An audited elementary algebra

C. J. SANGWIN

1. *Introduction*

Solving equations is one of the core reasons for elementary algebraic manipulation. As Chrystal wrote [1]:

> There are few parts of algebra more important than the logic of the derivation of equations, and few, unhappily, that are treated in more slovenly fashion in elementary teaching. No blind adherence to set rules will avail in this matter; while a little attention to a few simple principles will readily remove all difficulty.

The purpose of this paper is to show how many of the classic *problems*, *fallacies* or examples of *false reasoning* can be avoided by using an *audited elementary algebra*. The word *audited* refers to the explicit tracking of domains throughout the calculation. Auditing (i) eliminates many of the problems of spurious or missing solutions when solving equations, (ii) reveals where in a chain of reasoning such problems occur, and (iii) is a natural extension of algebra which does not introduce any artificial devices. Indeed, all the machinery of auditing, e.g. dealing with Boolean statements or systems of inequalities, occurs already in existing curricula. While auditing does not solve all problems, neither does it entail anything which needs to be forgotten or 'un-learned' later.

'Reasoning by equivalence' is a particularly important activity in elementary algebra. It is an iterative formal symbolic procedure where algebraic expressions, or terms within an expression, are replaced by an equivalent until a 'solved' form is reached. The point is that replacing an expression or a sub-expression in a problem by an equivalent expression provides a new problem having the same solutions. Judging the equivalence of individual expressions, e.g. $x^2 - y^2$ and $(x - y)(x + y)$, is different from judging the validly of moving from one equation to the next. Asserting that two expressions A and B are equivalent means that evaluating for all values of all the variables in some domain (often implicit) we have $A = B$. Asserting that two *equations* $A = 0$ and $B = 0$ are equivalent is subtly but significantly different. It means that the solutions of $A = 0$ are precisely the solutions of $B = 0$, i.e. those *particular values* of the variables coincide. Operating on equations, *equational reasoning*, includes substituting equivalent expressions within part of an equation. However, other forms of reasoning are also used such as operating on both sides of an equation or splitting a single equation into cases: $AB = 0 \Leftrightarrow A = 0$ or $B = 0$. In this paper $\Leftrightarrow$ should be read as 'is equivalent to', just as $=$ is read as 'is equal to'. 'Solving an equation' often means transforming it into a conventional form to make the solution 'clear'. This is usually taken to be *explicit*, e.g. in

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

solving a linear equation in x we would expect the solution to be written as $x = \ldots$. However, manipulating an equation to an equivalent standard form is often sufficient. For example, the form $(x-a)^2 + (y-b)^2 = r^2$ represents a circle, and in this form we may readily identify the centre and radius. One key purpose of algebraic form is to enable someone to recognise/interpret the equation.

2. *Audited expressions*

An audited elementary expression contains information about the domain for each of the variables. For example

$$x^2 = 4$$

is a traditional (un-audited) expression, whereas

$$x^2 = 4 \wedge x > 0$$

is audited. Variables restricted by inequalities are assumed to be real. Therefore, this is equivalent to

$$x^2 = 4 \wedge x > 0 \wedge x \in \mathbb{R}.$$

The explicit purpose of auditing is to combine traditional algebraic expressions with propositional logic. An audited expression is a *single expression* represented by a tree structure. The Boolean logical connectives 'and' $\wedge$ and 'or' $\vee$ are commutative associative binary operators (sometimes called 'n-ary operators'), just like $\times$ and $+$. The whole point of carrying around the auditing attached to the equational parts of the expression is that it cannot get lost. This is particularly important in a computer algebra system, or communicating an equation from one person to another.

This is subtly different from assigning each variable a *type* as would be the case in computer code. `x:pos real` is a perfectly respectable type, but, as we shall see in some examples, the work in sorting out the auditing is equal in computational difficulty and status to the work on the original expression or equation.

2.1 *Natural domain of definition*

The natural domain of an expression A is the maximal set of points at which A can be evaluated. A given problem context usually provides an underlying field X such as $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Since division by zero is always forbidden this may be a subset of X. For example, the natural domain of $\frac{1}{x}$ is $X \setminus \{0\}$.

The natural domain of a rational expression $\frac{A}{B}$ is the set for which $B \neq 0$.

Rational expressions sometimes have common factors in the numerator and denominator, e.g.

$$\frac{x^2 - 4}{x - 2}.$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

This expression has a natural domain of definition $x \neq 2$ so that the proposal we discuss here is that the audited expression should be written

$$\frac{x^2 - 4}{x - 2} \wedge x \neq 2.$$

Traditionally we could choose to factorise and cancel, and so write

$$\frac{x^2 - 4}{x - 2} = \frac{(x - 2)(x + 2)}{x - 2} = x + 2.$$

As elements of an algebraic field, e.g. $\mathbb{Q}[x]$, both $\frac{x^2-4}{x-2}$ and $x + 2$ are in the same equivalence class, and in this context are often considered to be the same. A rational expression in lowest terms, with suitable conventions for writing polynomials in the numerator and denominator, gives a canonical representative for the equivalence class. In analysis, when we are using an algebraic expression to represent a function, or when solving an equation, the domain is important. The audited form contains a 'memory' of any restriction which often needs to remain with the expression. That is, working at a higher level of detail,

$$\begin{aligned}\frac{x^2 - 4}{x - 2} \wedge x \neq 2 &\Leftrightarrow \frac{(x - 2)(x + 2)}{x - 2} \wedge x \neq 2\\ &\Leftrightarrow \frac{x - 2}{x - 2}(x + 2) \wedge x \neq 2\\ &\Leftrightarrow x + 2 \wedge x \neq 2.\end{aligned}$$

Note that

$$x + 2 \wedge x \neq 2$$

is a different audited expression from $x + 2 \wedge x \in \mathbb{R}$.

