On some definitions in matrix algebra

Jan R. Magnus

Department of Econometrics & Data Science, Vrije Universiteit Amsterdam,
and Tinbergen Institute, Amsterdam, The Netherlands

November 6, 2023

AMS Subject Classifications: 1501, 15A57, 15A63.

Key words: Definitions, symmetry, orthogonality, square root, Hermitian and quadratic forms, matrix derivatives.

Abstract: Many definitions in matrix algebra are not standardized. This note discusses pitfalls associated with some definitions, and deals with central concepts like symmetry, orthogonality, square root, Hermitian and quadratic forms, and matrix derivatives.

Corresponding author: Jan R. Magnus, Department of Econometrics & Data Science, Vrije Universiteit Amsterdam, The Netherlands. E-mail: jan@janmagnus.nl.

1 Introduction

Science works bottom-up, not top-down: we develop theory in a simple framework (real variables, univariate), then in a more general framework (complex variables, multivariate). When we generalize from $n = 1$ to $n > 1$, for example when we generalize the univariate normal distribution to the multivariate normal distribution, an obvious minimum requirement is that when we set $n = 1$ in the multivariate definition we obtain the univariate case. But this minimum requirement is not sufficient to define the generalization. What we need to do is map the 'essence' of the univariate case to the multivariate case.

In many (perhaps most) cases there is no controversy as to what constitutes the essence. The multivariate normal distribution is a specific generalization of the univariate normal distribution, and it is generally accepted. Similarly, when we move from scalar derivatives ($f(x) = x^2$ implies $\mathrm{D}f(x) = 2x$) to vector derivatives ($f(x) = Ax$ implies $\mathrm{D}f(x) = A$), then again all textbooks agree.

But this is not always the case. For example, when we move from vector derivatives to matrix derivatives, then many generalizations exist, but there is only one which maintains the essence of the concept of 'derivative', as I tried to point out in Magnus (2010).

Some men of genius, such as Euclid or Newton, anticipate generalizations. In the *Elements* (c. 300 BCE), Euclid introduces five unprovable but intuitive principles known as axioms, but he only introduces each axiom when he actually needs it, thus implicitly revealing an ordering of the axioms based on how intuitive he thinks the axiom is. The fifth axiom, the so-called parallel postulate, states that 'for any line L and point p not on L there exists a unique line through p not meeting L.' Euclid must have felt that this axiom was the least intuitive, but it was not until the second half of the 19th century that Gauss and others began to experiment with this postulate, eventually arriving at non-Euclidean geometries.

In a similar vein, Newton's second law of motion (late 17th century) is usually written as $F = ma$: force equals mass times acceleration. Since $a = dv/dt$ where v denotes velocity, we can also write $F = m\, dv/dt$. But Newton writes $F = d(mv)/dt$, even though m is assumed constant, thus anticipating Einstein by allowing the possibility that m also depends on t.

Euclid and Newton are, however, exceptions. It is much more common that when we generalize we discover that we have to make adjustments. A typical example is the Heine–Borel theorem in real analysis (late 19th century), which states that a set S in the Euclidean space $\mathbb{R}^n$ is closed and bounded if and only if every open cover of S has a finite subcover. If either of these conditions holds then we say that S is 'compact'. Since the concepts 'closed' and 'bounded' are easier and more transparent than covers and subcovers, most textbooks define S to be compact when it is closed and bounded, and then prove that every open cover of S has a finite subcover. However, when we move to more general topological spaces, then the Heine–Borel theorem is no longer true, and we realize that we made the wrong choice: compactness should have been defined in terms of covers and subcovers.

Many definitions in matrix algebra, even of such central concepts as symmetry and orthogonality, are not standardized. Thus, some authors define a symmetric matrix to be one that satisfies $A' = A$, where A may be real or complex; others (as we do) require that A is real. In the former case, only the properties of *real* symmetric matrices follow from those of Hermitian matrices. Similarly, some authors define an orthogonal matrix as a square matrix satisfying $A'A = I$, irrespective of whether A is real or complex.

