of B. The parts of each column of B represent vectors in subspaces of this intermediate space found by **projecting** the vector represented by the column into those subspaces. The whole vector can be reconstructed as the **sum** of those projections. Hence the application of A to the whole vector will be the sum of the applications to the projections. The application to a projection is given by selecting columns of A since **the subspace has been given a basis that is a subset of the basis of the intermediate space**. This shows that the columns of C are the sum of the products of the parts of A and parts of B obtained from our consistent partitions of columns of A and rows of B. The resulting expression **looks like** matrix multiplication of the arrays of blocks, with the product of blocks given by ordinary matrix multiplication. This is useful in proofs where a partition of bases is more significant that the individual basis vectors.

3. Schur's Theorem

Here is the statement of the theorem.

Schur's Unitary Triangularization Theorem. *Given an n by n matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$ in any prescribed order, there is a unitary n by n matrix U such that $T = U^{\mathrm{H}}AU$ is upper triangular and the diagonal elements $t_{ii} = \lambda_i$. Furthermore, if the entries of A and its eigenvalues are all real, U may be chosen to be real orthogonal.*

Note: There is no claim that the matrices U and T are unique. Indeed, the proof is constructive, and we will see that there will be many choices in following its steps.

If the characteristic polynomial of A has multiple roots, the λ_i should contain each zero of the characteristic polynomial as many times as the factor $(\lambda - \lambda_i)$ appears in the characteristic polynomial. These equal values may appear anywhere in the list of λ_i, but the total number of appearances (called the **algebraic multiplicity**) of the eigenvalue is fixed by the characteristic polynomial of A.

The dimension of the nullspace of $A - \lambda I$ (called the **geometric multiplicity**) of the eigenvalue will not play a role in the proof. At each stage, we need **only one** eigenvector for the eigenvalue being considered.

A triangular form shows the eigenvalues and allows simple determination of eigenvectors, so it is not surprising that such information about A will be needed in the construction of U. This theorem gives no special insight into the computation of eigenvalues; its main value is as a link to theoretical results.

Although Schur's theorem appears as result 5R in the textbook, the treatment here follows Roger A. Horn & Charles R. Johnson, *Matrix Analysis*, Cambridge, 1985. That book, with its companion, **Topics in Matrix Analysis**, provide a good source for more details on topics introduced in this course, along with some more advanced topics. The treatment is more theoretical than our textbook, but still down-to-earth. The book also has many references, some remarkably recent. Schur's theorem appears in section 2.3, with consequences filling the next several sections. Since U is unitary, $U^{-1} = U^{\mathrm{H}}$, and we use the latter expression in formulating the theorem. The proof will be by induction on n.

4. The basis for the induction

If $n = 1$, the matrix A looks just like a scalar, and that scalar is its only eigenvalue, so it is already in triangular form. Thus, $T = A$ and $U = I$ satisfies the conditions of the theorem. While T is necessarily unique in this case, U could be any complex number of absolute value 1 (or ± 1 in the real case).

5. Householder matrices

Before giving the induction step, we describe the special matrices used in that step. The construction will concentrate on producing U. The given information will lead to identifying the first column of U, which must then be extended to the whole matrix. The approach taken in elementary courses to construct a unitary matrix whose first column has a given direction is the **Gram-Schmidt method**: form a matrix M consisting of the column you want followed by n more columns forming a basis of $\mathbb{C}^n$ (you can use the standard basis if you like), giving an n by $n+1$ matrix of rank n whose first column is the given vector (if you use the standard basis, the remaining n columns form an identity matrix). Then, construct a QR factorization of this matrix (where Q is a **unitary** matrix, since we are working with complex matrices and using a hermitian inner product). In any factorization of M, the first factor will have n rows and the second factor will have $n+1$ columns. In the form of the QR factorization that we will construct, Q will be a square matrix with n columns, so R must have n rows, and R has the same shape as the matrix being factored. This differs from the usual form of the Gram-Schmidt process because the columns of the matrix being factored are **not** linearly independent, so we must **extract a basis**. We do this as part of the Gram-Schmidt process instead of using a separate computation.

