Tarski's Fixed Point Theorem for Power Sets[*]

Robert Harper

Spring, 2020

1 Introduction

Tarski's theorem states that a monotone function on a complete lattice has a complete lattice of fixed points, in particular a least and greatest. A useful class of special cases are powerset lattices ordered by inclusion.

2 Tarski's Theorem

Let X be a set, not necessarily non-empty, and let $\wp X$ be the set of all subsets of X. The set $\wp X$ forms a complete lattice under set inclusion, with meets given by intersection and joins given by union. That is, if $\mathcal{X} \subset \wp X$, then $\bigcap \mathcal{X}$ is its meet (greatest lower bound) and $\bigcup \mathcal{X}$ is its join (least upper bound). The least element is the join of the empty set, namely $\emptyset$, and the greatest element is the meet of the empty set, namely X.

A function $F : \wp X \to \wp X$ is *monotone* if it preserves inclusion: if $A \subseteq B \subseteq X$, then $F(A) \subseteq F(B) \subseteq X$. For monotone F on $\wp X$, a *pre-fixed point* of F is a set $A \subseteq X$ such that $F(A) \subseteq A$, and a *post-fixed point* of F is a set $A \subseteq X$ such that $X \subseteq F(X)$. A pre-fixed point of F is also said to be *F-closed* and a post-fixed point of F is said to be *F-consistent.* A *least pre-fixed point* of a monotone F is the *smallest* (with respect to containment) F-closed set, and a *greatest post-fixed point* of F is the *largest* (with respect to containment) F-consistent set. Viewing the lattice as a (skinny) category, a monotone function F on it is a *functor*, a pre-fixed point of it is an *F-algebra* and a post-fixed point of it is an *F-coalgebra.* Thus, a *least* pre-fixed point of F is an *initial F-algebra* and a *greatest* post-fixed point of F is a *final F-coalgebra.*

Every monotone $F : \wp X \to \wp X$ has a (unique) least pre-fixed point and (unique) greatest post-fixed point, given by the equations

$$\mu(F) = \bigcap \{\, A \subseteq X \mid F(A) \subseteq A \,\}$$
$$\nu(F) = \bigcup \{\, A \subseteq X \mid A \subseteq F(A) \,\}$$

It is evident that $\mu(F)$ is contained in all pre-fixed points of F, it being their intersection. In fact $\mu(F)$ is itself a pre-fixed point of F, $F(\mu(F)) \subseteq \mu(F)$, and it is thereby the least pre-fixed point. To see this, it is enough to show that if $F(A) \subseteq A$, then $F(\mu(F)) \subseteq A$. But if $F(A) \subseteq A$, then $\mu(F) \subseteq A$ by definition of $\mu(F)$, and so, by monotonicity, $F(\mu(F)) \subseteq F(A) \subseteq A$, as required. Then

[*] Copyright © Robert Harper. All Rights Reserved.

again by monotonicity $F(F(\mu(F))) \subseteq F(\mu(F))$, which is to say that $F(\mu(F))$ is a pre-fixed point of F, and therefore $\mu(F) \subseteq F(\mu(F))$. That is, $\mu(F)$ is a fixed point of F, and, because any fixed point is a pre-fixed point, it is the least such. Dually, $\nu(F)$ contains all post-fixed points of F, it being their union, and, arguing dually to the preceding, $\nu(F)$ is itself a post-fixed point of F. It is thus the greatest post-fixed point, and hence a fixed point. (Categorially, this is Lambek's Lemma, which states that the initial F-algebra and final F-coalgebra are isomorphisms.)

