The Teaching of Linear Algebra

I recently withdrew my own text in Linear Algebra. So for some time I have been looking some time for a replacement. I have examined a number of American books from large publishing houses, and turned away in bitter disappointment.

I will detail my frustration in this article. Hopefully my comments will be of some use to someone devising or developing a course in linear algebra. Or writing a book!

In general terms the most important problem is organization. I realize the dilemma. A very tightly knit structure may appear overly abstract and frustrating; the course builds and builds and everything happens towards the end. A looser structure, however, may turn the subject into a confusing array of fragments. I claim that prejudice, and impatience with the qualitative nature of learning, has produced too much of the latter. Short-term "simplifications" may alleviate initial difficulties at the cost of reaching no goals at all.

One instance of poor structure is that far too many theorems are stated, and too few are proved. "The eigenvalues of a triangular matrix can be read off from the diagonal!" Is that a theorem? It's an observation, and students will make it.

In my view, linear algebra is a synthesis of geometry and equation solving. Its most natural goals are least-squares and various uses of eigenvectors. My contacts in both electric and mechanical engineering confirm this view.

As for least-squares, it's important to understand the geometric idea. All efficient numerical algorithms start from that, not from the normal equations. In the latter case it's important to have good examples from applications, e.g., discrete heat conduction or discrete oscillations. In these, eigenvalues and eigenvectors *mean something*. For instance, appeal to the physical model explains why the eigenvalues in either case must be non-positive; also when, and why, zero eigenvalues appear.

The so-called classification of quadric surfaces contributes nothing to the students' understanding of these concepts. Its merits (if any) lie elsewhere.

There are other, quite valid, goals. However, in my experience, one must concentrate on these two, else confusion and fragmentation will prevail. Given the time, the Singular Value Decomposition would be the most natural climax, interrelating these two goals. Books should give it, along with applications.

I now turn to my list.

Geometry

For some reason vectors are usually identified with coordinates right away. My colleagues in Mechanics are extremely critical of this fact. Students (who have been exposed to American textbooks) know how to subtract and cross multiply but don't know where the resulting vectors point. I believe vectors should be introduced (informally) as equivalence classes, directed segments which "may" be translated. There should be many very simple problems without coordinates, or problems inviting the student to introduce his own.

One recent favorite of mine is, determine the dot product of two vectors from a diagram. What exactly needs to be measured?

Of course, geometric vectors are there to prepare for more abstract things. Therefore, the treatment of linearity properties never ceases to amaze me. In most treatments, the coordinate rule for dot products is deduced from the Law of Cosines (not wanting proof?). Linearity is read *from the expression*.

The natural procedure is to deduce the linearity properties from those of orthogonal projection. This is a higly non-trivial fact, requiring similarity. The insistence on identifying vectors with coordinates conceals this idea.

Most books introduce the cross product by way of coordinates. This raises serious invariance questions.

Their geometric characterization is proven by computation, except orientation, which is too hard, in this approach. It "may" be proved, or is "beyond the scope of our presentation".

In contrast, starting with geometry offers the golden opportunity to introduce several important examples of linear mappings. The cross product, one factor fixed, is composed of a projection, a rotation and a dilation. Orientation is then no problem at all.

If rigor demands the coordinate approach, then why not simply define orientation by the sign of a determinant? This could actually be justified by some kind of continuity argument.

I once lectured this three-step idea to students using Howard Anton's text. I didn't care to illustrate the linearity of each step, instead I referred the students to the text for that. Only afterwards did I realize that Anton (like Edwards-Penney, like D. Norman, but unlike most other texts) *doesn't have a single diagram illustrating the single most important concept of the whole subject*!!!

By the same token, some authors presuppose the addition laws of Trigonometry in order to determine the matrix of a rotation. Again the natural procedure would be to derive the laws from the linearity of the mapping. Linearity should be presented as a powerful unifying tool, that produces, and interrelates, some very striking, non-trivial, results.

Systems of Equations

So many pages, so many fancy words. Four simple examples are enough to explain row elimination and obstructions to it. We need no more. The formal apparatus introduced in many books conceals the elementary nature of the operations. To quote J W S Cassels, from another context, we should treat these things like the triviality they are.

Matrices and their Inverses

I don't know of one single American textbook explaining matrix inverses, and the condition for their existence, in a natural manner. The simple theory is imbedded in a pompous discussion on elementary matrices, their products and factorizations.

Why not do the obvious? The equation $AX = I$ *immediately* translates into a set of systems of equations, all having the same left members. They are uniquely solvable if $\det A \neq 0$. If $\det A = 0$ row operations will produce a row of zeros in the left member, but not in all of the right members, since $\det I = 1 \neq 0$.

And, in the first case, $XA = I$ because both members satisfy $AY = A$, which is again uniquely solvable.

