The Secret Life of the $ax + b$ Group

Linear functions $x \to ax + b$ are prominent if not ubiquitous in high school mathematics, beginning in, or now before, Algebra I. In particular, they are prime exhibits in any discussion of functions. It is perhaps strange, then, that the topic of composition of linear functions does not receive more attention. Especially, the associated group, the affine group of the line, also known as the "$ax + b$" group, along with its group law, is pretty universally ignored. This note will explore some ways in which the $ax + b$ group surfaces unacknowledged in high school mathematics, and some extensions of standard topic which are suggested by this point of view.

1. Translations, dilations, and affine transformations.

Consider the function of translation by a constant a:

$$T_a : x \to x + a.$$

Here a is any fixed number, and x is a real variable.

A main property of functions is that they can be composed. We compute

$$(T_b \circ T_a)(x) = T_b(T_a(x)) = T_a(x) + b = (x + a) + b = x + (a + b) = x + (b + a) = T_{b+a}(x). \quad (Transcomp)$$

In other words, the composite of T_b and T_a is T_{b+a}. The family of translations is closed under composition, and for translation functions, composition mimics addition of real numbers.

We may also consider functions of dilation, aka homogeneous linear functions (or what in more modern mathematical parlance would be called simply linear functions)

$$D_s : x \to sx.$$

Here s is a fixed non-zero real number, and x again is a real variable. If we compose dilations, we get another one:

$$(D_t \circ D_s)(x) = D_t(D_s(x)) = tD_s(x) = t(sx) = (ts)x = D_{ts}(x). \qquad (Dilcomp)$$

Thus, dilations are another family of functions which is closed under composition, and for dilations, composition mimics multiplication of (non-zero numbers.

We may also compose a translation and a dilation. This gives a new kind of function, whose traditional name in high school algebra is "linear function", but which we will try to call *affine function,* or affine mapping or affine transformation. These are the mappings

$$A_{s,a}(x) = T_a \circ D_s(x) = T_a(D_s(x)) = sx + a.$$

We can compose two affine mappings:

$$(A_{s,a} \circ A_{t.b})(x) = A_{s,a}(A_{t,b}(x)) = sA_{t,b}(x) + a = s(tx + b) + a = atx + sb + a = A_{st,a+sb}(x).$$

This shows us that the collection of all affine mappings is closed under composition, and the composition of affine transformations is described by the rule (aka "group law")

$$A_{s,a} \circ A_{t,b} = A_{st,a+sb}. \qquad (Affcomp)$$

In this connection, we note that $A_{1,0}$ is the identity map, and therefore that the composition formula lets us check that

$$(A_{s,a})^{-1} = A_{\frac{1}{s},\frac{-a}{s}}. \qquad (AffInv)$$

Application 1: Solve the linear equation

$$sx + a = c. \tag{$L1$}$$

Solution: In terms of affine mappings, this equation says that $A_{s,a}(x) = c$, so we can find x by inverting $A_{s,a}$:

$$x = (A_{s,a})^{-1}(c) = A_{\frac{1}{s},\frac{-a}{s}}(c) = \frac{c}{s} - \frac{a}{s} = \frac{c-a}{s}.$$

Remark: The equations $x + a = c$ and $sa = c$ are often referred to in introductory algebra texts as "one-step" equations, whereas the equation $sx + b = c$ is called a "two-step" equation, reflecting the fact that we construct $A_{s,a}$ as a composition of two transformations, one dilation and one translation.

2. Conjugation and Fixed Points.

Since the affine transformations do not commute with each other, they act on themselves non-trivially by conjugation. Explicitly, we have

$$A_{s,a} \circ A_{t,b} \circ (A_{s.,a})^{-1} = A_{st,a+sb} \circ A_{\frac{1}{s},\frac{-a}{s}} = A_{t,a+sb-ta} = A_{t,(1-t)a+sb}. \tag{$Conj1$}$$

In particular,

$$T_a \circ D_t \circ T_{-a} = A_{t,(1-t)a}. \tag{$Conj2$}$$

We are interested in the fixed points of affine transformations. Fixed points of transformations behave in a very simple way under conjugation.

