INCORPORATING PRINCIPLES IN TEACHING LINEAR ALGEBRA

David Easdown
School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia,
david.easdown@sydney.edu.au

KEYWORDS: linear algebra, mathematics education, education innovation

ABSTRACT: In this article we explore and illustrate some general principles that teachers of linear algebra might consider, in building frameworks or scaffolds for imparting ideas and techniques that engage with students.

INTRODUCTION

Teaching and learning linear algebra is problematic for many reasons (see, for example, Cronin and Stewart (2022)). Linear algebra may be the first university mathematics that students encounter involving abstract mathematics, where technical details run the risk of either seeming too far removed from the relevance of everyday life or experience, or becoming tedious or dull, especially when tasks appear to be unmotivated. This article suggests drawing attention to some underlying principles, which might be employed by the instructor and woven into the preparation, scaffolding and delivery of materials. These principles arise by stepping back from the mathematics, zooming out at several possible different stages of the learning process. At the most general level, which may apply to any mathematics learning environment, we discuss a warning of Hanna Neumann (1972) about the delicate role of proofs in teaching mathematics, and the successful approach of Francis Su (2010), utilizing Fun Facts and cultivation of the *mathematical yawp*. We mention and illustrate *The Plateau Principle*, introduced in Easdown (2011), which advises us to exploit a variety of explicit starting points, “plateaus”, for investigations. This is analogous to “standing on the shoulders of giants”, as noted by Isaac Newton (1675), something that is ubiquitous in research mathematics, when relying on theorems, and used extensively, though often only implicitly, in teaching undergraduate mathematics. More specific to linear algebra, and used in Easdown (2023), we discuss *The Linear Algebra Principle*, which advises us to move in straight lines whenever we can, and *The Conjugation Principle*, which provides a general template for obstacle avoidance, used to solve difficult problems, motivated by the notion of a *conjugate* in group theory. This leads to the use of *commutators*, which measure how far operations are from commuting, explaining how a cat manages to land on its feet, and for restoring Rubik’s cube. All examples below are well tested in classes given by the author, and suitable for introductory courses in linear algebra, providing the instructor with physical, hands-on illustrations of key ideas and concepts. The materials on commutators are also suitable for an introductory course on group theory, especially when blended with a first or second course in linear algebra. Many of the examples relate to common, everyday experiences, and can easily be mimed by the instructor, as if doing tasks in the kitchen or on the street, perhaps with the exception of the example explaining why a cat lands on its feet. The cat example, involving conservation of angular momentum, can be modified and simulated using a person sitting on a swivel chair, with the aim of rotating a full circle with feet not touching the ground. This requires a fair degree of physical dexterity, to achieve the full commutator, possibly beyond the ability of the instructor, and makes for an engaging class if one or more of the athletic students can be persuaded to attempt the manoeuvre on a swivel chair under instructions from the teacher.

CULTIVATING THE MATHEMATICAL YAWP

Many of us would find the following statement about the practice of mathematics as being self-evident:

> ***Meta-axiom:*** *Proof is sacrosanct in mathematics.*

However, the eminent mathematician and educator Hanna Neumann (1972) wrote that

> *One can teach it [mathematics] and insight by informal methods, but proofs are useless or worse.*

Of course, she was not suggesting that one should violate the sanctity of proof in mathematics, but, rather, was speaking about teaching or communicating difficult concepts and results to students in a way that engages them and facilitates their learning. The context of the quote was the topic in one of her classes: the surprising and difficult theorem that the row and column rank of a matrix coincide. She gives an anecdote where one of her best students was put off:

> *But then you proved it [the row and column rank coincide] and that spoilt it all.*

Of course, a valid proof cannot "spoil" a result whose truth it establishes beyond doubt, but it can, if not handled carefully or sensitively, damage the perception of the mathematics in the mind of the learner. If the proof does not contribute towards a sense of excitement or discovery, or pique the student's curiosity, then, as Neumann advises, then it may be useless or worse. In the last section of this article, we discuss in some detail the Cayley-Hamilton Theorem, and contrasting ways in which it may be approached, both to illustrate the result and to "prove it". The method is based on principles that we develop in the preceding sections.

Francis Su (2010) emphasizes the importance of intuition and quotes Henri Poincaré (1914):

> *It is by logic that we prove, but by intuition that we discover.*

Su then goes on to explain how he cultivates a learning environment that enables students to make discoveries. He employs a device, which he refers to as the *mathematical yawp*, which, like an "aha moment", is an expression of surprise or delight at discovering the beauty of a mathematical idea or argument. Note that Su is not excluding mathematical proof, just that it should become a source of surprise or delight, or celebration of mathematical beauty. For Su, the principal aim in teaching is to cultivate the mathematical yawp and help transform it into poetry. The word "yawp" appears in a poem by Walt Whitman (1855), and is quoted in the movie *Dead Poets Society* (Shulman, 1989), by the teacher, John Keating, as he inspires students to create poetry and live extraordinary lives:

> *I sound my barbaric yawp over the roof(tops) of the world!*

Su writes about nurturing the yawp through Fun Facts, which may have nothing to do with the topic of the lecture, but whose main point was to broaden students' perspectives and whet their appetite for learning more. Through extensive student feedback, it occurred to Su that nurturing the yawp was more important than "getting through the material". He acknowledges however that it is possible for the yawp and the syllabus to be in harmony. The remaining sections of this paper are written in the spirit of bringing together content, intuition and inspiration, in the teaching and learning of linear algebra. Rather than offering possibly disconnected or disparate fun facts, the modus operandum for nurturing the yawp is to first enunciate some general principles and then instill them through increasingly sophisticated examples and illustrations.