We move through M finding the unique expression of each column as a linear combination of previous columns plus a vector orthogonal (in the Hermitian sense) to the previous columns. If this orthogonal vector is not zero, write it as the next column of Q, and form a column of R such that the product of Q with this column gives the expression for the column of M as a linear combination of the columns of Q that have already been written, with zeros in the lower part of the column to force later columns of Q to be ignored. If the orthogonal vector is the zero vector, do not change Q. However, you must still form a column of R expressing the current column of M in terms of the columns of Q that you already have, since R has the same shape as M. Since the columns of Q will be a basis for the column space of M at the end of the process and M was constructed to have all of $\mathbb{R}^n$ as its column space, we must find n independent vectors to form the columns of Q. If M is a real matrix, Q can be constructed to be a real matrix. So far, we have not been concerned about the lengths of the vectors forming the columns of Q, except to make sure that the length isn't zero. However, to arrive at a unitary matrix, we need to modify Q so that all columns have length 1. To divide each column by its length, we form the matrix QD^{-1}, where D is the matrix whose diagonal entries are the lengths of the columns of Q in order, and whose off-diagonal entries are zero. Then $QR = (QD^{-1})(DR)$ and $(QD^{-1})^{\mathrm{H}}(QD^{-1}) = I$, so that (QD^{-1}) is a unitary matrix. Note that DR is obtained from our original R by multiplying the rows by the length of the vector in corresponding column of the original Q. Also note that the normalization process forces the columns of QD^{-1} to be unit vectors.

Programmers often say, "First make it work, then make it fast". Now that we are sure that there are unitary matrices with any given unit vector as first column, **we can try to write one without doing so much work**. This can be used both as an alternate to the Gram-Schmidt method for finding any QR factorization, not just the one used in the construction proving Schur's theorem. A geometric interpretation of the desired matrix is that it gives a rigid motion of $\mathbb{C}^n$ (or $\mathbb{R}^n$ in the real case) that takes the first vector in the standard basis into a unit vector that has the same direction as the vector that forms the first column of the given matrix. In the real case, an easy way to do this is to reflect in the $(n-1)$-dimensional space that is the perpendicular bisector of the segment joining these points (the heads of the vectors with tails at the origin if you think of vectors as arrows). Our intuition suggests that this is always possible if the vectors have the same length, and we will give a proof. The complex case will have an extra complication that we will give after describing the form of the matrix we seek. Take e_1 as the first member of the **standard basis** and v as the desired **unit vector**, let $u = v - e_1$. If $u = 0$, there is noting to do, so the identity is our unitary matrix. Otherwise, let

$$H_u = I - 2\frac{uu^{\mathrm{H}}}{u^{\mathrm{H}}u},$$

where the numerator of the last term is an n by n matrix and the denominator is a scalar that is the square of the length. Direct calculation shows that $H_u^{\mathrm{H}} = H_u$ and $H_u^2 = I$, so H_u is simultaneous unitary and Hermitian. It remains to show that v can be chosen so that $H_u e_1 = v$ (and, since $H_u^2 = I$, ${}_u v = e_1$).

Since H_u is unitary, we have, for all vectors w,

$$(H_u w)^{\mathrm{H}}(H_u w) = w^{\mathrm{H}} H_u^{\mathrm{H}} H_u w = w^{\mathrm{H}} w.$$

That is, H_u preserves length. Thus, we will need to choose v to be a unit vector. In the real case, this suffices, but there is an extra condition in the complex Hermitian case.

Since H_u is Hermitian, we have, for all vectors w,

$$\begin{aligned}\left(w^{\mathrm{H}}(H_u w)\right)^{\mathrm{H}} &= (w^{\mathrm{H}} H_u^{\mathrm{H}})w \\ &= w^{\mathrm{H}} H_u w \\ &= w^{\mathrm{H}}(H_u w).\end{aligned}$$

That is, $\langle w, H_u w\rangle$ is real. Thus a vector can be taken to the first vector in the standard basis only if it has length 1 and a real first entry.

