Exploring Mathematical Theory Building Through Canonical Matrix Forms

J. Sylvestre
Assistant Professor, Mathematics
University of Alberta, Augustana Campus

Introduction. The purpose of this vignette is to share how the investigation of canonical matrix forms can be used to highlight a rich variety of mathematical theory building philosophies and techniques at a relatively early stage of the undergraduate mathematics curriculum. Focusing on the matrix version instead of the equivalent linear transformation version helps make this topic accessible in a second course in linear algebra.

Taking a *problem-first* approach as opposed to the traditional *theory-then-applications* approach allows the students to see theory as arising naturally out of the study of problems. Furthermore, carrying the study of a single problem over an entire unit (perhaps three to four weeks in duration) helps students appreciate the process of mathematical investigation as a sustained, multifaceted endeavour, as compared to the *solve-this-homework-problem-then-on-to-the-next-one* nature of the typical activities assigned in lower level undergraduate mathematics courses.

The exposition of canonical forms given in the *Outline* section below is in no way novel. But what I believe is of interest is the variety of problem-solving philosophies and heuristics that can be exposed through this topic, as well as the pedagogical choice to bring these to the fore while allowing examples and discovery to lead to development of theory.

The problem. For a given n, the similarity relation is an equivalence relation on the set of $n \times n$ matrices over $\mathbb{C}$. What is the "simplest" representative member of any given similarity class? And what should "simplest" even mean?

The process. In carrying out a step-by-step investigation to attempt to provide an answer to this problem, students discover that the answer can take on many forms. Along the way, they are exposed to many common mathematical problem-solving and theory-building philosophies and techniques, such as

- initial consideration of special cases before tackling the full problem,
- generalizations of previous results,
- attempting to apply old methods to new situations and adapting as necessary,
- decomposing an object to analyze it by pieces,
- special analysis of "cusp" cases, on the boundary between two types of behaviour,
- using indirect methods of gathering information, and
- collecting results into a general theory encompassing all cases,

all in one self-contained unit.

Motivation. Augustana Campus is a small, liberal arts, residential, undergraduate-focused teaching campus of the University of Alberta, located in rural Camrose, Alberta, approximately 100 kilometres southeast of the main campus (Edmonton). Due to the size of the campus (approximately one thousand students), Augustana offers a combined Mathematics and Physics (MAP) program, rather than standalone programs in each of these two disciplines. As well, attracted to the liberal arts model, Augustana students bring a diversity of interests to their studies. One way in which this diversity manifests itself is in the many combinations of major and minor programs of study an instructor can encounter in his or her class list. It is common to have a number of history or music or psychology or biology or drama majors enrolled in second- and third-year mathematics courses as part of a mathematics minor.

For these two reasons — the applied mathematical slant to the MAP program and the diversity of student interests — a dry, rigorous, theory-first approach to linear algebra is not suitable. Even more so since only an extremely small number of our students will go on to pursue graduate studies in mathematics.

On the other hand, the structure of the MAP program at Augustana is such that the courses *Linear Algebra II* and *Introduction to Group Theory* (both at the second-year level) represent the end-of-the-road for the study of abstract algebra at Augustana Campus, aside from the possibility of an individual directed reading course. This is true for MAP majors and mathematics minors alike. It is mainly for this reason that I initially chose to develop a unit on canonical matrix forms for *Linear Algebra II*, as I felt this topic to a be a particularly rich one, offering an invitation to many beautiful and abstract, though surprisingly accessible, concepts in linear algebra. However, my first attempt at delivering this topic did not go well. [*See:* the *Student experience* section below.] After reflecting upon this initial failure and choosing not to give up on the topic, I decided to try a more problem-centred approach. As I reworked the course material, I discovered the wonderful opportunity that this topic allows to bring problem solving philosophies and heuristics to the fore, where examples lead to theory. I believe that this approach is ultimately more valuable to the education of my students.

Inspiration. Some of the inspiration for developing this approach came from experiences in teaching two other courses.

