PRIMUS

Problems, Resources, and Issues in Mathematics Undergraduate Studies

ISSN: 1051-1970 (Print) 1935-4053 (Online) Journal homepage: http://www.tandfonline.com/loi/upri20

Advanced Linear Algebra: A Call for the Early Introduction of Complex Numbers

Stephan Ramon Garcia

To cite this article: Stephan Ramon Garcia (2017) Advanced Linear Algebra: A Call for the Early Introduction of Complex Numbers, PRIMUS, 27:8-9, 856-868, DOI: 10.1080/10511970.2016.1235644

To link to this article: https://doi.org/10.1080/10511970.2016.1235644

Published online: 12 Dec 2016.

PRIMUS, 27(8–9): 856–868, 2017

ISSN: 1051-1970 print / 1935-4053 online
DOI: 10.1080/10511970.2016.1235644

Advanced Linear Algebra: A Call for the Early Introduction of Complex Numbers

Stephan Ramon Garcia

Abstract: A second course in linear algebra that goes beyond the traditional lower-level curriculum is increasingly important for students of the mathematical sciences. Although many applications involve only real numbers, a solid understanding of complex arithmetic often sheds significant light. Many instructors are unaware of the opportunities afforded by the early introduction of complex arithmetic. Most elementary properties of complex numbers have immediate matrix analogues and many important theorems can be deduced, or at least postulated, from the basics of complex arithmetic alone.

Keywords: Linear algebra, complex arithmetic, SVD, polar decomposition, positive semidefinite, unitary, self-adjoint, normal, Hermitian

1. INTRODUCTION

A second course in linear algebra, which goes beyond the traditional lower-level curriculum (e.g., linear systems, row reduction, determinants, eigenvalues), is increasingly important for students of the mathematical sciences. Not only in pure mathematics, but also in applied mathematics, physics, computer science, engineering, operations research, and statistics, a fluency with more advanced matrix methods (e.g., QR decomposition, singular value decomposition, discrete Fourier transforms) is required. In a big-data world, students must be prepared to face large matrices and the procedures required to deal with them.

Although many applications involve only real numbers, a solid understanding of complex arithmetic often sheds significant light. Some instructors are unaware of

Address correspondence to Stephan Ramon Garcia, Department of Mathematics, Pomona College, 610 N College Ave, Claremont, CA 91711, USA. E-mail: stephan.garcia@pomona.edu; http://pages.pomona.edu/~sg064747/.

Color versions of one or more figures in the article can be found online at www.tandfonline.com/upri.

the opportunities afforded by the early and vigorous introduction of complex arithmetic. Most elementary properties of complex numbers have immediate matrix analogues. Moreover, many important theorems can be deduced, or at least postulated, from the basics of complex arithmetic alone.

Ideally, students will enter a second linear algebra course having already mastered complex arithmetic. If this is not possible, the instructor should spend time at the beginning of the course on the subject. Students should be fluent with complex arithmetic and the geometry of the complex plane. No knowledge of complex *analysis*, however, is required.

This philosophy pertains to both advanced undergraduate courses and introductory graduate courses in the subject. We focus here on material that is more typical of a second course in linear algebra. A discussion of the standard, lower-level material would lead us too far astray and into well-trodden territory. For instance, there is no need to discuss the role of complex numbers in the consideration of second order, constant coefficient differential equations. Similarly, we do not consider sophisticated topics that straddle the border with functional analysis. For example, we do not mention Banach spaces or Fourier analysis.

In what follows, $\mathrm{M}_{m\times n}(\mathbb{F})$ denotes the set of all $m \times n$ matrices with entries in the field $\mathbb{F}$; we consider only $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$. If $m = n$, then we write $\mathrm{M}_n(\mathbb{F})$ instead. We use capital letters to denote matrices and lower-case letters to denote complex numbers. Since the lower-case letters z, w are somewhat standard notation for generic complex numbers, in order to hammer home some analogies, we often employ the decidedly non-standard letters Z, W to denote matrices.