The purpose of this note is to point out a number of these deviations, to propose what we believe are the 'right' definitions, and to highlight the dangers of not following these definitions. We aim for a common-sense viewpoint without idiosyncracies. In generalizing a concept we try and preserve its essential characteristics. To specialize we simply give a name to an important subclass. For example, in generalizing a symmetric matrix from real to complex, the essential characteristics are only preserved by a Hermitian matrix. In specializing, we call a real Hermitian matrix symmetric. The class of complex matrices satisfying $A' = A$ does not fit in this common-sense view. It is neither the right generalization (because the essential characteristics of real symmetric matrices are lost) nor is it the right specialization (because this class is not a subclass of the Hermitian matrices and therefore does not share its properties).

We write a complex number z as $z = a + ib$, where a and b are real. The complex conjugate of z is $z^* = a - ib$. For a complex matrix $U = A + iB$, the conjugate transpose is $U^* = A' - iB'$, where we notice the transpose in the generalization. For real matrices, the conjugate transpose is the transpose. Some care is required because the transpose enjoys some properties which the conjugate transpose does not have. In particular, while $\operatorname{tr}(A') = \operatorname{tr}(A)$ for real matrices, it is not generally true that $\operatorname{tr}(U^*) = \operatorname{tr}(U)$. Also, while $x'A'x = x'Ax$ for real A and x, it is not generally true that $z^*U^*z = z^*Uz$.

The following five definitions are without controversy. A square matrix U is said to be:

Hermitian if $U^* = U$,
skew-Hermitian if $U^* = -U$,

unitary if $U^*U = I$,
normal if $U^*U = UU^*$, and
idempotent if $UU = U$.

If the matrix happens to be real, we call the first three matrices symmetric, skew-symmetric, and orthogonal, respectively, while the last two matrices continue to be called normal and idempotent. A symmetric matrix is simply a real Hermitian matrix, and all properties of Hermitian matrices apply to symmetric matrices. Although real idempotent matrices are often symmetric, one should *not* require that an idempotent matrix is Hermitian, and hence a real idempotent matrix is not necessarily symmetric.

In Sections 2 and 3 we discuss symmetry and orthogonality, respectively, and present examples of why these concepts should only apply to real matrices. In Section 4 we define the square root of a matrix. In Section 5 we discuss Hermitian and quadratic forms. A lot of unnecessary confusion still exists about the definition of matrix derivatives, and this confusion is simply resolved in Section 6. Section 7 concludes.

2 Symmetry

Any matrix A satisfying $A^* = A$ is called *Hermitian*. Such a matrix is necessarily square. A real Hermitian matrix is called *symmetric*. Hence, a symmetric matrix is a *real* square matrix satisfying $A' = A$.

Many authors define A to be symmetric when it satisfies the property $A' = A$, irrespective of whether A is real or complex. This is undesirable as we shall argue below. If we wish to give a name to complex matrices satisfying $A' = A$ we could call them *complex-symmetric*. Symmetric matrices are Hermitian and therefore share all properties of Hermitian matrices — an important and useful fact — but complex-symmetric matrices are not Hermitian and do not necessarily share these properties.

To demonstrate our point, consider the matrix

$$A = \begin{pmatrix} 1 & i \\ i & \alpha \end{pmatrix},$$

where α is real. This is a complex matrix satisfying $A' = A$, hence not Hermitian. The eigenvalues of A are

$$\lambda_{1,2} = \frac{\alpha + 1 \pm \sqrt{(\alpha+1)(\alpha-3)}}{2},$$

and both eigenvalues are complex for $-1 < \alpha < 3$. When $\alpha = -1$, both eigenvalues are 0 and the rank is $r(A) = 1$. Hence the rank of A is not necessarily equal to the number of its nonzero eigenvalues. Finally, the matrix A is not normal (it does not satisfy $A^*A = AA^*$) unless $\alpha = 1$. Since a matrix can be diagonalized if and only if it is normal (Abadir and Magnus, 2005, Exercise 7.71), the matrix A cannot be diagonalized unless $\alpha = 1$.

We conclude that a complex-symmetric matrix need not have real eigenvalues, that its rank is not necessarily equal to the number of its nonzero eigenvalues, and that is not necessarily possible to diagonalize the matrix. Basically, none of the attractive properties of symmetric matrices holds for complex-symmetric matrices. It seems better, therefore, to let all symmetric matrices be real by definition.

