Notes on composition of maps

1. Some notational conventions. There are four common notations for the image of an element x under a function f:

(1) Left multiplicative notation: fx.

(2) Standard functional notation: $f(x)$.

(3) Right multiplicative notation: xf.

(4) Exponential notation: x^f.

Consider the following rule of thumb:

Convention 1. *Suppose we are given functions* $X \xrightarrow{f} Y \xrightarrow{g} Z$. *Then if* f *and* g *belong to a class of functions written on the* left *of their arguments* (*as in notation* (1) *or* (2) *above*), *write their composite as* gf *or* $g\circ f$. *If, on the other hand,* f *and* g *belong to a class of functions written on the* right *of their arguments* (*as in* (3) *or* (4)), *call their composite* fg *or* $f\circ g$.

Recall that whichever order of composition one uses, composition of functions is associative. The advantage of Convention 1 is that it creates, by definition, a ‘‘second associative law’’ for functions and their arguments. That is, in notations (1) and (3),

(5) $$(fg)x = f(gx), \qquad \text{respectively} \qquad x(gf) = (xg)f.$$

In either case, this associativity makes it possible to drop parentheses, and write fgx or xgf.

Essentially all writers follow Convention 1 in a ‘‘weak sense’’; namely, if they write most functions to the left of their arguments, they write composites of such functions in the ‘‘lefthand’’ manner, while if they write most functions to the right of their arguments, they use ‘‘righthand’’ composition. However, most authors try to keep all function symbols on the same side of their arguments, and if they occasionally give in to special circumstances and write functions on the other side, they often stick to their original definition of composition, and thus have ‘‘scrambled associative laws’’ such as $(fg)x = g(f(x))$ holding in these situations.

I claim, however, that in many situations, especially linear algebra (i.e., module theory), it is useful to work both with functions written on the left *and* and with functions written on the right, and to compose each type of function according to the appropriate rule.

Let us make clear the price. First, we have to apply contradictory definitions of composition to different functions occurring in the same discussion. As a consequence, we cannot uniformly make statements like ‘‘a left-invertible function is one-to-one’’. Rather, this statement applies to classes of functions that we write on the left of their arguments, while left-invertible members of classes of functions written on the right of their arguments are *surjective*. However, since in reading different authors, we have to deal sometimes with functions written on one side and sometimes with functions written on the other, it is not that hard to adjust to different notations being used within one text, as long as the motivation for the difference is made clear.

Secondly, Convention 1 gives us no way of denoting the composite of a function f, g etc. written on the left with another function a, b etc. written on the right. It is easy to denote the action of such a composite on an element x, e.g., $(fx)a$ or $f(xa)$; but if we need a symbol for the composite itself, we must resort to some artifice; e.g., we might introduce new symbols $\tilde{a}$, $\tilde{b}$, etc., denoting the same functions as a, b etc., but written to the left of their arguments. We can then write composites $\tilde{a}f$, $f\tilde{a}$ etc., without violating our Convention. (Note that if we do this, then though $\tilde{a}, \tilde{b}, \widetilde{ab}$ name the same functions as a, b, ab, we have $\widetilde{ab} = \tilde{b}\,\tilde{a}$.)

But why would we want to write some maps to the left of their arguments, and others to the right? Because of another very useful principle:

Convention 2. *If one class of functions is characterized as commuting with* (''*respecting*'') *the functions in another class, write members of the two classes on opposite sides of their arguments.*

Thus, if f belongs to a class of functions written on the left, and a to a class of functions respecting members of that class, and written on the right, then the fact that f and a respect one another becomes a third ''associative law'':

(6) $$f(xa) = (fx)a,$$

(again allowing us to drop parentheses and write fxa if we wish).