Proposition 2.1: Let Φ and Ψ be transformations of some set X. Let F_Ψ denote the set of fixed points for Ψ: $F_\Psi = \{x \in X : \Psi(x) = x\}$. Then

$$F_{\Phi\circ\Psi\circ\Phi^{-1}} = \Phi(F_\Psi).$$

This is proved by a simple calculation.

Of course, the identity transformation leaves all points fixed. It is easy to check that non-trivial translations have no fixed points. Non-trivial dilations have the origin as a unique fixed point: if $D_s(x) = sx = x$ for any $x \neq 0$, then $(s-1)x = 0$, whence $s - 1 = 0$, or $s = 1$. Combining this with the formula above for the conjugate of a dilation by a translation (or just by direct computation), we conclude that

Proposition 2.2: If $s \neq 1$, the affine transformation $A_{s,a}$ has the unique fixed point $\frac{a}{1-s}$.

Corollary 2.3: Assuming that $s \neq 1 \neq t$, the affine transformations $A_{s,a}$ and $A_{t.b}$ commute with each other if and only if they have the same fixed point.

Proof: Two transformations Φ and Ψ commute with each other if and only if the conjugate of Ψ by Φ is Ψ itself. Applying this to $A_{s,a}$ and $A_{t,b}$, and using formula (Conj 1), we see that $A_{s,a}$ and $A_{t,b}$ commute with each other if and only if $(1-t)a + sb = b$, or $(1-t)a = (1-s)b$, or $\frac{a}{1-s} = \frac{b}{1-t}$. But the two sides of the last equation are exactly the fixed points of $A_{s,a}$ and of $A_{t,b}$ respectively.

Application 2: Solve the (general one-variable linear) equation

$$sx + a = tx + b. \tag{$L2$}$$

Solution: In terms of affine mappings, this equations says that $A_{s,a}(x) = A_{t,b}(x)$. Proceeding as before, and inverting $A_{s,a}$, we get $x = (A_{s,a})^{-1} \circ A_{t,b}(x)$. Since the right hand side of the equation also involves x, we have not arrived at a solution. However, we have done something. We can interpret this equation as saying that x is a fiixed point for the transformation $(A_{s,a})^{-1} \circ A_{t,b} = A_{\frac{t}{s},\frac{-a}{s}+\frac{b}{s}} = A_{\frac{t}{s},\frac{b-a}{s}}$. According to Proposition 2.2, this fixed point is

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

Remark: This result can of course be found easily by manipulation of the original equation. However, equation (L2) is known to be considerably harder for students to deal with than equation (L1). Equation (L1) can be solved by "backtracking", or undoing the indicated operations in succession. In transformational terms, this means inverting $A_{s,a}$ by first applying $(T_a)^{-1}$, and then applying $(D_s)^{-1}$. This strategy fails for equation (L2); one must add and subtract terms from both sides of the equation to isolate x on one side, and one must consolidate the terms in x by using the Distributive Rule. The main point of the discussion above is to point out the way in which equation (L2) differs from (L1). Instead of calling for inverting a function, it calls for finding a fixed point, a very different kind of problem.

3. Iteration and Difference Equations.

It is often of interest to iterate transformations. From the equation (TransComp) describing the composition of two translations, it is easy to conclude that for a translation T_a, we have

$$(T_a)^n = T_{na}. \qquad (TransIt)$$

It is just as easy to pass from the formula (Dilcomp) to an expression for the iteration of a dilation:

$$(D_s)^n = D_{s^n}. \qquad (DilIt)$$

Using formula (Affcomp) to find the iteration of a general affine transformation is not as straightforward. However, if we take adapt formula (Conj2) by seting $(1-t)a = b$, then we can write

$$A_{\frac{b}{1-t}} \circ D_t \circ (A_{\frac{b}{1-t}})^{-1} = A_{t,b}.$$

Since conjugation preserves composition, this allows us to generalize formula (DilIt) to conclude that

$$(A_{t,b})^n = A_{\frac{b}{1-t}} \circ (D_t)^n \circ (A_{\frac{b}{1-t}})^{-1} = A_{t^n, \frac{b(1-t^n)}{1-t}}. \qquad (AffIt)$$