The least fixed point of a monotone F on $\wp X$ affords the following *induction principle*: to show that $\mu(F) \subseteq A$, it suffices to show that $F(A) \subseteq A$, which is to say that A is F-closed. Similarly, the greatest post-fixed point, $\nu(F)$, of F affords the following *coinduction principle*: to show that $A \subseteq \nu(F)$, it suffices to show that $A \subseteq F(A)$, which is to say that A is F-consistent. Re-phrased in terms of predicates and implication, the least fixed point of a monotone F is the *strongest* property A of elements of X such that if $x \in F(A)$, then $x \in A$. Thus, to show that $x \in \mu(F)$ implies $x \in A$, it is enough to show that if $x \in F(A)$, then $x \in A$. Dually, the greatest fixed point of a monotone F is the *weakest* property A of elements of X such that if $x \in A$, then $x \in F(A)$.

As a case in point there are two proofs that $\mu(F) \subseteq \nu(F)$, one using the minimality of $\mu(F)$, the other using the maximality of $\nu(F)$. Because $\mu(F)$ is the least pre-fixed point of F, it is itself F-closed, and because $\nu(F)$ is the greatest post-fixed point of F, it is itself F-consistent. Thus, to show the containment it suffices to show either that $\nu(F)$ is F-closed, $F(\nu(F)) \subseteq \nu(F)$, or that $\mu(F)$ is F-consistent, $\mu(F) \subseteq F(\mu(F))$. But these are exactly the converse containments that were obtained earlier to show that $\mu(F)$ and $\nu(F)$ are fixed points of F.

Yet another perspective on the least pre-fixed point and greatest post-fixed point of a monotone F is provided by the following visualization of inductive and coinductive proofs. To show that every element of $\mu(F)$ is also in some set A representing a property of interest, it is enough to show that $\mu(F) \cap A$ is closed under F. For if it is, then $\mu(F)$ is contained in this intersection, and hence in A. The intersection $\mu(F) \cap A$ is *a priori* smaller than $\mu(F)$, consisting only of those elements of $\mu(F)$ that are "good enough" to have property A. But if the intersection is not so restrictive as to not be closed under F, then in fact the intersection is no restriction at all, it being that $\mu(F) \cap A = \mu(F)$. Dually, to show that some collection $A \subseteq X$ of elements is contained in $\nu(F)$, it is enough to boldly assert that they are by forming $\nu(F) \cup A$, which is *a priori* larger than $\nu(F)$. But if the union is consistent with F, then the assertion of A is indefeasible, and hence $\nu(F) \cup A \subseteq \nu(F)$. The union is not, in fact, larger than $\nu(F)$ after all—the elements of A were present from the get-go.

It is also possible to show that the meet and join of any set of fixed points of a monotone function is itself a fixed point, but this seems not to be as useful as the construction of a least and greatest. Because the development relies only on the universal properties of intersection and union, their being the meet and join, respectively, with respect to set containment, it is straightforward to obtain Tarski's Theorem in full, which is stated for complete lattices, those pre-orders for which all subsets have meets and joins.

3 Bekić's Lemma

It sometimes arises that two sets (properties) are to be *simultaneously* inductively defined because the definition of each depends on the other. This situation can be expressed by considering two monotone operators $F, G : \wp X \times \wp X \to \wp X$ in the sense that if $A \subseteq A'$ and $B \subseteq B'$, then $F(A, B) \subseteq F(A', B)$ and $G(A, B) \subseteq G(A, B')$. The operator $(F, G)(A, B) \triangleq (F(A, B), G(A, B))$ is therefore monotone with respect to the componentwise ordering, $(A, B) \subseteq (A', B')$ iff $A \subseteq A'$ and

$B \subseteq B'$. By arguments analogous to those given above[1] it has a least fixed point, $\mu(F, G)$, a pair (A_0, B_0) of subsets of X such that $F(A_0, B_0) = A_0$ and $G(A_0, B_0) = B_0$, each "cross-referencing" the other as intended.

