HOW TO HANDLE PARTIAL TRANSFORMATIONS

CHRISTOPHER D. HOLLINGS

A number of concepts in modern algebra have arisen as abstract versions of systems of functions of one type or another, the most famous example of course being groups: the abstractions of systems of bijections or permutations. Going further, the articulation in the 1920s and 1930s (see, for example, [11]) of the fact that non-invertible transformations are no less ubiquitous than bijective ones, led to the abstraction of systems of arbitrary transformations and the beginning of the theory of semigroups [5].[1] Various other types of functions have also been studied in a similar manner. One particularly interesting instance, though one that is arguably lesser known, is that of *partial transformations*. In this article, I give a taste of the study of partial transformations and the corresponding abstract theory.

Simply put, a *partial transformation* α on a set X is a mapping between subsets of X. For convenience, we typically regard a partial transformation α as mapping from its domain $\operatorname{dom}\alpha$ to its image $\operatorname{im}\alpha$. Thus, any partial transformation is automatically regarded as being surjective. Partial transformations abound in mathematics, and we can easily construct (often slightly artificial) examples by tweaking the domain of a given function. Thus, to take an example from number theory, Euler's totient function φ is a fully-defined transformation of the integers, but only a *partial* transformation when considered on the reals. A more interesting and considerably less contrived example (and one that we will revisit below) involves transformations between open sets of a topological space: these are clearly *partial* transformations of the space as a whole.

The issue remains of how to compose two partial transformations, but this is easily dealt with: we compose them (from left to right[2]) on the largest domain upon which it makes sense to do so, namely, given partial transformations α, β on a set X, we form their composition $\alpha \circ \beta$ on the domain

$$\operatorname{dom}(\alpha \circ \beta) = (\operatorname{im}\alpha \cap \operatorname{dom}\beta)\,\alpha^{-1}, \tag{1}$$

where α^{-1} denotes the preimage under α. It is clear that this domain consists of all those elements in $\operatorname{dom}\alpha$ that are mapped into $\operatorname{dom}\beta$ by α, and to which

Date: June 7, 2016.

[1]A *semigroup* is a set with an associative binary operation.

[2]We do this not to be perverse but because it sits more easily with the upcoming material on binary relations. When reading compositions from left to right, it is also more convenient to write functions on the right of their arguments: thus, our 'xf' denotes what would more usually be written '$f(x)$'.

we may then apply β; the situation is illustrated in Figure 1. As one might expect, the result of applying the composite partial transformation $\alpha \circ \beta$ to any $x \in \operatorname{dom}(\alpha \circ \beta)$ is $x(\alpha \circ \beta) = (x\alpha)\beta$.

