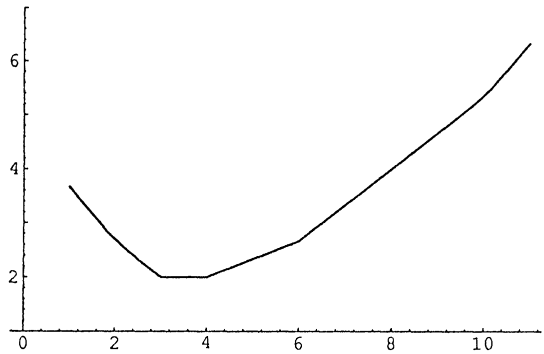

Figure 1

References

1. Kumar Joag-Dev, MAD property of a median: A simple proof, *American Statistician* 43 (1989) 26–27.
2. Neil C. Schwertman, A. J. Gilks, and J. Cameron, A simple noncalculus proof that the median minimizes the sum of absolute deviations, *American Statistician* 44 (1990) 38–39.

A Geometric Approach to Linear Functions

Jack E. Graver, Syracuse University, Syracuse, NY 13244-1150

Instead of drawing the traditional graphs, we will visualize linear functions as transformations of the real line $\mathscr{L}$. For example, the linear function $f(x)=\frac{4}{5}x+1$ is illustrated by the 1-dimensional picture in Figure 1. In this picture, the coordinates of the points on $\mathscr{L}$ are listed below the line with the function values listed directly above. The action of this function on the line is indicated by the arrows: 0 is mapped to 1, so an arrow originates at 0 and extends to 1; 2 is mapped onto $\frac{13}{5}$, an increase of $\frac{3}{5}$, so the arrow from 2 has length $\frac{3}{5}$ and extends to the right; and so on. From this picture, we perceive a very nice geometric description for the action of f: It is the contraction by a factor of $\frac{4}{5}$ about the point 5 (the center of the contraction). Geometrically, it is clear that f is uniquely determined by its slope $\frac{4}{5}$ and its center 5. We will start our investigation by considering this geometric observation in algebraic terms. We will then use our algebraic results to study the geometry of the line, and finally we will use both our algebraic and our geometric results to consider linear difference equations.

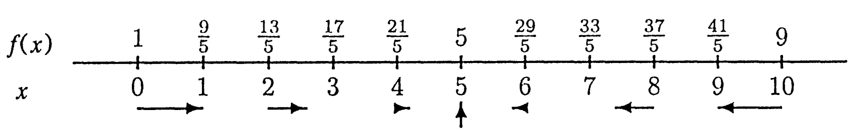

Figure 1

The slope-center form of a linear function. Many elementary courses start with a quick review of linear equations and their graphs. In these courses, we discuss the two-point form, the point-slope form and the slope-intercept form of the equation of a line. Our geometric approach motivates another useful form, the *slope-center* form. Consider the linear function $f(x)=ax+b$. Our first task is to discover if f has a center, or fixed point, and, if it does, to identify that point. Suppose x is a fixed point of f; then $x=ax+b$, or $(1-a)x=b$. Clearly:

- If $a = 1$ and $b = 0$, then f is the identity function and all points are fixed.
- If $a = 1$ and $b \neq 0$, then f has no fixed points.
- If $a \neq 1$, then $b/(1-a)$ is the unique fixed point of f.

If $a = 1$, then $f(x) = x + b$, so f shifts each point by the distance $|b|$—to the right when b is positive, to the left when b is negative—and we call f a *translation*. If f is not a translation, we call its unique fixed point the *center* of f. It is convenient to say that a translation other than the identity has its center at ∞, and that the identity function has each real number and ∞ as a center. With these definitions we have the following result.