Application 3: a) Find a general formula for numbers $\{x_n\}$ defined by the iterative scheme (aka, *difference equation*)

$$x_{n+1} = x_n + b. \qquad (TrDiff)$$

b) Find a general formula for numbers $\{x_n\}$ defined by the iterative scheme

$$x_{n+1} = tx_n. \qquad (DilDiff)$$

c) Find a general formula for numbers $\{x_n\}$ defined by the iterative scheme

$$x_{n+1} = tx_n + b. \qquad (AffDiff)$$

Solution: These difference equations are respectively equivalent to

$$x_{n+1} = T_b(x_n), \qquad (TrDiff')$$

$$x_{n+1} = D_t(x_n), \qquad (DilDiff')$$

and

$$x_{n+1} = A_{t,b}((x_n). \qquad (AffDiff')$$

Consequently, they are solved by

$$x_n = (T_b)^n(x_o) = x_o + nb, \qquad (TrDiffSoln)$$

$$x_n = (D_t)^n(x_o) = t^n x_o, \qquad (DilDiffSoln)$$

and

$$x_n = (A_{t,b})^n(x_o) = A_{t^n, \frac{b(1-t^n)}{1-t}}(x_o) = t^n x_o + \frac{b(1-t^n)}{1-t} = t^n(x_o - \frac{b}{1-t}) + \frac{b}{1-t}. \qquad (AffDiffSoln)$$

Remarks: i) Of course, the third equation (AffDiff) includes both of the others, so the corresponding solution also includes both of the others.

ii) These equations receive a fair amount of attention in some algebra courses, and the difference scheme (TrDiff) is entailed in many pre- or proto-algebraic investigations of "patterns".

iii) The solutions (TrDiffSoln) and (DilDiffSoln) are relatively easy to come by. However, the general formula (AffDiff) presents more obstacles. Whereas the first two applications clearly are simply refined (some might say overly) interpretations of simple mathematics, the solution of (AffDiff) may be a place near where the lines cross, and the transformational solution is in some ways simpler than a direct approach, which requires the summation of a geometric series. Alternatively, this approach can be thought of as showing that the summation of a geometric series, like so much else in mathematics, is bound up with symmetry.

iv) The three difference schemes can be characterized as follows:

a) A sequence $\{x_n\}$ satisfies an equation of the form (TrDiff)
if and only if $x_{n+1} - x_n = b$ (constant differences).
b) A sequence $\{x_n\}$ satisfies an equation of the form (DilDiff)
if and only if $\frac{x_{n+1}}{x_n} = t$ (constant ratios).
c) A sequence $\{x_n\}$ satisfies an equation of the form (AffDiff)
if and only if $\frac{x_{n+1}-x_n}{x_n-x_{n-1}} = t$ (constant ratios of differences).

4. Action on Polynomials.

So far, we have been looking at the composition of affine functions with themselves. However, we can compose them with any functions from **R** to **R**. We can compose them on the right (precomposition) or on the left (postcomposition). The results of these operations show even more dramatically than the group law for the ax_b group the difference order makes when composing functions.

Let

$$P(x) = \alpha_n x^n + \alpha_{n-1}x^{n-1} + \alpha_{n-2}x^{n-2} + \quad . \quad . \quad . + \alpha_1 x + \alpha_o$$

be a polynomial of degree n. Then

$$P \circ T_a(x) = \alpha_n(x+a)^n + \alpha_{n-1}(x+a)^{n-1} + \alpha_{n-2}(x+a)^{n-2} + \quad . \quad . \quad . + \alpha_1(x+a) + \alpha_o,$$

and

$$P \circ D_s(x) = \alpha_n s^n x^n + \alpha_{n-1}s^{n-1}x^{n-1} + \alpha_{n-2}s^{n-2}x^{n-2} + \quad . \quad . \quad . + \alpha_1 sx + \alpha_o.$$

On the other hand

$$T_a \circ P(x) = \alpha_n x^n + \alpha_{n-1}x^{n-1} + \alpha_{n-2}x^{n-2} + \quad . \quad . \quad . + \alpha_1 x + \alpha_o + a,$$

and

$$D_s \circ P(x) = \alpha_n s x^n + \alpha_{n-1}sx^{n-1} + \alpha_{n-2}sx^{n-2} + \quad . \quad . \quad . + \alpha_1 sx + \alpha_o s.$$