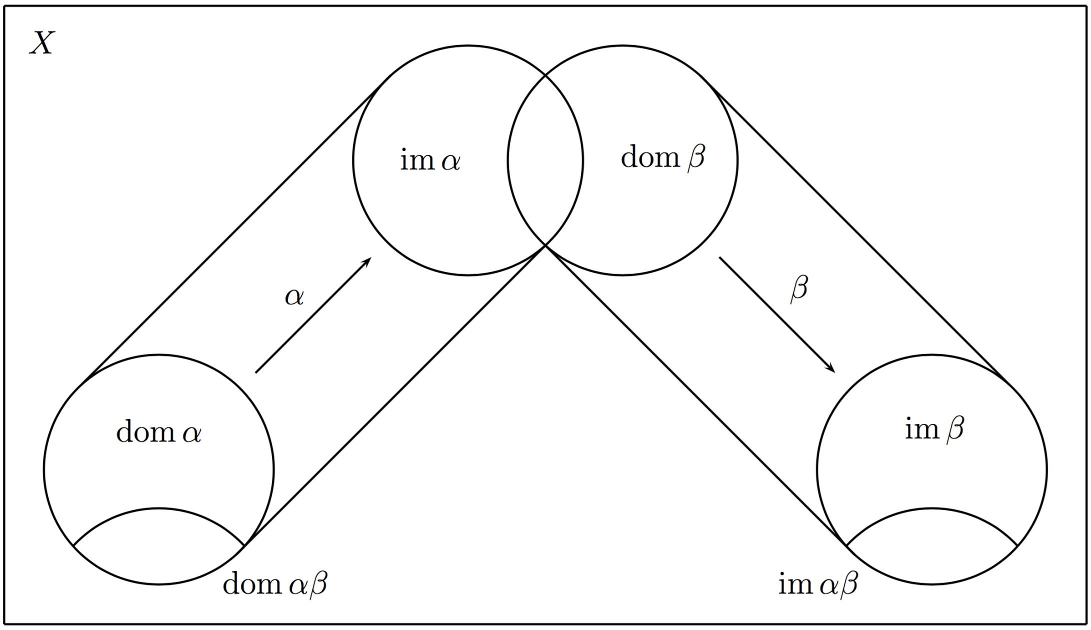

FIGURE 1. Composition of partial transformations

In fact, the above comment that the composition of partial transformations is "easily dealt with" was rather glib, for the issue of how to compose two partial transformations is one that came with certain conceptual difficulties that vexed mathematicians for a couple of decades. An early example of the use of partial transformations may be found in Oswald Veblen and J. H. C. Whitehead's *The foundations of differential geometry* of 1932 [12]: Veblen and Whitehead sought to devise a structure that would serve as a tool in differential geometry by analogy with the uses to which groups may be put in, say, Euclidean geometry. Their solution was the notion of a *pseudogroup*: a collection of homeomorphisms between open sets of a topological space such that the collection is closed under inversion of mappings and under composition. Speaking very loosely, the objects of interest in differential geometry would then be those structures that remain invariant under the action of a pseudogroup (cf. invariance under the action of a group in other geometries). But what was the composition that Veblen and Whitehead adopted for their pseudogroup? It wasn't that given in (1) but rather the rule that we may only compose partial transformations α, β (in that order) if $\operatorname{im}\alpha = \operatorname{dom}\beta$. The pseudogroup was thus endowed only with a *partial* composition.

Although by no means unknown in mathematics by this stage (see, for example, [1, 7]) the notion of a partial composition is one that appears to have caused some consternation amongst mathematicians. Certainly, when handling partial compositions, one must be very careful about where and when composite functions are defined, with the result that the corresponding mathematics can be rather fiddly. Well-behaved and fully-defined compositions appear therefore to have been preferred in general, and this was most certainly the case in connection with the pseudogroup. Efforts were made to extend the operation in a pseudogroup to a fully-defined one, although it was not immediately obvious how this might be achieved. The Dutch differential geometers J. A. Schouten and J. Haantjes, for example, extended the operation in such a way that $\alpha \circ \beta$ exists whenever $\operatorname{im} \alpha \cap \operatorname{dom} \beta \neq \emptyset$ [9]. A couple of years later, a Polish mathematician, Stanisław Gołąb, developed the beginnings of a broader theory of pseudogroups, using a composition whereby $\alpha \circ \beta$ exists whenever $\operatorname{im} \alpha \subseteq \operatorname{dom} \beta$ [3]. We see immediately, however, that although the compositions adopted by Schouten and Haantjes, and by Gołąb, were less restrictive than that of Veblen and Whitehead, they still failed to be fully-defined operations, for they did not take account of the possibility that we might have $\operatorname{im} \alpha \cap \operatorname{dom} \beta = \emptyset$. The composition of any such α, β simply was not defined. With the unfair benefit of hindsight, we can see that the detail that all of these authors were missing was the notion of the *empty transformation*. Often denoted by ε, this is the partial transformation on X with $\operatorname{dom} \varepsilon = \emptyset$. To return to the composition given in (1), we see in fact that the admission of the empty transformation is essential to this being a fully-defined composition: the presence of ε enables us to handle those cases where $\operatorname{im} \alpha \cap \operatorname{dom} \beta = \emptyset$. Indeed, the collection of all partial transformations on a set X (denoted $\mathcal{PT}_X$) is closed under the composition (1), and ε serves as the zero element of this set, i.e., $\varepsilon \circ \alpha = \varepsilon = \alpha \circ \varepsilon$, for all $\alpha \in \mathcal{PT}_X$.

