COMPOSITION OF LINEAR POLYNOMIALS

EDWARD J. BARBEAU

Abstract. In order for students of science, technology, engineering and mathematics to have a proper algebraic foundation, the secondary curriculum must go beyond memorizing results and technicalities. Algebra is a powerful tool that can be wielded only if its user has a sophisticated appreciation of its value as a language, its structure and its role in proving results. In this paper, we look at an example that is accessible to material in the early high school syllabus in which students are encouraged to deepen their grasp of the subject.

1. Introduction

Algebra is the core of the secondary syllabus, both for every educated citizen and for those who need a solid preparation in mathematics for study or employment. Students, particular those headed for a STEM program in university or college, need to appreciate the role of algebraic notation and process as a language for expressing relationships, as a tool for setting up and solving equations and as a powerful means of disposing of a broad range of problems.

STEM students require not only a broader and more technical syllabus, but tasks that shed varied light on the concepts and practices of algebra. There are many subtle distinctions that can only be understood by exposing them to a range of examples that bring out salient features. A helpful viewpoint is to regard an algebraic expression as a bearer of information, of which some mathematical facts can be easily read off and some are latent. Technical manipulation of an algebraic expression changes its form so that the particular information needed is readily visible.

In order for this side of algebra to be real for students, it should be presented in a context that means something to them. As argued in earlier chapters, this context might relate to their life and experience. But it can be purely mathematical, if it has its own coherence and is attractive on its own terms. At the secondary level, it is not easy to find applications that are both accessible and deal with aspects of mathematical usage and analysis that students should be exposed to. One might hope that courses like physics, chemistry and geography are sufficiently mathematical as to fulfil some of this necessity.

However, it is in mathematics courses themselves that one should develop skills of investigation, analysis and reasoning. STEM students in particular need to be attuned to the subtleties of notation, usage and reasoning that will make them confident and competent users of the discipline. Accordingly, in following the syllabus, one should avoid the sense of just doing one topic after another, and build in questions to be answered and goals to be reached. The example in this paper

Date: September 28, 2017.

is purely mathematical, and builds on one of the very first topics in the secondary syllabus, namely the study of linear functions and equations.

A linear function has the form $f(x) = ax + b$, where a and b are constants. Students can be asked to create and then evaluate their own linear functions at different values of x. It is helpful to think of a linear function as a transformation, $x \to ax+b$, that takes the real number line to itself. It then becomes natural to ask whether any number x is taken to itself under this transformation. This involves finding the values of x to satisfy the linear equation

$$ax + b = x.$$

Such a value is said to be a *fixed point* for f. This can be investigated using numerical values of a and b as well as by sketching the graph of the equation $y = f(x)$ and seeing if and where it intersects the graph of the equation $y = x$.

The important intellectual step here is to now think of a and b, as *parameters* that stand in for the numerical coefficients, and to express the fixed point in terms of these parameters. It turns out that there are essentially three situations. If $a = 1, b = 0$, then every point is fixed. If $a = 1, b \neq 0$, then there are no fixed points. This can be seen geometrically since the mapping $x \to x + b$ represents a translation of the line by a nonzero quantity. Finally, whenever $a \neq 1$, there is exactly one fixed point.

We now bring in a second strand that will be related to what we have just discussed. If we introduce a second linear function, $g(x) = cx + d$, then we can evaluate one of the two functions at a particular point and then evaluate the second at the result of the first evaluation. In mathematical terms, we consider $f(g(x))$ and $g(f(x))$, the *compositions* of the two functions.

A natural question is whether the order of composing the functions makes a difference to the final result. Students can once again investigate particular cases and perhaps be led to some tentative hypotheses. At the end, we can treat a, b, c, d as parameters and set up and solve the linear equation $f(g(x)) = g(f(x))$.

The idea of composition should not be completely new to the student, expecially if we regard a linear polynomial as implementing a transformation of the real line. Some students will have studied geometrical transformations such as reflections, rotations, translations and dilatations in elementary school and already experienced situations in which the transformation do not commute under composition. Indeed, we can describe $x \to ax + b$ as a dilatation with factor a followed by a translation by distance $|b|$ to the right or left depending on sign.