These formulas are all expressed in terms of powers of x, except for the formula for $P \circ T_a$. We expand this explicitly for quadratic and cubic polynomials. If $Q(x) = \alpha_2 x^2 + \alpha_1 x + \alpha_o$, then

$$Q \circ T_a(x) = \alpha_2 x^2 + (\alpha_1 + 2\alpha_2 a)x + (\alpha_o + \alpha_1 a + \alpha_2 a^2).$$

Similarly, for a cubic polynomial, $C(x) = \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_o$, then

$$C \circ T_a(x) = \alpha_3 x^2 + (\alpha_2 + 3\alpha_3 a)x^2 + (\alpha_1 + 2\alpha_2 a + 3\alpha_3 a^2)x + (\alpha_0 + \alpha_1 a + \alpha_2 a^2 + \alpha_3 a^3).$$

Given that we can transform polynomials in this explicit fashion by pre- and post-composing with affine transformations, it is natural to ask whether we can use transformations to simplify polynomials and polynomial equations. The answer is a resounding yes! for quadratic polynomials, and yes also for cubics.

A first step in simplification of P is to eliminate the x^{n-1} term. This can in fact be done for a polynomial of any degree: we take

$$a_o = -\frac{\alpha_{n-1}}{n\alpha_n}.$$

Then we find that the x^{n-1} term always vanishes.

For a quadratic polynomial, we get

$$\tilde{Q}(x) = Q \circ T_{a_o}(x) = \alpha_2 x^2 + (\alpha_o + \alpha_1(-\frac{\alpha_1}{2\alpha_2}) + \alpha_2(-\frac{\alpha_1}{2\alpha_2})^2) = \alpha_2 x^2 + (\alpha_o - \frac{\alpha_1^2}{2\alpha_2} + \frac{\alpha_1^2}{4\alpha_2})$$

$$= \alpha_2 x^2 + (\alpha_o - \frac{\alpha_1^2}{4\alpha_2}) = \alpha_2 x^2 + (\frac{4\alpha_o\alpha_2 - \alpha_1}{4\alpha_2}),$$

The simplest thing to do next to simplify $\tilde{Q}$ is to post-compose with $D_{\frac{1}{\alpha_2}}$. This gives

$$D_{\frac{1}{\alpha_2}} \circ \tilde{Q}(x) = x^2 + (\frac{4\alpha_o\alpha_2 - \alpha_1^2}{4\alpha_2^2}) = x^2 + \gamma_o.$$

Finally, if we postcompose translation with $-\gamma_o = -\frac{4\alpha_o\alpha_2 - \alpha_1^2}{4\alpha_2^2}$, we obtain the monomial x^2. We record this result.

Proposition 4.1: Every quadratic polynomial Q is the transform of $Q_o = x^2$ under precomposition by an appropriate translation, and postcomposition by an appropriate affine transformation:

$$Q = A_{s,a} \circ Q_o \circ T_b,$$

for appropriate a, b and s.

Precisely, $\alpha_2 x^2 + \alpha_1 x + \alpha_o = A_{s,a} \circ x^2 \circ T_b$,, where

$$s = \alpha_2, \qquad a = \frac{\alpha_1}{2\alpha_2}, \qquad \textbf{and} \qquad b = \alpha_o - sa^2 = \frac{4\alpha_2\alpha_o - \alpha_2^2}{4\alpha_2}.$$

Application 4: Solve the equation $Q(x) = 0$ for a given quadratic polynomial Q.

Solution: If we have a quadratic equation, $Q(x) = 0$, then Proposition 3.1 says that this is equivalent to

$$A_{s,a} \circ Q_0 \circ T_b(x) = 0,$$

with a, b and s as specified just above. Transforming by $A_{s,a} * -1$, this equation becomes $Q_o \circ T_a(x) = (A_{s,a})^{-1}(0) = -\frac{a}{s}$. Since $Q_o(x) = x^2$, this means that $T_a(x) = \pm\sqrt{-\frac{a}{s}}$. Hence we finally conclude that

$$x = T_{-b}(\pm\sqrt{-\frac{a}{s}}) = -b \pm \sqrt{-\frac{a}{s}}.$$

Remark: If we rewrite this formula in terms of the coefficients α_j of Q, it becomes the quadratic formula. Thus, the quadratic formula may be thought of as a practical reflection of the fact that all quadratic functions can be reduced to the simplest possible one, $Q_o = x^2$, by composing appropriately with affine transformations. In other words, if we know how to find square roots, we can solve any quadratic equation. This is another perspective on the quadratic formula: it reduces the solution of any quadratic equation to the problem of taking square roots.