Conversely, if v is such a vector and $u = v - e_1$,

$$\begin{aligned}u^{\mathrm{H}}u &= v^{\mathrm{H}}v - v^{\mathrm{H}}e_1 - e_1^{\mathrm{H}}v + e_1^{\mathrm{H}}e_1 \\ &= 2 - 2v^{\mathrm{H}}e_1,\end{aligned}$$

since the terms at the end are both 1 and the middle terms are equal.

In practice, there will be minor modifications of this description. Given any convenient vector v in the direction we want (for Schur's theorem, it will be an eigenvector of λ_1), the above description calls for us to first scale it by multiplying by the **complex conjugate** of the first entry, and then divide the resulting vector by its length. Alternatively, we can scale e_1 by a factor **equal to** the first entry of v (in each case, this step should be skipped if the first entry of v is zero) and then multiply each vector by the length of the other. There is no need for any concern about the length of u sind H_u has been defined to depend only on the direction of u.

From this, it is easy to see that $H_u e_1 = v$ and $H_u v = e_1$.

Matrices of the form H_u are called **Householder matrices**, except by Householder, who called them "elementary matrices" although that term is now used for **different matrices** — the matrices expressing the simplest row operations in Gaussian elimination.

If the vector v is a rational vector with unit length, then $u = v - e_1$ is a rational vector. The matrix H_u depends only on the direction of u, so we can scale this vector to have integer entries while computing H_u. On the other hand, if v has a rational direction, but is not a rational **vector**, then the computation of H_u will involve irrational quantities. Its first column will be the irrational unit vector in the direction of v.

6. Example

Lets use Householder's construction to produce a real symmetric matrix whose first column is in the direction of $(1, 1, 1)$. This vector has length $\sqrt{3}$, so we take $u = (1 - \sqrt{3}, 1, 1)$. Then $u^{\mathrm{T}}u = (1-\sqrt{3})^2 + 1^2 + 1^2 = 6 - 2\sqrt{3}$, so $2/(u^{\mathrm{T}}u) = 1/(3-\sqrt{3})$. Thus,

$$\begin{aligned}H_u &= \frac{1}{3-\sqrt{3}}\left(\begin{bmatrix} 3-\sqrt{3} & 0 & 0 \\ 0 & 3-\sqrt{3} & 0 \\ 0 & 0 & 3-\sqrt{3}\end{bmatrix} - \begin{bmatrix} 4-2\sqrt{3} & 1-\sqrt{3} & 1-\sqrt{3} \\ 1-\sqrt{3} & 1 & 1 \\ 1-\sqrt{3} & 1 & 1\end{bmatrix}\right) \\ &= \frac{3+\sqrt{3}}{6}\begin{bmatrix} 1+\sqrt{3} & -1+\sqrt{3} & -1+\sqrt{3} \\ -1+\sqrt{3} & 2-\sqrt{3} & -1 \\ -1+\sqrt{3} & -1 & 2-\sqrt{3}\end{bmatrix} \\ &= \frac{1}{6}\begin{bmatrix} 2\sqrt{3} & 2\sqrt{3} & 2\sqrt{3} \\ 2\sqrt{3} & 3-\sqrt{3} & -3-\sqrt{3} \\ 2\sqrt{3} & -3-\sqrt{3} & 3-\sqrt{3}\end{bmatrix}\end{aligned}$$

7. The inductive step

To return to the proof of Schur's theorem, consider λ_1. Since this is an eigenvalue, it must have at least one eigenvector. Let v be an eigenvector for λ_1, normalized to have length 1 and real first entry. Now, we take U_1 to be a unitary matrix whose first column is v. The first column of AU_1 is $\lambda_1 v$, so if A_1 is defined by $AU_1 = U_1 A_1$, the first column of A_1 gives the expression of $\lambda_1 v$ in terms of a basis formed by the columns of U_1. However, this expression must be λ_1 times v plus 0 times each other vector in the basis. The first column of A_1 thus has λ_1 in the first position and zero everywhere else.