Our final, and perhaps most important Convention concerns the situation where we have not merely a family of maps on a set, but an abstract structure of some sort ''acting on'' a set. (For instance, to anticipate the subject of the next section, if M is a module over a ring R, then the elements of R act on the underlying additive group of M.) In such situations, there is often a ''multiplication'' operation on the object that is acting, and we apply

Convention 3. *Suppose a set R has an operation denoted multiplicatively, $(f, g) \mapsto fg$, and we have an ''action'' of R on a set X, i.e., a map $R \times X \to X$. Then write this action as left or right ''multiplication'' if the action on X of a ''product'' of elements in R is given by the composite of the actions of the respective elements, in the order specified by Convention 1 for operations written on that side.*

The point of this Convention is clear. Let us now examine a particularly important class of applications.

2. Modules (and bimodules). Suppose a ring R acts on an additive group M in such a way that on applying to an element $x \in M$ an element $r \in R$, and then applying to the result another element $s \in R$, the final result is the same as that of applying $rs \in R$ to x. Then by Convention 3, it is natural to denote this action by writing elements of R on the right of elements of M. Assuming that the action is bilinear and that the identity element of R acts as the identity endomorphism of M, the resulting structure on M is that of a *right R-module*. If, instead, R acts so that the action of r followed by that of s is equivalent to the action of sr, then we write elements of R on the left of elements of M, and M becomes a *left R-module*. (Incidentally, one sometimes signals that an object M is a right or left R-module by writing it M_R, respectively ${}_R M$.)

Thus the essential difference between right and left modules is in the way multiplication of elements of R relates to composition of the associated operations on M. The difference in notation (xr vs. rx) is merely a corollary of this, though it gives the names to these two types of modules.

If M is a *right* R-module, then by Convention 2, a system of R-module *endomorphisms* of M should be written on the *left* of M. Such a system, if it is closed under composition and addition and contains the identity endomorphism, will form a ring. If a ring S acts on the left of a right R-module M by such module endomorphisms – or expressed more symmetrically, if a ring S acts on the left, and a ring R acts on the right, on the same abelian group M, and these actions respect one another – then the resulting structure is what is called an (S, R)-*bimodule* ${}_S M_R$. A familiar example is when M is the set ${}^m R^n$ of all $m\times n$ matrices over a ring R; this can easily be seen to form an $({}^mR^m, {}^nR^n)$-bimodule.

3. Multiplication and composition. The ''additional associativity laws'' (5) and (6) to which our conventions lead can be of more than just aesthetic and mnemonic value. A common way of constructing models of mathematical structures is as sets of *functions* – rings of operators, vector spaces of continuous functions on a topological space, etc.. In such cases, one often defines the operations of these objects by *composition* of functions. Since composition of functions is associative, we can only hope to realize a structure by such a model if it satisfies an appropriate associative law. Of course, if a structure satisfies a scrambled associative law such as $(f*g)x = g(fx)$, it may be possible to realize it by a system of functions with operation $*$ defined by $f*g = g\circ f$. But it is clearly easiest to see our way if we write things so that true associativity holds.

Since matrices are a model of linear maps among free modules, the matrix example discussed above is a model of the definition of bimodule, based on functional composition. Here is a related class of examples.

Suppose R is a ring, and M and N are right R-modules. Recall that the set of right R-module homomorphisms $M\to N$ forms an additive group, but (for R not commutative), does not in general form an R-module. (If we attempt to multiply a homomorphism h by an element $r\in R$, the result is not in general an R-module homomorphism, because multiplication by r itself is not; i.e., r does not in general commute with other elements of R.) However, if N is not only a right R-module but an (S,R)-bimodule for some ring S, then we can left-compose R-module homomorphisms $M\to N$ with the actions of elements of S (which, we have seen, are R-module endomorphisms of N) to get new R-module homomorphisms $M\to N$. In this way, the set of these right R-module homomorphisms becomes a left S-module. If, instead, we are given a (T,R)-bimodule structure, on M we can compose R-module homomorphisms $M\to N$ on the right with actions of elements of T, and our hom-set becomes a right T-module. If we have both kinds of structures, then our hom-set, $\mathrm{Hom}_R({}_TM_R, {}_SN_R)$, becomes an (S,T)-bimodule.

