CIRJE-F-476

On some definitions in matrix algebra

Jan R. Magnus
Tilburg University

Karim M. Abadir
Imperial College London

February 2007

On some definitions in matrix algebra

Jan R. Magnus
Department of Econometrics & OR, Tilburg University
and Department of Economics, University of Tokyo

Karim M. Abadir
Tanaka Business School, Imperial College London

February 21, 2007

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 some of the pitfalls associated with undesirable or wrong 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 & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-3092, fax: +31-13-466-3066, e-mail: magnus@uvt.nl.

1 Introduction

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 (correctly in our view) 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 too many idiosyncracies. To generalize a concept we try and preserve the essential characteristics of the concept. 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 u as $u = a + \mathrm{i}b$, where a and b are real. The complex conjugate of u is $u^* = a - \mathrm{i}b$. For a complex matrix $U = A + \mathrm{i}B$, the complex conjugate is $U^* = A' - \mathrm{i}B'$, where we notice the transpose in the generalization. For real matrices, the complex conjugate is the transpose. Some care is required because the transpose enjoys some properties which the complex conjugate does not have. In particular, while $\mathrm{tr}(A') = \mathrm{tr}(A)$ for real matrices, it is not generally true that $\mathrm{tr}(U^*) = \mathrm{tr}(U)$. 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, both for symmetric and nonsymmetric real matrices. In Section 5 we discuss Hermitian and quadratic forms. A lot of confusion still exists about the definition of matrix derivatives, and this is taken up in Section 6. Section 7 concludes.

2 Symmetry

Any matrix 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. We shall say that a complex matrix satisfying $A' = A$ is *complex-symmetric*. Symmetric matrices are Hermitian and therefore share all properties of Hermitian matrices — an important and useful fact. Symmetric matrices are also complex-symmetric and therefore share all properties of complex-symmetric matrices, but this of little use because complex-symmetric matrices hardly possess properties of interest.

To demonstrate our point, consider the matrix

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

where α is real. This is a complex matrix satisfying $A' = A$ and hence it is complex-symmetric. Its eigenvalues are found from the equation

$$(1 - \lambda)(\alpha - \lambda) = -1,$$

so that

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

The eigenvalues are both complex for $-1 < \alpha < 3$ and they are both real otherwise. Hence, the eigenvalues of a complex-symmetric matrix are not necessarily real.

The determinant is $|A| = \alpha + 1$ so that A is nonsingular unless $\alpha = -1$. In the special case where $\alpha = -1$, both eigenvalues are 0 and $\mathrm{rk}(A) = 1$. Hence, the rank of a complex-symmetric matrix is not necessarily equal to the number of its nonzero eigenvalues.

We further notice that

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

and

$$AA^* = \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & \alpha \end{pmatrix} \begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & \alpha \end{pmatrix} = \begin{pmatrix} 2 & -\mathrm{i}(1-\alpha) \\ \mathrm{i}(1-\alpha) & 1+\alpha^2 \end{pmatrix}.$$

Hence, 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, 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. Therefore, in papers and especially in textbooks it seems more logical and less error-prone to let all symmetric matrices be real by definition. The alternative is to refer to (almost) every symmetric matrix as "real symmetric."

3 Orthogonality

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

There also exist complex square matrices satisfying $B'B = I$. We will call such matrices *complex-orthogonal.* Many authors call any square matrix satisfying $B'B = I$ orthogonal, which leads to many errors and is better avoided.

The following example demonstrates the importance of letting the class of orthogonal matrices be a subclass of the class of unitary matrices. Consider the complex matrix

$$B := \begin{pmatrix} \beta & \mathrm{i} \\ -\mathrm{i} & \beta \end{pmatrix},$$

where β is a real number. We notice that

$$B^*B = \begin{pmatrix} \beta & \mathrm{i} \\ -\mathrm{i} & \beta \end{pmatrix} \begin{pmatrix} \beta & \mathrm{i} \\ -\mathrm{i} & \beta \end{pmatrix} = \begin{pmatrix} \beta^2+1 & 2\mathrm{i}\beta \\ -2\mathrm{i}\beta & \beta^2+1 \end{pmatrix},$$

and that

$$B'B = \begin{pmatrix} \beta & -\mathrm{i} \\ \mathrm{i} & \beta \end{pmatrix} \begin{pmatrix} \beta & \mathrm{i} \\ -\mathrm{i} & \beta \end{pmatrix} = \begin{pmatrix} \beta^2-1 & 0 \\ 0 & \beta^2-1 \end{pmatrix}.$$