2. THE ADJOINT AND TRANSPOSE

The *adjoint* of a matrix $Z = [z_{i,j}] \in \mathrm{M}_{m\times n}(\mathbb{C})$ is the conjugate transpose $Z^* = [\overline{z_{j,i}}] \in \mathrm{M}_{n\times m}(\mathbb{C})$ of Z.[1] In physics, $Z^\dagger$ is used instead of Z^*. Here the term “adjoint” encompass both the complex case (conjugate transpose) and the real case (transpose).

The transpose is commonly introduced in lower-level courses with little, if any, motivation. Students are required to memorize properties such as $(A^T)^{-1} = (A^{-1})^T$ and $(A^T)^T = A$ with no payoff; the punchline is often missing. Consequently, they are often confused by the transpose and the adjoint. What is the point of it? What is so special about “flipping” a matrix with respect to the main diagonal?

Students should be strongly encouraged to recognize that the adjoint is a higher-dimensional generalization of complex conjugation. For instance, if $Z = [z]$ is a 1×1 complex matrix, then $Z^* = [\overline{z}]$. Apart from the matrix

[1] The term *adjoint* is occasionally used to denote the transpose of the matrix of cofactors of a given matrix, although the much-preferred terms *adjugate* and *classical adjoint* have gained wide acceptance.

brackets, the adjoint and complex conjugation are fundamentally identical in this case.

The complex number $z = a + bi$, in which $a, b \in \mathbb{R}$, is often represented using the real matrix

$$M_z = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

The matrix that represents $\bar{z} = a - bi$ is $M_z^T = M_z^*$. This is another clear reminder of the relationship between the adjoint and conjugation.

Each formula involving complex conjugation has a matrix analogue. For instance, $\overline{z + w} = \bar{z} + \bar{w}$ suggests that $(Z + W)^* = Z^* + W^*$. Similarly, $(\bar{z})^{-1} = \overline{z^{-1}}$ for $z \neq 0$ suggests that $(Z^*)^{-1} = (Z^{-1})$ for invertible Z. The formula $\overline{zw} = \bar{z}\bar{w}$ requires only a little more imagination to generalize: $(ZW)^* = W^* Z^*$. Since there is a similar identity for matrix inversion, students are rarely surprised to see the order reversed.

The analogy between the adjoint and complex conjugation can be pushed much further, as many important classes of matrices are defined in terms of the adjoint. We consider this perspective in the following several sections.

3. UNITARY MATRICES

A *unitary* matrix is an invertible square matrix Z that satisfies $Z^{-1} = Z^*$. Motivation for considering complex unitary matrices, as opposed to focusing solely on real orthogonal matrices (i.e., real square matrices for which $Z^{-1} = Z^T$), can be found throughout engineering and physics. For instance, the bustling field of quantum information theory is built upon the arithmetic of complex unitary matrices. In many computational sciences, the discrete Fourier transform (DFT) plays a central role. It is the linear transformation on $\mathbb{C}^n$ induced by the unitary matrix

$$\frac{1}{\sqrt{n}} \left[\omega^{(j+1)(k+1)}\right]_{j,k=1}^{n},$$

in which $\omega = \exp(-2\pi i/n)$. Various fast implementations of it (which go by the name of "fast Fourier transforms") arise in signal processing and even in large-integer arithmetic. Needless to say, the student must have a firm grasp of complex arithmetic before working with the DFT, as its very definition involves nth roots of unity.

The structural parallels between complex numbers and unitary matrices provide an opportunity to encourage students to discover the properties of unitary matrices. The equation $z^{-1} = \bar{z}$ implies that $\bar{z}z = 1$, so that $|z| = 1$. This suggests that unitary matrices should be seen as matrix analogues of complex numbers of unit modulus. What does this analogy suggest?