It is also possible to manage the interdependencies by an iterated process of taking least fixed points of monotone operators on $\wp X$. This reduction of simultaneous to iterated least fixed points is known as *Bekić's Lemma*.[2] Given F and G as above, define their curried forms by $F_B(A) \triangleq F(A, B)$, which fixes the B argument of F, and $G^A(B) = G(A, B)$, which fixes the A argument of G. These are both monotone, and hence admit least fixed points, $\mu(F_B) = F_B(\mu(F_B)) = F(\mu(F_B), B)$, which solves for A parametrically in B, and $\mu(G^A) = G^A(\mu(G^A)) = G(A, \mu(G^A))$, which solves for B parametrically in A. The maps $A \mapsto F(A, \mu(G^A))$ and $B \mapsto G(\mu(F_B), B)$ are also monotone, and hence have least fixed points. Bekić's Lemma states that the simultaneous fixed point can be separated into a pair of iterated fixed points:

$$\mu(F, G) = (\mu(A \mapsto F(A, \mu(B \mapsto G(A, B)))), \mu(B \mapsto G(\mu(A \mapsto F(A, B)), B))).$$

The proof of Bekić's Lemma is not difficult, but it is easy to get lost in the morass of fixed points. As suggested in *loc. cit.* it is helpful to separate the pattern of diagonalization, and then instantiate it for the proof of Bekić's Lemma. Suppose that $H : \wp X \times \wp X \to \wp X$ is monotone in each argument. Define $\Delta_H(A) = H(A, A)$, the diagonalization of H. The least fixed point of the diagonal satisfies the equation $\mu(\Delta_H) = H(\mu(\Delta_H), \mu(\Delta_H))$, the least simultaneous fixed point of H. The diagonal lemma states that the simultaneous and iterated fixed points coincide,

$$\mu(A \mapsto \mu(B \mapsto H^A(B))) = \mu((A, B) \mapsto \Delta_H(A, B)) = \mu(B \mapsto \mu(A \mapsto H_B(A))),$$

or, more concisely,

$$\mu(A \mapsto \mu(H^A)) = \mu(\Delta_H) = \mu(B \mapsto \mu(H_B)).$$

The proof of the first of these equations proceeds by showing that each side is contained in the other using the universal property of the least pre-fixed point. (The proof of the second proceeds analogously.) For the left-to-right containment, it suffices to show that $\mu(H^{\mu(\Delta_H)}) \subseteq \mu(\Delta_H)$:

$$\begin{aligned} H^{\mu(\Delta_H)}(\mu(\Delta_H)) &= H(\mu(\Delta_H), \mu(\Delta_H)) \\ &= \Delta_H(\mu(\Delta_H)) \\ &= \mu(\Delta_H). \end{aligned}$$

For the opposite containment it suffices to show that $\Delta_H(\mu(A \mapsto \mu(H^A))) \subset \mu(A \mapsto \mu(H^A))$. Let $H'(A) = \mu(H^A)$, and calculate:

$$\begin{aligned} \Delta_H(\mu(H')) &= H(\mu(H'), \mu(H')) \\ &= H^{\mu(H')}(\mu(H')) \\ &= H^{\mu(H')}(\mu(H^{\mu(H')})) \\ &= \mu(H^{\mu(H')}) \\ &= \mu(H'). \end{aligned}$$

[1] This is where the (mild) generalization of Tarski's Theorem to complete lattices is helpful.

[2] The formulation given here is adapted from Davey and Priestley (2002) Chapter 8, exercises 8.30 and 8.31. The original article is reproduced as Bekić (1984).

To prove Bekić's Lemma, simply apply the Diagonal Lemma to $\Delta_{(F,G)}$, and note that its left and right equivalents are equal to the desired forms. The trick is to separate variables and follow the dependencies. First, specialize the left iterated form to $H = (F, G)$, and obtain

$$\mu((A, B) \mapsto \mu(A', B') \mapsto (F(A, B), G(A', B'))).$$

Let $L(A, B) \triangleq \mu(L'(A, B))$, where $L'(A, B)(A', B') = (F(A, B), G(A', B'))$, so that the displayed formula can be written more concisely as $\mu(L)$, where $L(A, B) = \mu(L'(A, B))$. Calculating,