SOME PRINCIPLES IN TEACHING AND LEARNING LINEAR ALGEBRA

The principle below is core to the business of linear algebra, and captures the idea that complex mathematics might be related to or reduced to the simple arithmetic involved in working with lines, their equations and generalisations:

> ***Linear Algebra Principle:*** *Move in straight lines whenever you can.*

Of course, the word "linear" appears as an adjective, so it is natural to enunciate a principle of this kind, recognising that linear algebra has something to do with lines, for which there should be nothing to fear! Straight lines are the simplest curves in geometry, and do not "curve" at all (though the notion of curvature is interesting, for example, on the surface of a sphere, or moving through space in the vicinity of a black hole, and already there are opportunities to pique students' imagination and curiosity). The invention of calculus stems from the idea of approximating a complicated curve by a tangent line at a given point, which generalises to many dimensions. The Linear Algebra Principle, discussed early on, points to the fact that the two pillars of first year mathematics, linear algebra and calculus intertwine, even though presented typically as separate disciplines.

The next principle is generic in nature, and applies to all branches of mathematics:

> ***The Plateau Principle:*** *Look for and be prepared to use a variety of plateaus as starting points for an investigation.*

Here *plateau* is a metaphor for a mathematical result or technique that stands out, sitting high above the mathematical landscape or jungle. Because of its height above the scenery, by reaching the plateau, one is empowered, surrounded by the extensive view and panorama of possibilities. How does one reach the plateau, or the soft powder snow on top of a rugged and crevasse-ridden glacier? We can use a "helicopter", in our mathematical imagination. We don't need to work through details of proofs, clambering through dangerous rock or navigate around ice hazards and fissures. The statement of a theorem or description of a technique or algorithm can behave like a plateau, which we can land on and use. A mathematician invokes the Plateau Principle every time he or she uses a theorem without thinking about why the theorem is true. When one drives a car, or uses a computer, one does not need to think about the details of what lies under the bonnet, or the circuitry, or the programming language behind the software. There may be the time or circumstances where it is necessary to delve inside, but mathematicians develop an ability to change focus, and, when it suits, operate at a higher level, which takes advantage of intelligent packaging of information. This principle is often used implicitly in mathematics teaching, especially in linear algebra and calculus, where we avoid too much detail or rigour, and focus on heuristics and intuition. An example is given in the final section, exploring the Cayley-Hamilton Theorem, by combining the Linear Algebra and Plateau Principles, with the next principle, which describes a general form of "obstacle avoidance":

> ***The Conjugation Principle:*** *To do something difficult, change position so that things get easier, and then return.*

Shortly, we will convert this statement in words to a symbolic equation involving operators or operations. "Getting easier" in the context of linear algebra, might be interpreted as "going straight ahead", or "following the simple path of a line", so that the Conjugation and Linear Algebra Principles work in tandem. In the spirit of John Keating's exhortation, and Francis Su's desire to transform the mathematical yawp into poetry, we offer the following poem:

Obstacle Avoidance:

If you wish to make a cake,
Then find an oven in which it may bake.
But if, on the hand, you prefer ice-cream
Then the refrigerator may be the answer to your dream.

The difference between the foolish and the wise
Is sometimes the ability to compromise,
For, if the feelings of others you need to placate,
The solution may be simply to conjugate!

The second verse might seem cryptic, but the underlying idea is illustrated below. In Figure 1, the candidate goes for a job interview but is unsuccessful. In Figure 2, the candidate is successful, realizing before entering the interview room, it might be prudent to dress appropriately to impress, and puts on a tie. Having succeeded in getting the job, the tie can be removed. The only difference in the diagrams is the insertion and deletion of the tie, before and after the job interview, with a totally different outcome.

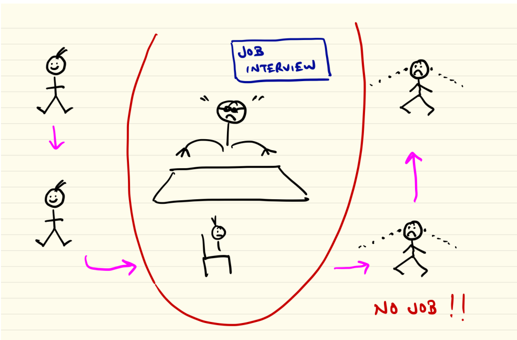

Figure 1: The difficult task of getting a job

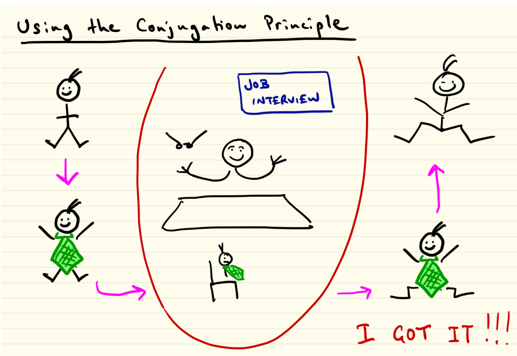

Figure 2: Using the Conjugation Principle to solve a difficult task

Symbolically, we can represent the difficult task of getting a job by Z, and the act of putting on a tie by X. The easier task of getting the job appropriately dressed may be represented by Y. Then, the equation $Z = XYX^{-1}$ describes the solution to the task, which then becomes a general abstract symbolic representation of the Conjugation Principle:

The Conjugation Principle: $Z = XYX^{-1}$ *where* Z *is difficult,* Y *is easy, and* X *is an invertible transformation that changes position or circumstances.*

Here, as usual, X^{-1}denotes the inverse process to X, that is, the result of "undoing X". The expression XYX^{-1} on the right-hand side is known as a *conjugate* (of Y) in group theory. This can be interpreted as actions that avoid or circumvent an obstacle: suppose you want to get from point A to point B, but there is an obstacle in the way. One could first step to the side from point A, using a transformation X, say, so that one can then move straight ahead (as advised by the Linear Algebra Principle!), by performing Y, say, which is easy, and then, at an appropriate moment, undo X, that is perform X^{-1}, in order to arrive at point B. The entire movement becomes Z, ostensibly difficult, but broken down into the easy sequence XYX^{-1}.