If $|z| = 1$, then

$$|zw|^2 = zw\overline{zw} = \bar{z}zw\bar{w} = w\bar{w} = |w|^2.$$

Consequently, $|zw| = |w|$, so multiplication by a unimodular constant preserves lengths; that is, it is an *isometry.* Almost the same proof works for unitary matrices

$$\| Z\mathbf{w} \|^2 = \langle Z\mathbf{w}, Z\mathbf{w}\rangle = \langle Z^*Z\mathbf{w}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{w}\rangle = \| \mathbf{w} \|^2.$$

Much more is true. Multiplication by a unimodular constant is a rigid motion of $\mathbb{C}$; it preserves both lengths and angles. The appropriate generalization is $\langle Z\mathbf{x}, Z\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$, in which $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Conversely, the complex-linear isometries of $\mathbb{C}$ that fix $\mathbf{0}$ are precisely the multiplications by unimodular constants. The complex-linear isometries of $\mathbb{C}^n$ that fix $\mathbf{0}$ are precisely the $n \times n$ unitary matrices.

4. SELF-ADJOINT MATRICES

A *self-adjoint* (or *Hermitian*) matrix is a square matrix Z that satisfies $Z = Z^*$. The fundamental axioms of quantum mechanics require the self-adjointness of the underlying Hamiltonian operator. At the undergraduate level, finite-dimensional "toy" physics problems are often posed using self-adjoint matrices. The adjacency matrices and graph Laplacians that arise in network theory are all self-adjoint. Positive semidefinite matrices, a special subclass of self-adjoint matrices, arise all the time in statistical applications (see Section 7).

The condition $Z = Z^*$ is analogous to the equation $z = \bar{z}$, which characterizes real numbers. Consequently, we expect that self-adjoint matrices enjoy certain properties reminiscent of real numbers; they should be analogous to "higher-dimensional real numbers." For example, a self-adjoint matrix has only *real* eigenvalues.

The proof of the following theorem is a trivial modification of the corresponding decomposition $z = a + ib$ for complex numbers. Here $a = \operatorname{Re} z = \frac{1}{2}(z + \bar{z})$ and $b = \operatorname{Im} z = \frac{1}{2i}(z - \bar{z})$.

Theorem (Cartesian decomposition). *If $Z \in \mathrm{M}_n(\mathbb{C})$, then there exist unique self-adjoint $A, B \in \mathrm{M}_n(\mathbb{C})$ so that $Z = A + iB$. Moreover,*

$$A = \frac{1}{2}(Z + Z^*) \quad \textit{and} \quad B = \frac{1}{2i}(Z - Z^*).$$

For students who will only ever deal with real matrices, the complex perspective still leads to the following purely “real” result.

Theorem. *If* $Z \in \mathrm{M}_n(\mathbb{R})$, *then there exist a unique symmetric* $A \in \mathrm{M}_n(\mathbb{R})$ *and a unique skew-symmetric* $B \in \mathrm{M}_n(\mathbb{R})$ *so that* $Z = A + B$. *Moreover,*

$$A = \frac{1}{2}(Z + Z^T) \quad \textit{and} \quad B = \frac{1}{2}(Z - Z^T).$$

The sum of two real numbers is real. In a similar manner, the sum of two self-adjoint matrices is self-adjoint. Multiplication is more troublesome; this is to be expected since one of the major ways in which matrix arithmetic differs from complex arithmetic is the noncommutativity of multiplication. However, if one “balances” things appropriately, there are many ways to multiplicatively combine two self-adjoint matrices; the texts [1, 4] contain a wealth of information on the topic.

5. THE SPECTRAL THEOREM

A *normal* matrix is a square matrix Z that commutes with its adjoint: $Z^*Z = ZZ^*$. Many standard and familiar classes of matrices are normal. For instance, unitary, self-adjoint, and positive semidefinite matrices are all normal (positive semidefinite matrices and their properties will be discussed in Section 7).

Most of the standard algebraic properties of a complex number z can be expressed, in one way or another, in terms of z and $\bar{z}$ alone. For instance, we have

$$\mathrm{Re}\, z = \frac{1}{2}(z + \bar{z}), \quad \mathrm{Im}\, z = \frac{1}{2i}(z - \bar{z}), \quad |z| = (\bar{z}z)^{1/2}, \quad \frac{1}{z} = \frac{\bar{z}}{|z|^2},$$

and so forth. A matrix that commutes with its adjoint behaves as much like a complex number as we can reasonably expect from a matrix. That is the content of the Spectral Theorem: up to an orthonormal change of basis, a normal matrix is just a diagonal matrix (i.e., a direct sum of complex numbers) in disguise.

