


THEORETICAL PEARL

An example of goal-directed, calculational proof

ROLAND CARL BACKHOUSE 

School of Computer Science, University of Nottingham, Nottingham NG8 1BB, UK
(e-mail: roland.backhouse@nottingham.ac.uk)

WALTER GUTTMANN 

Computer Science and Software Engineering, University of Canterbury, Christchurch, New Zealand
(e-mail: walter.guttman@canterbury.ac.nz)

MICHAEL WINTER 

Department of Computer Science, Brock University, St. Catharines, Ontario, Canada
(e-mail: mwinter@brocku.ca)

Abstract

An equivalence relation can be constructed from a given (homogeneous, binary) relation in two steps: first, construct the smallest reflexive and transitive relation containing the given relation (the “star” of the relation) and, second, construct the largest symmetric relation that is included in the result of the first step. The fact that the final result is also reflexive and transitive (as well as symmetric), and thus an equivalence relation, is not immediately obvious, although straightforward to prove. Rather than prove that the defining properties of reflexivity and transitivity are satisfied, we establish reflexivity and transitivity *constructively* by exhibiting a starth root—in a way that emphasises the creative process in its construction. The resulting construction is fundamental to algorithms that determine the strongly connected components of a graph as well as the decomposition of a graph into its strongly connected components together with an acyclic graph connecting such components.

1 Introduction

Given a (homogeneous, binary) relation R , the relation R^* is the smallest reflexive and transitive relation containing R , and $R \cap R^{\cup}$ is the largest symmetric relation that is included in R . By applying these two constructions in order, the resulting relation $R^* \cap (R^*)^{\cup}$ is obviously symmetric; less obvious is that it is also reflexive and transitive, that is, the relation is an equivalence relation.

Backhouse *et al.* (2022, Theorem 139) prove that, for all relations R , the relation $R^* \cap (R^*)^{\cup}$ can be reformulated using the identity:

$$R^* \cap (R^*)^{\cup} = (R \cap (R^{\cup})^*)^* . \quad (1.1)$$

In words, $R \cap (R^\cup)^*$ is a *starth root* of the equivalence relation $R^* \cap (R^*)^\cup$. By proving the property (1.1), one establishes *constructively* that the relation $R^* \cap (R^*)^\cup$ is indeed reflexive and transitive, and thus an equivalence relation.

(In general, a starth root of a reflexive–transitive relation U is a relation V such that $U = V^*$. Since U is reflexive and transitive, $U = U^*$, every reflexive–transitive relation is a starth root of itself. The importance of a starth root becomes evident when it has particular properties, such as some form of minimality.)

Because of its constructive nature, the identity (1.1) plays a significant role in algorithms that exploit the decomposition of a finite graph into an acyclic graph together with a collection of strongly connected components. (In this application, the relation R corresponds to the edge relation on nodes defined by the graph, and $R^* \cap (R^*)^\cup$ is the relation that holds of two nodes when they are both in the same strongly connected component. Readers unfamiliar with the notation and/or property are referred to Section 2 for a brief summary.) However, as observed in Backhouse *et al.* (2022), the proof left a lot to be desired since it used the definition of the star operator (reflexive–transitive closure) as a sum of powers of R together with a quite complicated induction property. Attempts we had made to apply fixed-point fusion had failed.

Recently, Guttman formulated a proof using the inductive definition of R^* in point-free relation algebra. Winter made some improvements to Guttman’s proof.

Originally, the Guttman–Winter proof was presented in the traditional mathematical style: a bottom-up proof that miraculously ends in the final step with the desired property. In this note, the proof has been rewritten in a way that emphasises the heuristics that were used to construct the proof. Some comments on how to present difficult proofs follow the calculations.

2 Relation algebra

In the proof, we use a number of properties without specific mention. These properties will be known to readers well versed in relation algebra but for others may not be so. For this reason, we give a very brief summary of the relevant properties.

Variables R , S and T in the proof all denote homogeneous binary relations of the same type. The set notation we use (“ \subseteq ”, “ \cap ” and “ \cup ”) has its standard meaning, and we do assume familiarity with the properties of the set operators. A property that may be less familiar is that, for all S , the function $\cap S$ has an upper adjoint, which we denote by $S \rightarrow$. That is, for all R , S and T ,

$$R \cap S \subseteq T \equiv R \subseteq S \rightarrow T . \quad (2.1)$$

The property is a consequence of the universal distributivity of set-intersection over set-union. We call it the “Heyting Galois connection” because it is essentially the same as the adjunction between $\wedge p$ and $p \Rightarrow$ (for all propositions p) in intuitionistic logic, the formalisation of which is generally attributed to Heyting.