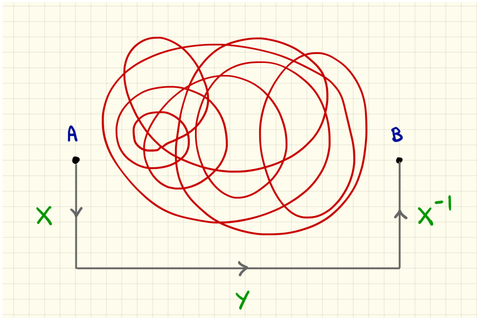

Figure 3: Using the Conjugation Principle to avoid an obstacle

When one becomes aware of this formula, one starts to see it everywhere. In the first verse of the poem, reference is made to making a cake. It is difficult for a cake to bake, left on the preparation bench. By putting it in the oven, using $\boldsymbol{X}$, it is then easy for the cake to bake, inside the oven, say $\boldsymbol{Y}$, and then, finally the cake can be taken out of the oven, returned to the bench using $\boldsymbol{X^{-1}}$. Thus, if the entire operation of baking a cake is $\boldsymbol{Z}$ then we have the decomposition $\boldsymbol{Z} = \boldsymbol{XYX^{-1}}$. The same principle applies to making ice-cream, though we use a refrigerator rather than an oven. Another example, easy to mime in a classroom, is travelling on a train, in order to go between stations. It is difficult to get between stations if one stands still on the platform. When the train arrives, one can shift position with $\boldsymbol{X}$, say, which is to step from the platform onto the train. It is easy then to stand still while the train transports you to the next platform, by $\boldsymbol{Y}$, say, and then one performs $\boldsymbol{X^{-1}}$, to step off the train onto the destination platform, the entire operation being $\boldsymbol{XYX^{-1}}$.

The process of diagonalising a matrix M, part of a final topic on eigentheory in a first course in linear algebra, is a prototypical application of the conjugation principle. If M is diagonalisable, then there is a matrix, P, say, of eigenvectors such that

$$MP = PD, \qquad (*)$$

where D is the diagonal matrix with corresponding eigenvalues down the diagonal. Equation $(*)$, has the matrix product MP on the left, where M may be a messy or complicated matrix. We can pull P through to the other side, to get the equivalent matrix product PD, where D is diagonal and has been "straightened out", as though M has been passed through a wringer. The Linear Algebra Principle is being invoked; indeed, the underlying motivation of eigentheory is to "move in the direction of straight lines" as much as possible. Equation $(*)$ rearranges to become

$$M = PDP^{-1}, \qquad (**)$$

applying the Conjugation Principle with M, P, D for Z, X, Y respectively. As students discover, otherwise complicated manipulations involving M, such as forming high powers of M, dissolve using the simplicity of diagonal matrices, when exploiting equation $(**)$.

The Conjugation Principle is empowering, and students start to see it everywhere. It enhances the comprehension of apparently difficult or complicated calculations or manipulations. For example, the proof of associativity of matrix multiplication is difficult and typically left as an exercise, or for further reading. Yet, one can present the entire proof (without going through the steps), by highlighting the underlying structure, which exploits the Conjugation Principle. The associative law for matrix multiplication, assuming A, B, C are matrices of compatible sizes says that

$$(AB)C \;=\; A(BC).$$

This equation follows by showing that typical corresponding entries are equal, which amounts to the long calculation in Figure 4 below. Students don't need to go through the steps (though they can later if they wish as an exercise), but realise that the transformation X exploits expansion of brackets, using distributivity (so that X^{-1} involves undoing this by rebracketing). The easy step Y in the middle, invokes associativity of the underlying field elements:

$$a_{ij}\,(b_{jk}\,c_{kl}) = (a_{ij}\,b_{jk})\,c_{kl}\,.$$

Students see, at a glance, the underlying architecture of the proof, and also see that the argument works without invoking field inversion, so that associativity holds in more general settings, where elements of a matrix need not come from a field. (A useful nontrivial application is where elements of a matrix are themselves matrices.)