Theorem (Spectral Theorem). *Let* $Z \in \mathrm{M}_n(\mathbb{C})$ *be normal. Then there exist a unitary* $U \in \mathrm{M}_n(\mathbb{C})$ *and diagonal* $\Lambda \in \mathrm{M}_n(\mathbb{C})$ *so that* $Z = U\Lambda U^*$.

The decomposition $Z = U\Lambda U^*$ above is called the *spectral decomposition* of Z. Various important classes of normal matrices are characterized entirely by their eigenvalues and are easily treated using the Spectral Theorem. For instance:

- if Z is normal and has only real eigenvalues, then Z is self-adjoint;
- if Z is normal and has only eigenvalues of unit modulus, then Z is unitary;
- if Z is normal and has eigenvalues only in $[0, \infty)$ (resp., $(0, \infty)$), then Z is positive semidefinite (resp., positive definite);
- if Z is normal and has eigenvalues only in $\{0, 1\}$, then Z is an orthogonal projection.

The power of the Spectral Theorem truly emerges only when it is combined with polynomial algebra. If

$$p(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_0$$

is a complex polynomial, then we let

$$p(Z) = c_n Z^n + c_{n-1} Z^{n-1} + \cdots + c_0 I.$$

If $Z = U\Lambda U^*$ is the spectral decomposition of a normal matrix Z, then

$$p(Z) = Up(\Lambda)U^*, \tag{1}$$

for any complex polynomial p. This *polynomial functional calculus* permits us to define functions of a normal (or, more generally, diagonalizable) matrix in a simple, unified manner. For instance, if p is a polynomial for which $p(\lambda) = e^{\lambda}$ for each eigenvalue λ of a diagonalizable matrix Z, then $p(Z) = \exp Z$ (the existence of such a polynomial is justified by the Lagrange Interpolation Theorem; see Section 6). Consequently, matrix exponentials can be approached entirely through polynomial interpolation; infinite series of matrices (and the associated hand-waving) are unnecessary.

6. EIGENVALUES AND POLYNOMIALS

Even though matrices that arise in the "real world" often contain only real entries, complex numbers are essential to their study. A "typical" real, square matrix has many complex eigenvalues; see Figure 1. As another example, complex eigenvalues are expected to arise in the consideration of "typical" discrete-time Markov chains, which have many states. The theory of random matrices can make all of these assertions precise [2, 5].

Since eigenvalues and polynomials go hand-in-hand, fluency with the basic properties of complex polynomials yields benefits in the study of matrices and their eigenvalues. The following theorem should be in every mathematician's arsenal.

Theorem (Lagrange Interpolation). *Let $n \geq 1$, let $z_1, z_2, \ldots, z_n$ be distinct complex numbers, and let $w_1, w_2, \ldots, w_n \in \mathbb{C}$. There is a unique polynomial*