In particular, if M is a right R-module, and we form the set $\mathrm{Hom}_R(M,R)$, then since R is an (R,R)-bimodule, its left module structure induces a structure of left R-module on this set; this left R-module is called the *dual* of the right R-module M. Likewise, the dual of a *left* R-module, i.e., the set of left linear functionals on it, has a natural structure of *right* R-module.

This duality gives a contravariant functor $^\vee$ from right R-modules to left R-modules, and vice versa. If one writes all morphisms on the same side of their arguments, this contravariance takes the form $(fg)^\vee = g^\vee f^\vee$. However, if we adjust the sides on which morphisms are written according to Convention 2, and compose them accordingly, we get $(fg)^\vee = f^\vee g^\vee$; though we must still regard $^\vee$ as contravariant! If there is no danger of ambiguity, we can even drop the symbol ''$^\vee$'' from our morphisms, applying a homomorphism $f: M\to N$ to the left of elements of M to get elements of N, but to the right of elements of $N^\vee$ to get (by composition) elements of $M^\vee$.

It is not hard to see that the dual of a free right R-module of finite rank is a free left R-module of the same rank. If we regard the set ${}^mR^n$ of all $m\times n$ matrices as representing the set of module homomorphisms from the standard free right R-module on n generators to the standard free right R-module on m generators, then the above observation shows that the free left R-module on n generators can be identified with the $m=1$ case of this hom-set, i.e., the space row vectors of length n. Clearly, the free right R-module on m generators can itself be identified with the $n=1$ case, i.e., the space column vectors of height m. Thus, the free modules have been incorporated right into the system of matrices designed to represent their homomorphisms, and this system is symmetric, with free left modules and free right modules now in completely analogous roles. Suppressing superscripts ''^{1}'', this leads to the notation ${}^{m}R$ for the free right R-module on m generators, and R^{m} for the corresponding free left R-module, often used by ring-theorists.

Matrices also illustrate strikingly our earlier observation that if we write some maps on the right and some on the left, the properties of ''left'' and ''right'' invertibility will depend on the sort of morphisms we are looking at. A left-invertible $m\times n$ matrix represents both a *surjective* linear map from column vectors of height n to column vectors of height m, and a *one-to-one* linear map from row vectors of length m to row vectors of length n.

4. Assorted remarks.

4.1. Superscript notation is most frequently used for automorphisms. In particular, the inner automorphisms of a group are often written $h^{-1}gh = g^{h}$. This notation is consistent with Convention 3, since $(g^{h})^{h'} = g^{hh'}$, while if we defined g^{h} to mean hgh^{-1} (also often called the ''conjugate of g by h''), this would fail. A way to remember the correct choice is to note that g^{h} should have the form ''$...gh$'', so that when we compose two such operations we will get $(g^{h})^{h'} = ...ghh' = g^{hh'}$; if we defined the conjugate of g by h to have the form ''$hg...$'' it would, instead, behave like a left-multiplication operator. Group-theorists sometimes do use this definition, and write this conjugate ${}^{h}g$.

4.2. Suppose we are given an abelian group M, and an additive group R of endomorphisms of M closed under composition, and containing the identity endomorphism. It is amusing to note that until we decide on the notation we are going to use – writing these endomorphisms on the left and composing them accordingly, or writing them on the right and composing them accordingly – we can't formally say what their ring structure is, and whether M is a right or left module over this ring.

Similarly, if in a paper one makes a brief remark about functions in a context where it is not completely clear whether these would be written on the right or on the left of their arguments, there can be danger of ambiguity if one says something like ''composing f on the right with an appropriate automorphism α, ...''. I sometimes avoid this problem by saying ''precomposing f ...'', i.e., I use ''precompose'' and ''postcompose'' to mean, respectively, that the other map is to act before or after f.

4.3. If M is a right or left R-module, then its double dual $M^{\vee\vee}$ is a module of the same kind, and, just as in elementary linear algebra, there is an obvious natural module homomorphism $M \to M^{\vee\vee}$. If this map is an isomorphism, M is called *reflexive*. We see from previous remarks that free modules of finite rank are reflexive.