Relation composition and converse are denoted by “ \circ ” and “ \smile ”, respectively, and the identity relation is denoted by I . All of intersection, union, composition and converse are monotonic with respect to the subset ordering.

Converse is defined by the Galois connection, for all R and S ,

$$R^{\cup} \subseteq S \equiv R \subseteq S^{\cup}$$

together with the distributivity property, for all R and S ,

$$(R \circ S)^{\cup} = S^{\cup} \circ R^{\cup}$$

and the property that

$$I^{\cup} = I .$$

The modularity rule (aka the Dedekind rule) is used in both its forms: for all R, S and T ,

$$R \circ S \cap T \subseteq R \circ (S \cap R^{\cup} \circ T)$$

and its symmetric counterpart

$$R \cap S \circ T \subseteq (R \circ T^{\cup} \cap S) \circ T .$$

The rule is important because composition does not distribute over intersection: it gives a handle on expressions involving both operators where the intersection is on the lower side of a set inclusion.

R^* denotes the reflexive, transitive closure of R . The inductive definition of R^* used here¹ is the combination of the two properties:

$$I \cup R \circ R^* \subseteq R^*$$

and, for all T ,

$$R^* \subseteq T \Leftarrow I \cup R \circ T \subseteq T .$$

That is, R^* is the least prefix point of the function mapping T to $I \cup R \circ T$. We do not directly use the fact that R^* is a fixed point of this function, but we do use the (derived) property that, for all R ,

$$I \subseteq R^* \wedge (R^*)^* = R^* \wedge R^* \circ R^* = R^* \wedge (R^{\cup})^* = (R^*)^{\cup} .$$

We also use the fact that the star operator is monotonic with respect to the subset ordering.

It is mentioned in the introduction that the identity we have proved is central to a number of algorithms that exploit graph theory. In such algorithms, the relation R is the edge relation on nodes of a finite directed graph: specifically, two nodes u and v are related by R iff there is an edge in the corresponding graph from u to v . Conversely, two nodes u and v are related by R^{\cup} iff there is an edge in the graph from v to u . The graph corresponding to R^{\cup} is thus the graph obtained by reversing the edges of the graph corresponding to R . Nodes u and v are related by R^* iff there is a path from u to v in the graph, and by $(R^{\cup})^*$ iff there is a path from u to v in the graph formed of reversed edges. Equivalently, u and v are related by $(R^{\cup})^*$ iff there is a path from v to u in the graph. Formally, the equivalence is expressed by the identity $(R^{\cup})^* = (R^*)^{\cup}$.

The relation $R^* \cap (R^{\cup})^*$ holds between nodes u and v iff there is both a path from u to v and a path from v to u in the corresponding graph. Thus, $R^* \cap (R^{\cup})^*$ is the relation that

¹ An alternative fixed-point definition—alluded to in the text—is the direct formalisation of the property that R^* is the least reflexive, transitive relation that contains R .

holds between nodes u and v when both are in the same strongly connected component of the graph; moreover, since $(R^U)^* = (R^*)^U$, it equals $R^* \cap (R^*)^U$.

The relation $R \cap (R^U)^*$ holds between nodes u and v iff there is an edge from u to v and a path from v to u . The identity (1.1) thus states that nodes u and v are strongly connected iff there is a path from u to v in the graph corresponding to this relation. This insight is fundamental to algorithms that determine the strongly connected components of a graph as well as the decomposition of a graph into its strongly connected components together with an acyclic graph connecting such components.

3 The proof

As stated previously, for arbitrary relation R , the relation $R^* \cap (R^*)^U$ is an equivalence relation; it is, thus, reflexive and transitive. Our goal is to establish reflexivity and transitivity *constructively* by calculating a starth root of the relation. That is, we aim to calculate a relation T such that $R^* \cap (R^*)^U = T^*$.

The form of this goal suggests that the fixed-point fusion theorem is applicable: R^* is a least fixed point and, for all S , the function $\cap S$ is a lower adjoint in a Galois connection (the Heyting connection mentioned above). These are precisely the circumstances in which fusion is applicable: the theorem provides sufficient conditions on R and S under which R^* and $\cap S$ can be “fused” into a fixed point of the form T^* . However, our efforts to achieve the goal in this way failed: the conditions required by the fusion theorem are just too strong.