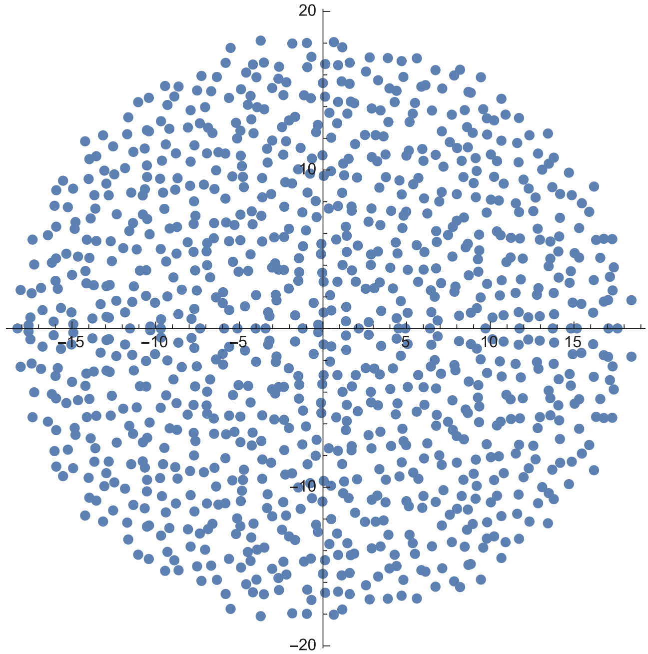

Figure 1. Plot in the complex plane of the eigenvalues of a randomly generated 1000×1000 real matrix. The matrix entries are drawn independently from the uniform distribution on $[-1, 1]$. Readers who are interested in an explanation for the distinctive appearance of this plot are invited to consult [2, 5].

p of degree at most $n-1$ such that $p(z_i) = w_i$ for $i = 1, 2, \ldots, n$. If the data $z_1, z_2, \ldots, z_n, w_1, w_2, \ldots, w_n$ are real, then p is a real polynomial.

This basic fact about complex polynomials is phenomenally useful in linear algebra. Moreover, this relationship is reciprocal because the best proofs of the interpolation theorem involve linear algebra! One proof employs a convenient basis for the n-dimensional space of polynomials of degree at most $n-1$. For $j = 1, 2, \ldots, n$, the polynomials

$$\ell_j(z) = \prod_{\substack{1 \leq k \leq n \\ k \neq j}} \frac{z - z_k}{z_j - z_k}$$

are of degree $n-1$ and satisfy $\ell_j(z_k) = \delta_{jk}$. Therefore, $p(z) = \sum_{j=1}^n w_j \ell_j(z)$ has degree at most $n-1$ and satisfies $p(z_k) = \sum_{j=1}^n w_j \ell_j(z_k) = w_k$. Another well-known approach can be based on the invertibility of the associated *Vandermonde matrix*

$$V_n = \begin{bmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^{n-1} \\ 1 & z_2 & z_2^2 & \cdots & z_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & z_n^2 & \cdots & z_n^{n-1} \end{bmatrix}.$$

A nice consequence of the Lagrange Interpolation Theorem is the following seminal result that underpins diagonalization.

Theorem. *Suppose that* $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ *are eigenvectors of* $Z \in \mathrm{M}_n(\mathbb{C})$ *corresponding to distinct eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_r$. *Then the vectors* $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ *are linearly independent.*

The proof is elegant and simple. If $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_r\mathbf{w}_r = 0$, then for any polynomial p

$$\begin{aligned} 0 &= p(Z)(c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_r\mathbf{w}_r) \\ &= c_1 p(Z)\mathbf{w}_1 + c_2 p(Z)\mathbf{w}_2 + \cdots + c_r p(Z)\mathbf{w}_r \\ &= c_1 p(\lambda_1)\mathbf{w}_1 + c_2 p(\lambda_2)\mathbf{w}_2 + \cdots + c_r p(\lambda_r)\mathbf{w}_r. \end{aligned} \tag{2}$$

For $i = 1, 2, \ldots, r$, select polynomials p_i so that $p_i(\lambda_j) = \delta_{i,j}$ for $j = 1, 2, \ldots, r$. Now substitute $p = p_i$ into equation (2) to see that each $c_i = 0$.

This polynomial interpolation technique can also be used to prove the existence of a positive semidefinite square root of a positive semidefinite matrix; see Section 7. Before proceeding, we cannot resist a brief discussion about Geršgorin's Theorem, which elegantly combines linear algebra and complex arithmetic. Further information can be found in [4]; a treatise on the theorem and its generalizations is [6].

Theorem (Geršgorin's Theorem). *If* $n \geq 2$, *then every eigenvalue of* $A = [a_{ij}] \in \mathrm{M}_n$ *is contained in*

$$G(A) = \bigcup_{k=1}^{n} G_k(A),$$

where $G_k(A) = \{z \in \mathbb{C} : |z - a_{kk}| \leq R'_k(A)\}$ *denotes the closed disk in* $\mathbb{C}$ *whose center is at the point* a_{kk} *and whose radius is given by the deleted absolute row sum* $R'_k(A) = \sum_{j \neq k} |a_{kj}|$.