Next, consider solving

$$\frac{1}{x + 1} + \frac{1}{x - 1} = \frac{2}{x^2 - 1}.$$

Rationalising both sides we have

$$\frac{2x}{x^2 - 1} = \frac{2}{x^2 - 1}$$

from which it is tempting to conclude $2x = 2$ and so $x = 1$. Since $x \neq \pm 1$ in the domain of the original expression we have no real solutions. This is just the kind of hidden loss of domain information we seek to avoid.

In a particular problem context we may choose to permit a domain enlargement back to the whole real line, or complex plane. This *choice* should be conscious and explicit in the written working, not an accident of algebraic cancelling. In many problems a point which has been excluded will be precisely the interesting one.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

The variables which occur in a formula are implicitly recorded in tracking the domain. For example, when auditing the expression

$$\frac{a^2}{(a-b)(a-c)} + \frac{b^2}{(b-c)(b-a)} + \frac{c^2}{(c-a)(c-b)}$$

we have $a \neq b \wedge b \neq c \wedge c \neq a$. The equivalent expression 1 does not contain any of these variables, but these domain restrictions will remain in the equivalent form, i.e.

$$1 \wedge a \neq b \wedge b \neq c \wedge c \neq a.$$

Situations such as this are prime candidates for domain enlargements by dropping conditions such as $a \neq b$ for variables which no longer occur in the expression.

Example 2.1 (taken from [2, p. 81]):

By ordinary long division we have, for any value of x,

$$\begin{aligned}
\frac{x-1}{x-1} &= 1 \\
\frac{x^2-1}{x-1} &= x+1 \\
\frac{x^3-1}{x-1} &= x^2+x+1 \\
\frac{x^4-1}{x-1} &= x^3+x^2+x+1 \\
\dots &= \dots \\
\frac{x^n-1}{x-1} &= x^{n-1}+\dots+x^2+x+1.
\end{aligned}$$

Now in all of these identities we let x have the value 1. The right-hand sides then assume the values 1, 2, 3, 4, …, n. The left-hand sides are the same. Hence $1 = 2 = 3 = 4 = \dots = n$.

The above has a silent domain enlargement to the whole of $\mathbb{R}$. In the audited version, we cannot evaluate the right-hand side, because $x = 1$ is excluded from the domain. Preventing evaluation of an expression outside its domain of definition avoids the fallacy.

In many problems we will be working over the real numbers, in which case we may restrict the domain of an expression to ensure we remain real. For example, the natural real domain of $x^{1/2}$ is $x \geqslant 0$.

Algebraic roots are particularly troublesome. If working in the real domain $\sqrt{x}$ is traditionally taken to be the positive square root of the positive

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

number x. Therefore any expression with surd quantities should be audited to track the domain, i.e. rather than write $\sqrt{x}$ we should always used the audited form

$$\sqrt{x} \wedge x \geqslant 0$$

$$\sqrt{x} \wedge x \in \mathbb{R} \Leftrightarrow \sqrt{x} \wedge x \geqslant 0.$$

In this paper $\sqrt{x}$ and $x^{1/2}$ are considered to be different notation for the same thing. This is not always the case. Not all authors do this. For example, in [3] we read:

> Having two symbols to indicate the n th root of a, namely $\sqrt[n]{a}$ and $a^{1/n}$, we shall employ the first in the simple arithmetical sense, and the second to denote any one of the algebraical roots, that is, any one we please, unless some particular root be specified. Thus $\sqrt{4}$ is 2, without reference to sign; but $4^{1/2}$ may be either +2 or −2.

If n is even then $x^{1/n}$ also has natural domain $x \geqslant 0$. If n is odd then $x^{1/n}$ is defined on the whole real line. Often we don't know if n is odd or even, which restricts our ability to audit domains or apply particular rewrite rules which rely on a constraint such as $x \geqslant 0$ without error. In such cases we audit as needed.

Points which evaluate to 0^0 are also excluded from the natural domain.

There are many notational ways to express a set of numbers. For example, the positive real numbers can be written as any of the following.

$$x > 0, \qquad \mathbb{R}_{>0}, \qquad (0, \infty), \qquad \{x \mid x > 0\}.$$

When auditing an expression the author should chose notation best suited to their particular circumstance. Sometimes this will be a system of inequalities which is amenable to calculation, and sometimes it will be using sets written with interval notation.

3. *Solving equations*

An essential process in solving equations is operating on both sides, i.e. 'doing the same thing to both sides'. When doing this we must ensure that given expressions A and B and the operator f we have $A = B \Leftrightarrow f(A) = f(B)$. This is the definition of an injective function. It is not necessary that f is surjective, just that we restrict the domains. Thus

$$A = B \text{ on } X \Leftrightarrow f(A) = f(B) \text{ on } f(X).$$

Given a number a, the function $x \mapsto a + x$ is injective. Hence, we can add the same thing to both sides of an equation.

Multiplication is *not injective* in general. In particular multiplication by zero is not injective. Normally we encounter this problem when dividing both sides of an equation by an expression. However, when asserting equivalence multiplication and division are taken together. Hence *audited* forms of the rules are needed.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

$$CA = CB \Leftrightarrow A = B \vee C = 0. \quad (1)$$

$$CA = CB \wedge C \neq 0 \Leftrightarrow A = B \wedge C \neq 0. \quad (2)$$

$$A = B \Leftrightarrow \big(CA = CB \wedge C \neq 0\big) \vee A = B = 0. \quad (3)$$

The first form is familiar. In (2), the condition $C \neq 0$ appears odd on the right hand side as there is no C elsewhere in the formula $A = B$. The last is almost never seen explicitly in elementary algebraic derivations. The omission of the auditing in the last of these hides the potential introduction of spurious solutions in the transformation of $A = B$ into the new equation $CA = CB$.

