Comments from Thirty Years of Teaching Matrix Algebra to Applied Statisticians[1]

BU-1411-MA

January 1999

Shayle R. Searle

Departments of Biometrics and Statistical Science
Cornell University, Ithaca, NY 14853

Abstract

A few simple ideas gleaned from teaching matrix algebra are described. For teaching beginners how to prove theorems a useful mantra is "Think of something to do, do it and hope."

[1]Contributed paper at Seventh International Workshop on Matrices and Statistics. Fort Lauderdale, Florida, December 11-14, 1998. Paper BU-776 in the Department of Biometrics.

Introduction: hooked on matrix algebra

As a Master's student of mathematics in New Zealand, I had a course on matrices from Jim Campbell who had done his doctorate under A.C. Aitken at Edinburgh. That one could have something like $\mathbf{AB} = \mathbf{0}$ without $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$ fascinated me – and I was hooked. Eight years later, as a graduate student at Cornell, I gave an informal 1957 summer course on matrix algebra; and after returning to Cornell on faculty that course developed in 1963 into a regular fall semester offering, Matrix Algebra, in the Biometrics Unit. I taught it until 1995, when I retired; and it still continues with 20-30 students a year.

One may well ask "why matrix algebra when linear algebra is widely taught in Mathematics?" Several reasons: I'm no geometer. Dropping a perpendicular from 8-space to 4-space helps my intuition not one bit. I cannot illustrate that perpendicular on a blackboard, and I cannot use my fingers to do so either. I have many kindred spirits in this regard. They come from students minoring in statistics for use in their down-to-earth majors such as agronomy, agricultural economics, plant breeding, pomology and animal breeding. They find no joy in geometry, but they can get the hang of algebra – even to the stage of enjoying it, despite the 8 a.m. hour for the course! That lasted 25 years, whereupon I changed it to 9 a.m., giving the college curriculum committee the reason "After 25 years, I'm tired of 8 a.m.!"

Features of the course

Each lecture began with my immediately launching into the day's topic, saving the end of the hour, when all of a day's attendees would be there, for announcements. And so no repetition was needed.

An 8 a.m. class inevitably generates late arrivals. That never worried me. Mathematical subjects demand using today's new knowledge for tomorrow's work, so that I insisted that late arrival was far better than not coming at all. And with the classroom having a rear door, which I heartily recommend, entry for laggards caused no interruption, except on the day a chronically 30-minute late arrival was greeted with the 8:30 a.m. salutation "Good afternoon".

The course always had homework – every Wednesday, a help session the following Monday and hand it in two days later. By that time, one hoped it was all correct. Each week's work was graded

2, 1 or 0 for satisfactory, mediocre or hopeless, respectively. No student ever argued about this holistic grading. And a whole term's homework counted for no more than 10% of the term's grade; in some years nothing. Yet there was notivation: an early handout explained that the term grade would be F if there was any failure to do homework, no matter how much help had been given. To me, homework is for students to learn to *do* mathematics; it is not inquisitorial for assessing how much they have learned. That is the purpose of exams – and students were told that exam questions would be similar to homework problems – indeed, most exams included at least one of the homework problems.

For a number of years the course was split into two back-to-back 7-week modules, each of two credits and two exams. This came about because agricultural economics graduate students asserted the second part of the course was too theoretical. So with the split they did only the first part and got no further than inverting a matrix (not even to rank and linear independence). So their attendance dwindled almost to zero, coinciding with increasing enrolment of statistics undergraduates. For them I soon learnt that two modules involving four exams was too weighty a presentation, and after 10 years the course returned to a regular 3-lecture, 14-week routine. That was less intensive, more enjoyable for the students, and under less pressure I believe they learnt more, and more easily.

Teaching how to do mathematics

How do we do mathematics? How do we learn how to do mathematics? It seems to me that we don't do a good job of teaching this – of how to do algebra, for example. Many students have difficulty in learning this: they find even simple methods of proof difficult to assimilate. This arises, one must assume, from high school and mandatory college courses in mathematics paying insufficient attention to these matters. Students seem to have little idea as to how mathematicians think, how they go about solving problems, and especially how they *start* trying to solve a problem. So the matrix algebra course attempted to give students ideas about some of the thought processes which are helpful in proving already-established results. One would hope that in learning such techniques students would subsequently find them very useful, both for checking on the validity of results they will find in their own subject-matter research literature, and also for developing their own results *ab initio*. And part of the necessary learning is how do mathematicians motivate themselves to develop the step-by-step process which leads to final and useful results? The nature of this motivation can be illustrated in the proving of a simple, well known matrix result.