Despite the attractiveness of this topic, it does not appear to have received much attention in the educational literature. The references include two articles in the National Council of Teachers journal, *The Mathematics Teacher*, and one a quarter century ago in an expository mathematics journal.

Before discussing pedagogical issues any further, I will devote a section just to the presentation of the mathematics. The reader is invited to work through the material and consider how it might be framed for a secondary class. In any situation where an extra topic is introduced, there is a trade-off to be made. The first reaction might be that it is quite enough to get through the prescribed curriculum without introducing any side issues. On the other hand, one should always ask whether there is any value added by having students go through a mathematical experience that may improve their skills and insights to the extent that later topics can be handled more expeditiously.

2. Commuting linear functions: the mathematical issues.

Suppose that we have two linear polynomials, say $f(x) = 3x-2$ and $g(x) = 2x+5$. We can compose them in two ways to form two new functions: $f \circ g(x) \equiv f(g(x)) = 3(2x+5)-2 = 6x+13$ and $g \circ f(x) \equiv g(f(x)) = 2(3x-2)+5 = 6x+1$. As you can see, the order of composing makes a difference to the result. Is it possible to find linear functions for which the result of the two options is the same? In other words, under what conditions do two linear polynomials commute under composition?

Investigation of this question leads us to distinguish two separate roles for letters in algebra, that of *parameters* that play the role of constants and represent for example numerical coefficients, and that of *variables* which represent values in the domain of some function or expression and allow us to link numbers in the domain and range.

To deal with linear polynomials in general, we can denote them by $f(x) = ax+b$ and $g(x) = cx+d$, with a, b, c, d being the parameters. Composing them in both orders leads to

$$f(g(x)) = a(cx+d)+b = acx+ad+b;$$
$$g(f(x)) = c(ax+b)+d = acx+bc+d.$$

At this point, we should make clear what sort of equality is at stake. One perspective is to see $f \circ g$ and $g \circ f$ as new functions created by performing an operation on the pair f and g. Then we may ask under what conditions are the functions $f \circ g$ and $g \circ f$ the same? For this to occur, they must have the same domain and take the same value at each point of the domain: $f(g(x)) = g(f(x))$ for every real number x. (This is usually expressed by saying that $f(g(x)) = g(f(x))$ is an *identity* in x.) Here we are looking for conditions on the parameters a, b, c, d.

A second perspective is to consider the functions f and g to be given (*i.e.*, a, b, c, d represent a particular choice of numerical coefficients), and ask for which values of x we have $f(g(x)) = g(f(x))$. (In other words, $f(g(x)) = g(f(x))$ is a conditional equation for x.)

The equation $f(g(x)) = g(f(x))$ is equivalent to $ad+b = bc+d$. This is notable in that there is no dependence on x, What is the signficance of this? It means that $f(g(x)) = g(f(x))$ for *some* value of x implies that $f(g(x)) = g(f(x))$ for *all* values of x. For a geometrical take on this, observe that the graphs of $y = f(g(x))$ and $y = g(f(x))$ have the same slope, so that they are either distinct and parallel, or they coincide. Thus they have either no point in common or every point in common. A consequence is that if we want to check commutativity of two linear polynomials under composition, we just have to check it for one value in the domain.

The condition $ad+b = bc+d$ that f and g commute is not very informative as it stands, so we transform it so that we can dig out other information more readily. For example, it can be rewritten as

$$(a-1)d = (c-1)b.$$

This is helpful, because we are in a position to collect to one side of the equation terms that pertain only to a and b (*i.e.* to the function f) and to the other side the terms that pertain only to c and d. However, we can do this only if we are sure we are not dividing by 0, so we must take care of that possibility first.

If $a = c = 1$, then the condition is satisfied and we can check that the functions $f(x) = x+b$ and $g(x) = x+d$ commute and that the composite in either order is $x + b + d$. Geometrically, each function represents a translation of the real line. Their composition represents a translation through the sum of the distances of its components.

If $a = 1$ and $b = 0$, then $f(x) = x$, and this commutes with any function. In fact $f \circ g = g \circ f = g$ so that f is an *identity* function. The case $c = 1$ and $d = 0$ is similarly handled.