Theorem 1. *Two linear functions commute under composition if and only if they have a common center.*

Proof. Let $f(x) = ax + b$ and $g(x) = a'x + b'$ be two linear functions. Then f and g commute if and only if

$$a(a'x + b') + b = a'(ax + b) + b',$$

that is, if and only if

$$(1 - a')b = (1 - a)b'.$$

Both sides of this last equation are zero if and only if one of f and g is the identity, both are translations, or both are nontranslations with center 0. Both sides are equal and nonzero if and only if neither f nor g is a translation and they have the same nonzero center. ■

If $a \neq 1$, the linear function $f(x) = ax + b$ can be written in slope-center form:

$$f(x) = ax + (1 - a)c, \qquad \text{where} \quad c = \frac{b}{1 - a} \quad \text{is the center of } f.$$

In general, this form is most important for understanding the geometric properties of the function; occasionally it is the natural form for interpreting an application. For example, a linear function popular with students expresses their final average as a function of their (yet to be taken) final exam percentage. The center of this function is the average at the end of the course work: If 79% is the course work average and 79% is the exam score, the final average will be 79%. Thus the final average is given by $f(x) = ax + (1 - a)c$, where x is the score on the final exam (as a percentage), c is the present average (as a percentage), and a is the weight of the final exam (as a fraction of the total possible points). Suppose that, before the exam, a student has earned 237 points out of a possible 300 (79%) and that the exam is worth 200 points. Then $f(x) = \frac{2}{5}x + \frac{3}{5}(79)$ gives the student's final average (as a percentage), where x is the student's exam score (as a percentage).

Linear functions as transformations. We now wish to find a geometric description for an arbitrary linear function $f(x) = ax + b$. In the case $a = 1$, we have already seen that $f(x) = x + b$ can be interpreted as a translation. We adopt the notation $t_{[b]}$ for this translation. By Theorem 1, any two translations commute and we easily verify the following equations both geometrically and algebraically:

$$t_{[b]} \circ t_{[b']} = t_{[b']} \circ t_{[b]} = t_{[b+b']}.$$

Next consider the function $f(x) = -x + b$. In slope-center form, it becomes $f(x) = -x + 2c$, where $c = b/2$. One easily checks that this is the reflection of $\mathscr{L}$ about the center c:

$$f(c \pm y) = c \mp y, \qquad \text{for all } y.$$

We adopt the notation $r_{[c]} = -x + 2c$ for the reflection with center c. Distinct reflections have different centers and hence do not commute. We see geometrically and easily verify algebraically that

$$r_{[c]} \circ r_{[c']} = t_{[2(c-c')]} \qquad \text{while} \qquad r_{[c']} \circ r_{[c]} = t_{[2(c'-c)]}.$$

Recall that a function $\phi: \mathscr{L} \to \mathscr{L}$ is a *similarity* of the line if there is a positive real number m (called the *magnification* of ϕ) so that, for all points (real numbers) x and y on the line, $|\phi(x) - \phi(y)| = m|x - y|$. In other words, the distance between the images of x and y is m times the distance between x and y. We easily verify that *a linear function $f(x) = ax + b$ is a similarity with magnification $|a|$*:

$$|f(x) - f(y)| = |(ax+b) - (ay - b)| = |a(x-y)| = |a|\,|x-y|.$$

The natural question is: Do the linear functions account for all of the similarities of the line? The answer is: Yes. To see why, we first argue that a similarity is uniquely determined by its effect on any two distinct points or, equivalently, that two similarities that agree at two distinct points are, in fact, equal. Let ϕ and γ be two similarities and let x and y be two distinct points, so that $\phi(x) = \gamma(x)$ and $\phi(y) = \gamma(y)$. Clearly ϕ and γ have the same magnification $m = |\phi(x) - \phi(y)|/|x-y| = |\gamma(x) - \gamma(y)|/|x-y|$. Thus for any point z, $|\phi(z) - \phi(x)| = m|z-x| = |\gamma(z) - \gamma(x)|$; from which we conclude that the point $\phi(x) = \gamma(x)$ is equidistant from the points $\phi(z)$ and $\gamma(z)$. By a similar argument the point $\phi(y) = \gamma(y)$ is also equidistant from the points $\phi(z)$ and $\gamma(z)$. If $\phi(z) \neq \gamma(z)$, then $\phi(x)$ and $\phi(y)$ both equal the midpoint of the segment joining $\phi(z)$ and $\gamma(z)$. Since $\phi(x)$ and $\phi(y)$ are distinct, this is impossible. We conclude that $\phi(z) = \gamma(z)$ and, since z was arbitrary, we conclude that $\phi = \gamma$.

Now it is easy to see that any similarity ϕ of the real line is given by a linear function. Let $a = \phi(1) - \phi(0)$ and $b = \phi(0)$. Then the linear function $f(x) = ax + b$ is a similarity with $f(0) = \phi(0)$ and $f(1) = \phi(1)$, so $\phi = f$.