We turn now to cubic polynomials. If we translate by a_o to eliminate the x^2 term from C, we get and

$$\tilde{C}(x) = C \circ T_{a_o}(x) = \alpha_3 x^3 + (\alpha_1 + 2\alpha_2(\frac{-\alpha_2}{3\alpha_3}) + 3\alpha_3(\frac{-\alpha_2}{3\alpha_3})^2)x + (\alpha_o + \alpha_1(\frac{-\alpha_2}{3\alpha_3}) + \alpha_2(\frac{-\alpha_2}{3\alpha_3})^2 + \alpha_3(\frac{-\alpha_2}{3\alpha_3})^3)$$

$$= \alpha_3 x^3 + (\alpha_1 - \frac{1}{3}\frac{\alpha_2^2}{\alpha_3})x + (\alpha_o - \frac{\alpha_1\alpha_2}{3\alpha_3} + \frac{2}{27}\frac{\alpha_2^3}{\alpha_3^2}) = \beta_3 x^3 + \beta_1 x + \beta_o.$$

If it should happen that at this point we have $\beta_1 = 0$, then we can proceed as we did with the quadratic, and solve $c(x) = 0$ by taking a cube root. From now on, we will assume that we are not in this relatively simple case, and that both β_1 and β_3 are non-zero. Thus we have two terms involving x, the terms x andx^3. We can adjust both of them by the same scalar via postcomposition with a dilation. Precomposition with a dilation adjusts them by different powers of the dilation factor. It turns out that we can adjust their ratio by any positive factor. It turns out to be convenient to make β_1 three times the magnitude of β_3 Specifically, let us set

$$s = \pm|\frac{\beta_1}{3\beta_3}|^{\frac{1}{2}},$$

where the sign is the sign of β_3, so that $\beta_3 s^3 > 0$. We then compute

$$\tilde{Q} \circ D_s(x) = \beta_3 s^3 x^3 + \beta_1 s x + \beta_o = \frac{1}{3}\frac{|\beta_1|^{\frac{3}{2}}}{|3\beta_3|^{\frac{1}{2}}}(x^3 \pm 3x) + \beta_o.$$

This formula implies an analog of Proposition 3.1.

Proposition 4.2: Every cubic polynomial C is the transform by pre- and postcomposition by affine transforms of one of three specific cubic polynomials $C_\epsilon(x) = x^3 + \epsilon 3x$, where ϵ is equal to 1, 0, or -1:

$$C(x) = A_{t,b} \circ C_\epsilon \circ A_{s,a},$$

for appropriate affine transformations $A_{t,b}$ and $A_{s,a}$.

Precisely, $\alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_o = T_{s,a} \circ (x^3 + \epsilon x) \circ T_{t,b}$, where

$$t = \alpha_3\sqrt{\frac{3}{\delta}}, \quad b = \frac{\alpha_2}{3\alpha_3} = \alpha_2\sqrt{\frac{1}{3\delta}}, \quad s = \frac{\alpha_3}{t^3} = \frac{\sqrt{3}\delta^{\frac{3}{2}}}{9\alpha_3^2}, \quad a = \alpha_o + \frac{2\alpha_2^3 - 9\alpha_1\alpha_2\alpha_3}{27\alpha_3^2}.$$

In these formulas, $\epsilon = sgn(3\alpha_1\alpha_3 - \alpha_2^2)$, and $\delta = |3\alpha_1\alpha_3 - \alpha_2^2|$.