In view of this, we are obliged to substantially weaken our goal. The inclusion

$$(R \cap S)^* \subseteq R^* \cap S^* .$$

(for all R and S) is very easily proved and, since $(R^U)^* = (R^*)^U$, it immediately follows that

$$(R \cap (R^U)^*)^* \subseteq R^* \cap (R^*)^U .$$

(The full details are given at the end of this section.) This means that to prove (1.1) it suffices to prove the converse inclusion:

$$R^* \cap (R^*)^U \subseteq (R \cap (R^U)^*)^* .$$

This, in turn, suggests the weaker goal: we try to determine conditions on R , S and T such that

$$R^* \cap S \subseteq T^* .$$

The calculation is guided by the fact that the condition on S must be satisfied by $(R^U)^*$ and the condition on T by $R \cap (R^U)^*$, but we may be lucky and find weaker conditions. (In fact, we don’t—but it is worth a try.)

Let us begin the calculation:

$$\begin{aligned} & R^* \cap S \subseteq T^* \\ = & \quad \{ \text{Heyting Galois connection} \} \\ & R^* \subseteq S \rightarrow T^* \\ \Leftarrow & \quad \{ \text{fixed-point definition of } R^* \} \end{aligned}$$

$$\begin{aligned}
 & I \cup R \circ (S \rightarrow T^*) \subseteq S \rightarrow T^* \\
 = & \quad \{ \text{Heyting Galois connection} \} \\
 & (I \cup R \circ (S \rightarrow T^*)) \cap S \subseteq T^* \\
 = & \quad \{ \text{distributivity} \} \\
 & (I \cap S) \cup (R \circ (S \rightarrow T^*) \cap S) \subseteq T^* \\
 = & \quad \{ \text{Galois connection defining “}\cup\text{”} \} \\
 & I \cap S \subseteq T^* \wedge R \circ (S \rightarrow T^*) \cap S \subseteq T^* \\
 = & \quad \{ I \subseteq T^* \} \\
 & R \circ (S \rightarrow T^*) \cap S \subseteq T^* .
 \end{aligned}$$

Summarising, we have proved that, for all R, S and T ,

$$R^* \cap S \subseteq T^* \iff R \circ (S \rightarrow T^*) \cap S \subseteq T^* . \tag{3.1}$$

So far the steps taken have been relatively routine. The next steps are less so: we seek a condition on S that enables the elimination of “ $S \rightarrow$ ”. To this end, we calculate

$$\begin{aligned}
 & R \circ (S \rightarrow T^*) \cap S \\
 \subseteq & \quad \{ \text{modularity rule} \} \\
 & R \circ (S \rightarrow T^* \cap R^{\cup} \circ S) \\
 \subseteq & \quad \{ \text{introduce assumption as prelude to cancellation:} \\
 & \quad \bullet R^{\cup} \circ S \subseteq S \} \\
 & R \circ (S \rightarrow T^* \cap S) \\
 \subseteq & \quad \{ \text{(Heyting Galois connection) cancellation,} \\
 & \quad \text{and monotonicity of composition} \} \\
 & R \circ T^* .
 \end{aligned}$$

In this way, we have derived the property that, for all R, S and T ,

$$R \circ (S \rightarrow T^*) \cap S \subseteq R \circ T^* \iff R^{\cup} \circ S \subseteq S . \tag{3.2}$$

Note that the condition $R^{\cup} \circ S \subseteq S$ is indeed satisfied by $S = (R^{\cup})^*$.

We now continue the calculation that led to (3.1).

$$\begin{aligned}
 & R \circ (S \rightarrow T^*) \cap S \subseteq T^* \\
 = & \quad \{ \text{the hardest step in the calculation: as a prelude to applying (3.2),} \\
 & \quad \text{we exploit the idempotency of set-intersection} \} \\
 & R \circ (S \rightarrow T^*) \cap S \cap S \subseteq T^* \\
 \Leftarrow & \quad \{ \text{(3.2) and monotonicity} \} \\
 & R^{\cup} \circ S \subseteq S \wedge R \circ T^* \cap S \subseteq T^* \\
 \Leftarrow & \quad \{ \text{aiming for fixed-point definition of } T^*, \text{ use modularity rule} \} \\
 & R^{\cup} \circ S \subseteq S \wedge (R \cap S \circ (T^*)^{\cup}) \circ T^* \subseteq T^*
 \end{aligned}$$