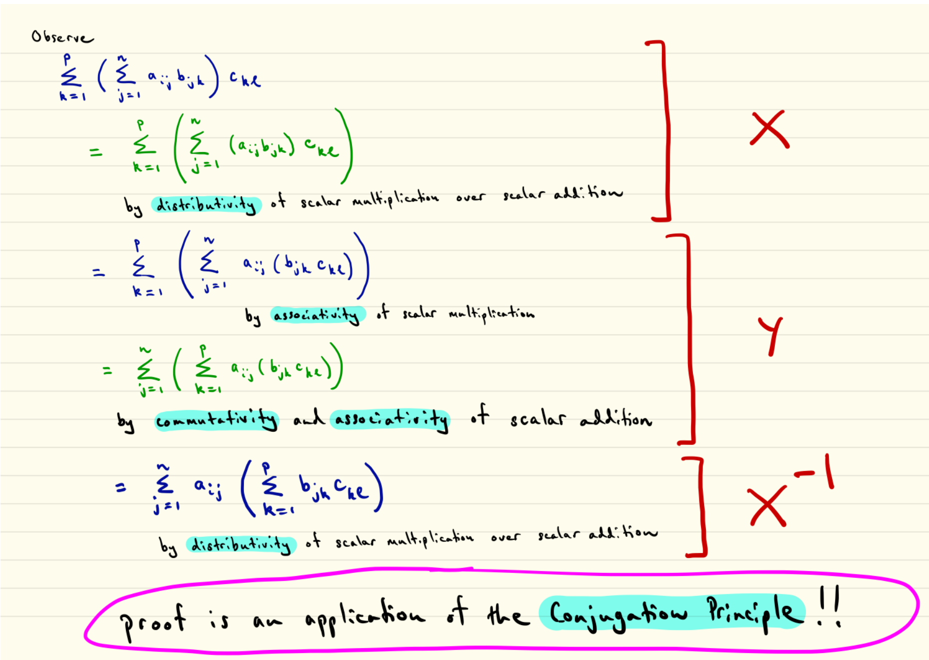

Figure 4: Using the Conjugation Principle to prove associativity of matrix multiplication

As a further, more elevated or abstract, use of the Conjugation Principle, matrices correspond to linear transformations. If X now takes matrices to linear transformations, then Y can be the (obvious) observation that composition of linear transformations is associative, which then converts, under X^{-1} to the associative law for matrix multiplication. Of course, this is only a sketch, as there are technical details involved in setting up the correspondence X, but at least, through the Conjugation Principle, students can understand relationships between mathematical objects at a broad brush-stroke conceptual level.

Another example is the usual recipe book solution to first order linear differential equations of the form

$$y' + p(x)\,y \;=\; r(x)\;,$$

where $y \;=\; y(x)$ is a differentiable function of x and $p(x)$ and $r(x)$ are some particular functions of x. The recipe tells us to create an integrating factor

$$F = F(x) = e^{\int p(x)dx},$$

after which, multiplying through the original equation by F, rearranging the information, including an integration, the general solution can be expressed in the following form:

$$y = \frac{1}{F(x)} \int F(x)r(x)\;dx\;.$$

There is an underlying motivation, relating to using the integrating factor to exploit the Product Rule of differentiation. However, for inexperienced students, the method looks strange or abstruse. At a conceptual level, one can explain it using the Conjugation Principle. At first sight, the original equation looks difficult to solve. However, one may think of multiplying $\boldsymbol{r(x)}$ through by $\boldsymbol{F(x)}$ as applying a "side-stepping" transformation $\boldsymbol{X}$; then one can think of integration as applying a simple transformation $\boldsymbol{Y}$, in fact, the most direct way of "unravelling" a derivative; and then, finally, dividing through by $\boldsymbol{F(x)}$ is an application of the inverse transformation $\boldsymbol{X^{-1}}$. The solution is obtained, using the conjugate $\boldsymbol{XYX^{-1}}$, and the formula above starts to look natural and intuitive.

An elementary but difficult exercise to set students is to find the equation of the line resulting by rotation of a given line, with equation $ax + by = c$, by an angle θ about a given point $P(x_0, y_0)$. The equation of the rotated line turns out to be the following:

$$\begin{aligned}(a\cos\theta - b\sin\theta)x + (a\sin\theta + b\cos\theta) \\ = c + a(x_0(\cos\theta - 1) + y_0\sin\theta) - b\big(x_0\sin\theta + y_0(1-\cos\theta)\big).\end{aligned}$$

There is no chance anyone could reasonably guess this equation, having such complicated mixtures of constants. However, one can work it out systematically and relatively painlessly, by applying the Conjugation Principle. The transformation X translates the point $P(x_0, y_0)$ to the origin $O(0,0)$. One then performs Y, which is rotation about the origin, executed easily using a rotation matrix (or the rule associated with a rotation matrix). By undoing X, that is performing X^{-1}, translating back from the origin to the point $P(x_0, y_0)$, we obtain the final rotated line, and we are done. The reader is invited to work through the details. The Conjugation Principle simplifies an otherwise difficult problem.

CONJUGATIONS IN COMBINATION

The cake baking example has some subtlety. As above, we let X denote the placement of the uncooked cake in the oven, Y denote the act of leaving the cake in the oven for the time specified by the recipe. In the earlier example, we simply considered taking the cake out of the oven to be the inverse operation X^{-1}. However, the cake is hot and care is required to "undo X" We really need to put on oven mitts, to protect our hands, with an operation W, say enabling us to perform X^{-1} safely. Having taken the cake out of the oven, without injury, we then take the oven mitts off, that is, perform W^{-1}. The entire operation Z of making the cake is described by the following equation involving a "conjugate embedded within a conjugate":

$$Z = XYWX^{-1}W^{-1},$$

A chameleon and certain types of large molluscs survive by changing their skin colours. Their survival strategy may be thought of as a cascade of iterated applications of the Conjugation Principle, employing one conjugate after another, doing and undoing skin colours, determined by the environment in which they find themselves.