$$\begin{aligned}
\mu(L) &= L(\mu(L)) \\
&= L(L_1, L_2) \\
&= (F(L_1, L_2), G(\mu(L'(L_1, L_2)))),
\end{aligned}$$

where $L_1 \triangleq \mu(L) \cdot 1$ and $L_2 \triangleq \mu(L) \cdot 2$ are the two components of $\mu(L)$. Similarly define $L_1' \triangleq \mu(L'(L_1, L_2)) \cdot 1$ and $L_2' \triangleq \mu(L'(L_1, L_2)) \cdot 2$, and calculate to obtain these equations:

$$\begin{aligned}
L_1 &= F(L_1, L_2) \\
L_2 &= G(L_1', L_2') \\
L_1' &= F(L_1, L_2) \\
L_2' &= G(L_1', L_2').
\end{aligned}$$

Then $L_2 = G(F(L_1, L_2), G(L_1', L_2')) = G(L_1, L_2)$, which is to say that $L_2 = \mu(B \mapsto G(L_1, B))$. Then $L_1 = F(L_1, \mu(B \mapsto G(L_1, B)))$, which is to say that $L_1 = \mu(A \mapsto F(A, \mu(B \mapsto G(A, B))))$, the desired left component. A similar calculation completes the proof.

Here is Bekić's formulation and proof of his eponymous lemma:

Lemma 1 (Bekić)**.** *For monotone $F, G : \wp X \times \wp X \to \wp X$,*

$$\mu(F, G) = (A_0, B_0), \text{ where } A_0 = \mu(A \mapsto F(A, \mu(G^A))) \text{ and } B_0 = \mu(G^{A_0}).$$

Proof. First, note that $A_0 = F(A_0, \mu(G^{A_0})) = F(A_0, B_0)$ and $B_0 = G(A_0, B_0)$, so $(F, G)(A_0, B_0) = (A_0, B_0)$ is a pre-fixed point of (F, G). Then, suppose that $(F, G)(A, B) \subseteq (A, B)$ is another pre-fixed point of (F, G), and show that $(A_0, B_0) \subseteq (A, B)$. Expanding, $(F, G)(A, B) = (F(A, B), G(A, B))$, and so $G^A(B) = G(A, B) \subseteq B$, and therefore $\mu(G^A) \subseteq B$, and so by monotonicity $F(A, \mu(G^A)) \subseteq F(A, B) \subseteq A$. But A_0 is the least such set, so $A_0 \subseteq A$. Moreover, $B_0 \subseteq B$ because $\mu(G^{A_0}) \subseteq \mu(G^A) \subseteq B$ by monotonicity, and B_0 is the least such set. □

4 Assertions and Rules

A typical application of the fixed point constructions is to justify the definition of one or more *assertions*, or *formal judgments*, by a collection of *rules*. The idea is that the rules constitute an inductive definition of the mentioned assertions. For example, the following rules define the judgment $n\,\mathsf{nat}$, stating that n is a natural number:

$$\text{ZERO}\quad \frac{}{\mathsf{zero}\,\mathsf{nat}} \qquad\qquad \text{SUCC}\quad \frac{n\,\mathsf{nat}}{\mathsf{succ}(n)\,\mathsf{nat}}$$

Similarly, the even and odd numbers may be simultaneously defined by the following rules:

$$\overset{\text{ZERO-EVEN}}{\frac{}{\texttt{zero}\,\mathsf{even}}} \qquad \overset{\text{SUCC-ODD}}{\frac{n\,\mathsf{even}}{\texttt{succ}(n)\,\mathsf{odd}}} \qquad \overset{\text{SUCC-EVEN}}{\frac{n\,\mathsf{odd}}{\texttt{succ}(n)\,\mathsf{even}}}$$

In both cases the subjects of the assertions are abstract binding trees in the sense of Harper (2016), among which are those used above.