We commented in the preceding section that in writing the dual $f^{\vee}$ of a module homomorphism $f: M \to N$, one might drop the superscript ${}^{\vee}$ if there is no danger of ambiguity. The case where this is so is essentially that in which M and N are reflexive. In that case, it is easy to show that the functor ${}^{\vee}$ gives a bijection between right module homomorphisms $M \to N$ and left module homomorphisms $N^{\vee} \to M^{\vee}$; in other cases, the map in question may fail to be one-to-one and/or onto, and one needs to distinguish between the two hom-sets. It is the fact that free modules of finite rank *are* reflexive that allows us to use matrices both for maps among free left modules and maps among free right modules, in such a carefree fashion.

4.4. Here is an example of considerations like those of Convention 3, but in an important *nonassociative* context. A *Lie algebra* L over a field k is a k-vector space given with a bilinear operation called the ''Lie bracket'', not in general associative, and not written ab but $[a,b]$, which satisfies identities modeled on the properties of the *commutator* operation $[a,b] = ab - ba$ in an associative k-algebra. A ''module'' over the Lie algebra L is a vector space M given with a ''multiplication'' $(a,x) \mapsto ax$

($a \in L$, $x \in M$). Here one does not require the result of multiplying an element x by a bracket $[a, b]$ to coincide with the result of applying a and b to x in either of the two orders, $a(bx)$ or $b(ax)$. Instead, one requires that $[a, b]x = a(bx) - b(ax)$; this turns out to be the ''right'' definition for the subject.

4.5. One objection that could be raised to the idea of using two different notations for composition of maps on the same objects is that it would make for great notational confusion in considering the *category* of those objects. However, note that if we are considering, say, right modules over a ring R, and writing their homomorphisms on the left, following Convention 2, it is only these homomorphisms, and not the actions of elements of R, that occur as morphisms in the module category; so in considering that category, our Convention causes no problem. It is only if we want to regard these modules as objects in the category of abelian groups, and work both with the module homomorphisms and with the endomorphisms arising from the action of R, that we get a conflict. In that situation, it is probably best to set Convention 2 aside; though it is conceivable that if we wanted to work extensively with this sort of situation in a general category $\mathbf{C}$, we could set up conventions that tell us to write some morphisms in $\mathbf{C}$, and others in the opposite category, $\mathbf{C}^{\mathrm{op}}$.

4.6. A similar problem arises with homomorphisms of *bimodules*. If M, N are (S, T)-bimodules, and $\varphi: M \to N$ a bimodule homomorphism, i.e., both an S-module and a T-module homomorphism, then since it commutes with both the action of S on the left and the action of T on the right, Convention 2 gives us contradictory instructions as to where to write it. Here, obviously, we cannot fully follow that Convention! (If our writing system did not encourage linearity, we might write φ above or below the element to which it is to be applied. The condition that our homomorphism respect the actions of both S and T would again allow us to drop ''parentheses'' in such expressions; but otherwise the idea seems to have little to say for it.)

In this context I will mention that ring-theorists sometimes find it useful to note that an (S, T)-bimodule is equivalent to an abelian group with *left* actions of the rings S and T^{op} which commute with one another elementwise; equivalently, to a left module over the ring $S \otimes T^{\mathrm{op}}$, and they investigate properties of the latter ring as a tool in studying such bimodules.

5. Commutative rings, and rings with involution. The opposite to the sort of problem described in the last remark arises when we consider modules over a commutative ring R, since there is no distinction between saying that multiplication by $r \in R$ followed by multiplication by $s \in R$ acts like multiplication by rs, and saying that it acts like multiplication by sr.[1] We can only speak of objects on which this composite behaves like the common value $rs = sr$, which are, of course, called *R-modules*. Here we can decide arbitrarily to write the actions of elements of R on one side, say the right, and homomorphisms on the other; in particular, we can decide arbitrarily to identify a free R-module M of rank n with the standard left module of column vectors of height n; in that case the dual module (often called M^*, but called by Lang $M^{\vee}$) will be identified with the standard left module of row vectors. Although modules of row and of column vectors are isomorphic, if they arise differently in a given context, this formal distinction can actually help us keep track of things; e.g., distinguish modules that arise covariantly or contravariantly from given data.