The proof of Geršgorin's Theorem uses only the definition of eigenvalues and complex arithmetic. However, its power is stunning. For example, consider the matrix

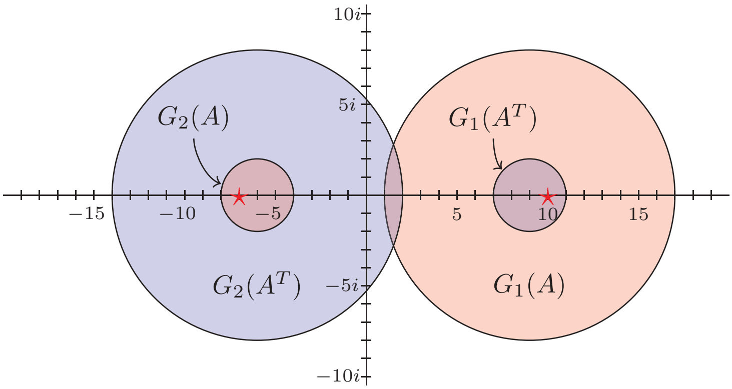

Figure 2. The intersection of $G(A)$ and $G(A^T)$, which is the union of the two small disks and the lens-shaped region, contains the eigenvalues $z = 10$ and $z = -7$ of A.

$$A = \begin{bmatrix} 9 & 8 \\ 2 & -6 \end{bmatrix},$$

whose eigenvalues are -7 and 10. We have $R_1'(A) = 8$, $R_2'(A) = 2$, $R_1'(A^T) = 2$, $R_2'(A^T) = 8$, and hence

$$G_1(A) = \{z \in \mathbb{C} : |z - 9| \leq 8\}, \ G_1(A^T) = \{z \in \mathbb{C} : |z - 9| \leq 2\},$$
$$G_2(A) = \{z \in \mathbb{C} : |z + 6| \leq 2\}, \ G_2(A^T) = \{z \in \mathbb{C} : |z + 6| \leq 8\}.$$

Since A and A^T always have the same eigenvalues, the eigenvalues of A lie in $G(A) \cap G(A^T)$; see Figure 2. This example is from [3].

7. POSITIVE SEMIDEFINITE MATRICES

We say that $Z \in \mathrm{M}_n(\mathbb{C})$ is *positive semidefinite* (resp., *positive definite*) if $\langle Z\mathbf{w}, \mathbf{w}\rangle \geq 0$ (resp., $\langle Z\mathbf{w}, \mathbf{w}\rangle > 0$) for all $\mathbf{w} \in \mathbb{C}^n$. Statisticians deal with positive semidefinite matrices, such as covariance and correlation matrices, all the time. Orthogonal projections, which are useful in approximation theory, are examples of positive semidefinite matrices. These matrices are characterized by the two conditions $Z = Z^*$ (self-adjointness) and $Z^2 = Z$ (idempotence).

What are positive semidefinite matrices an analogue of? For a complex number z, asserting that $zw\bar{w} \geq 0$ for all $w \in \mathbb{C}$ is another way of saying that $z \geq 0$; that is, z is non-negative. This suggests that positive semidefinite matrices should enjoy properties that are reminiscent of non-negative real numbers. For instance, a complex number z is non-negative if and only if $z = p^2$ for some non-negative p. The same is true for matrices.

Theorem. *If $Z \in \mathrm{M}_n(\mathbb{C})$ is positive semidefinite, then there exists a positive semidefinite matrix $P \in \mathrm{M}_n(\mathbb{C})$ so that $Z = P^2$.*

The matrix P is uniquely determined and it is usually denoted $Z^{1/2}$. In fact, there exists a real polynomial p so that $P = p(Z)$; simply let $p(\lambda) = \sqrt{\lambda}$ for each (necessarily real) eigenvalue λ of Z. The existence of such a polynomial is guaranteed by the Lagrange Interpolation Theorem; that $P = p(Z)$ is positive semidefinite and squares to Z follows easily from the polynomial functional calculus (1).