All matrices in this step of the construction will be described using blocks formed by separating the first row and and column from everything else.

Since we only want to find a **triangular** matrix equivalent to A (or A_1), we can ignore the rest of the first row of A_1. Rows and columns 2 through n describe a mapping from the subspace V_1 orthogonal to v to itself. Restricting the original linear transformation to this subspace gives a linear transformation whose matrix, using the rest of the image of the standard basis under U_1, can be taken to be the result of dropping the first column of A_1. Here, the columns of U_1 except the first are a basis for the domain, but all columns of U_1 are a basis for the codomain. To ignore the first row as well means that we follow this linear transformation by the projection that sends v to zero and is the identity on V_1.

The induction hypothesis applies to the composite linear transformation of V_1 to itself, represented by rows 2 through n of A_1, described above. The $n-1$ by $n-1$ U and T matrices expressing this result are extended to n by n matrices by putting an appropriate quantity — 1 for U and λ_1 for A and T — in the $(1,1)$ position and zero elsewhere in the first row and column — a process we will call **bordering**. This applies a block structure on n by n matrices in which the first row is separated from all remaining rows and the first column is separated from all remaining columns. If you multiply two such block matrices, the result has the block structure, as described in the **Block** section. Denote these bordered matrices by A_-, U_-, and T_-. Then $A_- U_- = U_- T_-$.

The λ_i were defined to be the roots of the characteristic polynomial of A with multiple roots counted according to their multiplicities. As part of the induction, we note that A and A_1 have the same characteristic polynomial because they are similar. The characteristic polynomial of A_1 can be calculated by expanding $\det(A_1 - \lambda I)$ by its first column. This gives $(\lambda_1 - \lambda)$ times the characteristic polynomial of the matrix in rows and columns 2 through n of A_1. This is independent of the first row of A_1. The eigenvalues of this block are thus the same as those of A with the multiplicity of λ_1 reduced by 1. (This holds for **any** partial triangularization. For example, you can use a basis extending an independent set of known eigenvectors to give an **immediate** proof that the algebraic multiplicity is no smaller than the geometric multiplicity.)

To recover A_1 from A_- add a matrix T_0 that is zero except for its upper right block, which is taken from A_1, so $A_1 = A_- + T_0$. Now, introduce $T_1 = T_0 U_-$. This is also zero except for its upper right block, which is the product of the upper right block of T_1 with the $n-1$ by $n-1$ matrix playing the role of U. Since U_- was formed by bordering with 1, $U_- T_1 = T_1$. Hence, $T_0 U_- = U_- T_1$ and $A_1 U_- = U_-(T_- + T_1)$. Let $T = T_- + T_1$ and $U = U_1 U_-$. Then

$$AU = AU_1 U_- = U_1 A_1 U_- = U_1 U_- T = UT.$$

If A has real entries and real eigenvalues, we can construct real eigenvectors, allowing this construction to be done using real matrices. A unitary matrix with real entries is called an **orthogonal** matrix.

With the completion of the inductive step, the proof of Schur's theorem is finished. If you are uncomfortable with the abstract form of this inductive proof, you should work an example. Take $n = 3$ and let A be an n by n matrix whose eigenvalues you know. Begin the induction step. This will introduce a 2 by 2 matrix to triangularize. Begin the inductive step for this matrix. You now must triangularize a 1 by 1 matrix. This is done by the **basis** of the induction. You can now complete the inductive step for the 2 by 2 matrix. The result of that is used to complete the inductive step for the 3 by 3 matrix, which gives the desired factorization of the original matrix.

8. A classical result

The construction of the unitary change of basis in Schur's theorem can also be used to write a **given** unitary matrix as a **product of reflections**. As in Schur's theorem, the matrix is reduced to the identity one column at a time. Since an n by n matrix has n columns, at most n steps are needed. This shows that an orthogonal transformation on $\mathbb{R}^n$ is a product of at most n reflections. Since reflections have determinant -1, the number of factors will be even or odd depending on whether the determinant of the given transformation is $+1$ or -1. In particular, in $\mathbb{R}^3$ an orthogonal transformation of determinant $+1$ must be the product of 0 or 2 reflections. The former case gives the identity, while the latter case shows that the intersection of the fixed planes of the reflections will give a fixed line. If the transformation is not the identity, it will have a fixed line, which requires it to be a rotation.