We don't even need elementary matrices to explain the product rule. Just look at the row operations turning $AX = AB$ into $X = B$; the determinants of both members change by the same factor. The case $\det A = 0$ admits a simple separate proof.

Determinants

"Down with determinants" wrote Sheldon Axler a while back. I would rather say, down with the "standard" treatment, by induction. This is supposed to be a simplification, because permutations and their signs are too hard on the students.

The test is proofs. The same books that devote pages of induction proofs to the relatively simple linearity properties have great trouble with antisymmetry. The proofs are long and unnatural, not given at all, or relegated to an Appendix.

The definition by permutations puts antisymmetry right under your nose. First deal with adjacent rows, then the general case, in an odd number of steps. The sign of a permutation is best explained by counting "negative slopes" as in Shilov's classical textbook.

I've been using that approach for decades, and it has always worked. For instance, it turns the statement $\det A = \det A^t$ into a triviality. A half turn does not change the sign of a slope.

In many cases, where certain phenomena can be explained more directly - such as uniqueness (independence) implying existence (spanning a space) - it seems that determinants help fix the ideas. One may deplore this fact; Axler seems to. I challenge him and everybody else to prove the following theorem without using determinants:

"Let the two real n/n–matrices A and B be connected by a complex change of basis: $A = T^{-1}BT$, T complex. T can then be replaced by a real matrix" (the issue is invertibility).

Which spaces?

There has been much discussion whether to introduce "general" vector spaces, or just $\mathbf{R}^n$, possibly along with its subspaces. If we stick to just $\mathbf{R}^n$ most of the theory becomes pointless. E.g., the theorem stating that linear independence is equivalent to spanning, if the number of elements is "right", produces nothing, since one condition is just as easy to check as the other.

And if we do introduce subspaces we need the general defintion to explain that the full space and its subspaces are examples of the same thing. Really, the most important question is what examples to introduce besides these. The simple rule is: abstraction must produce something. Spaces of matrices, and traceless matrices, whatever, seem rather unproductive. By contrast, applying the general theory to polynomial spaces or solution spaces of differential equations produces insight into various concrete problems. Interpolation, existence of particular solution, etc.

Linear Mappings. Eigenvectors. Quadratic Forms.

A few decades ago linear mappings disappeared. They were replaced by their matrices altogether, because matrices are more "elementary" or "concrete". This, of course, is the common fallacy of confusing vectors and coordinates. Also it represents the misconception of computation being more concrete than vision.

A careful analysis has convinced me that the main trouble students have with linear mappings is the *passage* from mappings to matrices. The reason is that they are used to practising algorithms after being told exactly what to do. They are not used to distinguishing between definitions and theorems, for instance, a fact that turns elementary ideas into utter confusion.

The punishment for emphasizing the matrix over the mapping is immediate. We can no longer explain, let alone illustrate, the concept of eigenvectors. A rotational axis is no longer an eigenvector of a rotation. Its coordinates (in some God-given basis) are, and they again belong to a matrix, not to the geometric object.

But the coordinates of one given eigenvector can be exactly anything but zero!

It also becomes logically impossible to discuss diagonalization. We no longer simplify the description of one given geometric object. We change a matrix into something "similar". The students are expected to swallow the idea of an eigenvector being a GL_n–orbit of columns compatible with a GL_n–action (by conjugation) on a square matrix!

Of course, many important mappings are not geometric at all. An important situation is the shift or differentiation operator acting on the solution space of a difference or differential equation. Where's the matrix? It depends not only on the equation but on a choice of basis.

There are at least three equally natural choices! The Jordan basis, the basis derived from the impulse response by shifts or differentiation, and the basis belonging to "standard" initial values. These are the bases producing the most important "normal forms" of matrices: the Jordan form, and the two rational "companion" forms. I believe this example is by far the most important.

Another case in point is quadratic forms. The "concrete" approach is to represent them as $Q = X^tAX$, A symmetric. We find an orthogonal matrix T, with "eigenvectors" for columns, diagonalizing A: $T^tAT = D$. The substitution $X = TY$ transforms X^tAX into Y^tDY. Really!

Over the last 25 years I've asked students to explain (in a numerical example) what *happens* when they compute AT and TA^tT. They don't see it. They don't realize how to read the matrix mutliplications, the first time column-wise, the next time as dot products.

The "abstract" way does not present this difficulty; it also displays the two expressions as different descriptions *of the same thing*.

The quadratic form should be conceived of as a function of a vector agument $\mathbf{u}$ First write $Q(\mathbf{u}) = F(\mathbf{u})\cdot\mathbf{u}$ where F is a symmetric *mapping*; then write $\mathbf{u}$ in terms of an orthogonal eigen-basis. The rest is automatic and the students usually see it. I've tried both approaches by way of an exercise, and the "abstract" (rather, "vector") approach is always the easiest.