$$\begin{aligned}
 &\Leftarrow \{ \text{fixed-point definition of } T^* \} \\
 &R^\cup \circ S \subseteq S \wedge R \cap S \circ (T^*)^\cup \subseteq T \\
 &\Leftarrow \{ \text{aiming for } S = (R^\cup)^*, \text{ rewrite } (T^*)^\cup \text{ as } (T^\cup)^* \\
 &\quad \text{and introduce conditions } (T^\cup)^* \subseteq S \text{ and } S \circ S \subseteq S \} \\
 &R^\cup \circ S \subseteq S \wedge (T^\cup)^* \subseteq S \wedge S \circ S \subseteq S \wedge R \cap S \subseteq T \\
 &\Leftarrow \{ \text{1st conjunct: fixed-point definition of } (R^\cup)^* \\
 &\quad \text{4th conjunct: reflexivity of } \subseteq \\
 &\quad \text{3rd conjunct: transitivity of } (R^\cup)^* \\
 &\quad \text{2nd conjunct: } R \cap S \subseteq R \text{ and monotonicity of converse and star } \} \\
 &S = (R^\cup)^* \wedge T = R \cap S .
 \end{aligned}$$

Summarising the calculation, we have proved that, for all R, S and T ,

$$R \circ (S \rightarrow T^*) \cap S \subseteq T^* \Leftarrow S = (R^\cup)^* \wedge T = R \cap (R^\cup)^* . \tag{3.3}$$

Combining (3.1) and (3.3), we get

$$R^* \cap (R^\cup)^* \subseteq (R \cap (R^\cup)^*)^* . \tag{3.4}$$

As mentioned earlier, the opposite inclusion is easy to prove:

$$\begin{aligned}
 &(R \cap (R^\cup)^*)^* \subseteq R^* \cap (R^\cup)^* \\
 &= \{ \text{Galois connection defining intersection} \} \\
 &(R \cap (R^\cup)^*)^* \subseteq R^* \wedge (R \cap (R^\cup)^*)^* \subseteq (R^\cup)^* \\
 &\Leftarrow \{ \text{1st conjunct: star is monotonic} \\
 &\quad \text{2nd conjunct: } (R^*)^* = R^* \text{ (with } R := R^\cup \text{) and star is monotonic} \} \\
 &R \cap (R^\cup)^* \subseteq R \wedge R \cap (R^\cup)^* \subseteq (R^\cup)^* \\
 &= \{ \text{Galois connection defining intersection} \} \\
 &R \cap (R^\cup)^* \subseteq R \cap (R^\cup)^* \\
 &= \{ \text{reflexivity of } \subseteq \} \\
 &\text{true .}
 \end{aligned}$$

The identity (1.1) now follows from the antisymmetry of the subset relation and the fact that $(R^\cup)^* = (R^*)^\cup$.

4 Specific comments

Before making more general remarks, some comments on the calculation are in order.

The central problem in the initial calculations is how to deal with the occurrence of the intersection operator (“ \cap ”) on the lower side of an inclusion (“ \subseteq ”).

The first calculation is quite straightforward and relatively self-evident: R^* is by definition a least fixed point, and it is very common to use fixed-point induction to establish less obvious properties. (Formally, fixed-point induction is the rule that a least fixed point is a least prefix point. In this case, the rule used is that, for all R and S ,

$$R^* \subseteq S \Leftarrow I \cup R \circ S \subseteq S .$$

There is a choice of which fixed-point definition of R^* to use should the calculation fail.) The combination of fixed-point induction with the use of a Galois connection is also very common. In this case, the ‘‘Heyting’’ Galois connection is, for all R, S and T ,

$$R \cap S \subseteq T \equiv R \subseteq S \rightarrow T .$$

The problem of the intersection operator is resolved by simply ‘‘shunting’’ it out of the way and then ‘‘shunting’’ it back. The remaining steps are relatively self-evident.

The issue that must be resolved in the second calculation is that ‘‘ $S \rightarrow$ ’’ has been introduced on the left side of an inclusion. It is vital that this is eliminated. The Heyting Galois connection suggests a line of attack. Specifically, we have the cancellation rule: for all S and T ,

$$(S \rightarrow T) \cap S \subseteq T .$$

Aiming to apply cancellation, the calculation begins by applying the modularity rule. In this way, (3.2) is easily derived.