9. Multiple eigenvalues

If the eigenvalue taken as λ_1 is a multiple eigenvalue of A, it will also be an eigenvalue of the lower right block of A_1. If we take an eigenvector of this block corresponding to λ_1, and consider it as a vector v^* of the n dimensional space V, the component of $(A - \lambda_1 I)v^*$ in V_1 will be zero, which means that, in V, it is a multiple of the original eigenvector v. Because v is an eigenvector for **the same** eigenvalue, this multiple is the same for all vectors $v^* + cv$. If v^* is an eigenvector in the whole space, the multiple will be zero, but we have examples where λ_1 is a multiple eigenvalue with no eigenvectors other than the multiples of v. The vectors v^* found in this way are sometimes called **generalized eigenvectors**. They can be used to produce a space on which some power of $A - \lambda I$ is zero whose dimension is the algebraic multiplicity of λ. As with ordinary eigenvalues, elements of such spaces for different eigenvalues are linearly independent, so a basis can be found consisting of eigenvectors and generalized eigenvectors.

Our computation of the characteristic polynomial of an n by n matrix A was described using computations with matrices, but a change of basis simply conjugates **all** matrices appearing in the computations. That is, the operations could have been described abstractly in terms of vector spaces and linear transformations. However, in most cases, a basis would need to be chosen in order to be able to compute the coefficients of the characteristic polynomial. Schur's Theorem allows us to assume that all matrices appearing in this calculation are upper triangular.

If zero is an eigenvalue of multiplicity k, we may also assume that the first k entries on the diagonal are zero. The other diagonal entries are the remaining eigenvalues, all of which are nonzero. In the calculation of the characteristic polynomial, all coefficients after step $n - k$ are zero, although the matrices arising may be nonzero. The matrix obtained (in the bottom row of our tabular description) at the end of this **first phase** of the computation is a polynomial in A of degree $n - k$ with nonzero constant term, so the **first** k places on the diagonal contain this nonzero constant.

The rules for generating these matrices shows that each remaining step **only** involves multiplying by A. For the part of the matrix below row k, this has the effect of multiplying by a nonsingular matrix. Since we end up with a zero matrix, these rows must be zero at the start of this **second phase**. Thus, the matrix obtained at the end of the first phase has rank k. The columns of this matrix span the nullspace of A^k, so their span includes the nullspace of A, which are the eigenvectors of A for $\lambda = 0$. Action of A takes this subspace to a subspace of itself, so all the matrices in the second phase of the computation have column spaces contained in the nullspace of A^k. The last nonzero matrix in the computation will have columns that belong to the nullspace of A. However, the whole nullspace of A may not survive that long. The matrix

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

provides a simple example.

If A is not a triangular matrix, this remarkable behavior will not be visible in the calculation. However, **those properties that depend on the linear transformation rather than the basis will continue to hold.** In particular, the matrix appearing at the end of the first phase will have rank k and its column space will be the nullspace of A^k. Such vectors are often called **generalized eigenvectors**. Every matrix gives rise to a basis of generalized eigenvectors. Additional conditions may be placed on this basis to get the **Jordan canonical form** described in **Appendix B** of the textbook.

10. Normal matrices

Result **5U** of the text characterizes matrices that can be diagonalized by unitary matrices. The treatment is very brief with key results given only as exercises, but the result is important, so it is appropriate to elaborate on it.

First, **normal matrices** are defined to be complex matrices N that commute with their conjugate-transpose, i.e., $NN^{\mathrm{H}} = N^{\mathrm{H}}N$. This class includes Hermitian, skew-Hermitian and unitary matrices. This definition can be tested by matrix multiplication, and so is computationally easy.