This approach also conforms much better to applications in rigid body and solid mechanics where the quadratic form *and* its associated linear mapping have natural interpretations.

Diagonalization doesn't replace energies by simpler energies. It *simplifies their expressions*!

The Spectral Theorem, and its proof

The deepest, and most impressive, theorem in standard courses is the Principal Axes, or Spectral, Theorem. Some books, e.g., Anton, omit its proof. Some deal only with the case of simple characteristic roots. Others use partitioned matrices or appeal to the concept and theory of orthogonal complements.

Again, reference to the *mapping* proves superior. The important point is that the symmetry of the matrix is an orthogonal invariant. Therefore, the equivalence of the matrix property and a mapping property $F(\mathbf{u})\cdot\mathbf{v} = \mathbf{u}\cdot F(\mathbf{v})$, must be proved. The mapping identity holds very little intuitive meaning but it's easy to verify in the case of, e.g., ortogonal projections, reflections and some mappings of physical interest, such as the "inertia tensor".

A proof in 3 dimensions will now contain the necessary ingredients. Those so inclined will supply the general case, by induction.

Find one normalized eigenvector $\mathbf{e}_1$ (existence is easy in 3 dimensions, highly non-trivial in higher dimension). Incorporate it into a suitable ON-basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. By the definition of egeinvectors, and the symmetry of the mapping we arrive at:

$$F(\mathbf{e}_1 = \lambda_1 \mathbf{e}_1$$
$$F(\mathbf{e}_2) = \qquad p\mathbf{e}_2 + q\mathbf{e}_3$$
$$F(\mathbf{e}_3) = \qquad q\mathbf{e}_2 + r\mathbf{e}_3$$

We now have F acting on a two-dimensional space. Find another eigenvector and repeat the process!

Of course, the two-dimensional case admits several instructive proofs which could be given as exercises.

Change of Basis

This is a simple topic usually rendered abstruse by confusing notation and the confusingly arbitrary use of the prepositions "to" and "from". When I learned the subject more than 35 years ago I was fortunate to find some notes by Olov Hanner which presented the formulae in a concise and unified manner. The idea is to introduce row matrices of basis vectors. We can then write a vector as a matrix product

$$\mathbf{u} = \underline{\mathbf{e}} X_e$$

where the first factor (a row) contains the basis vectors and the second, column, factor contains the coordinates. A change of basis may be written

$$\underline{\mathbf{f}} = \underline{\mathbf{e}} T$$

Writing

$$\mathbf{u} = \underline{\mathbf{f}} X_f = \underline{\mathbf{e}} X_e$$

we immediately infer $TX_f = X_e$. Invertibility is immediate from the mere existence of an inverse relationship $\underline{\mathbf{e}} = \underline{\mathbf{f}} U$.

And, automatically, a matrix and its inverse commute!

The real gain is the transition formula for matrices of linear mappings. I leave it to the reader to interpret the following manipulations:

$$F(\underline{\mathbf{e}} X_e) = \underline{\mathbf{e}} A_e X_e$$
$$= F(\underline{\mathbf{f}} X_f) = \underline{\mathbf{f}} A_f X_f$$

whence

$$A_f = T^{-1} A_e T$$

by substitution.

I've given this question in several exams. It was clear from the solutions that the students didn't know the proof by heart. They found it.

Most texts don't give any revealing examples or exercises. A nice example to try is the equation and normal vector of a line in a plane. One basis could be orthonormal, the other oblique. It is easy to check the equation of a line in either basis by intersecting it with the axes. A neatly drawn diagram will convince the students that the equation and the normal vector do not transform the same.

Other nice examples are, in which points does a given line intersect a tetrahedron; how many sides of a box are visibly from an exterior point; does a given point lie in the shadow of a given triangle (a light source given at some other point).

Rank and Nullity

This is the first topic I would simply drop if I didn't have enough time or bright enough students. I know of no other topic that lends itself so naturally to meaningless and incomprehensible juggling with numbers, sometimes row operations sometimes columns operations, sometimes both, please, when do we do which, what are the rules?

The subject cries out for applications, most of them inaccessible by the matrix approach. This is the natural point to introduce other spaces than $\mathbf{R}^n$ and its subspaces.

The rank-and-nullity, or dimension, theorem, is there to prove existence, for instance of interpolation formulae for polynomials or particular solutions to differential equations. If there is no time for such illustrations, then I don't see at all the point of dwelling on this item. Maybe it should be saved for last, especially if some of the students are dragging.

This finishes my list. I believe that textbooks, and lecturers, incorporating at least some of my suggges-tions will go a long way towards turning students into doers and thinkers, rather than slow and capricious computers.

Peter Hackman
Lecturer,
Applied Math,
Linköpings Tekniska Högskola,
SE 581 39 Linköping, SWEDEN
email: pehac@mai.liu.se