Application 5: Solve the cubic equation

$$C(x) = 0.$$

Solution: Proposition 3.2 provides a reduction of this problem to the problem of solving $C_\epsilon(x) = y$. More specifically, if C is as in Proposition 3.2, then the equation $C(x) = 0$ is equivalent to $A_{t,b} \circ C_\epsilon \circ A_{s,a}(x) = 0$, which is equivalent to $C_\epsilon \circ A_{s,a}(x) = A_{t,b}^{-1}(0) = -\frac{b}{t}$. If we can invert C_ϵ, then we can turn this equation to $A_{s,a}(x) = C_\epsilon^{-1}(-\frac{b}{t})$, or

$$x = A_{s,a}^{-1}(C_\epsilon^{-1}(-\frac{b}{t}) = \frac{1}{s}(C_\epsilon^{-1}(-\frac{b}{t}) - a).$$

If $\epsilon = 0$, then computing C_ϵ^{-1} amounts to taking a cube root. Otherwise, there is no standard notation which describes c_ϵ^{-1}. If $\epsilon = 1$, then C_ϵ is strictly monotone, and so is globally invertible as a function on $\mathbf{R}$. Approximate values can be computed for $C_\epsilon^{-1}(y)$ by means of a computational scheme based on the Newton-Raphson method. If $\epsilon = -1$, then C_ϵ is not globally invertible. For values of y between -2 and 2, there are three possible roots x of $C_1(x) = y$. One will be in the interval $(-1, 1)$, one will be greater than 1, and one will be less than -1. Iterative schemes to compute these roots precisely are fairly easy to construct. This would be an approach to solving cubic equations parallel to the quadratic formula for quadratics. It is not nearly as elegant, because of the need for cases, but it is nevertheless effective.

5. One parameter subgroups.

We hope the reader will agree that the applications sketched above of the $ax + b$ group in school mathematics are interesting, and in fact, connect the $ax + b$ groups with key parts (with the exception of normal forms of cubics) of the high school curriculum. However, the connections we will now discuss are in some sense more fundamental. They give insight into the nature of the functions commonly studied in high school, and help to explain why we study these particular functions.

The transformations $A_{s,a}$ depend on two parameters, the s and the a. We are interested in knowing the *one-parameter subgroups*, the subgroups which can be parametrized by a single variable. Formally, we call a set ϕ_t of transformations a one-parameter group if the ϕ_t have multiplication law which mirrors addition of numbers. That is a one-parameter group is a collection of transformations ϕ_s such that

$$\phi_r \circ \phi_s = \phi_{r+s}. \qquad OnePar$$

We should also require that ϕ_s depend continuously on s.

We note that, as defined, a one-parameter group is not simply a collection of transformations, it comes with a parametrization $s \to \phi_s$. This means that if we rescale the group by sending $s \to cs$ for some constant c, then we will get the same set of transformations, but labeled by different numbes. We remark that such rescalings are the only way we can change parametrizations, subject to the conditiion $(1P)$. For if $s \to \phi_{\gamma(s)}$ is another parametrization of the T_s which gives a one-parameter group, that is, which satsifies $(1P)$, then $\gamma(s) = cs$ for some constant c. To see this, we note that the group law implies that $\phi_{\gamma(r+s)} = \phi_{\gamma(r)} \circ \phi_{\gamma(s)} = \phi_{\gamma(r)+\gamma(s)}$. Assuming that the ϕ_s are all distinct, this implies that $\gamma(s+t) = \gamma(r) + \gamma(s)$. It is well-known, and we will not rehearse the argument here, that the only continuous solutions to this functional equation are $\gamma(s) = cs$ for some constant c. In the case the parametrization of ϕ_s is not unique, the argument is slightly more elaborate. One must analyze the possibilities for redundancy. We will also not do this here. The one-to-one case is the only one that will affect us here.

An important fact about a one-parameter group is that all the operators in it necessarily commute. This follows directly from the definition: the group law in a one-parameter group is modeled on the group law of $\mathbf{R}$ under addition, and this is of course commutative.

This fact allows us almost immediately to determine the one-parameter subgroups of the $ax + b$ group; for we have seen that if two (non-identity) transformations commute, they are either both translations, or they have the same fixed point (Corollary 2.3). This together with Proposition 2.2, the equation (Conj2) of §2, and the discussion above of reparametrization implies the following result.

