A VERIFICATION OF ASSOCIATIVITY

Somewhere in a first complex analysis course, one encounters the formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}(z) = \frac{az+b}{cz+d},$$

and one hears various assertions about whatever is happening here, in particular that it is associative, meaning that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}(z)\right) = \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}\right)(z).$$

That is,

matrix **of** (*matrix of number*)   equals   (*matrix* **times** *matrix*) *of number*.

The verification of associativity is typically left as an exercise for the reader, or it is carried out in a fashion that seems ad hoc or vague. But the associativity in question has not arisen by flukish accident, and so a complicated, case-laden verification of it cannot be to the point. Such a verification cannot be good mathematics. The purpose of this writeup is to explain that *from an appropriate perspective, there is nothing to verify*.

Let S be any set. The collection of maps of S to itself (called *endomorphisms* of S) is denoted

$$\mathrm{End}(S).$$

The composition of two endomorphisms of S is another endomorphism of S, and the composition of endomorphisms of S is associative though not necessarily commutative. To verify the associativity, compute that for any $f, g, h \in \mathrm{End}(S)$ and any $s \in S$,

$$((f \circ g) \circ h)(s) = (f \circ g)(h(s)) = f(g(h(s))) = f((g \circ h)(s)) = (f \circ (g \circ h))(s).$$

Since $s \in S$ is arbitrary, $(f \circ g) \circ h = f \circ (g \circ h)$.

The collection of bijective endomorphisms of S (called *automorphisms*) is denoted

$$\mathrm{Aut}(S).$$

Essentially by definition, $\mathrm{Aut}(S)$ forms a group. That is, the composition of two automorphisms of S is another automorphism of S (this is not automatic from the corresponding fact about endomorphisms), the composition of automorphisms of S is associative (but this is), the identity map on S is an automorphism, and the inverse of an automorphism of S exists (this need not hold for endomorphisms) and is another automorphism of S. If the group structure is not familiar then it should be verified not only for its general importance but also in order to see that the verification is utterly natural. The very sparseness of the environment forces one to think about the things that are relevant and disallows one from self-obfuscating by thinking about anything else instead.

As a special case of a set S, let V be any vector space over $\mathbb{C}$, perhaps infinite-dimensional. Now we have the notion of a $\mathbb{C}$-linear map of V to itself. The collection of $\mathbb{C}$-linear automorphisms of V is denoted

$$\mathrm{GL}(V).$$

Again we have a group. Here the only issues—now suitably separated from the more general issues of the previous paragraph, and now clearer since they are suitably separated—are that the composition of linear maps is again linear, and the inverse of an invertible linear map is again linear. And the verification is intrinsic rather than involving coordinates, since we have not yet introduced coordinates. Again the absence of spurious detail forces one to discover that thinking in better ways is in fact easier.

Let n be a positive integer, and let $V = \mathbb{C}^n$. The standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{C}^n$ gives a natural way to describe each linear map L of $\mathbb{C}^n$ to itself by an n-by-n matrix m_L with complex entries,

$$m_L = [a_{ij}]_{n\times n},$$

where specifically the entries a_{ij} of the matrix are obtained from the linear map via the standard basis vectors and the usual inner product on $\mathbb{C}^n$,

$$a_{ij} = \langle e_i, L(e_j)\rangle \quad \text{for all } i,j \in \{1,\dots,n\}.$$

Matrix-by-vector multiplication is *defined* to be compatible with the action of linear maps. That is, for any linear map L with matrix m_L,

$$m_L v = L(v) \quad \text{for all } v \in \mathbb{C}^n.$$

And matrix-by-matrix multiplication is *defined* to be compatible with the composition of linear maps. That is, for any linear maps K and L with matrices m_K and m_L,

$$(m_K m_L)v = (K \circ L)(v) \quad \text{for all } v \in \mathbb{C}^n.$$

One should *derive* the familiar index-laden formulas for matrix-by-vector multiplication and matrix-by-matrix multiplication from their defining compatibility properties. Instead, one typically is first given the formulas as weird rote processes, and only later is one pleasantly surprised to see their compatibilities with the natural intrinsic mapping operations that they are designed to describe, or just vaguely baffled by this, or never fully aware of the connection at all. In any case, the group of n-by-n invertible matrices with complex entries, with matrix-by-matrix multiplication as its operation, is denoted

$$\mathrm{GL}_n(\mathbb{C}).$$

The compatibility of this structure with the linear isomorphism group $\mathrm{GL}(\mathbb{C}^n)$ means—with no further reference to formulas—that there is a commutative diagram