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

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

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

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

Next let $\beta = \sqrt{2}$, so that B is complex-orthogonal, and consider its eigenvalues. The determinant of B is one; in fact the determinant of any complex-orthogonal matrix equals ± 1. But the eigenvalues of B 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 hold for complex-orthogonal matrices. To call complex-orthogonal matrices "orthogonal" is bound to lead to errors, and many textbooks — also good ones — provide ample demonstrations of this point.

4 Square root

In mathematics (but not in high school mathematics) it is not uncommon to say that any number y such that $y^2 = x$ is a square root of x. However, the *function* $\sqrt{x}$ is the inverse function of $x = y^2$ for $y \geq 0$. Hence there is a difference between the phrase "square root of x" (not unique) and the symbol $\sqrt{x}$ (unique). If we accept this distinction, then any nonnegative real number x has two square roots, but only one nonnegative square root. This square root is called the principal square root, is unique, and can thus be written as $\sqrt{x}$.

We shall not adopt this distinction which can easily lead to errors. We consider the square root $y := \sqrt{x}$ as a unique real-valued function defined for all $x \geq 0$ with values $y \geq 0$. This is the most common usage. We show first how the scalar square root function extends unambiguously to positive semidefinite matrices. Next we consider nonsymmetric real matrices.

We start with the diagonal matrix

$$\Lambda = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} \equiv \operatorname{diag}(4, 9).$$

There are four matrices which satisfy the equation $D^2 = \Lambda$, namely

$$\begin{aligned} D_1 &:= \operatorname{diag}(-2, -3), \quad & D_2 &:= \operatorname{diag}(-2, 3), \\ D_3 &:= \operatorname{diag}(2, -3), \quad & D_4 &:= \operatorname{diag}(2, 3), \end{aligned}$$

but only D_4 is the square root of Λ compatible with our one-dimensional definition given above. Hence, the matrix function $\Lambda^{1/2}$ is uniquely defined for any diagonal matrix with nonnegative elements on the diagonal.

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, we define $A^{1/2} = S\Lambda^{1/2}S'$. But $\Lambda^{1/2}$ is only defined as before if 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. If A is positive definite (hence nonsingular), then the notation $A^{-1/2}$ denotes the inverse of $A^{1/2}$ or the square root of A^{-1} — they are the same.

Can we also define a unique square root for nonsymmetric real matrices? Yes, this is possible provided the eigenvalues are all positive. Consider a real $n \times n$ matrix A with positive eigenvalues $\lambda_1, \dots, \lambda_n$, not necessarily distinct. Denote by $J_m(\lambda)$ a Jordan block, that is, an $m \times m$ matrix of the form

$$J_m(\lambda) := \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}.$$

(For $m = 1$ we let $J_1(\lambda) := \lambda$.) Jordan's decomposition theorem then guarantees the existence of a nonsingular $n \times n$ matrix T such that $T^{-1}AT = J$, where

$$J := \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & J_{n_2}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & J_{n_k}(\lambda_k) \end{pmatrix}$$

with $n_1 + n_2 + \cdots + n_k = n$.

The square root of A can now be defined through its Jordan representation $A = TJT^{-1}$ as

$$A^{1/2} := TJ^{1/2}T^{-1},$$

where $J^{1/2} := \operatorname{diag}\left(J_{n_1}^{1/2}(\lambda_1), \dots, J_{n_k}^{1/2}(\lambda_k)\right)$ and

$$J_{n_i}^{1/2}(\lambda_i) := \begin{pmatrix} \lambda_i^{1/2} & g_1(\lambda_i) & \dots & g_{n_i-1}(\lambda_i) \\ 0 & \lambda_i^{1/2} & \dots & g_{n_i-2}(\lambda_i) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_i^{1/2} \end{pmatrix} \tag{1}$$

for $i = 1, \dots, k$. The functions g_j are scaled derivatives of $f(\lambda) := \lambda^{1/2}$ (see Abadir and Magnus, 2005, Exercise 9.18), and specialize here to