Proposition 5.1: Up the conjugation and reparametrization, the one-parameter subgroups of the $ax + b$ group are

$$s \to T_s \qquad \text{and} \qquad s \to D_{e^s}$$

6. Functions invariant under symmetries of both axes.

We ask the following question: suppose that we have a one-parameter group of affine transformations acting on each axis:

$$x \to \phi_r(x), \qquad \text{and} \qquad y \to \psi_s(y).$$

What functions Φ, if any, are compatible with these actions, in the sense that

$$\Phi(\phi_r(x)) = \psi_r(\Phi(x))? \qquad (Inv)$$

We can think about the equation (Inv) in the following way. The graph of Φ is the curve consisting of the points

$$\begin{bmatrix} x \\ \Phi(x) \end{bmatrix} : x \in \mathbf{R}$$

in the cartesian plane. We can combine the one-parameter groups ϕ_s and ψ_s to get a one-parameter group of (affine) transformations of the plane:

$$(\phi \times \psi)_r(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} \phi_r(x) \\ \psi_r(y) \end{bmatrix}. \qquad (\phi \times \psi)$$

The compatibility condition (Inv) simply amounts to asking that the graph of Φ be transformed into itself by all transformations $(\phi \times \psi)_r$, i.e., by the one-parameter group $\phi \times \psi$.

What functions are singled out by the conditions (Inv)? Evidently, this depends on the one-parameter groups ϕ and ψ. We know that, up to conjugation, there are two types of one parameter groups, translations and dilations. This gives us four cases to consider:

TT) Both ϕ and ψ are groups of translations.

TD) The group ϕ consists of translations and the group ψ consists of dilations.

DT) The group ϕ consists of dilations and the group ψ consists of translations.

DD) Both groups ϕ and ψ consist of dilations.

We will consider these cases in succession.

Case TT. Suppose that $\phi_s(x) = x + as$ and $\psi_s(y) = y + bs$. Then the condition (Inv) amounts to $\Phi(x+as) = \Phi(x) + bs$. Setting $x = 0$ and $s = \frac{t}{a}$, this can be rewritten as $\Phi(t) = \Phi(0) + \frac{b}{a}t$. In other words, Φ is itself just an affine function (aka *linear function*, in the language of K-12), with dilation factor (aka *slope*) equal to the ratio $\frac{b}{a}$ of the scalings of the x-translation group and the y-translation group.

Case TD. Suppose again that $\phi_s(x) = x + as$, but that now $\psi_x(y) = e^{bs}y$. Then the condition (Inv) becomes $\Phi(x + as) = e^{bs}\Phi(x)$. Again setting $x = 0$ and $s = \frac{t}{a}$, we find that $\Phi(t) = e^{\frac{b}{a}t}\Phi(0)$. Thus in this case, Φ is a multiple of (or equivalently, a horizontal translation of) an exponential function.

Case DT. Here we reverse the roles of dilation and translation from the previous case. We set $\phi_s(x) = e^{as}x$ and $\psi_s(y) = y + bt$. Then the condition (Inv) reads $\Phi(e^{as}x) = \Phi(x) + bs$. If we set $x = 1$ and $e^{as} = t$, so that $s = \frac{\ln(t)}{a}$, then we get the equation $\Phi(t) = \Phi(1) + \frac{b}{a}\ln(t)$. Thus, Φ is an affine transform of the natural logarithm function. The factor $\frac{a}{b}$ can be absorbed into the logarithm by using a different base, and the constant could also be absorbed by translation Φ.

Case DD. Here both ϕ and ψ involve dilations: $\phi_s(x) = e^{as}x$, and $\psi_s(y) = e^{bs}y$. The compatibility condition (Inv) now reads $\Phi(e^{as}x) = e^{bs}\Phi(x)$. Again setting $x = 1$ and $e^{as} = t$, we see that $\Phi(t) = e^{bs}\Phi(1)) = (e^{as})^{\frac{b}{a}}\Phi(1) = t^{\frac{b}{a}}\Phi(1)$. Thus, in this case, Φ is a multiple of a power function.

In summary, we see that the basic functions elementary functions - "linear" functions (aka affine or first order functions), exponential and logarithm functions, and power functions - which occupy so much attention in algebra/precalculus are singled out by symmetry considerations, specifically, symmetry with respect to affine transformations. The other functions which receive substantial attention in the late high school curriculum are trigonometric functions. These, especially sine and cosine, are also deeply involved with symmetry, specifically with rotational symmetry.