$$\begin{array}{ccc} \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C}) \times \mathbb{C}^n & \longrightarrow & \mathrm{GL}_n(\mathbb{C}) \times \mathbb{C}^n \\ \big\downarrow & & \big\downarrow \\ \mathrm{GL}_n(\mathbb{C}) \times \mathbb{C}^n & \longrightarrow & \mathbb{C}^n \end{array}$$

where the mappings are

$$\begin{tikzcd}
(m,n,v) \arrow[r, mapsto] \arrow[d, mapsto] & (m,nv) \arrow[d, mapsto] \\
(mn,v) \arrow[r, mapsto] & (mn)v = m(nv).
\end{tikzcd}$$

Next, increment n to $n+1$. The group

$$\mathbb{C}^\times I_{n+1} = \{cI_{n+1} : c \in \mathbb{C}^\times\}$$

is readily seen to be a normal subgroup of $\mathrm{GL}_{n+1}(\mathbb{C})$. Thus the quotient group

$$\mathrm{PGL}_{n+1}(\mathbb{C}) = \mathrm{GL}_{n+1}(\mathbb{C})/\mathbb{C}^\times I_{n+1}$$

is well defined. In more concrete terms, each element of the quotient group consists of all nonzero scalar multiples of a matrix,

$$\mathrm{PGL}_{n+1}(\mathbb{C}) = \{\mathbb{C}^\times m : m \in \mathrm{GL}_{n+1}(\mathbb{C})\},$$

and what is well defined is the multiplication law $(\mathbb{C}^\times m)(\mathbb{C}^\times n) = \mathbb{C}^\times mn$. The natural homomorphism from the matrix group to the quotient group is denoted P,

$$P : \mathrm{GL}_{n+1}(\mathbb{C}) \longrightarrow \mathrm{PGL}_{n+1}(\mathbb{C}), \qquad m \longmapsto \mathbb{C}^\times m,$$

and the quotient group is called the projective linear group. We incremented to $n+1$ in this discussion because projecting to the quotient conceptually loses a dimension, returning us to dimension n.

In a similar vein, consider the set $\mathbb{C}^{n+1} - \{0\}$ consisting of $(n+1)$-dimensional complex Euclidean space punctured at the origin. This set is a disjoint union of punctured complex lines, each such line being the set of nonzero complex scalar multiples of any of its points,

$$\mathbb{C}^\times v = \{cv : c \in \mathbb{C}^\times\}.$$

The set of such lines is *complex n-dimensional projective space*,

$$\mathbb{P}^n(\mathbb{C}) = \{\mathbb{C}^\times v : v \in \mathbb{C}^{n+1} - \{0\}\},$$

and the natural map from punctured $(n+1)$-dimensional space to n-dimensional projective space is

$$\mathbb{P} : \mathbb{C}^{n+1} - \{0\} \longrightarrow \mathbb{P}^n(\mathbb{C}), \qquad v \longmapsto \mathbb{C}^\times v.$$

More conceptually, $\mathbb{C}^{n+1} - \{0\}$ is the space of ordered bases of 1-dimensional subspaces of $\mathbb{C}^{n+1}$ while $\mathbb{P}^n(\mathbb{C})$ is naturally viewed as the space of the subspaces themselves, each subspace represented by the union of its ordered bases. Clearly this construction scales up to k-dimensional suspaces, but here we need only $k = 1$.

There is a commutative diagram

$$\begin{tikzcd}
\mathrm{GL}_{n+1}(\mathbb{C}) \times (\mathbb{C}^{n+1} - \{0\}) \arrow[r] \arrow[d, "P\times\mathbb{P}"'] & \mathbb{C}^{n+1} - \{0\} \arrow[d, "\mathbb{P}"] \\
\mathrm{PGL}_{n+1}(\mathbb{C}) \times \mathbb{P}^n(\mathbb{C}) \arrow[r] & \mathbb{P}^n(\mathbb{C})
\end{tikzcd}$$

where the mappings are

$$\begin{tikzcd}
(m,v) \arrow[r, mapsto] \arrow[d, mapsto] & mv \arrow[d, mapsto] \\
(\mathbb{C}^\times m, \mathbb{C}^\times v) \arrow[r, mapsto] & \mathbb{C}^\times m\,\mathbb{C}^\times v = \mathbb{C}^\times mv.
\end{tikzcd}$$

That is, the action of the general linear group on the punctured space in dimension $n+1$ descends to an action of the projective general linear group on projective space in dimension n.