$$g_j(\lambda) := \frac{f^{(j)}(\lambda)}{j!} = \binom{1/2}{j} \lambda^{1/2-j}, \tag{2}$$

where the binomial coefficients are given as

$$\binom{1/2}{j} = \frac{\prod_{i=0}^{j-1} \left(\frac{1}{2} - i\right)}{j!},$$

and the empty product $\prod_{i=0}^{-1}$ equals one, by convention. For example, we have

$$\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}^{1/2} = \sqrt{3} \begin{pmatrix} 1 & 1/6 & -1/72 \\ 0 & 1 & 1/6 \\ 0 & 0 & 1 \end{pmatrix}.$$

Notice that the reason we may apply the result in Abadir and Magnus (2005) is that $f(\lambda) := \lambda^{1/2}$ has a power series expansion that is summable (but not necessarily convergent) for $\lambda \neq 0$.

This brings us to two further questions: what if $\lambda = 0$, and what if we extend the definition to include real $\lambda < 0$ or complex λ?

First, for $\lambda = 0$ the matrix $A^{1/2}$ may not exist. The simplest example is the matrix $J_2(0)$, which has no square root because the matrix equation

$$X^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

has no solution. More generally, the matrix $J_m(0)$ has no square root when $m > 1$, because the matrix is nilpotent and its index is equal to its dimension. (If a square matrix J of any dimension satisfies $J^{m-1} \neq 0$ and $J^m = 0$, then J is nilpotent of index m.) Clearly, the square root of $J_1(0) = 0$ exists. But the square root exists also if we can pair a block like $J_m(0)$ with another block $J_m(0)$ or $J_{m\pm1}(0)$ in the decomposition of A. We illustrate the effect of the latter pairing with

$$\operatorname{diag}(J_2(0), J_1(0)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix can be transformed by a permutation matrix P and its transpose (inverse) to $P \operatorname{diag}(J_2(0), J_1(0))P'$, thereby preserving the Jordan structure $A = TJT^{-1}$. Choosing

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

the result of the transformation is a nilpotent matrix of index 2 whose square root exists (since the index is less than the dimension) and is nilpotent of index 3. More precisely, we find

$$\left(P \begin{pmatrix} J_2(0) & 0 \\ 0 & J_1(0) \end{pmatrix} P' \right)^{1/2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{1/2} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = J_3(0),$$

and hence

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{1/2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Similarly, if we can pair two blocks $J_m(0)$ in the decomposition of A, then the square root exists too. If all Jordan blocks having $\lambda = 0$ can be paired in one of these two ways, with a leftover block (if any) of the form $J_1(0) = 0$, then the square root of the matrix will exist uniquely for all real $\lambda \geq 0$.

Second, is it possible to extend the definition to include real $\lambda < 0$ or complex λ? Yes, this is possible because the result of (1)–(2) is applicable to all complex $\lambda \neq 0$, by analytic continuation of the binomial expansion. One has to be careful here to choose the principal value of the square roots in (1)–(2). For example, the principal values of $\sqrt{4}$ and $\sqrt{-4}$ are 2 and 2i, respectively, and not -2 and -2i. This extension is applicable to complex matrices, as well as real matrices with complex eigenvalues.

5 Quadratic forms

For any complex matrix U and two conformable complex vectors x and y, an expression of the form $\varphi(x, y) := x^*Uy$ is called a sesquilinear form. The matrix U need not be Hermitian; in fact, it need not be square. If the matrix U is Hermitian, then the expression

$$\varphi(x, y) := x^*Hy$$

is called a Hermitian sesquilinear form, and we write H instead of U to emphasize the nature of H. A Hermitian sesquilinear form is, in general, a complex function. An important subclass of the Hermitian sesquilinear forms are the functions $\varphi(x, x) = x^*Hx$, and these are called Hermitian forms. Since H is Hermitian, we have $(x^*Hx)^* = x^*Hx$. Hence, Hermitian forms are real.