Topics in Geometry. This is a course on axiomatic geometry at the second-year level. Being not particularly well-versed in the subject, when I was first assigned to teach this course in the fall of 2009 I dutifully followed the presentation in the book I had chosen: a coil-bound set of course notes developed by colleagues at another institution. These notes were a rather dry compendium of axiomatic Euclidean geometry. Likely a more experienced instructor could make this material come alive in class, but I could not — part-way through the course I realized that I was even boring myself! With another crack at the course in the winter of 2012, I chose a different textbook, and was determined to find a way to make the material compelling. The manner in which the material was developed in the new book provided the means to achieve this. Neutral geometry was first developed as thoroughly as possible, and then the real fun started: Euclidean and non-Euclidean geometry were developed somewhat in parallel, with the relatively staid and intuitive Euclidean geometry acting as a foil for all the weird and wonderful aspects of non-Euclidean geometry.

Real Analysis. On two occasions I have supervised a directed reading course consisting of a basic introduction to real analysis at the third year level, a subject that does not appear in the Augustana course calendar. I was very much inspired in my own teaching by the presentation of the material in the book that I settled upon for this course. To motivate the development of theory and to justify the level of care and rigour taken in this development, the author opened each topic with an interesting, counter-intuitive example, the investigation of which would require new tools and techniques. Then the author ended each topic with a historical vignette on the original development of the theory, often highlighting the pitfalls of being too lax with rigour.

Both of these experiences inspired me to try to find more compelling and enticing presentations of the material in all my courses. The current vignette describes the result of an attempt to do this for one topic in linear algebra.

Student experience. In the winter term of 2011 I delivered the topic of canonical matrix forms to my *Linear Algebra II* students in the traditional fashion, dutifully running through all the prerequisite theory before getting to the actual topic itself. Aside from a handful of particularly strong students, the majority were lost, adrift in a sea of abstract facts that had no connection to anything of any relevance to them.

In the winter term of 2013 I changed tactics completely, delivering the topic in the manner outlined below, with very different results. Having a central motivating problem underlying the entire topic and returning to the central theme of similarity at the beginning of each subtopic kept the students engaged and focused. Another factor that contributed to the students' ability to engage the material was giving myself permission to drop generality in the theory as a goal, as this is not a goal that is shared by a typical second-year student. In organizing the material around a central motivating problem, I was able to streamline the theory to the purpose of attacking the problem, in the process making it more accessible to the students.

Here are two comments extracted from the end-of-term course evaluations.

> The supplementary notes are great...I found I learned those sections the best.
> — Anonymous student evaluation comment, *AUMAT 220 Linear Algebra II*, Winter 2013.

> I learned to think beyond the question. If that makes sense...
> — Anonymous student evaluation comment, *AUMAT 220 Linear Algebra II*, Winter 2013.

Outline. Here I will provide a brief outline of the subtopics of a unit on canonical matrix forms, as delivered in *Linear Algebra II* at Augustana Campus in the winter term of 2013.

Topic. Diagonalizable matrices.
Philosophy. Start with a simple case.

Students will suggest the zero matrix and the identity matrix as the simplest matrices they know, but the similarity classes of these particular two examples don't take very long to investigate! Soon the students will suggest diagonal matrices, the analysis of which leads to the theory eigenvalues and eigenvectors. The students have likely seen these concepts in *Linear Algebra I* but they are worth reviewing anew.

Topic. Similarity revisited.
Philosophy. Find the essence of what made the analysis of the simple case work.

While eigenvalues and eigenvectors are important in further cases, it is not readily apparent exactly how they will be of practical use. At this initial stage, the most important part of the analysis of the diagonal case is the algebra that *led* to the consideration of eigenvalues and eigenvectors. Repeating this analysis for general pairs of similar matrices, by rearranging $P^{-1}AP = B$ to $AP = PB$ and taking the linear combination point of view of matrix multiplication, we find that *A and B are similar via P if and only if the columns of B encode the action of A on the basis of $\mathbb{C}^n$ formed by the columns of P*. By *encode the action*, I mean that the coordinate vector of $A\mathbf{p}_j$ with respect to the basis $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is precisely the j^{th} column of B, where $\mathbf{p}_j$ is the j^{th} column of P.