For any complex number z, we have $\bar{z}z \geq 0$. Entirely analogous to this is the fact that for any $Z \in \mathrm{M}_{m\times n}(\mathbb{C})$, the matrix $Z^*Z \in \mathrm{M}_n(\mathbb{C})$ is positive semidefinite

$$\langle Z^*Z\mathbf{w}, \mathbf{w}\rangle = \langle Z\mathbf{w}, Z\mathbf{w}\rangle = \|Z\mathbf{w}\|^2 \geq 0.$$

As such, we may define $|Z| \in \mathrm{M}_n(\mathbb{C})$ to be the positive semidefinite square root $(Z^*Z)^{1/2}$ of Z^*Z. This opens the door to matrix generalizations of the polar decomposition of a complex number.

For the sake of simplicity, we now restrict our attention solely to square matrices. Both the polar decomposition and the singular value decomposition, which we discuss below, can be suitably generalized to the non-square case. Although these generalizations are important in many applications, they are more cumbersome to write down, and doing so would detract from our present course. Our main interest is demonstrating that complex arithmetic can be our guide through many of the crucial topics in a second linear algebra course.

Theorem (Polar decomposition). *Let $Z \in \mathrm{M}_n(\mathbb{C})$. There exists a unitary matrix $U \in \mathrm{M}_n(\mathbb{C})$ so that $Z = U|Z|$. If Z is invertible, then U is uniquely determined.*

From the complex perspective, the proof of the preceding theorem is (almost) trivial. If $z \neq 0$, we may write $z = u|z|$, in which $u \in \mathbb{T}$. So $u = z|z|^{-1}$. Now suppose that $Z \in \mathrm{M}_n(\mathbb{C})$ is invertible. Then the matrix $U = Z|Z|^{-1}$ is unitary since

$$U^*U = (Z|Z|^{-1})^*(Z|Z|^{-1}) = |Z|^{-1}|Z|^2|Z|^{-1} = I,$$

because $Z^*Z = |Z|^2$. Consequently, $Z = U|Z|$ as claimed. Of course, there are details to fill in if Z is not invertible. The point, however, is not to give a complete proof of the most general theorem possible. We simply wish to show that complex arithmetic suggests, in an intuitive manner, some of the fundamental results in advanced linear algebra.

As another example, consider the formula

$$z^{-1} = \bar{z}|z|^{-2}, \tag{3}$$

in which $z \in \mathbb{C}$ is nonzero. The same formula holds for an invertible $Z \in \mathrm{M}_n(\mathbb{C})$, with a minor adjustment to take noncommutativity into account:

$$Z^{-1} = |Z|^{-1} Z^* |Z|^{-1}. \tag{4}$$

To see why, write the polar decomposition of Z as $Z = U|Z|$, in which U is unitary. Taking inverses yields $Z^* = |Z|^{-1}U^*$; taking adjoints yields $Z^* = |Z|U^*$. Combining these last two equations leads to equation (4).

Perhaps the most important matrix decomposition, from the perspective of applications, is the singular value decomposition (SVD), which we state below in the square case.

Theorem (SVD). *Let $Z \in \mathrm{M}_n(\mathbb{C})$. Then there exist unitary matrices $U, V \in \mathrm{M}_n(\mathbb{C})$ and a diagonal matrix Σ with non-negative entries so that $Z = U\Sigma V^*$.*

Here is how to pass from the polar decomposition to the SVD. Write $Z = WP$, in which W is unitary and P is positive semidefinite. The Spectral Theorem provides a unitary V and a diagonal matrix Σ with non-negative entries so that $P = V\Sigma V^*$. Thus, $Z = U\Sigma V^*$, in which $U = WV$ and V are unitary.

Unlike the polar decomposition, the positive semidefinite matrix involved is *diagonal* in the SVD. In order to accomplish this, we must pay a price. Instead of a single unitary matrix, we must employ two unitary matrices. In light of the noncommutativity of matrix multiplication, this is natural. Indeed, we have already seen a similar "two-sided" phenomenon in equation (4), in which a simple formula (3) from complex arithmetic required a slight "two-sided" adjustment.