The projective action is again associative. That is, there is a commutative diagram

$$\begin{tikzcd}
\mathrm{PGL}_{n+1}(\mathbb{C}) \times \mathrm{PGL}_{n+1}(\mathbb{C}) \times \mathbb{P}^n(\mathbb{C}) \arrow[r] \arrow[d] & \mathrm{PGL}_{n+1}(\mathbb{C}) \times \mathbb{P}^n(\mathbb{C}) \arrow[d] \\
\mathrm{PGL}_{n+1}(\mathbb{C}) \times \mathbb{P}^n(\mathbb{C}) \arrow[r] & \mathbb{P}^n(\mathbb{C})
\end{tikzcd}$$

where, now letting $\overline{m}$ and $\overline{n}$ denote projective matrix classes and letting $\overline{v}$ denote a point of projective space, again the gist of the matter is that

$$(\overline{m}\,\overline{n})\overline{v} = \overline{m}(\overline{n}\,\overline{v}).$$

This is essentially automatic. Since the nonprojective action is associative, as in our first commutative diagram, and since the action commutes with projection, as in our second, the result is guaranteed. Basically, both sides of the equality are $\overline{mnv}$. For the reader who wants to say more about this point, one procedure is to compute

$$\begin{aligned}
(\overline{m}\,\overline{n})\overline{v} = (\mathbb{C}^\times m\,\mathbb{C}^\times n)\mathbb{C}^\times v &= \mathbb{C}^\times(mn)v \\
&= \mathbb{C}^\times m(nv) = \mathbb{C}^\times m(\mathbb{C}^\times n\,\mathbb{C}^\times v) = \overline{m}(\overline{n}\,\overline{v}).
\end{aligned}$$

For those who insist on making a whole Megillah of the matter, a second procedure involves a cube-shaped diagram

$$\begin{tikzcd}[row sep=small, column sep=small]
& P_{n+1}^2 \times \mathbb{P}^n \arrow[rr] \arrow[dd] & & P_{n+1} \times \mathbb{P}^n \arrow[dd] \\
G_{n+1}^2 \times (\mathbb{C}^{n+1} - 0) \arrow[ur] \arrow[rr, crossing over] \arrow[dd] & & G_{n+1} \times (\mathbb{C}^{n+1} - 0) \arrow[ur] & \\
& P_{n+1} \times \mathbb{P}^n \arrow[rr] & & \mathbb{P}^n \\
G_{n+1} \times (\mathbb{C}^{n+1} - 0) \arrow[ur] \arrow[rr] & & \mathbb{C}^{n+1} - 0. \arrow[ur] \arrow[from=uu, crossing over] &
\end{tikzcd}$$

Elementwise, the diagram is

$$\begin{tikzcd}
& (\overline{m}, \overline{n}, \overline{v}) \arrow[rr, mapsto] \arrow[dd] & & (\overline{m}, \overline{nv}) \arrow[dd, mapsto] \\
(m, n, v) \arrow[ur, mapsto] \arrow[rr, mapsto] \arrow[dd, mapsto] & & (m, nv) \arrow[ur, mapsto] \arrow[dd, mapsto] & \\
& (\overline{mn}, \overline{v}) \arrow[rr] & & \overline{mn} \cdot \overline{v} = \overline{m} \cdot \overline{nv} \\
(mn, v) \arrow[ur, mapsto] \arrow[rr, mapsto] & & (mn)v = m(nv). \arrow[ur, mapsto] &
\end{tikzcd}$$

We want the back square to commute. But

- The top, side, and bottom squares commute because the various products are compatible with projection.
- The front square commutes because the nonprojective action is associative.
- The map along the top of the left side square surjects.

Using these commutativities one can show that starting at the front top left corner, the map into the page and then one way around the back square equals the map into the page and then the other way around the back square. Since the map into the page surjects, this makes the back square commute as desired. But really, carrying all of this out more than once in one's life—or perhaps *even* once in one's life—is patently silly.

When $n = 2$, points of projective space $\mathbb{P}^1(\mathbb{C})$ (also called the *complex projective line*) are written $[z_1, z_2]$. The action of a matrix class $\overline{\left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]} \in \mathrm{PGL}_2(\mathbb{C})$ on a point $[z_1, z_2] \in \mathbb{P}^1(\mathbb{C})$ is

$$\boxed{[z_1, z_2] \longmapsto [a_{11}z_1 + a_{12}z_2, a_{21}z_1 + a_{22}z_2].}$$

This case-free notation is optimal, and because the action is projective-linear, it is innately associative.