Undoubtedly, the hardest step of all is the first step of the third calculation: the step in which idempotency of set-intersection is applied to replace ‘‘ $\cap S$ ’’ by ‘‘ $\cap S \cap S$ ’’. Effectively, instead of (3.2), the equivalent property

$$R \circ (S \rightarrow T^*) \cap S \subseteq R \circ T^* \cap S \Leftarrow R^U \circ S \subseteq S \tag{4.1}$$

has been applied. In fact, (4.1) can be further strengthened by replacing the inclusion on the consequent by an equality since, for all R, S and U ,

$$\begin{aligned} & R \circ U \cap S \subseteq R \circ (S \rightarrow U) \cap S \\ \Leftarrow & \quad \{ \text{monotonicity of composition and intersection} \} \\ & U \subseteq S \rightarrow U \\ = & \quad \{ \text{Heyting Galois connection} \} \\ & U \cap S \subseteq U \\ = & \quad \{ \text{property of intersection} \} \\ & \text{true} . \end{aligned}$$

Thus, by antisymmetry of the subset ordering together with (4.1),

$$R \circ (S \rightarrow T^*) \cap S = R \circ T^* \cap S \Leftarrow R^U \circ S \subseteq S . \tag{4.2}$$

Although the stronger property (4.2) is not used directly, its derivation provides a useful safety check: because we have derived an equality, we know that simplifying the expression ‘‘ $R \circ (S \rightarrow T^*) \cap S$ ’’ to ‘‘ $R \circ T^* \cap S$ ’’ does not incur any loss of information (so long as the condition $R^U \circ S \subseteq S$ is satisfied).

5 General comments

So much for the details of the calculation; now more general comments.

Since the earliest days of the development of “correct-by-construction” program design techniques, goal-directed reasoning has always been a central theme of “program calculation”. For example, “programming as a goal-oriented activity” was a specific topic in Gries’s textbook “The Science of Programming” (Gries, 1981, Chapter 14) and broadening the theme to mathematical proofs in general was the topic of Van Gasteren’s thesis (Van Gasteren, 1990). Goal-directed reasoning is also evident in many of Dijkstra’s “EWD”s (available from the University of Texas) and many other publications of the last 50 years.

In contrast, the standard mathematical style is “bottom-up”. That is evident from the fact that mathematicians almost always use *only-if* arguments (implication) as opposed to *if* arguments (follows-from). In our view, it is extremely important that the more challenging calculations are presented in a goal-directed way, as we have tried to do above. It is important because it helps to teach the creative process underlying the mathematics of program construction. Of course, when a new theory is being developed, the work often proceeds in a bottom-up fashion: one identifies the more straightforward properties and builds up to properties that are not so obvious. But each step in the process is an exploration. One seeks properties of a certain type (e.g., distributivity properties), but the exact form of the properties is not known at the outset. It is vital that we develop a style of calculation that exposes the creative process and that we communicate this process to our students.

Many calculations are, of course, straightforward and do not merit much discussion. Less interesting calculations are ones where each step *simplifies* the expression under consideration (in some sense of the word “simplify”). In contrast, the calculation above involves several *complication* steps. In particular, the step we have singled out as the hardest of all is a complication step: idempotency is used in the derivation of (3.2) to replace an expression of the form $X \cap S$ by $X \cap S \cap S$. Idempotency is normally presented as a simplification rule, whereby the number of occurrences of the operator in question is reduced. In order to foster creative calculation, it is also vital to avoid an undue bias in the presentation of equational properties; equality is after all a symmetric operator.

In summary, what we have presented is, in our view, a good example of a non-trivial calculation that deserves careful study. We hope that, in future, more effort is spent in research publications and textbooks on elucidating the process of creative calculation. Historically, one argument against calculations in the style above is the need to save space. But modern technology—the much reduced reliance on “hard copy”—makes this argument much less relevant.

Acknowledgements

The authors thank the anonymous referees for their detailed and helpful comments.

Competing interests

The authors report no conflict of interest.

References

- Backhouse, R., Doornbos, H., Glück, R. & van der Woude, J. (2022) Components and acyclicity of graphs. An exercise in combining precision with concision. *J. Logical Algebraic Methods Program.* **124**, 100730.
- Gries, D. (1981) *The Science of Programming*. Springer-Verlag.
- Van Gasteren, A. J. M. (1990) *On the Shape of Mathematical Arguments*. Lecture Notes in Computer Science, vol. 445. Springer-Verlag.