Note that $\Leftrightarrow$ is a symmetric, transitive relation. Since we read (English) left to right, these rules are written to emphasise the direction of use as a transformation (re-write rule), such as

$$CA = CB \wedge \neq 0 \rightarrow A = B \wedge C \neq 0.$$

Much of elementary algebra consists of *re-write* rules. E.g. $0 \times x \rightarrow 0$. Here the expression $0 \times x$ is rewritten to be 0. Rewrite rules can be used in both directions. However, they are most often used in the direction left to right. We match the pattern on the left and rewrite with the pattern on the right. This is symbol pushing.

It might be argued that *traditional practice* does not mix the logical connectives for 'and' $\wedge$ and 'or' $\vee$ with algebraic equations in this way. This is simply not true.

$$AB = 0 \Leftrightarrow A = 0 \vee B = 0. \quad (4)$$

$$AB > 0 \Leftrightarrow \big(A > 0 \wedge B > 0\big) \vee \big(A < 0 \wedge B < 0\big). \quad (5)$$

The whole point is that many operations on equations which are taken to be an equivalence, are in fact not equivalent. Multiplying or dividing by variables, in particular cancelling terms, and taking powers of both sides of an equation need to be qualified. Auditing makes this qualification explicit.

Example 3.1: An incorrect solution takes the square root of both sides

$$\begin{aligned} (2x - 7)^2 &= (x + 1)^2 \\ 2x - 7 &= x + 1 \\ x &= 8 \end{aligned}$$

In this case we may derive the correct solution using the difference of two squares.

$$\begin{aligned} (2x - 7)^2 &= (x + 1)^2 \\ \Leftrightarrow (2x - 7)^2 - (x + 1)^2 &= 0 \\ \Leftrightarrow (2x - 7 + x + 1)(2x - 7 - x - 1) &= 0 \\ \Leftrightarrow (3x - 6)(x - 8) &= 0 \\ \Leftrightarrow x = 2 \vee x &= 8. \end{aligned}$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Notice how using the difference of two square avoids taking roots of both sides of an equation.

Using the difference of two squares avoids the dilemma of 'when to use ±'. For example, when solving $x^2 = a^2$ we would like to take the square root of both sides. However, do we have ± on the left, right or both sides of this equation? Writing

$$\begin{aligned} x^2 = a^2 \Leftrightarrow x^2 - a^2 = 0 &\Leftrightarrow (x - a)(x + a) = 0 \\ &\Leftrightarrow x - a = 0 \vee x + a = 0 \\ &\Leftrightarrow x = a \vee x = -a \end{aligned}$$

entirely avoids this problem.

Auditing can be quietly dropped when dealing with pure numbers. For example, if in (3) we know $C = 2$ then $C \neq 0$ is true. We will *never* audit as follows

$$A = B \Leftrightarrow 2A = 2B \wedge 2 \neq 0.$$

This would be wasteful: the condition $2 \neq 0$ immediately evaluates to *TRUE* and so can be eliminated. The logical value *TRUE* is the identity element for $\wedge$, i.e. $A \wedge TRUE \Leftrightarrow A$. Wastefulness in algebra is to be avoided. After all, a primary purpose of algebraic symbolism is abbreviation and efficient computation. Auditing is only needed for variable quantities.

In the following examples auditing which is traditionally omitted is shown in **bold**.

Example 3.2: Assume

$$\begin{aligned} & a = b \\ \Leftrightarrow\ & (a^2 = ab \wedge \boldsymbol{a \neq 0}) \vee \boldsymbol{a = b = 0} \\ & \qquad \text{using (3) with } A = C = a \text{ and } B = b \\ \Leftrightarrow\ & (a^2 - b^2 = ab - b^2 \wedge \boldsymbol{a \neq 0}) \vee \boldsymbol{a = b = 0} \\ \Leftrightarrow\ & ((a - b)(a + b) = b(a - b) \wedge \boldsymbol{a \neq 0}) \vee \boldsymbol{a = b = 0} \\ \Leftrightarrow\ & [(a + b = b \wedge \boldsymbol{a \neq 0}) \vee (\boldsymbol{a - b = 0} \wedge \boldsymbol{a \neq 0})] \vee \boldsymbol{a = b = 0} \\ \Leftrightarrow\ & [(2b = b \wedge \boldsymbol{a \neq 0}) \vee \boldsymbol{a = b} \wedge \boldsymbol{a \neq 0})] \vee \boldsymbol{a = b = 0} \\ \Leftrightarrow\ & [(2 = 1 \wedge \boldsymbol{a \neq 0} \wedge \boldsymbol{b \neq 0}) \vee \boldsymbol{a = b} \wedge \boldsymbol{a \neq 0})] \vee \boldsymbol{a = b = 0}. \end{aligned}$$

The unaudited version appears to derive the contradiction $1 = 2$. The audited version concludes

$$[(2 = 1 \wedge a \neq 0 \wedge b \neq 0) \vee (a = b \wedge a \neq 0)] \vee a = b = 0.$$

The statement $2 = 1$ can be replaced by its Boolean evaluation *FALSE*, the

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

rule $FALSE \wedge A \to FALSE$ is used twice together with the rule $FALSE \vee A \to A$ to give

$$\begin{aligned}(a = b \wedge a \neq 0) \vee a = b = 0 &\Leftrightarrow (a = b \wedge a \neq 0) \vee (a = b \wedge a = 0)\\ &\Leftrightarrow a = b \wedge (a \neq 0 \vee a = 0)\\ &\Leftrightarrow a = b.\end{aligned}$$

The audited version shows this as nothing more than rewriting and obfuscating the hypothesis.

In future examples, for brevity, we may choose to show auditing only where it changes from line to line, just as in writing elementary algebra parts are not repeated where this causes clutter on the page. These are aesthetic judgements.