It is difficult for us to appreciate nowadays the conceptual difficulties that accompanied the admission of ε as a valid partial transformation, although perhaps we could draw parallels between this and the only-very-gradual admission of 0 as a valid number in centuries past. The realisation that the empty transformation ought to be employed in this way appears to have been made first by yet another differential geometer, the Russian V. V. Wagner, who also pointed to an easier way of viewing it [14]. The contribution of a lengthy appendix to the Russian translation of Veblen and Whitehead's text seems to have set Wagner thinking about pseudogroups and related concepts. His key observation in this connection was that the composition of partial transformations, as given in (1), is merely a special case of that of binary relations. A *binary relation* ρ on a set X is simply a subset of $X \times X$. We may compose two binary relations ρ, σ on the same set

X in the following manner:[3]

(2) $(x, y) \in \rho \circ \sigma$ if there exists $z \in X$ such that $(x, z) \in \rho$ and $(z, y) \in \sigma$.

Any partial transformation α on X may be viewed as a binary relation, namely

$$\{(x, y) \in \operatorname{dom} \alpha \times \operatorname{im} \alpha : y = x\alpha\}.$$

It is left as an exercise for the reader to verify that, upon adopting this new point of view, the composition rules (1) and (2) are in fact one and the same in the case of partial transformations. If one thus views partial transformations as particular subsets of $X \times X$, it no longer seems like such a conceptual leap to admit the empty transformation: this simply corresponds to the empty subset of $X \times X$.

The case that particularly interested Wagner was that of *injective* partial transformations; given the convention that any partial transformation is defined so as to be surjective, we see that an *injective* partial transformation on a set X is then a *bijection* between subsets of X, and is therefore often called a *partial bijection*, a term that we adopt here. Indeed, the transformations that made up Veblen and Whitehead's pseudogroup were assumed to be injective, and hence provided a description of 'local symmetries' within the space in question: a bijection from one subset to another expresses the fact that the two subsets 'look alike'.

Given a partial bijection α on a set X, we clearly have the inverse partial bijection $\alpha^{-1} : \operatorname{im} \alpha \to \operatorname{dom} \alpha$. However, as Wagner observed, something interesting happens when we apply (1) to a partial bijection and its inverse: we do not obtain the identity transformation on X as we would in the case of a group (indeed, this need not belong to the collection of partial bijections under consideration), but rather a *partial identity*. If we denote by I_A the identity transformation on $A \subseteq X$, then it is easy to verify that $\alpha \circ \alpha^{-1} = I_{\operatorname{dom} \alpha}$ and $\alpha^{-1} \circ \alpha = I_{\operatorname{im} \alpha}$. Applying (1) again, we may further deduce that $\alpha \circ \alpha^{-1} \circ \alpha = \alpha$ and $\alpha^{-1} \circ \alpha \circ \alpha^{-1} = \alpha^{-1}$. Indeed, these latter relationships proved to be key to Wagner's investigations, for not only did he seek to 'complete' Veblen and Whitehead's composition, but he also, to return to the theme of our opening paragraph, sought to derive the abstract version of a system of partial bijections [15, 16]. The abstraction in question was dubbed a *generalised group* by Wagner, although it is now known as an *inverse semigroup*, a name coined by the British mathematician G. B. Preston, who devised the same structure independently [8]. An *inverse semigroup* is a semigroup S in which every element s has a unique 'inverse' in the sense that there exists an element $s^{-1} \in S$ such that $ss^{-1}s = s$ and $s^{-1}ss^{-1} = s^{-1}$. We see from above that the collection of all partial bijections on a set X, denoted $\mathcal{I}_X$, forms an inverse semigroup under the composition (1); $\mathcal{I}_X$ is termed[4]

---

[3]It is because it seems more natural to read the composition of binary relations from left to right that we have adopted the same convention for composition of functions.

[4]It should be noted that the point of view presented here is very much the semigroup-theoretic one. Amongst combinatorialists and representation theorists, what we have termed the symmetric inverse semigroup is more commonly known, in the case of finite X, as the *rook*

the *symmetric inverse semigroup on* X because it plays a role in the theory of inverse semigroups that is analogous to that of the symmetric group in group theory, namely that any inverse semigroup may be embedded in some $\mathcal{I}_X$. Thus, to revisit the ideas noted in the preceding paragraph, inverse semigroups provide us with an abstract theory of 'local symmetries' [6].