The forms of assertion may be thought of as labels that distinguish one from another. Thus, the underlying set X of the inductive definition is the collection of *assertions* consisting of an abt together with a label drawn from the set of such forms. Each rule r is of the form

$$\frac{j_1 \cdots j_n}{j}$$

wherein the j_i's and j are assertions (elements of X). Each such rule r determines a monotone function $F_r : \wp X \to \wp X$ that applies rule r to a given set $A \subseteq X$ of assertions:

$$F_r(A) = \{\, j \in X \mid j_1, \ldots, j_n \in A \,\}.$$

A set R of rules induces a monotone function that collectively closes up under each rule in the set:

$$F_R(A) = \bigcup_{r \in R} F_r(A).$$

The assertions defined by the set of rules R are precisely those in $\mu(F_R)$. The principle of *rule induction* is simply the induction principle associated with the least fixed point of F_R.

What about the greatest fixed point $\nu(F_R)$? As remarked earlier, $\nu(F_R)$ contains $\mu(F_R)$, but is the containment strict? The answer depends on the choice of subjects for the assertions in X. For an object to be in the least fixed point of a rule set means that it must be forced to be so by applying rules. But to be in the greatest fixed point means only that if an assertion is in it, and if it arises as the conclusion of some rule in R, then the premises of that rule must also be present. This condition allows for circular reasoning in the presence of self-referential syntactic objects. For example, suppose that ω is the infinite stack of successors $\texttt{succ}(\texttt{succ}(\texttt{succ}(\ldots)))$, which is to say that $\omega = \texttt{succ}(\omega)$, it is its own successor.[3] Considering the greatest fixed point interpretation of the rules defining $-\,\mathsf{nat}$, it is the case that $\omega\,\mathsf{nat}$, because $\omega = \texttt{succ}(\omega)$ satisfies the requirement that if $\omega\,\mathsf{nat}$, then $\omega\,\mathsf{nat}$, which is tautologous.

Finally, a word about "side conditions" on rules is in order[4]. Often rules are not presented fully schematically, but with so-called side conditions that constrain their applicability. For example, in the typing rule for λ-abstractions it is usual to include the requirement that x is not already declared in Γ:

$$\frac{\Gamma, x{:}A \vdash M : B \qquad x \notin \Gamma}{\Gamma \vdash \lambda x{:}A.M : A \to B}$$

When terms are identified up to α-equivalence, the restriction can always be met by choosing an appropriate representative, because Γ declares only finitely many of the infinitely many choices for

[3] If you are wondering how this could come about, just consider that the syntactic objects are graphs, rather than trees, with cycles allowed.

[4] Avron (1991) calls these "impurities."

the bound variable. In that case the condition may be omitted from the rule, it being understood implicitly.

Whereas in the preceding case the restrictions on the applicability of the rule are benign, it is entirely possible, even common, to abuse the privilege to the point of absurdity. For example, it is entirely meaningless to state as a premise of a rule the *negation* of the judgment being defined, as if to say "this rule is applicable if the stated assertion is not derivable." Doing this, or things tantamount to it, ruins the inductive character of the intended definition, precisely because the associated operator is no longer monotone! And without that the rules cannot be said to define anything at all. For a particularly blatant case, consider the supposed inductive definition of an assertion j by the rule

$$\frac{\neg j}{j}$$

Were there a fixed point of the associated (non-monotone) operator, the assertion would be derivable iff it is not, so obviously it must not have one. More subtle examples are observed in the wild, so it is wise to be careful out there!

References

Arnon Avron. Simple consequence relations. *Information and Computation*, 92(1):105–139, 1991.

Hans Bekić. Definable operations in general algebras, and the theory of automata and flowcharts. In C. B. Jones, editor, *Programming Languages and Their Definition: H. Bekić (1936-1982)*, Lecture Notes in Computer Science, pages 30–55. Spring Verlag, Heidelberg, 1984.

B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Second edition, 2002.

Robert Harper. *Practical Foundations for Programming Languages*. Cambridge University Press, Cambridge, England, Second edition, 2016.