Characteristic functional equations

While it is debatable whether the above considerations need to be part of the high school curriculum, I would argue that teachers should understand these ideas. Furthermore, there is another aspect of these ideas which not only teachers should be aware of, but should be strongly considered for inclusion in normal precalculus courses. This is the matter of functional equations which are satisfied by, and which characterize, these basic functions. We will reprise the case-by-case discussion above to bring out these equations.

Case TT. Here the (Inv) relation is $\Phi(x + as) = \Phi(x) + bs$. If we set $as = x'$, then $bs = \Phi(x') - \Phi(0)$. Thus another was of stating the relation (Inv) is

$$\Phi(x + x') = \Phi(x) + \Phi(x') - \Phi(0). \qquad (AffChar)$$

We see then that the function $\tilde{\Phi} = \Phi - \Phi(0)$ which is the homogeneous linear function (aka, linear function) with the same slope as Φ satisfies the functional equation

$$\tilde{\Phi}(x + x') = \tilde{\Phi}(x) + \tilde{\Phi}(x'). \qquad (LinChar))$$

As we have remarked, the equation (LinChar) characterizes the functions $x \to cx$: they are the only (continuous) functions which satisfy (LinChar). Similarly the functions $x \to cx + d$ are characterized by the equation (AffChar). Indeed, as we have shown, Φ satisfies (AffChar) if and only if $\tilde{\Phi} = \Phi = \Phi(0)$ satisfies (LinChar).

Case TD. The (Inv) relation is $\Phi(x+as) = e^{bs}\Phi(x)$. Setting $as = x'$, we have that $e^{bs} = \frac{\Phi(x')}{\Phi(0)}$. Thus, in this case, Φ satisfies the functional equation

$$\Phi(x+x') = \frac{\Phi x)\Phi(x')}{\Phi(0)}. \qquad (ExpMultChar)$$

The corresponding equation for the normalized function $\tilde{\Phi} = \frac{\Phi}{\Phi(0)}$ is

$$\tilde{\Phi}(x+x') = \tilde{\Phi}(x)\tilde{\Phi}(x'). \qquad (ExpChar)$$

This label captures the essential feature of the exponential function- that it converts addition into multiplication. As its label suggests, it characterizes exponential functions: any (continuous) function satisfying it has the form $x \to c^x$ for some positive number c (the *base* of the exponents). Similarly, and by a reduction similar to the one in case TT, the equation (ExpMultChar) characterizes multiples of exponential functions.

Case DT. We will be somewhat briefer with the last two cases. The (Inv) relation here is essentially the reverse of the last situation: $\Phi(e^{as}x) = \Phi(x) + bs$. If we set $e^{as} = x'$, then $bs = \Phi(x') - \Phi(1)$, so we get the functional equation

$$\Phi(xx') = \Phi(x) + \Phi(x') - \Phi(1). \qquad (LogMultChar)$$

The associated normalized function is $\tilde{\Phi} = \Phi - \Phi(1)$. It satisfies the functional equation

$$\tilde{\Phi}(xx') = \tilde{\Phi}(x) + \tilde{\Phi}(x'). \qquad (LogChar).$$

Again the equation (LogCare) characterizes logarithm functions, and the equation (LogMultChar) characterizes multiples of logarithm functions.

Case DD. Finally, for the case DD, the functional equation derived in the manner of the ones above is

$$\Phi(xx') = \frac{\Phi(x)\Phi(x')}{\Phi(1)}. \qquad (MultPowerChar)$$

The associated "homogeneous" functional equation is

$$\tilde{\Phi}(xx') = \tilde{\Phi}(x)\tilde{\Phi}(x'). \qquad (PowerChar)$$

Again the only continuous solutions to equation PowerChar are the exponential functions. The equation (MultPowerChar) also allows multiples of exponential functions. We remark that, in the language of group theory, equation (PowerChar) is just the functional equation for homomorphisms from the multiplicative group of **R** to itself.

Thus, all the functions that are invariant under one-parameter affine groups which preserve the axes in the plane are characterized by simple functional equations. These equations are somewhat analogous to those characterizing the solutions to the affine difference equations analyzed in §4.