Finally, if $b = d = 0$, then $f(x) = ax$ and $g(x) = cx$. Geometrically, each function represents a dilatation of the real line, and the factor of the composite dilatation is the product of the factors of the component dilatations.

Excluding these cases, we can now carry out the division and get the condition for commuting in the form

$$\frac{b}{1-a} = \frac{d}{1-c}. \tag{1}$$

For example, if $f(x) = 5x + 2$, then $g(x)$ must have the form $(2d + 1)x + d$ for some real number d. It is straightforward to check that this works.

However, this is not the end of the matter, because the condition turns out to signify a striking relationship between the two functions. Solving the equation $x = f(x)$ for the fixed point of f leads to

$$x = \frac{b}{1-a}.$$

Similarly, we find that the fixed point of g is given by

$$x = \frac{d}{1-c}.$$

Thus equation (1) tells us that the two functions in question commute if and only if they have a common fixed point.

With some reflection, we realize that this is not surprising. If f and g have a common fixed point p, then $f(g(p)) = f(p) = p = g(p) = g(f(p))$, so that $f \circ g$ and $g \circ f$ take the same value at p. But we already noted that this implies they take the same value everywhere, and so are equal as functions.

On the other hand, suppose that $f(p) = p$ and $g(q) = q$. Exluding the possibility that either is the identity function x, we note that we obtain the fixed point by solving a linear equation which has exactly one solution. Thus, p and q are unique. What happens if we take on board the hypothesis that $f \circ g = g \circ f$? Then

$$g(f(q)) = f(g(q)) = f(q)$$

so that $f(q)$ is a fixed point of g. Therefore $f(q) = q$ by the uniqueness of the fixed point of g. But the fixed point of f is also unique, so that $p = q$.

3. Commuting linear polynomials: pedagogical issues

As students advance in mathematics, they absorb greater levels of abstraction, beginning with the notion of number itself. The symbols of algebra which first play the role of placeholders of numbers become entities in their own right, as do functions and polynomials in particular with their own structure. This process will continue as students take on board in their later education such things as vector spaces and groups. A similar evolution takes place in other areas, such as geometry and combinatorics, and the syllabus should be taught keeping the need for such maturation in mind.

The basic technical requirements for this example, involving as it does linear equations, is within the range of a Grade 9 course. But the level of sophistication is high, and teachers who take it on must first take ownership of it by working through the details on their own terms. Then follows a number of strategic decisions as to the goals to be achieved and how to get there. A careful and sensitive approach is needed that works out from what is familiar to the student and allows a steady intellectual progression towards a textured approach that will help them enjoy future success in mathematics. For the teacher has the task of not just securing technical proficiency, but also of fostering judgment as well as precision and critical thinking.

The teacher can be likened to a conductor of an orchestra. The syllabus and the topics in it are the score. It is the conductor who has to bring it to life, a task that requires scholarship, judgment, technical skill and empathy. The conductor has to know the context and bring a unifying vision to the whole, make sure that the level of skill of the orchestra is adequate and communicate his vision to the players and the audience. While the music performed should be true to its composer and its authenticity respected, each conductor has his own particular take that distinguishes the performance.

So it is with the teacher. The mathematics she teaches should be reliably and honestly presented, but she brings into the task her own experiences and analysis, a sense of context and a treatment appropriate to the students in front of her. Before the conductor meets with the orchestra, it is necessary to make a deep study of the score and decide what its essence is. The teacher is in a similar position. She has to live with the mathematics herself so that she gets her own feeling for what is significant and pleasing. Good teaching is done from a foundation of experience – the teacher is a witness to her subject.

From the get-go this example presents a challenge for a typical class. In my experience, the flow of discussion can vary considerably from one group of students to another. The following discussion is not intended to be prescriptive but only to indicate issues that might arise or directions that teachers may find productive. The composition of functions will likely be a new idea, as indeed is the idea of a function itself. Is there something in the student's previous experience that can be formulated in terms of functionality? It is hard to convey composition using words alone, and efficiency of communication demands that some notation be invented. But this is part of what algebra is all about. Even at its most basic level, we need to introduce variables to describe relationships, such as $A = \pi r^2$, and solve algorithmically word problems that would be difficult if left in the realm of arithmetic.