A similarity is an *isometry* (or *congruence*) if it has magnification 1. As we saw above, translations and reflections are the only isometries of the line. What about the remaining similarities, those with magnification $m \neq 1$? When $|a| \neq 1$, we adopt the notation $s_{[a,c]}$ for the linear function with slope a and center c and observe that, for any real number y:

$$s_{[a,c]}(c+y) = a(c+y) + (1-a)c = c + ay. \tag{1}$$

Thus for positive a, $s_{[a,c]}$ stretches the line away from c if $a > 1$ and shrinks it toward c when $0 < a < 1$. When $a > 0$ we call $s_{[a,c]}$ the *dilation* with magnification a and center c. The linear function pictured in Figure 1 is the dilation $s_{[\frac{4}{5},5]}$. When a is negative, $s_{[a,c]}$ simultaneously dilates the line from c by a factor of $|a|$ and reflects it through c. That is,

$$s_{[a,c]} = r_{[c]} \circ s_{[|a|,c]} = s_{[|a|,c]} \circ r_{[c]}, \qquad \text{when } a < 0,$$

as is easily verified algebraically.

To summarize, *the only similarities of the Euclidean line are the isometries (translations and reflections), the dilations, and the dilating reflections*. It is worth noting that the results in this section generalize immediately to the complex linear functions and result in a complete classification of the direct isometries and direct similarities of the plane. To extend the study to the opposite isometries and similarities, one must consider the conjugate linear functions, $f(z) = a\bar{z} + b$ [see H. S. M. Coxeter, *Introduction to Geometry*, 2nd ed., Wiley, New York, 1980, p. 145].

Linear difference equations. The ability to think about linear functions geometrically is particularly useful in considering linear difference equations. Such equations are now discussed in many elementary courses, most notably courses that include a section on the mathematics of finance. By a *(first-order) linear difference equation*, we mean an equation of the form

$$x_n = ax_{n-1} + b, \qquad \text{where } a \notin \{-1, 0, 1\}.$$

The sequence $x_0, x_1, \ldots, x_n, \ldots$ satisfying this equation for $n = 1, 2, \ldots$ is called the *solution* to the difference equation with *initial condition* x_0. Associated with this linear difference equation is the linear function $f(x) = ax + b$, and we see that the solution to the linear difference equation with initial condition x_0 is simply the trajectory of x_0 under successive iterations of f:

$$x_1 = f(x_0), \quad x_2 = f(f(x_0)) = f^2(x_0), \quad \ldots, \quad x_n = f^n(x_0).$$

The main result about linear difference equations is the formula for directly computing the nth term of the solution sequence:

$$x_n = a^n x_0 + \frac{(1 - a^n)b}{1 - a}. \tag{2}$$

Traditionally this formula is proved by applying the formula for the sum of the terms of a geometric progression. But the significance of the factor $b/(1 - a)$ in the second term of (2) is not at all apparent in this approach and is not even discussed in many texts.

Using our geometric approach, we not only have a better understanding of the structure of a solution sequence but we also get a simple derivation of the formula for the nth term. Rewriting the associated linear function in slope-center form, we have from equation (1)

$$s^n_{[a,c]}(c + y) = c + a^n y.$$

Thus,

$$x_n = s^n_{[a,c]}(x_0) = c + a^n(x_0 - c),$$

or

$$x_n = a^n\left(x_0 - \frac{b}{1-a}\right) + \frac{b}{1-a}.$$

When $|a| < 1$ the sequence $\{x_n\}$ converges to c, so the significance of c is apparent. But even in many applications with $|a| > 1$, c still has a useful interpretation. A prime example is the linear difference equation of a loan. The balance due on a loan is given by the solution to the linear difference equation

$$x_n = (1+I)x_{n-1} - R,$$

where x_0 is the amount of the loan, x_n is the balance due at the end of the nth payment period, I is the periodic rate (the interest rate per payment period), and R is the size of the payments (called the "rent"). We assume that the interest compounding periods and the payment periods coincide. Thus, if you owe x_{n-1} on the loan at the end of the $(n-1)$st period, you will owe x_{n-1} plus Ix_{n-1} (the interest that accrues) minus R (your next payment) at the end of the nth period.

