Journal of Multivariate Analysis 133 (2015) 377–381

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Note(s)

On the use of coordinate-free matrix calculus

Jan Brinkhuis

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Burg. Oudlaan 50, 3062PA Rotterdam, The Netherlands

HIGHLIGHTS

- A vector-space calculation of the derivative of the determinant function.
- All conventions of matrix calculus arise naturally in the vector-space approach.
- The vector-space approach clarifies the role of the Kronecker product in matrix calculus.
- The vector-space approach gives a quick access to the v-space in interior point methods.

ARTICLE INFO

Article history:
Received 1 August 2014
Available online 18 October 2014

AMS subject classifications:
30
40

Keywords:
Maximum likelihood estimators
Kronecker products
Matrix calculus
Derivative of the determinant
Coordinate-free

ABSTRACT

For a standard tool in econometrics, matrix calculus, an approach is illustrated in this note that is unusual in that context, a coordinate-free approach. It can help to eliminate the persistent use of non-standard conventions. The Kronecker product and its use can be better understood. The complications and pitfalls of defining twice differentiability by partial derivatives are avoided. Its use is demonstrated, for example by giving a coordinate-free determination of the derivative of the determinant.

0. Introduction

Matrix calculus has been used in econometrics for many decades in the determination of maximum likelihood estimators. In applied mathematics, there is an even longer tradition of a coordinate-free approach to matrix calculus, making use of tensor calculus and exterior algebra (see, for example, [9]). This approach has advantages, but almost all econometrics treatments do not make use of it.

In this note, we give a royal road that allows to profit from some advantages at the cost of only a small investment of time. In particular we try to argue that by concentrating one's attention to partial derivatives, one is liable to lose sight of the underlying mathematical structure, thereby running the danger of pursuing inappropriate definitions of the matrix derivatives. Therefore it is proposed that a proper understanding of the mathematical structure of matrix differential calculus can be achieved by concentrating attention on the associated mappings between vector spaces.

In the recent papers [8,4], the persistent use in wikipedia and by practitioners of non-standard conventions – notwithstanding the availability of good sources such as [6,5,1] – has been pointed out and it has been explained how this can lead to problems. It seems worthwhile to further emphasize these issues, which the present paper tries to do. The coordinate-free or

E-mail address: brinkhuis@ese.eur.nl.

http://dx.doi.org/10.1016/j.jmva.2014.09.019

vector-space approach gives the insight that the standard conventions are the only possibility given the standard convention for writing the transformation matrix of a linear map. We show that defining differentiability and the differential by means of coordinates, that is, by using partial derivatives, leads to complications, which are avoided in the vector-space approach.

Moreover, this paper tries to foster a better understanding of the Kronecker product of matrices and its use. The need to differentiate functions $Y = f(X)$ where X and Y are matrices leads to Kronecker products. Indeed, the differential of such a function taken in a matrix X is a linear map between matrix spaces and so it can be shown to be equal to the sum of maps of type $H \mapsto BHA'$ with A and B matrices—in applications one usually needs only a few terms, fortunately. The map $H \mapsto BHA'$ is the vector-space definition of the Kronecker product of A and B. The transition from differentials to derivatives leads to the need to write down the matrix of the linear map $H \mapsto BHA'$; this matrix is the usual Kronecker product $A \otimes B$ and its properties follow from evident properties of the maps $H \mapsto BHA'$. We make the link with the work in [6,7] where Kronecker products as special cases of tensor products have been analyzed in great detail.

Twice differentiability can be defined in a straightforward way by the condition that the differential of the differential, $d(df)$, exists. Indeed, taking the differential of $f : X \to Y$, where X and Y are finite dimensional vector spaces, gives a similar object, $df : X \to L(X, Y)$, where $L(X, Y)$ is the finite dimensional vector space of linear maps from X to Y. Therefore, one can take the differential again. If one defines twice differentiability in the usual way, using coordinates, that is, using second order partial derivatives – as is done for example in [5, pp. 116–117] – then there are some complications to be taken care of and pitfalls to be warned against.