3.1 *Squares and square roots of equations*

To take square roots of both sides of an equation we should formalise the difference of two squares:

$$A^2 = B^2 \Leftrightarrow A^2 - B^2 = 0 \Leftrightarrow (A - B)(A + B) = 0 \Leftrightarrow A = B \vee A = -B.$$

This can conveniently be remembered as a single rewrite rule

$$A^2 = B^2 \Leftrightarrow A = B \vee A = -B. \tag{6}$$

Example 3.3: Returning to Example 3.1, an alternative audited solution could be

$$\begin{aligned}&(2x - 7)^2 = (x + 1)^2\\ \Leftrightarrow\ &2x - 7 = x + 1 \vee \mathbf{2}\boldsymbol{x} - \mathbf{7} = -\boldsymbol{x} - \mathbf{1}\\ \Leftrightarrow\ &x = 8 \vee \boldsymbol{x} = \mathbf{2}.\end{aligned}$$

Squaring both sides of a real-valued equation can also be useful, but needs to be audited:

$$\text{If } A = B \text{ then } A^2 = B^2.$$

However, when solving an equation we need an equivalence, not a simple implication. Therefore the following audited version is necessary.

$$A = B \wedge A \in \mathbb{R} \wedge B \in \mathbb{R} \Leftrightarrow A^2 = B^2 \wedge \big((A \geqslant 0 \wedge B \geqslant 0) \vee (A < 0 \wedge B < 0)\big). \tag{7}$$

This rule looks complex, but it will prove to be particularly useful for removing square roots from equations.

In particular, we shall make use of the rules

$$\left(\sqrt{A}\right)^2 \wedge A \geqslant 0 \Leftrightarrow A \wedge A \geqslant 0. \tag{8}$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Note, that the domain condition $A \geqslant 0$ on the left-hand side of this is implied by the natural domain of the square root function. Now consider $\sqrt{A^2} = |A|$. This does not have the domain restriction $A \geqslant 0$, but it does not equal A. Notice that $(\sqrt{A})^2 \neq \sqrt{A^2}$. Written in index form $(A^2)^{1/2} \neq (A^{1/2})^2$, without some auditing. If $A \geqslant 0$ then equality holds, i.e.

$$\text{if } A \geqslant 0 \text{ then } (A^2)^{1/2} = (A^{1/2})^2 = A.$$

This is an important source of false reasoning.

The expressions which form the auditing are themselves mathematical expressions which can, and should, be manipulated to equivalent forms. This is done in the above argument by rewriting $a - b = 0$ as $a = b$. Where constraints evaluate to logical false (*FALSE*) (e.g. as in $2 = 1$) they should be eliminated. The following example is similar, but this time uses (6).

Example 3.4:

$$\begin{aligned}
&(n+1)^2 = n^2 + 2n + 1\\
\Leftrightarrow\ &(n+1)^2 - (2n+1) = n^2\\
\Leftrightarrow\ &(n+1)^2 - (2n+1) - n(2n+1) = n^2 - n(2n+1)\\
\Leftrightarrow\ &(n+1)^2 - (n+1)(2n+1) = n^2 - n(2n+1)\\
\Leftrightarrow\ &(n+1)^2 - 2(n+1)\left(\frac{2n+1}{2}\right) = n^2 - 2n\left(\frac{2n+1}{2}\right)\\
\Leftrightarrow\ &(n+1)^2 - 2(n+1)\left(\frac{2n+1}{2}\right) + \left(\frac{2n+1}{2}\right)^2 = n^2 - 2n\left(\frac{2n+1}{2}\right) + \left(\frac{2n+1}{2}\right)^2\\
\Leftrightarrow\ &\left((n+1) - \frac{2n+1}{2}\right)^2 = \left(n - \frac{2n+1}{2}\right)^2\\
\Leftrightarrow\ &(n+1) - \frac{2n+1}{2} = n - \frac{2n+1}{2}\\
&\qquad \vee\ \boldsymbol{(n+1) - \frac{2n+1}{2} = -n + \frac{2n+1}{2}}\\
\Leftrightarrow\ &n+1 = n\\
&\qquad \vee\ \boldsymbol{2n+2-2n+1 = -2n+2n+1}\\
\Leftrightarrow\ &1 = 0 \vee \mathbf{1 = 1.}
\end{aligned}$$

What appears to be the contradiction $1 = 0$ is in fact the tautology $1 = 1$, which evaluates to *TRUE* and is equivalent to the identity $(n+1)^2 = n^2 + 2n + 1$. In other words, deriving a tautology proves the identity.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

The following examples are taken from [4].

Example 3.5: Solve $\sqrt{3x+4} = 2 + \sqrt{x+2}, x \in \mathbb{R}$.

The unaudited version is as follows:

$$\begin{aligned}
\sqrt{3x+4} &= 2 + \sqrt{x+2} \\
3x + 4 &= 4 + 4\sqrt{x+2} + (x+2) \\
x - 1 &= 2\sqrt{x+2} \\
x^2 - 2x + 1 &= 4x + 8 \\
x^2 - 6x - 7 &= 0 \\
(x-7)(x+1) &= 0.
\end{aligned}$$

Note, the only real solution is $x = 7$.

First we audit with natural domains from the original expression

$$\sqrt{3x+4} = 2 + \sqrt{x+2} \wedge \mathbf{3x + 4 \geqslant} \wedge \boldsymbol{x} + \mathbf{2 \geqslant 0.}$$

Rewriting these domain conditions we have

$$\sqrt{3x+4} = 2 + \sqrt{x+2} \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \mathbf{-2} \leqslant \boldsymbol{x.}$$

Eliminating the redundant condition

$$\sqrt{3x+4} = 2 + \sqrt{x+2} \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x}.$$

Now we use (7)

$$\left(\sqrt{3x+4}\right)^2 = \left(2 + \sqrt{x+2}\right)^2$$

$$\wedge\left(\sqrt{\mathbf{3x+4}} \geqslant \mathbf{0} \wedge \mathbf{2} + \sqrt{\boldsymbol{x}+\mathbf{2}} \geqslant \mathbf{0}\right) \vee \left(\sqrt{\mathbf{3x+4}} < \mathbf{0} \wedge \mathbf{2} + \sqrt{\boldsymbol{x}+\mathbf{2}} < \mathbf{0}\right) \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x.}$$