There is no complete agreement in the literature about these terms. A sesquilinear (one-and-a-half times linear) form is linear in one argument and conjugate-linear in the other. Some authors, especially when working solely in a complex setting, refer to sesquilinear forms as bilinear forms, but a bilinear (two times linear) form should be linear in both arguments. Conventions also differ as to which argument should be linear in a sesquilinear form. We take the first argument to be conjugate-linear and the second to be linear, which appears to be the common convention in physics and matrix algebra. Note also that some authors call x^*Hy a Hermitian form rather than a Hermitian sesquilinear form.

If A is a real matrix and x and y are conformable real vectors, then the scalar function $x'Ay$ is called a bilinear form. For symmetric A, an expression $x'Ay$ is a symmetric bilinear form, and the special case $x'Ax$ is a quadratic form.

Now we have to decide whether symmetry of the matrix A is implicit in the definition of a quadratic form or not. There are three good reasons why symmetry should be included in the definition. First, we want a quadratic form to be a special case of a Hermitian form, so that all results on Hermitian forms apply to quadratic forms. Second, we want a quadratic form to be a special case of a symmetric bilinear form, just as a Hermitian form is a special case of a Hermitian sesquilinear form. In fact, and in contrast to the complex case, the symmetric bilinear forms are in one-to-one correspondence with (symmetric) quadratic forms, because

$$x'Ay = \frac{1}{2}\left((x+y)'A(x+y) - x'Ax - y'Ay\right).$$

Third, *all* theorems about quadratic forms concern the case where A is symmetric.

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 $x^*Ux \neq x^*U^*x$, 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 mean that A is symmetric unless stated explicitly otherwise.

A positive (semi)definite matrix, however, is *always* symmetric. This is because a Hermitian matrix is positive (semi)definite if and only if $x^*Hx > 0$ ($x^*Hx \geq 0$) and the requirement that H is Hermitian is essential; otherwise x^*Hx is not real for all x. Hence the first displayed matrix above is not positive definite, but the second (symmetric) matrix is, even though the quadratic forms are the same.

6 Matrix derivatives

If f is an $m \times 1$ vector function of an $n \times 1$ vector x, then the derivative (or Jacobian matrix) of f is the $m \times n$ matrix

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

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. All mathematics texts define vector derivatives in this way. Since we wish to call on results from the mathematics literature, we should not deviate from the standard definition.

Thus, when $y = Ax$, then $\partial y/\partial x' = A$ (when A is a matrix of constants). Also, for a scalar function $\varphi(x)$, the derivative $\partial\varphi(x)/\partial x'$ is a *row* vector, not a column vector.

The definition of matrix derivatives is not generally treated in mathematics texts, but it 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. Clearly, the derivative is a matrix containing all $mpnq$ partial derivatives. But how are these partial derivatives organized? We shall argue in this section that 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. There exist other stacking operators (for example, by organizing the elements row-by-row rather than column-by-column), and these could be used equally well as long as this is done consistently. A form of stacking, however, is essential in order to preserve the notion of "derivative." We note in passing that all stacking operations are in one-to-one correspondence with each other and connected through permutation matrices. For example, the row-by-row and column-by-column stacking operations are connected through the commutation matrix. There is therefore little advantage in developing the theory of matrix calculus for more than one stacking operation.

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 φ is a scalar function of a matrix X ($m = p = 1$), then $\mathrm{D}\varphi(X) = \partial\varphi(X)/\partial(\operatorname{vec} X)'$, a $1 \times nq$ row vector. The choice of ordering the partial derivatives is not arbitrary. For example, the derivative of the scalar function $\varphi(X) = \operatorname{tr}(X)$ is not $\mathrm{D}\varphi(X) = I_n$ (as is often stated), but $\mathrm{D}\varphi(X) = (\operatorname{vec} I_n)'$.

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, as the discussion and examples below will demonstrate.

Let us consider an alternative arrangement of the partial derivatives. If $F = (f_{ij})$ is an $m \times p$ matrix function of a scalar x, then one may define an expression

$$\frac{\delta F(x)}{\delta x} := \begin{pmatrix} \frac{\partial f_{11}(x)}{\partial x} & \dots & \frac{\partial f_{1p}(x)}{\partial x} \\ \vdots & & \vdots \\ \frac{\partial f_{m1}(x)}{\partial x} & \dots & \frac{\partial f_{mp}(x)}{\partial x} \end{pmatrix},$$