An important and useful way of combining or compressing two conjugates involving operations X and Y is to form the so-called *commutator*

$$C = XYX^{-1}Y^{-1}.$$

Notice that C is the result of intertwining or overlapping the conjugates XYX^{-1} and $YX^{-1}Y^{-1}$. The terminology arises because C can be thought of as a measure of how far the operations X and Y are from commuting, because of the following equation:

$$XY = (XYX^{-1}Y^{-1})YX = CYX.$$

If X and Y were to commute, then C would "disappear" (become the identity operation). It is common in mathematics (and in real life) for operations to be far from commuting, so that the commutator becomes nontrivial and interesting. This can be exploited to solve practical problems, and commutators are renowned for "the art of killing two birds with one stone", where a single action can apparently achieve two things at once. For example, a truck driver relies on commutators to make a living, in moving goods back and forth between destinations. In the following diagram, a truck driver picks up goods A from Shed 1 and delivers them to Shed 2. The truck switches platforms at Shed 2 in order to pick up goods B, which it then delivers to Shed 1 on the return journey. By switching platforms again, the truck is back at the platform in Shed 1 for goods A, in order to repeat this process. We can denote the entire operation starting and finishing at Shed 1 by Z. The action X_1 moves the truck to the front of whichever shed it happens to be in. The inverse action X_1^{-1} undoes this, interpreted as taking the truck from the front of a given shed to the platform with the goods labelled on the back of the truck, whatever that label happens to be. The action X_2 is the process of moving from the front of Shed 1, facing outwards to the front of Shed 2, again facing outwards. Let Y denote the action of switching from platform A to platform B, in whichever shed the truck happens to be. Thus, we have the following equation:

$$Z = (X_1X_2X_1^{-1})\,Y\,(X_1X_2^{-1}X_1^{-1})\,Y^{-1} = X\,Y\,X^{-1}\,Y^{-1},$$

where $X = X_1X_2X_1^{-1}$, so that Z is a commutator. If Z^n denotes the effect of composing n instances of Z, then, in executing this, the truck driver carts n instances of goods A from Shed 1 to Shed 2, whilst bringing back n instances of goods B from Shed 2 to Shed 1.

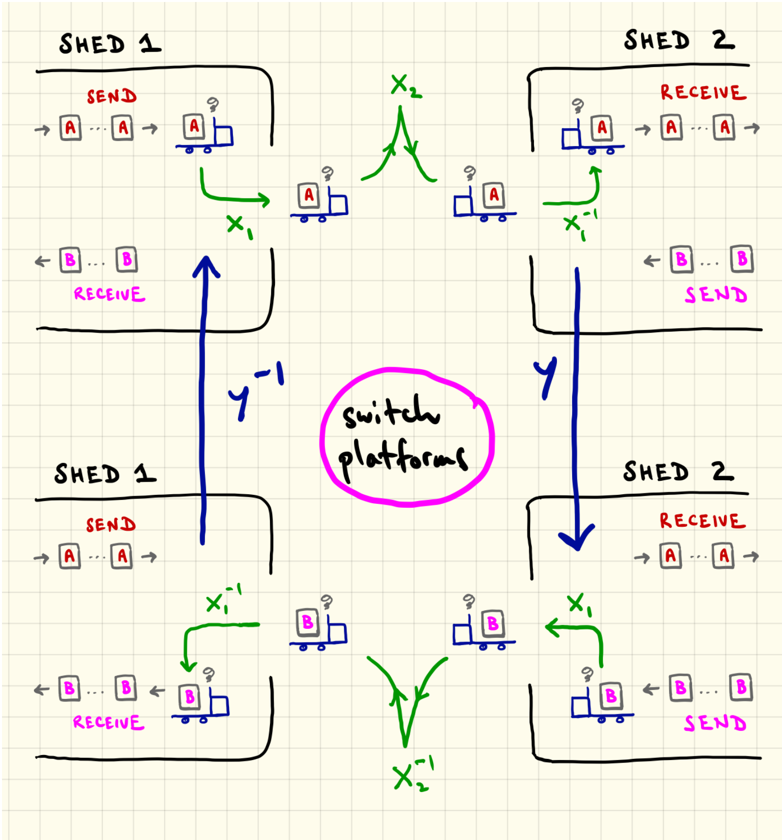

Figure 5: Using commutators to truck goods between destinations

Note the important role of "switching platforms" denoted by Y and Y^{-1} in the commutator Z. This is an important heuristic to keep in mind in applying this technique to other settings.

Let's adapt this idea to explain how a cat lands on its feet. It uses commutators and the conservation of angular momentum (exploited also by ballet dancers). In the diagram below a cat is falling, presently flat on its back. Notice that its front paws are held in close to its spine, its axis of rotation, whereas its back paws and tail are pushed out, as far away as possible. Suppose the cat rotates its spine so that the front part of the cat rotates clockwise ninety degrees. To compensate, in the preservation of angular momentum, the back of the spine rotates anticlockwise, but a lot less, say, in this example, half as much, forty-five degrees. Call this particular twist of the spine X. If the cat undid this move immediately, performing X^{-1}, then the cat would be falling exactly as before. To make progress, before performing X^{-1}, the cat "switches platforms", so to speak, pushing the front paws as far away as possible, and bringing the back paws and tail in as close as possible, to the axis of rotation. Call this movement Y. Now, by performing X^{-1}, undoing the twist of the spine, to conserve angular momentum, the front part of the cat only rotates forty-five degrees anticlockwise, whilst the back of the cat rotates ninety degrees clockwise. So far, the cat has performed the conjugate XYX^{-1}, which is significant progress. However, to get back to the original configuration of arms and legs, the cat has to "switch platforms again", by undoing Y, that is, performing Y^{-1}, which brings the front paws close in to the axis of rotation, and the back paws and tail far away. This completes the performance of the commutator $XYX^{-1}Y^{-1}$, and the cat has achieved a forty-five-degree clockwise rotation. Doing this four times in succession, the cat lands on its feet! The angles were chosen to illustrate the underlying mathematics, though the cat is so flexible that it can use far greater angles to land on its feet in fewer steps.