Notice, the complex conditions from (7) can be evaluated to be *TRUE* or *FALSE* immediately, so we have the equivalent expression

$$\left(\sqrt{3x+4}\right)^2 = \left(2 + \sqrt{x+2}\right)^2 \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x}.$$

Now we are in a position to use (8) on both sides, together with other algebraic manipulation.

$$\begin{aligned}
&3x + 4 = 4 + 4\sqrt{x+2} + \left(\sqrt{x+2}\right)^2 \quad \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x}. \\
\Leftrightarrow\ &3x + 4 = 4 + 4\sqrt{x+2} + x + 2 \quad \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x.} \\
\Leftrightarrow\ &x - 1 = 2\sqrt{x+2} \quad \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x.}
\end{aligned}$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Now we use (7) again

$$(x-1)^2 = \left(2\sqrt{x+2}\right)^2$$
$$\wedge\left(\boldsymbol{x-1\geqslant 0\wedge 2\sqrt{x+2}\geqslant 0}\right)\vee\left(\boldsymbol{x-1<0\wedge 2\sqrt{x+2}<0}\right)\wedge -\tfrac{4}{3}\leqslant \boldsymbol{x}.$$

We reduce the inequalities, then proceed as normal. Note the second to third line uses (8) on the right-hand side.

$$\begin{aligned}
(x-1)^2 &= \left(2\sqrt{x+2}\right)^2 \wedge \boldsymbol{x\geqslant 1}\wedge -\tfrac{4}{3}\leqslant \boldsymbol{x}\\
\Leftrightarrow\ x^2-2x+1 &= \left(2\sqrt{x+2}\right)^2 \wedge \boldsymbol{x\geqslant 1}\\
\Leftrightarrow\ x^2-2x+1 &= 4(x+2)\wedge \boldsymbol{\sqrt{x+2}\geqslant 0}\wedge \boldsymbol{x\geqslant 1}\\
\Leftrightarrow\ x^2-6x-7 &= 0\wedge \boldsymbol{x\geqslant 1}\\
\Leftrightarrow\ (x-7)(x+1) &= 0\wedge \boldsymbol{x\geqslant 1}\\
\Leftrightarrow\ (x=7\vee x=-1) &\qquad \wedge \boldsymbol{x\geqslant 1}\\
\Leftrightarrow\ x &= 7.
\end{aligned}$$

Notice how in the audited version we have the natural domain constraint which eliminates the spurious solution giving a complete and correct answer $x = 7$.

In this example the complex auditing of domains in (7) has been spelled out. In future we will not write out this in full, but will immediately simplify this in the way we have with some of the other (traditional) algebra.

Example 3.6: Without auditing we have the following:

$$\begin{aligned}
\sqrt{3x+4} &= 2-\sqrt{x+2}\\
3x+4 &= 4-4\sqrt{x+2}+(x+2)\\
x-1 &= -2\sqrt{x+2}\\
x^2-2x+1 &= 4x+8\\
x^2-6x-7 &= 0\\
(x-7)(x+1) &= 0.
\end{aligned}$$

The only real solution is $x = -1$.

Auditing with natural domains

$$\sqrt{3x+4} = 2-\sqrt{x+2}\wedge -\tfrac{4}{3}\leqslant \boldsymbol{x}.$$

Now we use (7)

$$\left(\sqrt{3x+4}\right)^2 = \left(2-\sqrt{x+2}\right)^2\wedge \boldsymbol{2\geqslant \sqrt{x+2}}\wedge -\tfrac{4}{3}\leqslant \boldsymbol{x}.$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Note, we need (7) on the domain constraint itself:

$$\begin{aligned} 2 \geqslant \sqrt{x+2} &\Leftrightarrow 4 \geqslant x+2 \\ \wedge\big((2 \geqslant 0 \wedge \sqrt{x+2} \geqslant 0) \vee (2 < 0 \wedge \sqrt{x+2} < 0)\big) &\Leftrightarrow 4 \geqslant x+2 \\ &\Leftrightarrow x \leqslant 2. \end{aligned}$$

Incorporating this side calculation, we have

$$\begin{aligned} 3x+4 = 4-4\sqrt{x+2}+x+2 \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow x-1 = -2\sqrt{x+2} \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow x^2-2x+1 = 4x+8 \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow x^2-6x+7 = 0 \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow (x-7)(x+1) = 0 \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow \big(x=7 \vee x=-1\big) \quad \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \wedge \boldsymbol{x} \leqslant \boldsymbol{2} \\ \Leftrightarrow x=-1. \end{aligned}$$

Hence $x = -1$ is the only solution.

Evaluating the constraints on the domain might be as difficult as in the original problem, simply pushing the work elsewhere. Including the auditing makes this clear, and gives these calculations equal status.

Example 3.7: This equation has no real solutions. Without auditing we appear to get two.

$$\begin{aligned} \sqrt{x+2} &= 2+\sqrt{3x+4} \\ x+2 &= 4+4\sqrt{3x+4}+(3x+4) \\ -x-3 &= 2\sqrt{3x+4} \\ x^2+6x+9 &= 4(3x+4) = 12x+16 \\ x^2-6x-7 &= 0 \\ (x-7)(x+1) &= 0 \end{aligned}$$

The audited version gives

$$\begin{aligned} \sqrt{x+2} &= 2+\sqrt{3x+4} \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \\ \Leftrightarrow x+2 &= 4+4\sqrt{3x+4}+(3x+4) \wedge -\tfrac{4}{3} \leqslant \boldsymbol{x} \end{aligned}$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

$$\Leftrightarrow -x - 3 = 2\sqrt{3x + 4} \wedge -\tfrac{4}{3} \leqslant x.$$

Trying to apply (7) we have the constraints

$$-x - 3 \geqslant 0 \wedge 2\sqrt{3x + 4} \geqslant 0 \wedge -\tfrac{4}{3} \leqslant x.$$

$$-3 \geqslant x \wedge -\tfrac{4}{3} \leqslant x.$$

This is the empty set, so that the original equation has no real solution.