which has the same dimension as F. Since this is not the derivative of F (which is an $mp \times 1$ column vector), we use a different notation: δ instead of ∂, and we call the expression not a derivative but a derisative. If $X = (x_{st})$ is a matrix of dimension $n \times q$, then we may extend this definition to the derisative of a matrix with respect to a matrix as

$$\frac{\delta F(X)}{\delta X} := \begin{pmatrix} \frac{\delta F(X)}{\delta x_{11}} & \dots & \frac{\delta F(X)}{\delta x_{n1}} \\ \vdots & & \vdots \\ \frac{\delta F(X)}{\delta x_{1q}} & \dots & \frac{\delta F(X)}{\delta x_{nq}} \end{pmatrix},$$

where we note the transposition of the indexing. The resulting matrix is of order $mq \times np$. The derisative $\delta f(x)/\delta x$ of a vector with respect to a vector is equal to the derivative $\mathrm{D}f(x) = \partial f(x)/\partial x'$. In fact, this is the reason why the indexing is transposed in the above formula. One might therefore think that both definitions generalize vector calculus to matrix calculus, and that the choice is a matter of taste. This, however, is not the case. Although both definitions contain all partial derivatives, their dimensions are different and their properties are completely different. Only one definition (the derivative) generalizes the concept of a derivative; the other just contains the partial derivatives, but has no useful properties. This fact is well documented; see Pollock (1985) and Magnus and Neudecker (1985, 1988), but the derisative still exists prominently. Since this is an important issue, we give below four reasons why the derisative is not a derivative.

First, consider the identity function $F(X) = X$. One would expect that the derivative of this function is the identity matrix, and of course we have $\mathrm{D}F(X) = I$. But the derisative of the identity function is not the identity matrix. For example, when X has dimension 2×2, we find

$$\frac{\delta F(X)}{\delta X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Second, the "product rule" for derisatives is often stated as

$$\frac{\delta (FG)(X)}{\delta X} = \frac{\delta F(X)}{\delta X} G(X) + F(X) \frac{\delta G(X)}{\delta X}.$$

Just looking at the dimensions of the matrices, it is immediately clear that this rule cannot hold in general: a matrix sum is only defined when the two matrices have the same dimensions, and a matrix product requires that the number of columns in one matrix equals the number or rows in the other. The product rule for derisatives is therefore only defined when X is a scalar. But then the result is in fact true. In general, whether the argument X is a scalar or a matrix, we have

$$\mathrm{d}(FG) = (\mathrm{d}\, F)G + F(\mathrm{d}\, G),$$

the product rule for differentials. Applying the vec-operator gives

$$\mathrm{d}\,\mathrm{vec}(FG) = (G' \otimes I_m)\,\mathrm{d}\,\mathrm{vec}\, F + (I_r \otimes F)\,\mathrm{d}\,\mathrm{vec}\, G,$$

assuming that F has m rows and that G has r columns. Then,

$$\frac{\partial\,\mathrm{vec}(FG)}{\partial(\mathrm{vec}\, X)'} = (G' \otimes I_m)\frac{\partial\,\mathrm{vec}\, F}{\partial(\mathrm{vec}\, X)'} + (I_r \otimes F)\frac{\partial\,\mathrm{vec}\, G}{\partial(\mathrm{vec}\, X)'},$$

so that we obtain the correct product rule for derivatives as

$$\mathrm{D}(FG)(X) = (G' \otimes I_m)\mathrm{D}F(X) + (I_r \otimes F)\mathrm{D}G(X).$$

Third, let us consider the chain rule, obviously an essential ingredient without which matrix calculus cannot exist. If F $(m \times p)$ is differentiable at X $(n \times q)$, and G $(l \times r)$ is differentiable at $Y = F(X)$, then the composite function $H(X) := G(F(X))$ is differentiable at X, and

$$\mathrm{D}H(X) = (\mathrm{D}G(Y))(\mathrm{D}F(X)).$$

This is the correct chain rule. By analogy, the "chain rule" for derisatives is typically stated as

$$\frac{\delta H(X)}{\delta X} = \frac{\delta G(Y)}{\delta Y}\frac{\delta F(X)}{\delta X}.$$