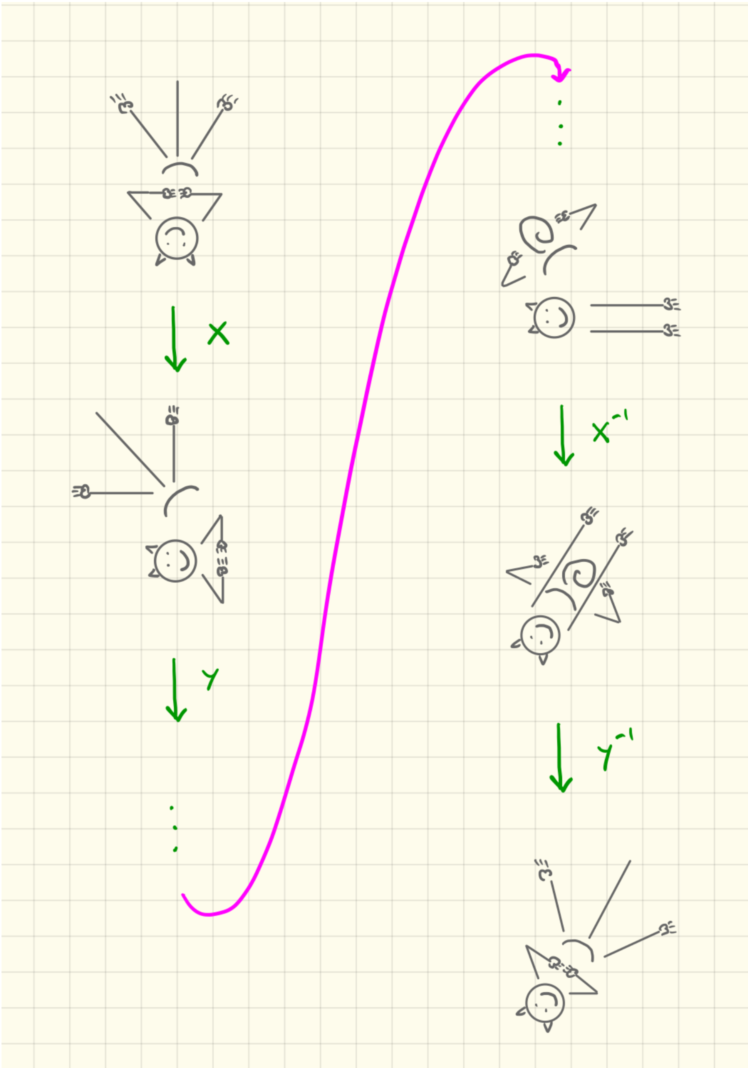

Figure 6: Using commutators to land on ones feet

Let's apply this technique to restore Rubik's cube, where the only thing needing fixing is that two edge cubies are in the wrong orientation. We need to (i) reorient the middle yellow/blue faces of one of the edge cubies (analogous to a truck driver carting goods A from Shed 1 to Shed 2) and (ii) reorient the middle yellow/red faces of another edge cubie (analogous to the truck driver bringing goods B back from Shed 2 to Shed 1). Suppose we can find a transformation X that achieves (i) at the expense of mixing up the bottom two layers. If we undo X immediately, that is perform X^{-1}, then we go back to where we were without making progress. So, before performing X^{-1}, we first perform Y, analogous to "switching platforms", which positions the other edge cubie in the place of the first edge cubie, so that undoing X, will reorient the second edge cubie, at the same time automatically restoring the bottom two layers of the cube. To get the final solution, it remains to undo Y, that is, perform Y^{-1}, to restore the original positions of the edge cubies on the top layer, and then the entire cube is fully restored. The entire solution is represented by the commutator $XYX^{-1}Y^{-1}$.

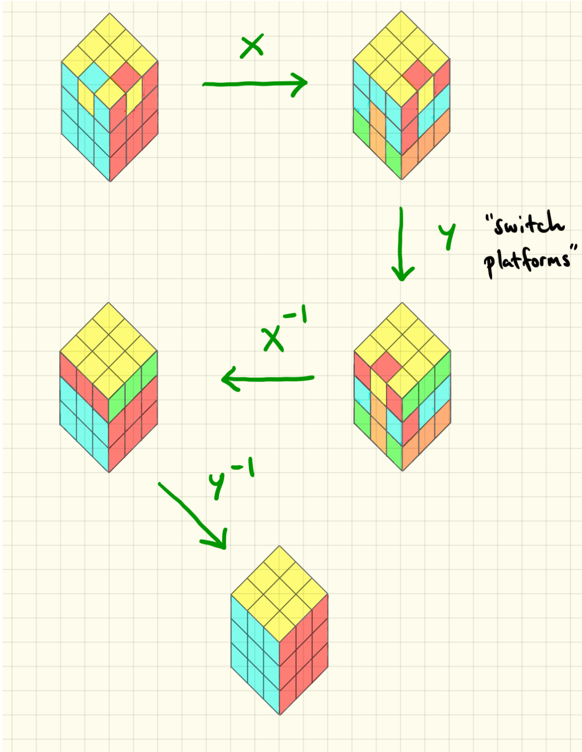

Figure 7: Using commutators to restore Rubik's cube

It remains to find the transformation X. This can be achieved using the Conjugation Principle twice, first to bring the edge cubie from the top layer to the bottom layer, without disturbing any other cubies in the top layer, and then secondly to put it back in the top layer in the correct orientation. (The details of finding these two conjugations are left as a nice exercise.) Note that there are two things needing fixing, and commutators "kill two birds with one stone". This is related to the fact that, when represented by permutations of faces of cubies, commutators become even permutations. (An odd permutation of cubie faces is not possible, the reason for which is a nice point of departure for a discussion with students, leading to interesting and nontrivial group theory.)