A source of similar examples is [5].

3.2 *n* th *powers and roots of equations*

We should, as far as possible, avoid taking the *n* th roots of both sides of an equation. Guided by the difference of two squares, it is much more direct to subtract, factorise and split into cases. For example

$$\begin{aligned} & A^3 = B^3 \\ \Leftrightarrow\ & 0 = A^3 - B^3 \\ \Leftrightarrow\ & 0 = (A - B)\left(A^2 + AB + B^2\right) \\ \Leftrightarrow\ & A = B \vee A^2 + AB + B^2 = 0. \end{aligned}$$

Depending on the underlying domain, $A^2 + AB + B^2$ may or may not factorise further. In particular if $\omega \neq 1$ is a primitive cube root of unity we have

$$A^2 + AB + B^2 = (A - \omega B)(A - \omega^2 B)$$

so we might, over the complex numbers, write this more elegantly as

$$A^3 = B^3 \Leftrightarrow \left(A - \omega B\right)\left(A - \omega^2 B\right)\left(A - \omega^3 B\right) = 0.$$

This generalises to *n* th roots.

For example, applying this reasoning to an irreducible quadratic

$$\begin{aligned} & x^2 = -4 \\ \Leftrightarrow\ & 0 = x^2 + 4 \\ \Leftrightarrow\ & 0 = (x - 2i)(x + 2i) \\ \Leftrightarrow\ & x = 2i \vee x = -2i. \end{aligned}$$

Working in this way we retain all the roots and none are 'lost'. Other authors have comments on this issue, e.g.

> It is by no means uncommon with algebraical authors, when they have led their readers through a process which terminates in an equation, to select that root which gives the answer they require, without explaining the signification of the other roots that are equally comprised in it; and this incomplete mode of solution, which is censurable from revealing only a part of the

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

> truth, had in some instances caused the most interesting circumstances attending a question to be entirely overlooked. [6].

Examples of such discussions include [6] and [7, Chapter XII].

4. *Further examples*

Inequalities can also be solved in the same style, using identical mathematical techniques we propose for auditing, e.g. using (5).

4.1 *Solving inequalities*

$$
\begin{aligned}
& 2x^2 + x \geqslant 6 \\
\Leftrightarrow\ & 2x^2 + x - 6 \geqslant 0 \\
\Leftrightarrow\ & (2x - 3)(x + 2) \geqslant 0 \\
\Leftrightarrow\ & \big((2x - 3) \geqslant 0 \wedge (x + 2) \geqslant 0\big) \vee \big((2x - 3) \leqslant 0 \wedge (x + 2) \leqslant 0\big) \\
\Leftrightarrow\ & \big(x \geqslant \tfrac{3}{2} \wedge x \geqslant -2\big) \vee \big(x \leqslant \tfrac{3}{2} \wedge x \leqslant -2\big) \\
\Leftrightarrow\ & x \geqslant \tfrac{3}{2} \vee x \leqslant -2.
\end{aligned}
$$

This differs considerably from a traditional rhetorical argument such as:

> To solve $2x^2 + x \geqslant 6$ we first solve $2x^2 + x = 6$, or $2x^2 + x - 6 = (2x - 3)(x + 2) = 0$. This has solutions $x = \frac{3}{2}$ and $x = -2$, hence $2x^2 + x = 6$ at these, and only these, values. Since $f(x) = 2x^2 + x - 6$ is continuous and changes sign only at $x = \frac{3}{2}$ and $x = -2$ by the intermediate value theorem we need only consider the algebraic sign of $f(x)$ at individual points x in three intervals. Since $f(0) \leqslant 0$ and $f(-3) \geqslant 0$ and $f(2) \geqslant 0$ we conclude that $2x^2 + x \geqslant 6$ for $x \in (-\infty, -2] \cup [\frac{3}{2}, \infty)$.

The more formal symbolic style reduces the rhetoric, replacing loosely connected sentences with precise logical symbols. This is an extension of the process of algebraic symbolism, started in the 1550s, to include more logical reasoning within algebraic calculation.

Another example is from the International Baccalaureate specimen questions collection:

Find all values of x that satisfy the inequality $\dfrac{2x}{|x - 1|} < 1$.

The mark scheme assumes students will sketch graphs and draw inferences from the graphical solution: a perfectly reasonable way to proceed. However, purely algebraic reasoning by equivalence is possible.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

$$\frac{2x}{|x-1|} < 1 \wedge \boldsymbol{x \neq 1}$$

$$\Leftrightarrow 2x < |x-1| \wedge \boldsymbol{x \neq 1}$$

$$\Leftrightarrow (x \geqslant 1 \wedge 2x < x-1) \vee (x < 1 \wedge 2x < -x+1) \wedge \boldsymbol{x \neq 1}$$

$$\Leftrightarrow (x \geqslant 1 \wedge x < -1) \vee (x < 1 \wedge 3x < 1) \wedge \boldsymbol{x \neq 1}$$

$$\Leftrightarrow x < \tfrac{1}{3} \wedge \boldsymbol{x \neq 1}$$

$$\Leftrightarrow x < \tfrac{1}{3}.$$

This illustrates the potential of auditing to deal with piecewise functions, such as $|x|$, in a natural way.

4.2 *Maxwell's fallacies*

The classic collection [8] defines a *fallacy* as leading *by* guile *to a* wrong *but plausible conclusion.* To what extent does auditing prevent these fallacies? Clearly auditing is primarily an algebraic technique, and so we omit the purely geometric fallacies.