Topic. Block diagonal form.
Philosophy. Generalize.

This form can be approached through geometric examples, as well as through the prototypical example of eigenspaces.

$$B = \begin{bmatrix} 3 & 1 & & \\ 1 & 1 & & \\ & & 2 & 1 \\ & & 1 & 2 \end{bmatrix} \qquad A = \begin{bmatrix} 6 & 4 & -1 & -3 \\ 7 & 10 & -2 & -6 \\ 0 & 6 & 0 & -3 \\ 14 & 12 & -3 & -8 \end{bmatrix} \qquad P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 1 & -2 \\ 2 & 4 & 2 & 3 \end{bmatrix}$$

Example 1. How is it that $P^{-1}AP = B$?

We can use the view of similarity from the preceding topic to answer the question posed in Example 1: the crux of the similarity of A and B is that $A\mathbf{p}_1$ and $A\mathbf{p}_2$ lie in $\text{Span}\{\mathbf{p}_1, \mathbf{p}_2\}$, while $A\mathbf{p}_3$ and $A\mathbf{p}_4$ lie in $\text{Span}\{\mathbf{p}_3, \mathbf{p}_4\}$. Thus, we are led to theory of *invariant subspaces* and *collections of independent subspaces*.

Here is an example of attempting to streamline the theory to the problem at hand. At this level and in this context, the concept of *direct sum* is needlessly abstract. The equivalent concept of collections of independent subspaces can be used instead, and is accessible to the students as a generalization of the concept of independent vectors.

Topic. Scalar-triangular form.
Philosophy. Try an incrementally more difficult case.

Students will likely propose (or at least agree) that a triangular form is the natural next step after diagonal. As an incremental step, consider triangular matrices with a single eigenvalue, which I call *scalar-triangular form*.

$$T = \begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & -1 \\ 0 & 0 & 3 \end{bmatrix} \qquad A = \begin{bmatrix} -2 & -14 & 5 \\ 1 & 6 & -1 \\ -2 & -5 & 5 \end{bmatrix} \qquad P = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$$

Example 2. How is it that $P^{-1}AP = T$?

Again, using the view of similarity developed previously, we can see in Example 2 that $\mathbf{p}_1$ must be an eigenvector of A, with eigenvalue $\lambda = 3$. However, $\mathbf{p}_2$ is less cooperative. It *would* be an eigenvector of A for $\lambda = 3$, except for the fact that $A\mathbf{p}_2$ has a nonzero $\mathbf{p}_1$ component. With a little algebra we see that although $\mathbf{p}_2$ does not satisfy the homogeneous linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$, it *does* satisfy $(\lambda I - A)^2\mathbf{x} = \mathbf{0}$. And similarly, we find that $\mathbf{p}_3$ satisfies $(\lambda I - A)^3\mathbf{x} = \mathbf{0}$. Thus, we are led naturally to the theory of *generalized eigenvectors*.

Topic. Triangular-block form.
Philosophy. Boldly forge ahead using whatever tools worked before.

What can be done when A has more than one eigenvalue? Generalized eigenvectors helped us out in the previous case, so why not throw them at this problem and see what comes of it. Carrying this out in examples, we quickly find that generalized eigenspaces are independent and invariant, and so we can put any matrix (over $\mathbb{C}$) into a block diagonal form, where each block is in scalar-triangular form. I call this *triangular-block form*.

Topic. Nilpotent matrices.
Philosophy. Break the problem apart.

We have a triangular form that can be achieved for any matrix. But can this form be "simplified" any further? We don't want to mess up the blocks we already have. However, it is not clear how to choose a basis for each generalized eigenspace to further "simplify" the individual blocks. The scalar part of each block in triangular-block form is as simple as it gets, so it is natural to consider just the messy upper triangular part. To get at the upper triangular part, we break a given block into a sum $\lambda I + N$, and so are naturally led to consider the special case of *nilpotent matrices*.