A simple theorem Suppose we want to prove the following little theorem: if $\mathbf{A}$ is idempotent, so is $\mathbf{I}-\mathbf{A}$. As we all know, the proof is easy:

$$(\mathbf{I}-\mathbf{A})^{\mathbf{2}} = \mathbf{I}-\mathbf{2A}+\mathbf{A}^{\mathbf{2}} = \mathbf{I}-\mathbf{2A}+\mathbf{A} = \mathbf{I}-\mathbf{A}; \qquad \text{Q.E.D}$$

"But", asks a student, "Where does this come from?" "Why?" asks another; and "So what?" says a third. These reactions, I suggest, stem from unfamiliarity of important steps in the process of proof, steps which are, it appears, seldom emphasized early enough in one's mathematical education. For example, two important steps for algebraic proof are (i) whenever possible convert text statements to algebraic statements and, in doing so, (ii) clearly label the statements as Given, or To Be Proved. In our case this yields

$$\begin{aligned} \text{Given} \quad &: \quad \mathbf{A}^{\mathbf{2}} = \mathbf{A}, \text{ the definition of idempotent.} \\ \text{To Be Proven} \quad &: \quad (\mathbf{I}-\mathbf{A})^{\mathbf{2}} = \mathbf{I}-\mathbf{A}. \end{aligned}$$

Now comes the difficult part. We know what has to be proven: we must show that $(\mathbf{I}-\mathbf{A})^{\mathbf{2}}$ can be reduced to $\mathbf{I}-\mathbf{A}$. On thinking about this one soon concludes that it seems (and indeed is) difficult to straightforwardly start at $\mathbf{I}-\mathbf{A}$ and get to $(\mathbf{I}-\mathbf{A})^{\mathbf{2}}$. Actually, one can do this, as follows. Beginning at $\mathbf{I}-\mathbf{A}$ we can observe that $\mathbf{A}(\mathbf{I}-\mathbf{A}) = \mathbf{A}-\mathbf{A}^{\mathbf{2}} = \mathbf{A}-\mathbf{A} = \mathbf{0}$. Therefore

$$\mathbf{I}-\mathbf{A} = \mathbf{I}-\mathbf{A}-\mathbf{A}(\mathbf{I}-\mathbf{A}) = \mathbf{I}-\mathbf{A}-\mathbf{A}+\mathbf{A}^{\mathbf{2}} = \mathbf{I}-\mathbf{2A}+\mathbf{A}^{\mathbf{2}} = (\mathbf{I}-\mathbf{A})^{\mathbf{2}}.$$

This is correct, but it is sort of gimmicky, and not obviously logical. It is much more logical to start from something complicated, in this case $(\mathbf{I}-\mathbf{A})^{\mathbf{2}}$, and try to reduce it to something simpler, $(\mathbf{I}-\mathbf{A})$. This is usually easier than the other way round. So we start with $(\mathbf{I}-\mathbf{A})^{\mathbf{2}}$ and hope to get to $\mathbf{I}-\mathbf{A}$. But now what? $(\mathbf{I}-\mathbf{A})^{\mathbf{2}} =?$

The following mantra, to be used iteratively, is what helps.

"Think of something to do: do it and hope."

What comes easily to mind? Expand:

$$(\mathbf{I}-\mathbf{A})^2 = (\mathbf{I}-\mathbf{A})(\mathbf{I}-\mathbf{A}).$$

Now what? Use the mantra again, and multiply out:

$$(\mathbf{I}-\mathbf{A})^2 = (\mathbf{I}-\mathbf{A})(\mathbf{I}-\mathbf{A}) = \mathbf{I}(\mathbf{I}-\mathbf{A}) - \mathbf{A}(\mathbf{I}-\mathbf{A}) = \mathbf{I}-\mathbf{A}-\mathbf{A}+\mathbf{A}^2 = \mathbf{I}-2\mathbf{A}+\mathbf{A}^2.$$

Now comes an important step in the process of developing results. (iii) Use what is given: in this case $\mathbf{A}^2 = \mathbf{A}$. Thus

$$(\mathbf{I}-\mathbf{A})^2 = \mathbf{I}-2\mathbf{A}+\mathbf{A} = \mathbf{I}-\mathbf{A}; \quad \text{Q.E.D.}$$

This process, especially the mantra, has helped numerous students to overcome their fear of the proof process and to go on to successfully tackle problems. True, the mantra is not much more than formalizing the trial-and-error process, but after all that is precisely how mathematical progress is often made. It seldom leads to the shortened proof, but who cares about that insofar as learning the process of proof is concerned. Supplementing the mantra there are, of course, a number of technical aids for helping with algebra, some of which are detailed in Searle (1977).