3 Orthogonality

Any matrix A for which $A^*A = AA^* = I$ is said to be *unitary*. Alternatively we can define A to be unitary if it is square and satisfies $A^*A = I$ (or $AA^* = I$). A real unitary matrix is called *orthogonal*. Hence, an orthogonal matrix is a *real* square matrix satisfying $A'A = I$ (or $AA' = I$).[1]

There also exist complex square matrices satisfying $A'A = I$. We may call such matrices *complex-orthogonal*. Many authors call any square matrix satisfying $A'A = I$ orthogonal which can lead to errors and is better avoided. Consider, for example, the complex matrix

$$A = \begin{pmatrix} \alpha & i \\ -i & \alpha \end{pmatrix},$$

where α is real, and notice that

$$A^*A = \begin{pmatrix} \alpha & i \\ -i & \alpha \end{pmatrix} \begin{pmatrix} \alpha & i \\ -i & \alpha \end{pmatrix} = \begin{pmatrix} \alpha^2+1 & 2i\alpha \\ -2i\alpha & \alpha^2+1 \end{pmatrix}$$

and

$$A'A = \begin{pmatrix} \alpha & -i \\ i & \alpha \end{pmatrix} \begin{pmatrix} \alpha & i \\ -i & \alpha \end{pmatrix} = \begin{pmatrix} \alpha^2-1 & 0 \\ 0 & \alpha^2-1 \end{pmatrix}.$$

Hence, A is unitary if and only if $\alpha = 0$, while A is complex-orthogonal if and only if $\alpha = \pm\sqrt{2}$.

The inner product of the two columns of A is given by

$$\begin{pmatrix} \alpha \\ -i \end{pmatrix}^* \begin{pmatrix} i \\ \alpha \end{pmatrix} = (\alpha, i) \begin{pmatrix} i \\ \alpha \end{pmatrix} = 2i\alpha,$$

[1]Of course, an orthogonal matrix should have been named 'orthonormal' instead, because the columns are not merely orthogonal to each other — they are also normalized. But the word seems too embedded in matrix language to change it now.

and the two columns are therefore orthogonal to each other (the inner product is zero) if and only if $\alpha = 0$, that is, if and only if A is unitary. Hence, the columns of a complex-orthogonal matrix are not necessarily orthogonal to each other.

Next let $\alpha = \sqrt{2}$, so that A is complex-orthogonal, and consider its eigenvalues. The determinant of A is one; in fact the determinant of any complex-orthogonal matrix equals ± 1. But the eigenvalues of A are $\sqrt{2} \pm 1$, and hence the eigenvalues of a complex-orthogonal matrix do not in general have modulus one.

We conclude that a complex-orthogonal matrix does not necessarily have columns that are orthogonal to each other, and need not have eigenvalues with modulus one. Basically, none of the attractive properties of orthogonal matrices holds for complex-orthogonal matrices. To call complex-orthogonal matrices 'orthogonal' is therefore misleading.

One further remark on orthogonal transformations. Let B be a real $m \times n$ matrix of full column rank n. Then $A = B'B$ is positive definite and symmetric, and we can decompose

$$A = S\Lambda S',$$

where Λ is diagonal with strictly positive elements and S is orthogonal. Suppose that our calculations would be much simplified if A were equal to the identity matrix. We can achieve this by transforming B to a matrix C, as follows:

$$C = BS\Lambda^{-1/2}T',$$

where T is an arbitrary orthogonal matrix. Then,

$$C'C = T\Lambda^{-1/2}S'B'BS\Lambda^{-1/2}T' = T\Lambda^{-1/2}\Lambda\Lambda^{-1/2}T' = TT' = I_n.$$

The matrix T is completely arbitrary, as long as it is orthogonal. It is tempting to choose $T = I_n$. This, however, implies that if $B = B(t)$ is a continuous function of some variable t, then $C = C(t)$ is *not* necessarily continuous; see De Luca et al (2018, Appendix B) and Magnus and Neudecker (2019, Section 8.8). There is only one choice of T that leads to continuity of C, namely $T = S$, in which case

$$C = BS\Lambda^{-1/2}S' = B(B'B)^{-1/2}.$$

4 Square root

In high-school algebra we are told that $\sqrt{4} = 2$ and not ± 2 even though $(-2)^2 = 4$. Formally, if $\mathbb{R}_+$ denotes the set of nonnegative real numbers, then we define the square root as the single-valued function defined on $\mathbb{R}_+$ with values in $\mathbb{R}_+$. This is the most common (but not the only) definition and it extends straightforwardly to positive semidefinite matrices.