We give a vector-space proof of the well-known formula for the derivative of a determinant. To begin with, one can immediately see what the derivative is in the identity matrix. Then one can transfer the outcome, by means of the chain rule, to an arbitrary invertible matrix. Finally one can extend this to arbitrary matrices by a continuity argument. Attempts to give a vector-space derivation have not been successful so far. We refer here to the sentences in [1] that precede the calculation: "We have emphasized several times that, when evaluating differentials, one should always try to work with matrices rather than with elements of matrices. This, however, is not always possible. Here is an example where, regrettably, we have to rely on the elements of X".

We give a vector-space determination of the stationarity conditions of two optimization problems. Analysis of these conditions leads to two valuable results. It gives the inequality of Hadamard and the construction of the v-space, a much used concept in interior point methods (due to [3] for the case of linear programming, and later extended, notably in [10], to semidefinite programming).

Finally, we outline how to extend matrix differential calculus to higher dimensions than two. So far such an extension has not been realized, despite efforts in this direction.

1. Unified definition of the derivative

It is necessary to define derivatives of functions $y = f(x)$ where x and y are matrices (and their special cases, that is, numbers, row vectors, column vectors) or symmetric matrices. To begin with, the *differential* $df(x)$ of a function $f : X \to Y$ at a point $x \in X$ – with X, Y finite-dimensional vector spaces (always provided with some length or norm) – is defined by splitting $f(x + h) - f(x)$ as the sum of a linear map of the variable vector h, $df(x)(h)$, and a term $r(h)$ that tends to zero faster than $\|h\|$: that is, the limit of the quotient $r(h)/\|h\|$ for $h \to 0$ is zero (such a term is called a *negligible* function). If such a splitting does not exist, then the function is said to be non-differentiable in x. The *derivative* $Df(x)$ is defined to be the transformation matrix of the linear map $df(x)$ with respect to suitably chosen bases of X and Y.

For the transition from differentials to derivatives one has to choose an ordered basis of $L(X, Y)$ (or equivalently, a isomorphism from $L(X, Y)$, the vector space of linear maps $X \to Y$, to a space of column vectors, called its *vectorization isomorphism*) in particular if $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$; then $L(X, Y) = M^{m\times n}$, the space of $m \times n$-matrices. Now we show how the standard convention for this choice arises in a natural way.

If x_i, $1 \le i \le n$ is a basis of a vector space X and y_j, $1 \le j \le m$ is a basis of a vector space Y, then one defines a basis of $L(X, Y)$ by taking for each pair (i, j) with $1 \le i \le n$, $1 \le j \le m$ the linear map $X \to Y$ that sends x_i to y_j and that sends all other elements of the basis of X to the zero-vector 0_Y. These vectors form a basis of $L(X, Y)$ and this basis is ordered by the lexicographical ordering on pairs (i, j). We call this the basis of $L(X, Y)$ that is *associated* to the chosen bases of X and Y. This defines a vectorization isomorphism $L(X, Y) \to \mathbb{R}^{mn}$ that is given by sending a linear map $X \to Y$ to its column of coefficients with respect to the associated basis.

The case of interest to us is $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ and so $L(X, Y) = M^{m\times n}$ (where an $m \times n$-matrix A corresponds to the linear map $x \mapsto Ax$). For this choice of X, Y one chooses the canonical bases ordered in the obvious way. That is, for $\mathbb{R}^p$ with $p = m$ or n one chooses $(1, 0, \ldots, 0)'$, $(0, 1, 0, \ldots, 0)'$, $\ldots$, $(0, \ldots, 0, 1)'$. This readily leads to the following explicit description of the vectorization isomorphism for $L(\mathbb{R}^n, \mathbb{R}^m) = M^{m\times n}$.

Proposition 1. *The vectorization isomorphism for $M^{m\times n}$ stacks the columns of a matrix into one long column,* $A \mapsto \mathrm{vec}(A) = (a_{11}, a_{21}, \ldots, a_{m1}, a_{12}, a_{22}, \ldots, a_{mn})'$.

Illustration 1. *The quadratic functions $f(x) = x'Ax + bx + c$, where $A \in \mathrm{Sym}_n$, $b \in (\mathbb{R}^n)'$, $c \in \mathbb{R}$ has differential $df(x)(h) = (2x'A + b)h\ \forall h$ and derivative $Df(x) = 2x'A + b$.*