THE CAYLEY-HAMILTON THEOREM

We complete the discussion by giving an extended example to illustrate how the principles mentioned above can be combined in a novel way in a teaching or learning activity. Consider the following celebrated theorem:

The Cayley-Hamilton Theorem: *Every square matrix is a root of its own characteristic polynomial.*

Symbolically, this is saying that if $\boldsymbol{M}$ is a square $\boldsymbol{n} \times \boldsymbol{n}$ matrix and $\boldsymbol{\chi}(\boldsymbol{\lambda}) = \mathbf{det}\,(\boldsymbol{\lambda I} - \boldsymbol{M})$ is the characteristic polynomial of $\boldsymbol{M}$, then

$$\chi(M) = 0, \qquad (\dagger)$$

where 0 denotes the $n \times n$ zero matrix, with all entries the zero scalar, where M is substituted for the indeterminate λ throughout and the $n \times n$ identity matrix I is used to replace the scalar 1.This is a difficult theorem to prove: the characteristic polynomial may be very complicated, so that evaluation of the left-hand side, using matrix arithmetic, is a formidable task. But nevertheless, students are challenged to find fault with the following short argument:

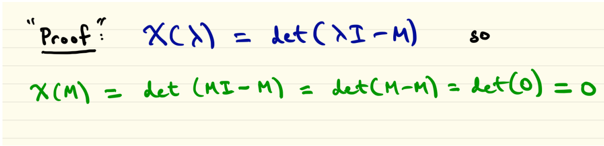

Figure 8: Fallacious one line proof of the Cayley-Hamilton Theorem.

The argument must be fallacious (except when $n = 1$), as the outcome is the zero scalar, whereas $\chi(M)$ is an $n \times n$ matrix. In fact, all the steps are valid, except for the first step, where the matrix is substituted for the indeterminate *before* forming the determinant. Substitution *after* forming the determinant is a complicated process, and the Cayley-Hamilton Theorem is far from obvious!

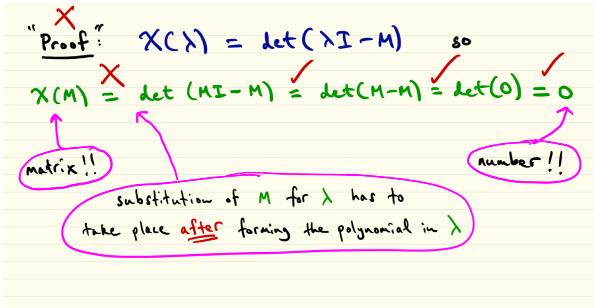

Figure 9: Pinpointing fault in the fallacious proof of the Cayley-Hamilton Theorem.

Here is a valid proof for $n = 3$, and it uses an elegant telescoping sum technique and properties of the adjugate matrix, relying on nontrivial properties of the determinant:

Proof: Write the characteristic polynomial of M as $\chi(M) = b_0 + b_1\lambda + b_2\lambda^2 + \lambda^3$, for some constants b_0, b_1, b_2 and put $B = \operatorname{adj}(\lambda I - M) = B_0 + \lambda B_1 + \lambda^2 B_2$, for some matrices B_0, B_1, B_2 whose entries do not contain any expressions involving λ. By properties of the adjugate matrix, we have

$$(\lambda I - M)(B_0 + \lambda B_1 + \lambda^2 B_2) \;=\; (b_0 + b_1\lambda + b_2\lambda^2 + \lambda^3)I\ .$$

Equating coefficients yields

$$-MB_0 = b_0 I\,,\quad B_0 - MB_1 = b_1 I\,,\quad B_1 - MB_2 = b_2 I\,,\quad B_2 = I\ .$$

But $\chi(M) = b_0 I + b_1 M + b_2 M^2 + M^3$, so that, from the above equations,

$$\chi(M) \;=\; -MB_0 + M(B_0 - MB_1) + M^2(B_1 - MB_2) + M^3 B_2 \;=\; 0\ ,$$

the zero matrix, and the proof of the Cayley-Hamilton Theorem is complete.

Figure 11: Formal proof of the Cayley-Hamilton Theorem when $n = 3$.

However, presenting such a proof in class would go down like a lead balloon, and probably be off-putting to students and lead to similar negative outcomes as described by Hanna Neumann, in her anecdote about proving that the row and column rank of a matrix coincide. A fastidious student can check the above argument step by step, and it is then a nice exercise to adjust it slightly to work for any positive integer n. However, simply checking the steps, in itself, is not likely to lead to *yawp* moment, even though the Cayley-Hamilton Theorem is a sophisticated and highly nontrivial result. There are many possible approaches to "explaining" this theorem in class, especially if one is prepared to use the Plateau Principle, looking for appropriate "plateaus" from which to launch an investigation. The first point to note, is that if we can conjugate M into a simpler matrix D, say

$$M = PDP^{-1} \qquad (\dagger\dagger)$$

for some invertible matrix P then D and M have the same characteristic polynomial $\chi(\lambda)$, which is a nice exercise (or may be considered as one of the "plateaus"). In this case, if one can show that