Assume first that $A = \operatorname{diag}(a_{11}, \ldots, a_{nn})$ is diagonal with nonnegative diagonal elements. Then there is a unique matrix $A^{1/2}$ compatible with our one-dimensional definition, namely $A^{1/2} = \operatorname{diag}(\sqrt{a_{11}}, \ldots, \sqrt{a_{nn}})$. Next consider a symmetric (hence real) matrix A. This matrix can be diagonalized so that we can write $A = S\Lambda S'$, where S is orthogonal and Λ is diagonal. In accordance with the usual definition of matrix functions for symmetric matrices (Abadir and Magnus, 2005, Chapter 9), we define $A^{1/2} = S\Lambda^{1/2}S'$. But $\Lambda^{1/2}$ is only defined when Λ has nonnegative diagonal elements. Hence, for any positive semidefinite matrix A there exists a *unique* positive semidefinite matrix B such that $B^2 = A$. This unique matrix is the square root of A. In particular, the often-seen statement 'any matrix B satisfying $B^2 = A$ is called a square root of A' is not correct.

Extending the definition, we can also define a unique square root for nonsymmetric real matrices, but only if all eigenvalues are positive. This extension is more complicated and involves the Jordan decomposition; see Abadir and Magnus (2005, Section 7.5) and Abadir (2012). If one of the eigenvalues is zero then no square root may exist. For example, when

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

then no matrix B exists such that $B^2 = A$.

Extending the definition in a different direction, we can include real $\lambda < 0$ or complex λ. In fact, the definition can be extended to all complex $\lambda \neq 0$ by analytic continuation of the binomial expansion. Care should however be taken on how to choose the principal value of the square roots. Both extensions are discussed in detail in Abadir (2012).

5 Quadratic forms and positive definiteness

If A is Hermitian and x is a conformable vector, then the expression x^*Ax is called a *Hermitian form*, and we have $(x^*Ax)^* = x^*Ax$ so that Hermitian forms are real. If A is symmetric (hence real) and x is also real, then $x'Ax$ is called a *quadratic form*. The symmetry of A is thus implicit in the definition of a quadratic form.

Most authors, however, do not include symmetry in the definition of a quadratic form. Perhaps they find it counterintuitive to say that an expression like

$$(x_1, x_2) \begin{pmatrix} 1 & -1 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is not a quadratic form. In practice, this is not a problem, because the matrix A can always be taken to be symmetric, due to the fact that

$$x'Ax = x' \left(\frac{A + A'}{2} \right) x.$$

(Note that this trick does not work for complex matrices, because $x^*Ax \neq x^*A^*x$, in general.) For example,

$$(x_1, x_2) \begin{pmatrix} 1 & -1 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 4x_1x_2 + 9x_2^2.$$

Because of this fact, the function $x'Ax$ is often called a quadratic form, even when the matrix A is not symmetric. Perhaps it is best to say that a quadratic form $x'Ax$ is understood to imply that A is symmetric unless stated explicitly otherwise.

A real positive (semi)definite matrix, however, is *always* symmetric. This is because a Hermitian matrix is positive (semi)definite if and only if $x^*Ax > 0$ ($x^*Ax \geq 0$) and the requirement that A is Hermitian is essential — otherwise x^*Ax is not real for all x. For example, the matrix

$$A = \begin{pmatrix} 1 & 3 \\ -6 & 9 \end{pmatrix}$$

satisfies $x'Ax > 0$ for all $x \neq 0$, but it is not symmetric and its eigenvalues are not real. This is *not* a positive definite matrix.

6 Matrix derivatives

Finally, let me briefly summarize my remarks in Magnus (2010).[2] If f is an $m \times 1$ vector function of an $n \times 1$ vector x, then the derivative of f is the $m \times n$ matrix

$$\mathrm{D}f(x) = \frac{\partial f(x)}{\partial x'},$$

[2] Some years ago I looked at the entry 'Matrix calculus' in Wikipedia, and found that almost every statement was wrong. I made one attempt to correct this, but without success. I now looked again and found that many changes had been made, and they even referred to Magnus (2010) stating that '...this Wikipedia article has been nearly completely revised from the version criticized in this article.' Unfortunately, it is still wrong.

the elements of which are the partial derivatives $\partial f_i(x)/\partial x_j$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. There is no controversy about this definition. In the simplest case where $y = Ax$ and A is a matrix of constants, we have $\partial y/\partial x' = A$. In particular, for a scalar function $f(x)$, the derivative $\partial f(x)/\partial x'$ is a *row* vector, not a column vector.