An important context where we *cannot* maintain such a distinction is when we are given a free module

[1] And the Lord said, ' ... should I not have pity on Nineveh, that great city, wherein are more than six score thousand persons that cannot discern between their right hand and their left hand, and also much cattle?' Jonah 4:10-11.

M over a commutative ring, together with an isomorphism with its dual. This is equivalent to a nonsingular bilinear map $M \times M \to R$. (An inner product on a real vector space is an example of such a map.) As an example of the difficulty and its consequences, consider a linear map $f: M \to N$ of R-modules. According to our Conventions, if we treat M and N as right R-modules, we want to write f to the left of its argument; and the dual map $N^{\vee} \to M^{\vee}$, given by composition with f, can be denoted by writing f to the right of members of $N^{\vee}$. But if M and N are given with isomorphisms that ''identify'' them with $M^{\vee}$ and $N^{\vee}$, we want to write homomorphisms between the latter modules on the left.

In fact, if we take the above induced map $N^{\vee} \to M^{\vee}$ and regard it as a map $N \to M$, the result is what is called the *transpose* ${}^t f$ of f; and the consequence of our switching it from the right to the left of its argument is the familiar equation ${}^t(fg) = {}^t g\, {}^t f$.

Some of these considerations on commutative rings can be generalized to not necessarily commutative rings R given with an *involution*, that is, a map $*: R \to R$ satisfying the identity $r^{**} = r$, which respects addition and *reverses* multiplication: $(rs)^* = s^* r^*$. If R is a ring with involution, then any right R-module M can be made a left R-module by defining, for $r \in R$, $x \in M$,

$$rx \;=\; xr^*.$$

Thus the dual of a left module can be turned back into a left module, so that we can again consider left modules given with isomorphisms with their duals, and again use duality to get, for every map $M \to N$ between modules given with such isomorphisms, a ''transpose'' map $N \to M$. In the case of standard free modules, made isomorphic with their duals by taking each member of the standard basis to the corresponding member of the standard dual basis, this ''transpose'' operation corresponds to transposing the matrix representing a module homomorphism, and applying $*$ to each entry.

Although I have brought this up as a way to extend certain ideas of commutative ring theory to the noncommutative case, the most familiar example is actually one in which R is a commutative ring, namely the complex numbers, and $*$ is complex conjugation. Then the above considerations yield the familiar ''conjugate transpose'' operation on matrices.

For the reader curious to see some noncommutative rings with involution, we can begin by biting our tail, and giving the example ${}^n R^n$, where R is any commutative ring, and $*$ is the transpose operation; a variant is ${}^n\mathbf{C}^n$ with $*$ the conjugate transpose operation. A somewhat related example is the division ring of quaternions, on which one can easily verify that $a+bi+cj+dk \;\mapsto\; a-bi-cj-dk$ is an involution. For a couple of examples a little farther afield, let k be any field, and form the group algebra kG of any group, or the free associative algebra $k\langle X\rangle$ on any set. In the former case, the map that is the identity on k and sends each $g \in G$ to g^{-1} is an involution; in the latter, the map that is the identity on k and reverses the order of factors in every monomial is one.

6. A final word. Even when no formal difficulties arise, I do not *always* follow, nor recommend that anyone *always* follow the system of Conventions given in §1 above. Sometimes one wishes to use ''standard functional notation'' $f(x)$ to make very clear what is function and what is argument, even if this leads to violations of Conventions 2 and 3; sometimes there are other reasons.

But in many situations, those Conventions are of great value, especially in understanding the structures of hom-sets of modules, and in working with matrices, so I urge the student to familiarize himself or herself with them.