8. NON-NORMAL MATRICES

As we have seen, normal matrices are the "best" matrices in the sense that their properties can largely be anticipated from elementary complex arithmetic. Fruitful analogies abound in their study and they hold few surprises for the attentive student. To a large extent, the Spectral Theorem reduces the study of normal matrices to that of diagonal matrices; questions about diagonal matrices boil down to simple arithmetic.

Diagonalizable matrices are already familiar to students, who are accustomed to diagonalizing matrices from a first course in linear algebra. What about non-diagonalizable matrices? Those must be grappled with at some point in a second linear algebra course. For instance, the $n \times n$ Jordan matrix J_n satisfies $J_n^n = 0$ and $J_n^k \neq 0$ for $k = 1, 2, \ldots, n-1$; that is, J_n is *nilpotent of order n*. Nilpotency does not occur in the arithmetic with which students are normally familiar. It certainly has no parallel in complex arithmetic.

Those who have followed along with our general philosophy should not be surprised. What our analogies tell us is that *normal* matrices are the most well-behaved matrices; their properties tend to mirror those of complex numbers. The farther one strays from the realm of normal matrices (or, more generally, diagonalizable matrices), the more pathological behavior one expects to find.

Fortunately, the matrices encountered most often in applications are normal. These include unitary matrices in quantum information theory, self-adjoint matrices in network theory, positive semidefinite matrices in statistics, orthogonal projections in approximation theory, and so forth. Students can safely swim in the shallow end at first, studying normal matrices and their applications for a while (armed with intuition gained from the complex perspective), before moving on to general matrices.

A major tool in the (theoretical) study of arbitrary square matrices is the Jordan canonical form. This standard topic requires complex arithmetic from the outset, as the roots and multiplicities of the characteristic polynomial must first be known (in addition to other information). An instructor hoping to reach Jordan canonical form by the end of the course must “face the music” at some point. The complex numbers *must* be introduced somewhere; we argue that this should be done as early as possible.

9. CONCLUSIONS

For the modern student of the mathematical sciences and allied fields, the importance of a second course in linear algebra is undeniable. As we have demonstrated, a firm understanding of complex arithmetic and the geometry of the complex plane provide the student with a solid foundation upon which to learn advanced linear algebra. This approach is relevant both to undergraduate and graduate courses on the subject.

Properties of important special classes of matrices (e.g., unitary, self-adjoint, positive semidefinite) are almost self-evident from this perspective. Fluency with complex polynomials opens the door to simple proofs of major theorems, whereas comfort with complex geometry yields qualitative insights into the behavior of matrices.

All of this suggests that an early introduction to complex arithmetic, either prior to or at the beginning of a second course in linear algebra, is greatly beneficial to the student. The instructor will also benefit from this approach, since it provides a unifying theme and an established framework upon which to build.

FUNDING

Partially supported by National Science Foundation Grant DMS-1265973.

REFERENCES

1. Bhatia, R. 2007. *Positive Definite Matrices*. Princeton, NJ: Princeton University Press.
2. Deift, P. 2009. *Random Matrix Theory: Invariant Ensembles and Universality*. Providence, RI: American Mathematical Society.
3. Garcia, S. R. and R. A. Horn. 2016. *A Second Course in Linear Algebra*. New York: Cambridge University Press (in production).
4. Horn, R. A. and C. R. Johnson. 2012. *Matrix Analysis, Second Edition*. New York: Cambridge University Press.
5. Tao, T. 2012. *Topics in Random Matrix Theory*. Providence, RI: American Mathematical Society.
6. Varga, R. S. 2004. *Gerschgorin and His Circles*. New York: Springer.

BIOGRAPHICAL SKETCH

Stephan Ramon Garcia grew up in San Jose, California before attending U.C. Berkeley for his B.A. and Ph.D. After graduating, he worked at U.C. Santa Barbara for several years before moving to Pomona College in 2006. He has earned three NSF research grants and five teaching awards from three different institutions. He was twice nominated by Pomona College for the prestigious US Professors of the Year Award. He is the author of two books and over 60 research articles in operator theory, complex analysis, matrix analysis, number theory, discrete geometry, and other fields.