This writeup should stop here, but instead we will now clutter up the case-free notation in order to relate it to the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Any point of the complex projective line satisfies

$$[z_1, z_2] = \begin{cases} [z_1/z_2, 1] & \text{if } z_2 \neq 0 \\ [1, 0] & \text{if } z_2 = 0. \end{cases}$$

This leads to the usual identification of the complex projective line with the Riemann sphere,

$$[z_1/z_2, 1] \longleftrightarrow z_1/z_2, \qquad [1, 0] \longleftrightarrow \infty$$

In fact this is an identification of topological spaces, not only of sets. The projective line inherits its topology from $\mathbb{C}^2$ via the projection, while the Riemann sphere inherits its topology from the round sphere S^2 via stereographic projection, and S^2 in turn inherits its topology from its ambient space $\mathbb{R}^3$. We do not show here that the identification respects topology. But the emerging idea here is that to facilitate thinking algebraically, we should trade in the Riemann sphere for the projective line. Doing so has no topological cost.

Returning to the formula

$$[z_1, z_2] \longmapsto [a_{11}z_1 + a_{12}z_2, a_{21}z_1 + a_{22}z_2],$$

we can note that the given cases are exclusive and exhaustive in the relation

$$\begin{aligned}[a_{11}z_1 + a_{12}z_2,&a_{21}z_1 + a_{22}z_2] \\ &= \begin{cases} \left[\dfrac{a_{11}(z_1/z_2) + a_{12}}{a_{21}(z_1/z_2) + a_{22}}, 1\right] & \text{if } z_2 \neq 0 \text{ and } a_{21}z_1 + a_{22}z_2 \neq 0, \\ \left[\dfrac{a_{11}}{a_{21}}, 1\right] & \text{if } z_2 = 0 \text{ and } a_{21} \neq 0, \\ [1, 0] & \text{if } z_2 \neq 0 \text{ and } a_{21}z_1 + a_{22}z_2 = 0, \\ [1, 0] & \text{if } z_2 = 0 \text{ and } a_{21} = 0. \end{cases}\end{aligned}$$

Classically, this formula is written

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}(z) = \begin{cases} \dfrac{az+b}{cz+d} & \text{if } z \neq \infty \text{ and } cz + d \neq 0, \\ \dfrac{a}{c} & \text{if } z = \infty \text{ and } c \neq 0, \\ \infty & \text{if } z \neq \infty \text{ and } cz + d = 0, \\ \infty & \text{if } z = \infty \text{ and } c = 0, \end{cases}$$

or more succinctly as the generic case, with the other cases being tacit,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}(z) = \frac{az+b}{cz+d}.$$

Then the associativity that we have verified is indeed that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}(z)\right) = \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}\right)(z).$$

If we didn't realize that the associativity has already been verified, few of us would have the fortitude to check all eight cases (!), instead checking only the generic case that doesn't involve ∞ anywhere, and that utterly predictably is a small distastefulness reducing to the associativity of 2-by-2 matrix multiplication and the fact that multiplying the numerator and the denominator of a quotient by the same factor preserves the quotient—precisely the mechanical manifestations of the issues that have been discussed here conceptually instead. Then one perhaps mutters something peevish about continuity, or about treating fractional linear transformations as formal expressions involving z as a variable rather than worrying about z as a value, in order to salve one's conscience regarding the other cases. This blurring is rather like the use of the word "series" to denote either a formal infinite sum or its actual value, thus sweeping convergence issues under the rug. But in any case, again, the real point here is that this all is as gratuitous as it is grim: since the action is a projective action descended from a linear action, it is is innately associative and there is nothing to check at all.

The reader may object that this writeup has served only to replace a morass of formula-crunching with a blizzard of jargon. However, the difference is that the jargon has *scope*. Working intrinsically rather than in coordinates is innately worthwhile because so much modern mathematics involves structures too elaborate to understand through the underpowered technology of formulas alone. Working in a projective environment rather than an affine one often makes the objects involved more coherent, analogously to how being aware that we live on a sphere rather than a plane clarifies observational phenomena. And more immediately, understanding that fractional linear transformations descend from linear maps means that further techniques of linear algebra can be brought to bear on complex analysis in the study of fractional linear transformations. This avoids the waste of labor and conception in continuing to work through the ideas *ad hoc* by methods that seem almost but not quite familiar. Mathematical thinking should not dismiss structure as abstract. Mathematical thinking should develop one's awareness that structure naturally plays a vivid role in illuminating explicit methods.