The study of partial transformations can of course be taken in many other directions. For instance, we might choose to drop the requirement of injectivity and thus return to the study of *arbitrary* partial transformations. Since these are not quite so well behaved as the injective ones, a slightly more elaborate theory is needed to describe the structure of $\mathcal{PT}_X$: for an outline of one approach to this, see [4]. On the other hand, we might instead follow the approach to pseudogroups taken by Charles Ehresmann [2] and try to overcome any prejudices that we might have against partial compositions by reverting to that of Veblen and Whitehead: $\alpha \circ \beta$ is defined only if $\operatorname{im}\alpha = \operatorname{dom}\beta$. The reader may already have observed that such a composition is redolent of that in a category, and so it is no coincidence that the modern theories of categories (in the special case of so-called *inductive groupoids*) and inverse semigroups share such rich interconnections (see, for example, [6]). To conclude, we note that one further direction in which we might take things would be to do as Wagner did and study partial transformations *between two distinct sets* A and B, i.e., mappings from subsets of A to subsets of B [13, 16]. Here, the theory has a rather different character, as we need to introduce a *ternary* operation into consideration, and that is another story entirely...

References

[1] Brandt, H. (1926) Über eine Verallgemeinerung des Gruppenbegriffes, *Math. Ann.*, vol. 96, pp. 360–366.

[2] Ehresmann, Ch. (1980–1983) *Oeuvres complètes et commentées*, supplements to *Cahiers de topologie et géométrie différentielle* (A. C. Ehresmann, ed.), Amiens.

[3] Gołąb, S. (1939) Über den Begriff der 'Pseudogruppe von Transformationen', *Math. Ann.*, vol. 116, pp. 768–780.

[4] Hollings, C. (2009) From right *PP* monoids to restriction semigroups: a survey, *Europ. J. Pure Appl. Math.*, vol. 2, pp. 21–37.

[5] Hollings, C. (2014) *Mathematics across the Iron Curtain: a history of the algebraic theory of semigroups*, Amer. Math. Soc.

[6] Lawson, M.V. (1998) *Inverse semigroups: the theory of partial symmetries*, World Scientific.

[7] Loewy, A. (1927) Über abstrakt definierte Transmutationssysteme oder Mischgruppen, *J. reine angew. Math.*, vol. 157, pp. 239–254.

[8] Preston, G.B. (1954) Inverse semi-groups, *J. London Math. Soc.*, vol. 29, pp. 396–403.

monoid [10]. For $|X| = n$, the rook monoid is typically viewed in the equivalent form of a semigroup of *partial permutation matrices*: $n \times n$ matrices whose entries are either 0 or 1, and in which each row and each column contains at most one non-zero entry. Such matrices correspond to all possible arrangements of non-attacking rooks on an $n \times n$ chessboard.

[9] Schouten, J.A. and Haantjes, J. (1937) On the theory of the geometric object, *Proc. London Math. Soc.*, vol. 42, pp. 356–376.
[10] Solomon, L. (2002) Representations of the rook monoid, *J. Algebra*, vol. 256, pp. 309–342.
[11] Sushkevich, A.K. (1937) *Theory of generalised groups*, DNTVU, Kharkov–Kiev (in Russian).
[12] Veblen, O. and Whitehead, J.H.C. (1932) *The foundations of differential geometry*, Cambridge Tract No. 29. Cambridge Univ. Press; Russian trans. with appendix by V. V. Wagner, Izdat. Inost. Lit., Moscow, 1949.
[13] Wagner, V.V. (1951) Ternary algebraic operations in the theory of coordinate structures, *Dokl. Akad. nauk SSSR*, vol. 81, pp. 981–984 (in Russian).
[14] Wagner, V.V. (1952) On the theory of partial transformations, *Dokl. Akad. nauk SSSR*, vol. 84, pp. 653–656 (in Russian).
[15] Wagner, V.V. (1952) Generalised groups, *Dokl. Akad. nauk SSSR*, vol. 84, pp. 1119–1122 (in Russian).
[16] Wagner, V.V. (1953) Theory of generalised heaps and generalised groups, *Mat. sb.*, vol. 32, pp. 545–632 (in Russian).

Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK

The Queen's College, Oxford, OX1 4AW, UK
E-mail address: christopher.hollings@maths.ox.ac.uk