Thus, the first task is to introduce the notation $f(x)$ and provide examples. Then one can discuss $f(g(x))$ and $g(f(x))$ for particular pairs, such as $(f(x), g(x)) = (3x - 4, 7x + 6)$. One should be sure to deal with the special case $f(x) = x$ and make the point that it is an identity, playing the role that 0 plays for addition of numbers and 1 for multiplication. While working out examples of composition can be a way of having the students do manipulative practice, the lesson can be given a bit of direction by noting that $f \circ g$ and $g \circ f$ seem to be generally different and asking students to seek examples for which they are the same. (The idea of a noncommutative operation will not be completely new, as students at this stage should have been exposed to exponentiation and be aware that, for example, 2^3 and

3^2 are different; however, the two directions of exponentiation can agree in special cases, as $2^4 = 4^2$.)

Having students search for examples is a wonderful teaching tool, as it forces then to pay attention to the details of the artefacts they are working with. If would be interesting, at this stage, to see whether pairs of the form $(f(x), g(x)) = (ax, cx)$ and $(f(x), g(x)) = (x + b, x + d)$ emerge. If so, this provides the opportunity to look at the geometric significance of the commutativity.

In general, it will probably be difficult for students to conjure up pairs, so a strategy will be needed to generate them. Why not take a particular example, such as $f(x) = 3x - 4$, and ask what it takes for c and d for g for commute with f? First, this is where the parameter-variable distinction weighs in, and it needs to be understood that we are looking for particular values of c and d to make the commutativity work.

Let us look at this particular example in more detail. The condition $f(g(x)) = g(f(x))$ leads to $3(cx+d)-4 = c(3x-4)+d$. This equation bears information, but it is not clear what we are supposed to do with it. To begin with, we can remove brackets just to see what happens. A small miracle occurs; the terms in x are the same on both sides of the equation, and we are left with $3d - 4 = -4c + d$, which in turn simplifies to $d + 2c = 2$, or $d = 2(1 - c)$. This is the only condition that has to be satisfied, so that there are infinitely many possibilities.

This illustrates a concept that is useful in physics, that of *degrees of freedom*. The two variables c and d represent potentially two degrees of freedom in that, without restriction, we can make independent choices of values for them. However, the equation $d = 2(1 - c)$ is a restriction that ties one to the other. So there is a net of one degree of freedom, and we can choose only one arbitrarily.

It is not a bad idea to check the answer. In particular, will any student realize that any function has to commute with itself, so that taking $(c, d) = (3, -4)$ should work? We know that the identity function x commutes with everything, so that $(c, d) = (1, 0)$ should be also satisfy the restriction. Students should be encouraged to find other numerical values for c and d and verify that they work, and then check $g(x) = cx+2(1-c)$ generally. This is a cheap way of getting them to practice their manipulative skills.

The equation $d = 2(1-c)$ is linear in c and d, so students should be encouraged to plot this line in the c-d plane and identify various points on it.

We are not done with this example. Consider $f(x) = 3x-4$ and a function that commutes with it, say $g(x) = -x + 4$, where the common composion is $-3x + 8$. (Each student can select his very own example other than $-x + 4$, $3x - 4$ and x, perform the following and then display the result before the whole class.) In the standard x-y plane, plot the graphs of the two equations $y = 3x-4$ and $y = -x+4$, or whatever the student picks. If will be noted that the two lines always intersect; ask them to name the point of intersection. In every case, the lines will pass through the point $(2, 2)$. What is the meaning of this? Since the coordinates are the same, the point lies on the line $y = x$. In fact $f(2) = 2$ and $g(2) = 2$, so this provides the incentive to introduce the notion of a fixed point. It is not hard to check that 2 is a common fixed point of $f(x) = 3x - 4$ and $g(x) = cx + 2(1 - c)$.

Further examples can be studied that lead towards the formulation of the conjecture that commuting of the functions $f(x) = ax + b$ and $g(x) = cx + d$ (when a and b are distinct from 1) is equivalent to their having a common fixed point.