Again, there is a problem with the dimensions, because the product is only defined when $p = q = r$. But even with this restriction the rule is wrong. Consider the case where all matrices are square of order two. Then the "chain rule" for derisatives reads

$$\begin{pmatrix} \frac{\delta H(X)}{\delta x_{11}} & \frac{\delta H(X)}{\delta x_{21}} \\ \frac{\delta H(X)}{\delta x_{12}} & \frac{\delta H(X)}{\delta x_{22}} \end{pmatrix} = \begin{pmatrix} \frac{\delta G(Y)}{\delta y_{11}} & \frac{\delta G(Y)}{\delta y_{21}} \\ \frac{\delta G(Y)}{\delta y_{12}} & \frac{\delta G(Y)}{\delta y_{22}} \end{pmatrix} \begin{pmatrix} \frac{\delta F(X)}{\delta x_{11}} & \frac{\delta F(X)}{\delta x_{21}} \\ \frac{\delta F(X)}{\delta x_{12}} & \frac{\delta F(X)}{\delta x_{22}} \end{pmatrix}.$$

In particular, we should have

$$\frac{\delta H(X)}{\delta x_{11}} = \frac{\delta G(Y)}{\delta y_{11}}\frac{\delta F(X)}{\delta x_{11}} + \frac{\delta G(Y)}{\delta y_{21}}\frac{\delta F(X)}{\delta x_{12}},$$

and specializing further:

$$\frac{\partial h_{11}}{\partial x_{11}} = \frac{\partial g_{11}}{\partial y_{11}}\frac{\partial f_{11}}{\partial x_{11}} + \frac{\partial g_{12}}{\partial y_{11}}\frac{\partial f_{21}}{\partial x_{11}} + \frac{\partial g_{11}}{\partial y_{21}}\frac{\partial f_{11}}{\partial x_{12}} + \frac{\partial g_{12}}{\partial y_{21}}\frac{\partial f_{21}}{\partial x_{12}}.$$

But this is not true. The correct chain rule is given by

$$\frac{\partial h_{11}}{\partial x_{11}} = \frac{\partial g_{11}}{\partial y_{11}}\frac{\partial f_{11}}{\partial x_{11}} + \frac{\partial g_{11}}{\partial y_{21}}\frac{\partial f_{21}}{\partial x_{11}} + \frac{\partial g_{11}}{\partial y_{12}}\frac{\partial f_{12}}{\partial x_{11}} + \frac{\partial g_{11}}{\partial y_{22}}\frac{\partial f_{22}}{\partial x_{11}}.$$

Finally, let us consider a transformation from a 2×2 matrix X to a 2×2 matrix $Y = F(X)$. We want to know whether this transformation is singular or nonsingular, and we want to know the Jacobian of the transformation, that is, the absolute value of the determinant of the derivative $\mathrm{D}F(X)$. Suppose that the derivative at a particular value of X is given by

$$C_1 := \frac{\partial \operatorname{vec} F(X)}{\partial (\operatorname{vec} X)'} = \begin{pmatrix} 4 & 2 & 2 & 1 \\ 1 & 1 & \gamma & 1 \\ 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix},$$

where we leave γ free for the moment. The corresponding derisative is then given by

$$C_2 := \frac{\delta F(X)}{\delta X} = \begin{pmatrix} 4 & 2 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 \\ \gamma & 0 & 1 & 2 \end{pmatrix}.$$

The determinants of C_1 and C_2 are $|C_1| = 6\gamma - 2$ and $|C_2| = -2\gamma$, respectively, and are therefore not equal (unless $\gamma = 1/4$). If $\gamma = 1/3$ then the transformation is singular ($|C_1| = 0$), but $|C_2|$ is nonzero, so that the derisative does not provide the correct information. If $\gamma = 0$, then $|C_2| = 0$, but the transformation is in fact nonsingular ($|C_1|$ is nonzero).

In summary: a derisative is not a derivative, and there exists not a single reason for using it.

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 and therefore must be avoided.

References

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.

Magnus, J.R. and H. Neudecker (1985). Matrix differential calculus with applications to simple, Hadamard, and Kronecker products, *Journal of Mathematical Psychology*, 29, 474–492.

Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons: Chichester/New York. Second edition (paperback) 1999.

Pollock, D.S.G. (1985). Tensor products and matrix differential calculus, *Linear Algebra and Its Applications*, 67, 169–193.