Topic. Elementary nilpotent form.
Philosophy. Analyze the "cusp" case.

Nilpotent matrices are not all created equal — they can be graded by their *degree of nilpotency*, the smallest positive integer so that $N^k = 0$. The maximum degree of nilpotency an $n \times n$ matrix can have is n, while the zero matrix is the unique matrix with minimum degree of nilpotency of 1. Thus, amongst nilpotent matrices, those with maximum degree of nilpotency could be considered to be furthest from the zero matrix, and on the *cusp* between nilpotency and non-nilpotency. For this reason, they are a natural first case of nilpotent matrix to consider. It turns out that all such matrices are similar, and that this similarity class contains a matrix with a particularly simple triangular form that I call *elementary nilpotent form*. An example of this form is the matrix N below. (Here we switch to a lower triangular form to make things work out more cleanly.)

$$N = \begin{bmatrix} 0&0&0&0\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0 \end{bmatrix} \qquad A = \begin{bmatrix} 1&2&0&-1\\ 3&7&-1&-4\\ -1&9&-2&-4\\ 6&10&-1&-6 \end{bmatrix} \qquad P = \begin{bmatrix} 1&1&1&1\\ 1&2&1&1\\ 0&0&1&-2\\ 2&4&2&3 \end{bmatrix}$$

Example 3. How is it that $P^{-1}AP = N$?

Once again applying our view of similarity, we see that the answer to the question in Example 3 is that the columns of P satisfy $\mathbf{p}_j = A^{j-1}\mathbf{p}_1$, and so we are led to the theory of *cyclic subspaces*.

Topic. Triangular-block form for nilpotent matrices.
Philosophy. Gather information indirectly.

Rather than dive headlong into the theory of *cyclic decomposition*, I prefer an indirect approach from this point on. Students should have no trouble believing that general nilpotent matrices can be put into a block form where each block is in elementary triangular form. If we insist that the blocks appear in order of decreasing size (or increasing, if you like), then any such matrix is completely determined by the number of blocks of each size, and a more detective-like approach can be taken: given a nilpotent matrix A, can we determine the corresponding block form matrix N without computing a corresponding P matrix?

Each block of N has rank one less than full, and taking increasing powers of N decreases these block ranks by one for each step increase in the exponent. Furthermore, the rank of a given block will first vanish in the power of N corresponding to the size of the block. These are all properties easily seen in examples of the form matrix.

From these observations, and since corresponding powers of N and A have the same rank, we can completely determine N by considering the sequence $\operatorname{rank} A, \operatorname{rank} A^2, \ldots, \operatorname{rank} A^{k-1}$, where k is the degree of nilpotency of A.

Topic. Jordan canonical form.
Philosophy. Develop a general theory encompassing all cases.

Finally, applying the results of the nilpotent case to the nilpotent part of each block in triangular-block form, we arrive at the Jordan canonical form. And similar to the general nilpotent case, we can use indirect methods to determine the Jordan form of a matrix A without having to calculate a corresponding P matrix, by accumulating the following information: the eigenvalues of A and their multiplicities, and the ranks of the powers of the matrices $A - \lambda_j I$ for each eigenvalue λ_j.

Conclusion. I believe that one of the most valuable contributions that we as educators can make to our students' intellectual growth is to provide opportunities to engage in many different modes of thought, investigation, and problem solving, including reflection on the *processes* of problem solving and theory building themselves. A unit on canonical matrix forms affords such an opportunity, and I am eager to engage with the mathematical education community to discover other topics in undergraduate mathematics that may do the same.

Note. This vignette is based on an article accepted for publication in *Problems, Resources, and Issues in Mathematics Undergraduate Studies (PRIMUS)*, titled "A Problem-Centred Approach To Canonical Matrix Forms." In particular, all the examples in this vignette are taken from this article.

The article is available to subscribing institutions at the following address.
http://www.tandfonline.com/doi/full/10.1080/10511970.2013.865691