A matrix that can be diagonalized by a unitary matrix is necessarily normal, since $M = UDU^{-1} = UDU^{\mathrm{H}}$ implies

$$\begin{aligned} MM^{\mathrm{H}} &= UDU^{\mathrm{H}}UD^{\mathrm{H}}U^{\mathrm{H}} = UDD^{\mathrm{H}}U^{\mathrm{H}} \\ M^{\mathrm{H}}M &= UD^{\mathrm{H}}U^{\mathrm{H}}UDU^{\mathrm{H}} = UD^{\mathrm{H}}DU^{\mathrm{H}} \end{aligned}$$

and these quantities are equal since diagonal matrices always commute. This allows one to construct normal matrices whose eigenvalues are arbitrary complex numbers, while the special types that we enumerated have strong limitations on their eigenvalues.

Similarly, if N is normal and $T = U^{-1}NU = U^{\mathrm{H}}NU$ with U unitary, then $TT^{\mathrm{H}} = T^{\mathrm{H}}T$. Schur's theorem says that U can be found for which T is upper triangular. If matrices that were both normal and upper triangular could only be diagonal, then Schur's theorem would actually diagonalize normal matrices. **This is exactly what happens**.

The proof can be organized in the same way as the induction argument used in the proof of Schur's theorem. The result clearly holds for 1×1 matrices, so we turn to the induction step.. We begin by looking at the $(1, 1)$ entry of $TT^{\mathrm{H}} = T^{\mathrm{H}}T$. The left side is the square of the length of the first row of T, while the right side is the square of the length of the first column. Since T is upper triangular, the first column has a simple form: its first entry is t_{11}, and the others are zero. Thus, its length is $|t_{11}|$. The first row is more complicated: its first entry is t_{11}, as before, but nothing is known about the other entries. However, the **length** of a vector has some nice properties: its square is the sum of nonnegative quantities, one of which is $|t_{11}|^2$. This means that the first row and first column can have the same length only if all entries in the first row other than the first are zero. The matrix T is then formed by bordering an $(n-1)\times(n-1)$ matrix T_1, that is easily seen to be normal and triangular if T is.

11. Real symmetric matrices

Since real symmetric matrices M have real eigenvalues, and the eigenvectors corresponding to real eigenvalues can always be chosen to have real entries, all of the steps of Schur's theorem can be done using real matrices, so there is a real orthogonal matrix such that $S^{-1}MS$ is triangular and symmetric, hence diagonal. Conversely, if M is a real matrix that can be diagonalized by a real orthogonal matrix, the diagonal matrix $S^{-1}MS$ must have all real entries. However, those entries are just the eigenvalues of M.

Other real normal matrices, like the orthogonal matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

can be diagonalized by a unitary matrix, but that matrix cannot be real. For this matrix, the columns of the diagonalizing matrix U must be proportional to $[1\ i]^{\mathrm{T}}$ and $[1\ -i]^{\mathrm{T}}$. For U to be unitary, these columns must be multiplied by complex numbers of norm $\sqrt{2}$. Any such multipliers can be used, and they can be chosen independently for the each column.

12. Exercises

1. Let

$$A = \begin{bmatrix} 1 & 0 & 5 & -3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Find AB. Use of **block matrices** with rows and columns partitioned as $\{1,2\}\cup\{3,4\}$ should be useful.

2. Find a real Householder matrix whose first column is

$$\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

3. Find a complex Householder matrix whose first column is

$$\begin{bmatrix} 2/3 \\ 2/3 + 1/3\,i \end{bmatrix}.$$

4. Use the fact that

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

is an eigenvector of

$$\begin{bmatrix} 1 & -9 \\ -3 & 7 \end{bmatrix}$$

to find a Schur form of this matrix and both eigenvalues. Don't bother finding the other eigenvector, although it will be easy to find at almost any stage of working with this matrix.

5. For the matrix

$$N = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

the first phase for the computation of the characteristic polynomial take **zero** steps, so the space of **generalized eigenvectors** is the column space of the identity matrix — all of $\mathbb{R}^3$. What are the eigenvectors of N? Which belong to the column space of the last nonzero matrix in the computation?