The definition of matrix derivatives should be a generalization of the vector case. Consider an $m \times p$ matrix function F of an $n \times q$ matrix of variables X. The derivative is a matrix containing all $mpnq$ partial derivatives, but how are these partial derivatives organized? This can only be done in one way, namely by stacking the elements of the function F and by stacking the elements of the argument X. Since the vec-operator is the commonly used stacking operator, we use the vec-operator. Thus we define

$$\mathrm{D}F(X) = \frac{\partial \operatorname{vec} F(X)}{\partial (\operatorname{vec} X)'},$$

which is an $mp \times nq$ matrix. The definition of vector derivative is a special case of the more general definition of matrix derivative, as of course it should. The definition implies that, if F is a function of a scalar x ($n = q = 1$), then $\mathrm{D}F(x) = \partial \operatorname{vec} F(x)/\partial x$, an $mp \times 1$ column vector. Also, if f is a scalar function of a matrix X ($m = p = 1$), then $\mathrm{D}f(X) = \partial f(X)/\partial(\operatorname{vec} X)'$, a $1 \times nq$ row vector. The choice of ordering the partial derivatives is not arbitrary. For example, the derivative of the scalar function $f(X) = \operatorname{tr}(X)$ is not $\mathrm{D}f(X) = I_n$ (as is often stated), but $\mathrm{D}f(X) = (\operatorname{vec} I_n)'$.

To define matrix derivatives correctly is important, because a derivative is not just a collection of partial derivatives. In particular, we want to be able to use a chain rule, we want to interpret the rank of a derivative, and we want to use its determinant in transformation theorems. This is only possible with a correct definition of matrix derivative.

7 Conclusion

In this note we have discussed a number of problems in defining standard concepts in matrix algebra. Apart from the fact that it is undesirable that standard concepts like symmetry and orthogonality are defined differently by different authors, we argue that some definitions are unnatural and error-prone, and thus better avoided. Some other definitions are simply wrong. In summary, consider the matrices

$$A = \begin{pmatrix} 1 & i \\ i & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \sqrt{2} & i \\ -i & \sqrt{2} \end{pmatrix},$$

and notice that the matrix A is not symmetric even though $A' = A$, and that the matrix B is not orthogonal even though $B'B = I$. Next consider the matrices

$$C = \begin{pmatrix} 1 & 3 \\ -6 & 9 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

The matrix C satisfies $x'Cx > 0$ for all $x \neq 0$, but it is not symmetric and hence it is not positive definite. The matrix D satisfies $D^2 = I$ but is not a square root of I. The matrix E is idempotent even though it is not symmetric. Finally, consider the functions

$$F(x) = \begin{pmatrix} x & x^2 \\ x^2 & x^3 \end{pmatrix}, \qquad f(X) = \operatorname{tr}(X).$$

The derivative of F is not

$$\begin{pmatrix} 1 & 2x \\ 2x & 3x^2 \end{pmatrix}$$

and the derivative of f is not the identity matrix I. Instead,

$$DF(x) = \frac{d \operatorname{vec} F(x)}{dx} = \begin{pmatrix} 1 \\ 2x \\ 2x \\ 3x^2 \end{pmatrix}, \qquad Df(X) = \frac{\partial \operatorname{tr}(X)}{\partial (\operatorname{vec} X)'} = (\operatorname{vec} I)'.$$

These two expressions are perhaps less compelling but at least they are the correct expressions.

Acknowledgement

I am grateful to Karim Abadir for discussions on notation, definitions, and many other things that helped clarify my mind.

References

Abadir, K. M. (2012). The square root of a matrix, *Journal of Time Series Econometrics*, 4(2), Article 4.

Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*, Volume 1 of *Econometric Exercises* (Eds. K. M. Abadir, J. R. Magnus and P. C. B. Phillips), Cambridge University Press: New York.

De Luca, G., J. R. Magnus and F. Peracchi (2018). Weighted-average least squares estimation of generalized linear models, *Journal of Econometrics*, 204, 1–17.

Magnus, J. R. (2010). On the concept of matrix derivative, *Journal of Multivariate Analysis*, 101, 2200–2206.

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, third edition. John Wiley and Sons: Chichester/New York.

Wikipedia (2023), Matrix calculus, last edited 2 November 2023, accessed 6 November 2023.