Example 4.1

This comes from [8, p. 37], with only essential auditing shown. Note when squaring both sides the conditions in (7) are redundant, and the square root is only applied to non-negative arguments, removing the need for natural domain constraints.

The fallacy that $4 = 0$.

$$\cos^2(x) = 1 - \sin^2(x)$$

$$\Leftrightarrow 1 + \cos(x) = 1 + \left(1 - \sin^2(x)\right)^{1/2} \vee \mathbf{1 + \cos(x) = 1 - \left(1 - \sin^2(x)\right)^{1/2}}$$

$$\Leftrightarrow (1 + \cos(x))^2 = \left(1 + \left(1 - \sin^2(x)\right)^{1/2}\right)^2$$

$$\vee \mathbf{(1 + \cos(x))^2 = \left(1 - \left(1 - \sin^2(x)\right)^{1/2}\right)^2}$$

Note that in using (7) when moving from the second to third line of the above, the additional auditing immediately simplifies to *TRUE* and hence is omitted.

Evaluating at $x = \pi$

$$\Rightarrow (1-1)^2 = \left(1 + (1-0)^{1/2}\right)^2 \vee \mathbf{(1-1)^2 = \left(1 + (1-0)^{1/2}\right)^2}$$

$$\Leftrightarrow 0 = 4 \vee \mathbf{0 = 0.}$$

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

When using trigonometric identities we also need auditing. Traditionally we write trigonometric identities without reference to the domain

$$\tan(A + B) = \frac{\tan(A) + \tan(B)}{1 - \tan(A)\tan(B)}.$$

Actually this requires auditing, using the set $T := \{\pi(\frac{1}{2} + k) | k \in \mathbb{Z}\}$.

$$\tan(A + B) \wedge A + B \notin T$$
$$= \frac{\tan(A) + \tan(B)}{1 - \tan(A)\tan(B)} \wedge A \notin T \wedge B \notin T \wedge \tan(A)\tan(B) \neq 1. \quad (9)$$

Example 4.2: This comes from [8, p. 42]. The fallacy that $+1 = -1$.

To solve the equation

$$\begin{aligned}
\cot(\theta) + \tan(3\theta) &= 0 \quad \textbf{Let } T := \{\pi(\tfrac{1}{2} + k) | k \in \mathbb{Z}\} \\
\cot(\theta) + \tan(\theta + 2\theta) &= 0 \quad \wedge \theta \neq k\pi \wedge 3\theta \notin T \\
\cot(\theta) + \frac{\tan(\theta) + \tan(2\theta)}{1 - \tan(\theta)\tan(2\theta)} &= 0 \quad \ldots \wedge \theta \notin T \wedge 2\theta \notin T \\
& \quad \wedge \tan(\theta)\tan(2\theta) \neq 1 \\
\cot(\theta) - \tan(2\theta) + \tan(\theta) + \tan(2\theta) &= 0 \\
\cot(\theta) + \tan(\theta) &= 0 \\
\tan^2(\theta) + 1 &= 0 \\
\tan^2(\theta) &= -1 \qquad (10)
\end{aligned}$$

Returning to the original equation $\tan(3\theta) = -\cot(\theta) = \tan(\theta + \frac{\pi}{2})$ so that

$$3\theta = \theta + \tfrac{\pi}{2} + n\pi$$

for $n \in \mathbb{Z}$. Thus $\theta = \frac{\pi}{4} + n\cdot\frac{\pi}{2}$. Taking $n = 0$ we evaluate (10) in the last line of the above derivation at $\theta = \frac{\pi}{4}$ to get Maxwell's fallacy

$$\tan^2\left(\frac{\pi}{4}\right) = 1 = -1.$$

However, looking at the auditing for (10) we have

$$\theta \neq k\pi \wedge \theta \notin T \wedge 2\theta \notin T \wedge 3\theta \wedge \tan(A)\tan(B) \neq 1.$$

The term $2\theta \notin T$ means $2\theta \neq \frac{\pi}{2} + k\pi$ for all $k \in \mathbb{Z}$ so that if $k = 0$ we see that $\theta = \frac{\pi}{4}$ has been excluded from the domain of the equation $\tan^2(\theta) = -1$, hence with the auditing (9) there is no fallacy.

Example 4.3: This comes from [8, p. 44]. The sum of the squares on two sides of a triangle is never less than the square on the third.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Take a triangle ABC with $(a > b)$. Then

$$(a\cos(C) > b\cos(C) \wedge \mathbf{cos(C) > 0}) \vee (\boldsymbol{a\cos(C) < b\cos(C)} \wedge \mathbf{cos(C) < 0}).$$

Now, by projecting points onto the segments a and b we have

$$b\cos(C) = a - c\cos(B), \qquad a\cos(C) = b - c\cos(A)$$

so that

$$\begin{aligned}&(b - c\cos(A) > a - c\cos(B) \wedge \mathbf{cos(C) > 0})\\ &\qquad \vee (\boldsymbol{b - c\cos(A) < a - c\cos(B)} \wedge \mathbf{cos(C) < 0})\\ \Leftrightarrow\ &((c\cos(B) - c\cos(A) > a - b \wedge \mathbf{cos(C) > 0})\\ &\qquad \vee (\boldsymbol{c\cos(B) - c\cos(A) < a - b} \wedge \mathbf{cos(C) < 0}).\end{aligned}$$

Multiply by $2ab \geqslant \mathbf{0}$ and use the cosine formula.

$$\begin{aligned}&(b(a^2 + c^2 - b^2) - a(b^2 + c^2 - a^2) > 2ab(a - b) \wedge \mathbf{cos(C) > 0})\\ &\vee (b(a^2 + c^2 - b^2) - a(b^2 + c^2 - a^2) < 2ab(a - b) \wedge \mathbf{cos(C) < 0}).\end{aligned}$$

Omitting the remaining legitimate steps this reduces to

$$\begin{aligned}&(a^2 + b^2 > c^2 \wedge \mathbf{cos(C) > 0})\\ &\vee (a^2 + b^2 < c^2 \wedge \mathbf{cos(C) < 0}).\end{aligned}$$

The case $\cos(C) = 0$ gives equality $a^2 + b^2 = c^2$.