$$\chi(D) = 0 \qquad (\dagger\dagger\dagger)$$

evaluates to the zero matrix, then

$$\chi(M) = \chi(PDP^{-1}) = P\chi(D)P^{-1} = P0P^{-1} = 0\,,$$

also evaluates to the zero matrix, and the conclusion of the Cayley-Hamilton Theorem will have been proved. Note that, in the middle step of the previous cascade of matrix equations, we are using the fact that conjugation can be brought outside a given polynomial expression of a given matrix D, using the same conjugating matrix P (which is a nice exercise, or regarded as another "plateau"). By the Conjugation Principle then, the proof of the Cayley-Hamilton Theorem has been reduced to establishing (††) and (†††). Now there are many possible starting points ("plateaus"). The simplest case would be that M is diagonalisable, where the characteristic polynomial is a product of linear factors and D is diagonal. The argument is a special case of the argument below, so we won't reproduce it here. In general, of course, matrices are not diagonalisable, but possess Jordan canonical forms. This is a difficult theorem (and usually proved using the Cayley-Hamilton Theorem!), and could be used as a "gigantic plateau" to investigate $\chi(D)$ in the case that D is a diagonal sum of Jordan blocks. This is an opportunity to introduce and discuss Jordan blocks and the issue of extending the field of scalars (say going from the field of real numbers $\mathbb{R}$ to the algebraically closed field of complex numbers $\mathbb{C}$), in order to factorise the characteristic polynomial, involving further "plateaus" (including the Fundamental Theorem of Algebra). A middle ground, say, could be to assume the characteristic polynomial has been fully factorized (whether or not one needs to appeal to something like the Fundamental Theorem of Algebra), and the following theorem (another possible "plateau"):

Theorem: *$M = PDP^{-1}$ for some invertible matrix P and some (upper) triangular matrix D.*

This theorem (related to the so-called *Schur decomposition* of M) can be taken as given, but it is also much easier to prove than the Jordan canonical form theorem (and one could if one wishes, depending on the level of the class, set it as an exercise or assignment task). Suppose then that we only need to consider an (upper) triangular matrix D. The eigenvalues of D (and hence also M) are just the elements down the diagonal, say $\lambda_1, \lambda_2, \ldots, \lambda_n$, which may have repetitions, and then

$$\chi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)\,,$$

so that

$$\chi(D) = (D - \lambda_1 I)(D - \lambda_2 I)\cdots(D - \lambda_n I),$$

which we assert evaluates to the zero matrix. Because of the triangular shapes, with successive matrices having a "zero" gap in successive places down the diagonal, it then becomes visually clear why $\chi(D)$ must evaluate to the zero matrix: as we move through the product from right to left, we see an extra row of zeros added at each step, so that at the final step, all the rows consist of zeros. Here is a visualization, where $n = 5$ and

$$D = \begin{bmatrix} \lambda_1 & * & * & * & * \\ 0 & \lambda_2 & * & * & * \\ 0 & 0 & \lambda_3 & * & * \\ 0 & 0 & 0 & \lambda_4 & * \\ 0 & 0 & 0 & 0 & \lambda_5 \end{bmatrix}$$

so that

$$\chi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)(\lambda - \lambda_4)(\lambda - \lambda_5)$$

with the eigenvalues appearing down the diagonal. Then, using asterisks to denote anything (zero or nonzero), and emphasising zero in bold appearing stepwise down the diagonal we have the following:

$$\begin{aligned}
\chi(D) &= (D-\lambda_1 I)(D-\lambda_2 I)(D-\lambda_3 I)(D-\lambda_4 I)(D-\lambda_5 I) \\
&= \begin{bmatrix} \mathbf{0} & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & \mathbf{0} & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & \mathbf{0} & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & \mathbf{0} & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & \mathbf{0} & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & \mathbf{0} & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & \mathbf{0} & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix}
\begin{bmatrix} * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

If we attempted to display the same calculation with M in place of D, all we would see is asterisks everywhere, except in the final zero matrix, which would appear to come out of nowhere. As the products resolve, one by one, the rank of the final matrix is guaranteed to diminish at least by one at each step, finishing at the final step with a rank zero matrix. This proves the conclusion of the Cayley-Hamilton Theorem for D and hence for M.

There is a pleasant symmetry in the calculation for D, noticing how the number of zero rows increases by one in the final matrix in the matrix product, at each step. If one performed the matrix calculations instead from left to right, the number of zero columns would increase by one in the first matrix occurring in each product, ending up at the last stage with all columns becoming zero. The number of zero rows appearing will match the number of zero columns, in each stage of the respective calculations, reflecting the fact that the row and column rank of a matrix coincide. This would be a further opportunity to reinforce that fact, and create another *yawp* celebratory moment, in something like Hanna Neumann's class discussion about rank of a matrix.

REFERENCES

Cronin, A. & Stewart, S. (2022). Analysis of tutors' responses to students' queries in a second linear algebra course at a mathematics support center. *The Journal of Mathematical Behaviour*, Volume 67, retrieved 8 July 2025: https://www.sciencedirect.com/science/article/abs/pii/S0732312322000554

Easdown, D. (2011). Introduction to Linear Algebra 3rd Edition. Pearson Australia. Frenchs Forest.

Easdown, D. (2023). *Introduction to Linear Algebra.* Coursera. University of Sydney. Retrieved 25 March 2025: https://www.coursera.org/learn/introduction-to-linear-algebra

Neumann, H. (1972). Teaching first year undergraduates: fads and fancies. *Australian Mathematics Teacher*, 29(1), 23-28.

Newton, I. (1675). Letter to Robert Hooke. Retrieved 25 March, 2025: https://digitallibrary.hsp.org/index.php/Detail/objects/9792

Poincarè, H. (1908). *Science et méthode*. E. Flammarion, Paris, 1908. English translation, F. Maitland, *Science and Method*, T. Nelson, New York, 1914.

Shulman, T. (1989). Dead Poet's Society (dir. Peter Weir), Walt Disney Studios Home Entertainment, Burbank, CA.

Su, F. (2010). Teaching research: encouraging discoveries. *The American Mathematical Monthly,* 117(9), 759-769.

Whitman, W. (1855). "Song of myself" from *Leaves of grass.* Collection of Poems. Self-published. Available online, Gutenberg Project, retrieved 25 March 2025: https://gutenberg.org/cache/epub/1322/pg1322-images.html