For this linear difference equation, the center is $-R/[1-(1+I)] = R/I$. The slope $1+I$ is greater than 1, so at each iteration the difference between the center and the current balance is magnified by the factor $1+I$ as illustrated in Figure 2 for a 12-month loan. If $x_0 < R/I$, the sequence decreases in larger and larger steps. Of course if $x_0 > R/I$ or equivalently $R < Ix_0$, the successive balances increase in larger and larger steps, since the payment R is not large enough to cover the interest that accrues each period.

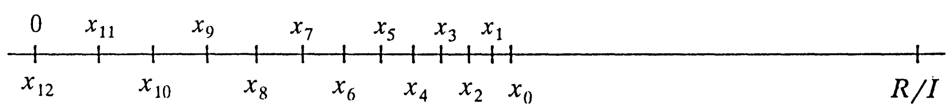

Figure 2

Given such a picture of any loan, a simple change of scale, multiplying by I, yields a geometric representation of the distribution of each payment between interest and debt reduction; see Figure 3. When the balance due on the loan is x_k, then, of the next payment of R dollars, Ix_k is the interest and the debt is reduced by $R - Ix_k$ dollars. Figure 3 helps to explain the difference between long-term and short-term loans. For a long-term loan, the distance between the center, R/I, and the amount borrowed, x_0, is relatively small, so the portion of the first few payments that goes to interest is high. But for a short-term loan, $(R/I) - x_0$ is large, so a relatively large part of the first few payments goes to reduce the debt.

The slope-center form for linear functions provides a good opportunity to introduce at the precalculus level the important idea of picturing a function $f(x)$

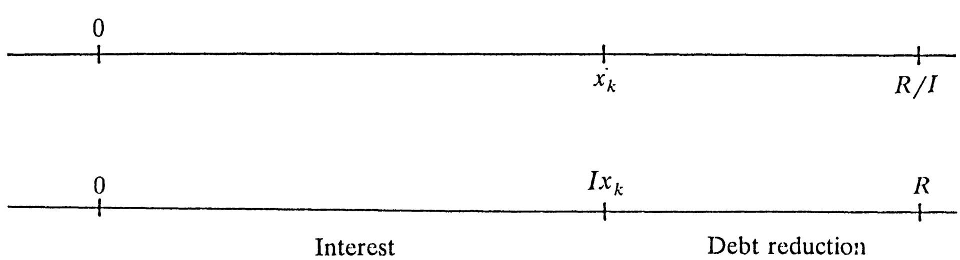

Figure 3

as a map of a number line onto itself. Students familiar with this way of visualizing linear functions will find it easy to understand the interpretation of the derivative $f'(x)$ of a nonlinear function as the local magnification of f at the point x, a point of view that makes the chain rule transparent. Because it gives additional insight into several topics while laying a good foundation for more advanced mathematics, this geometric approach and the slope-center form deserve to be more widely used in our classrooms.

Acknowledgment. This work was supported in part by a National Science Foundation Teacher Enhancement Grant TPE-9253298.

Cans and Calculus

To the Editor: In her Classroom Capsule, "Calculus in the Brewery" [*CMJ* 25:3 (1994) 226–227], S. Colley discussed the fact that certain types of beer cans (the tall, thin ones) appear to be larger than other cans with the same volume but a different shape (the short, fat ones). This reminded me of a classroom experience that I had when I had just begun teaching at a small college in Milwaukee. One day we were discussing the standard max/min problems in calculus and discovered that in order to minimize the amount of material used in constructing a can, the diameter of the can should be equal to its height. One of my students immediately pointed out that this was not the case in cans found in the grocery stores, because vegetable and soup cans were always taller and thinner than the prescribed shape. There were two local can manufacturing plants, so we decided to follow up on this. We contacted them and asked them if they knew the "right" way to construct cans. They replied that they did, indeed, know these facts, but that marketing surveys showed that customers, when given the choice between two cans of the same price, generally picked the tall, thin ones because they looked larger. Since that time vegetable cans of the "right" dimensions have appeared in the stores and have been accepted by the buying public. As a native of Milwaukee, it pained me to realize that Colley's article implied that beer drinkers might know less math than vegetable lovers.

—*Richard F. Maruszewski, United States Naval Academy*