5. *Conclusion*

Many of the classic fallacies in elementary algebra can be avoided by using *auditing*, i.e. explicitly tracking the domains throughout the calculation. This is particularly useful for real algebra problems.

Over the last four hundred years, algebraic symbolism has developed [9]. Purposes include abbreviation, e.g. Robert Recorde justified his invention of the equals sign as follows.

> ... to avoid the tedious repetition of these words: 'is equal to', I will set (as I do often in work use) a pair of parallels of one length (thus =), because no two things can be more equal. [10]

Another purpose is to aid efficient computation. For example, Newton explains his extension of meaning of x^n to fractional and negative indices.

> Since algebraists write a^2, a^3, a^4, etc., for aa, aaa, $aaaa$, etc., so I write $a^{1/2}$, $a^{3/2}$, $a^{5/2}$, for $\sqrt{a}$, $\sqrt{a^3}$, $\sqrt{a^5}$; and I write a^{-1}, a^{-2}, a^{-3}, etc., for $\frac{1}{a}$, $\frac{1}{aa}$, $\frac{1}{aaa}$, etc. [11].

He then uses this notation in his binomial formula for $(x + y)^{m/n}$. Euler also discusses fractional powers and the alternative surd notation.

> §200 We may therefore entirely reject the radical signs at present made use of, and employ in their stead the fractional

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

> exponents which we have just explained: but as we have been long accustomed to those signs, and meet with them in most books of Algebra, it might be wrong to banish them entirely from calculations; there is, however, sufficient reason also to employ, as is now frequently done, the other method of notation, because it manifestly corresponds with the nature of the thing. In fact we see immediately that $a^{1/2}$ is the square root of a, because we know that the square of $a^{1/2}$, that is to say $a^{1/2}$ multiplied by $a^{1/2}$ is equal to a^1, or a. [12].

Perhaps he lacked the courage of his conviction that fractional powers 'manifestly corresponds with the nature of the thing'; however there is little contemporary doubt that the design of this notation significantly aids computation. Unfortunately, $(a^n)^m = a^{mn}$ needs auditing.

The logical symbols as abbreviations, e.g. $\Leftrightarrow$ for 'if and only if', are well-established conventions. It is somewhat surprising that the use of such symbolism has not yet routinely and systematically been extended to writing complete arguments in many commonly occurring situations, particularly reasoning by equivalence. Proof by induction might be another candidate for formal or conventional written presentation. Using the two dimensional layout on the page, e.g. lining up in vertical columns, is also not standard practice in written work. It is unusual for contemporary students to write *any* logical connectives when reasoning by equivalence! Boole commented [13]:

> The theory of Logic is thus intimately connected with that of Language. A successful attempt to express logical propositions by symbols, the laws of whose combinations should be founded upon the laws of the mental processes which they represent, would, so far, be a step towards a philosophical language.

An expanded use of formal symbolism might have many of the same advantages as we see in algebra. This includes fewer mistakes, as demonstrated by the examples here. Notice that the symbol $\Leftrightarrow$ has been used in this article throughout the audited arguments. This connects individual equations into a complete logical argument, combining logical inference and algebraic calculation. Viewed in this way, individual equations or expressions, the auditing and the logic which links them become a *single mathematical entity*. The whole argument considered as a single mathematical entity could be manipulated by software, just as algebraic expressions can be manipulated by a computer algebra system. This include cut and paste of expressions which form part of an argument, retaining domain information. Proof assistants, proof verification, computer algebra and related tools would then combine both logic with computation.

https://doi.org/10.1017/mag.2015.38 Published online by Cambridge University Press

Acknowledgement: Many thanks to Richard Kaye for helpful comments on an early draft.

References

1. G. Chrystal, *Algebra, an elementary text-book for the higher classes of secondary schools and for colleges*, volume 1, Adam and Charles Black (3rd edn.) (1893) p. 290.
2. E. P. Northrop, *Riddles in mathematics: a book of paradoxes*, English Universities Press (1945) p. 21.
3. A. De Morgan, *Elements of algebra, preliminary to the differential calculus* (2nd edn.) Taylor and Walton (1837) p. 122.
4. M. H. A. Newman et al. *The teaching of algebra in sixth forms: a report prepared for the Mathematical Association*, G. Bell (1957) p. 8.
5. T. Lund, *The elements of algebra designed for the use of students in the university* (14th edn.) Longman, Brown, Green and Longmans (1852) p. 130.
6. C. Babbage, On notations, *Edinburgh Encyclopedia* Volume 15 (1830) pp. 394-399.
7. I. Todhunter, *An elementary treatise on the theory of equations*, MacMillan (1885).
8. E. A. Maxwell, *Fallacies in mathematics*, Cambridge University Press (1959).
9. J. A. Stedall, *A discourse concerning algebra: English algebra to* 1685, Oxford University Press, USA (2002).
10. R. Recorde, *The Whetstone of Witte*, I. Kyngston (1557).
11. S. Horsley, *Isaaci Newtoni opera*, Joannes Nichols (1782) p. 215.
12. L. Euler, *Elements of algebra* (3rd edn.) Longman, Hurst, Rees, Orme (1822). Translated from the French, with the notes of M. Bernoulli and the additions of M. de La Grange by J. Hewlett.
13. G. Boole, *The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning*, MacMillian, Barclay, MacMillan (1847) p. 5.

doi:10.1017/mag.2015.38

C. J. SANGWIN
Mathematics Education Centre, Ann Packer Building,
Loughborough University LE11 3TU
e-mail: *C.J.Sangwin@lboro.ac.uk*