Now we are in a position to tackle the general situation: under what conditions do $f(x) = ax + b$ and $g(x) = cx + d$ commute? Follow the analysis in the previous section to analyze the equation $ad + b = bc + d$. This is a place where one has to proceed cautiously to make sure that students maintain control over the connotations of the variables. We might think of $f(x)$ as a given example and see the problem as finding a corresponding restriction of c and d that ensure that f and g commute. Or we can see the matter more symmetrically as a mutual relationship between the functions f and g. It is vitally important first that the teacher think through the situation first on her own in order to, first, decide on her own mathematical perspective, and, secondly, to determine what preparation is needed for the members of her class to handle the situation. She also needs to be prepared for whatever ideas might come from the students themselves, whether they arise from misconceptions or from a competing viewpoint.

A benefit of this situation is that, because it involves a theorem, it takes us into the realm of proof (an area that is often seen only in the context of Euclidean geometry). Ideally, it would be nice if the students were led towards a conjecture about the common fixed point characterizing commuting functions, perhaps through checking out many examples of commuting and noncommuting pairs and making observations. Important things to emphasize include the need for a restrictive hypothesis (that a and c are not to be equal to 1) and the fact that the result is the equivalence of two properties.

Basically, we would begin by solving for the fixed point of each function. Then the structure of the reasoning would be as follows: *f and g commute if and only if* (1) *is true and if and only the expressions for the fixed points of f and g are equal.*

There is an alternative way of proceeding after one has discussed a particular case such as $f(x) = 3x - 4$ above. After looking at the intersections of the pairs of lines, one can move to the general case and look at where the graphs of two commuting functions intersect. As an exercise, students might be required to deal with the general situation: *Suppose that $f(x) = ax + b$ and $g(x) = cx + d$ commute under composition. Determine the intersection of the graphs of $y = f(x)$ and $y = g(x)$, and show that this point lies on the line with equation $y = x$.* This is not an easy question. The abscissa (first coordinate) of the intersection point is found by solving the equation $ax + b = cx + d$ to get

$$x = \frac{d - b}{a - c}.$$

Plugging this into $y = ax + b$ to get the ordinate (second coordinate) of the intersection point leads to

$$y = \frac{ad - bc}{a - c},$$

which does not look at all like the abscissa. But we have not yet fed in the hypothesis that the functions commute. The condition for this is that $ad + b = bc + d$, so $ad - bc = d - b$ and we find that the two coordinates are indeed equal. (As a check, the student might deal with $g(x)$ rather than $f(x)$.)

Again, this gets into the construction of a proof and the invoking of a hypothesis to lead to a desired conclusion.

4. Conclusion.

Since the foregoing represents a considerable investment of time, there are a number of questions that the teacher needs to settle. What group of students are likely to be receptive to it? For whom will it have value? The topic might not be presented to the whole class, but could be pursued by a group of students as a project or provided as enrichment in a mathematics club.

It will be argued that this topic is not on the list of expectations. But this depends on what sort of expectations we are considering. There are expectations of topic and expectations of practice and affect. The Ontario curriculum, for example, mentions several desiderata that are relevant: Problem Solving, Reflecting, Connecting, Critical Thinking, Reasoning and Proving. Such expectations cannot be inculcated in isolation, but are meant to inform the different mathematical topics to be covered and are best realized in a situation where there is particular program of investigation, discovery and proving. The task in this article allows students to "reason, connect ideas, make connections, apply knowledge and skills". It should provide the teacher with the opportunity to assess student understanding of concepts, and possibly provide the students with some enjoyment.

Even if it is not on the syllabus, in prosecuting the example, will it support the syllabus by requiring techniques and concepts that are part of the curriculum? What is the cost-benefit analysis? The cost is not only one of the time of setting up, but of taking students into fairly deep waters that they may not reap the value of until later in their algebraic life. The benefit might be a better connection with the mathematics that makes later learning easier. I argue that, with very simple materials, one raises matters of algebraic thinking and practice that will foster competence and fluency to a degree not normally seen among high school students. Here we touch on the idea of function and their combinations, the role of variables as unknowns, parameters and domain descriptors for functions, investigation and conjecturing, and at the end the proof of a rather interesting result. Above all, it teaches the important lesson that students must pay attention to details and meaning, and not approach algebra as an automaton.