Patterns and relationships We also need to teach students that mathematicians look for patterns and relationships. For example, in proving that

$$t = [a\ \ b\ \ c]\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1}\begin{bmatrix} p \\ q \\ r \end{bmatrix} = p,$$

without calculating the matrix inverse, it is the "pattern" of the row vector also being the first row of the matrix, combined with the thinking about the cofactors in the elements of the matrix inverse, that leads at once to

$$t = [1\ \ 0\ \ 0]\begin{bmatrix} p \\ q \\ r \end{bmatrix} = p.$$

Two kinds of thinking

This "thinking about the cofactors" brings us to what I feel are two kinds of thinking needed for doing mathematics. They can be called "automatic thinking" and "cogitative thinking". Automatic thinking is the kind of straightforward thinking involved in doing algebra, especially in simplifying algebraic expressions; e.g., $\mathbf{I} - \mathbf{2A} + \mathbf{A} = \mathbf{I} - \mathbf{A}$. The second type of thinking is that of cogitating; e.g., consider the statement

AB is a matrix of columns which are linear combinations of columns of **A**.

Automatic (e.g., algebraic) thinking is relatively easy, once one knows the rules of algebra, for example. In contrast, cogitative thinking is not so easy. As soon as we see that statement about **AB** we know it is true. But developing the statement requires cogitating (i.e., "thinking hard", according to the dictionary) about the operation of matrix multiplication. Then we hit on the statement about **AB**. Generally speaking, it is a much harder kind of thinking than that involved in $\mathbf{I} - \mathbf{2A} + \mathbf{A} = \mathbf{I} - \mathbf{A}$. And the real difficulty is to recognize when this kind of cogitating is required. The trick is to know when, in the presence of just the product symbol **AB**, will it be useful to think of that product as just described. At least a first step in being motivated towards this possibility is to recognize that there are occasions when the "automatic thinking" or "algebraic thinking" has to be replaced by the cogitating (hard thinking). Distinguishing between these two modes of thinking has proven to be one more aid in helping non-mathematics majors to learn a mathematical subject. A simple example involving both kinds of thinking is the following. Suppose

$$\begin{aligned}\mathbf{y}' &= [y_{11}\ \ y_{12}\cdots y_{1n}\ \ y_{21}\ \ y_{22}\cdots y_{2n}\cdots \quad \cdots y_{i1}\ \ y_{i2}\cdots y_{in}\cdots \quad \cdots y_{m1}\ \ y_{m2}\cdots y_{mn}] \\ &= \left\{_r\ \{_r\ \ y_{ij}\}_{j=1}^{m}\right\}_{i=1}^{m}.\end{aligned}$$

Then, with $y_{i\cdot} = \sum_{j=1}^{n} y_{ij}$,

$$\begin{aligned}(\mathbf{I}_m \otimes \mathbf{1}'_n)(\mathbf{I}_m \otimes \mathbf{J}_n)\mathbf{y} &= (\mathbf{I}_m \otimes n\mathbf{1}'_n)\mathbf{y} \\ &= n(\mathbf{I}_m \otimes \mathbf{1}'_n)\mathbf{y} \\ &= n\{_c\ \ y_{i\cdot}\}_{i=1}^{m}.\end{aligned}$$

The first two equations here stem from automatic (algebraic) thinking; but the last comes from cogitative thinking.

Doing mathematics is a game

Finally, there is nothing wrong in suggesting to students that doing mathematics is really a game: make the rules and play by them. For example, for scalars

$$ab = 0 \Rightarrow a = 0 \quad \text{and/or} \quad b = 0: \quad \text{and} \quad xy = yx \quad \text{always.}$$

But for matrices the rules are different:

$$\mathbf{AB} = \mathbf{0} \not\Rightarrow \mathbf{A} = \mathbf{0} \quad \text{and/or} \quad \mathbf{B} = \mathbf{0}; \quad \text{and} \quad \mathbf{XY} = \mathbf{YX} \quad \text{sometimes (not often).}$$

One must therefore always remember which game is being played.

Hopefully some of these ideas may make a contribution to reducing the fear of mathematics that we see in so many students and thence to improving their mathematical literacy!

Reference

Searle, S.R. (1977) Proof. *International Journal of Mathematics Education in Science and Technology* **8**, 195-202.