A rich situation like this prefigures aspects of algebra that might not be important at the moment but will emerge as students mature in the subject. In the appendix, I look at how quadratic functions, a later topic in the syllabus, might impinge.

Modern education is often criticized because many students do not know the "basics" and are maladept at any sort of technical task. To be sure, there is truth to this, but it is not the whole story. A more fundamental reason for student difficuly is their misconception of what mathematics is. They seem to feel that it is a body of fixed automatic processes that will lead inevitablly to an answer, and so is something that can be learned solely by rote. Rather it is a way of thinking that requires one to pay attention to structure, be careful about details and check for accuracy, reasonableness and consistency. There is strength in this consistency; while one may look at a mathematical situation in many different ways, each of them supports and enriches the others. Anything that fosters flexible critical thinking in the classroom will strengthen the student; anything that encourages a mindless formulaic approach will work against the student.

5. Appendix

Students should not be deceived into thinking that, for functions f and g in general, the truth of $f(g(x)) = g(f(x))$ for one value of x implies its truth for all x. A simple counterexample that might be possible for Grade 9 students and certainly possible when students learn about quadratic functions is to see what commutes with the square function. If we let $f(x) = ax + b$ and $h(x) = x^2$, and ask for the circumstance under which $f(h(x)) = h(f(x))$, we are led to the condition

$$(a^2 - a)x^2 + 2abx + (b^2 - b) = 0. \tag{2}$$

In this case, the condition that $f \circ h = h \circ f$ requires that the quadratic equation (2) in x is satisfied for all x.

This highlights an important point about polynomials that does not arise in the normal course of events when quadratic equations are taught, and that is what sort of polynomial equation will be satisfied for all values of x. Write the quadratic equation in the form

$$px^2 + qx + r = 0.$$

There are different ways of looking at the situation. If this quadratic is to vanish for *all* x, then it must vanish for *each particular* x. If we make three substitutions for x, then we obtain three homogeneous linear equations for the three variables p, q and r. It turns out that the only solution for this system is $(p, q, r) = (0, 0, 0)$. For example, if $x = 0$, $x = 1$ and $x = -1$, we get

$$r = 0;$$

$$p + q + r = 0;$$

and

$$p - q + r = 0.$$

It is easily found that the three equations are satisfied simultaneously only by $(p, q, r) = (0, 0, 0)$.

Another way of looking at the situation is to note that if not all the coefficients p, q, r vanish, then we have a nontrivial polynomial equation of degree not exceeding 2. We can always solve such an equation and find that there are at most two solutions.

Returning to the requirement that the equation (2) should be satisfied for all x leads to

$$0 = a(a - 1) = 2ab = b(b - 1),$$

which is satisfied only by $(a, b) = (1, 0)$ and $(a, b) = (0, 1)$. The first possibility leads to $f(x) = x$, the identity function which commutes with every function. The second possibity leads to $f(x) = 1$, a constant function that takes the value 1 everywhere. Indeed, $f(h(x)) = 1 = h(f(x)) = 1^2$.

It is however possible that, for particular values of a and b that (2) has two solutions. For example, let $f(x) = 2x + 1$ and $h(x) = x^2$. Then $f(h(x)) = 2x^2 + 1$ and $h(f(x)) = (2x + 1)^2$, two different functions. However, $f(h(0)) = 1 = h(f(0))$ and $f(h(-2)) = 9 = h(f(-2))$.

References

[1] Shmuel Avital and Edward Barbeau *Commutativity of polynomials* College Mathematics Journal **23:5** (November, 1992), 386-395
[2] Walter Ehrenpreis and Michael Iannone *Commutativity of linear functions* Mathematics Teacher **65:3** (March, 1972), 217-220
[3] Charles P. Seguin *Commuting linear functions and fixed points* Mathematics Teacher **58:8** (December, 1965), 702-704

Department of Mathematics, University of Toronto, Toronto, ON M5S 2E4