Extraneous and Lost Roots

In solving equations (for a variable in the equations) one often "does the same thing" to both sides of the equations. The things that can be done to both sides of the equation can sometimes be thought of as the application of a function. For example, in solving the equation

$$3x - 7 = 5$$

one might think of the process as being to first add 7 to both sides, and then to divide both sides by 3. Thus the functions $f(x) = x + 7$ and $g(x) = \frac{x}{3}$ were applied in succession, which is the same thing as applying $g \circ f(x) = \frac{x+7}{3}$ to both sides.

If we write the equation being solved in a generic form, such as $a(x) = b(x)$, then applying a function *f* to both sides of the equation amounts to concluding that $f(a(x)) = f(b(x))$. In this way, we are relying on a fundamental requirement of a function, that it must be single-valued (*a=b* implies *f(a)=f(b)*). Thus if we are solving an equation involving the variable *x* for that variable, and the equation we are given is thought of as a statement about that variable, then the equation we get by applying a function to both sides of the equation must follow from that statement and must also be a true statement about that variable. Whether applying that function to both sides of the equation appears to be a step in a useful direction toward solving for x is a separate question.

Some of the things that are commonly done to both sides of an equation are not so readily viewed as an application of a function to both sides of the equation. For example, simply adding *x* to both sides of an equation might not be an application of a function to both sides. Generally many of the operations that are performed on equations can be viewed as adding a function to both sides of an equation or multiplying both sides of an equation by a function. For example, to solve the equation

$$x^2 + 3x = 8 + x$$

one might add $h(x) = -x - 8$ to both sides of the equation to obtain

$$x^2 + 2x - 8 = 0$$

and to solve

$$\frac{3}{x-1} = x - 3$$

one might multiply both sides of the equation by $k(x) = x - 1$ to obtain

$$3 = (x-3)(x-1)$$

Note what is often referred to as "cross multiplication" can also be viewed as an example of multiplying both sides by a function. In general, to solve an equation of the form $\frac{f(x)}{g(x)} = \frac{h(x)}{k(x)}$, one might multiply both sides of the equation by $g(x)k(x)$ to obtain $f(x)k(x) = h(x)g(x)$.

Thus we have two general ways of viewing the golden rule of algebra, as application of a function to both sides of an equation and as addition or multiplication of both sides of an equation by a function. In many cases, the operation that is done to both sides of the equation can be thought of in either way (e.g., adding 7 to both sides of an equation can be an application of $f(x) = x + 7$ to both sides or as addition of $h(x) = 7$ to both sides).

Extraneous Roots

Sometimes in solving an equation, after performing correctly doing the same thing to both sides of the equation, an equation results that has a root or roots that are not solutions to the original equation. These are usually called extraneous roots.

If a sequence of functions was applied to both sides of the equation, then there must have been a first equation in the sequence of resulting equations at which an extraneous root appeared. Suppose that the equation just before the extraneous root appeared is written generically as *a*=*b*, and after the function *f* is applied the equation becomes $f(a) = f(b)$. Then when these two equations are evaluated at the extraneous root, the first equation *a*=*b* is not true but the second equation $f(a) = f(b)$ is true. That is, we have $a \neq b$ but $f(a) = f(b)$.

For example, in solving the equation $\sqrt{x+6} = x$, if both sides of the equation are squared, then the equation $x + 6 = x^2$ results, for which x=-2 is a root, although it is not a root to the original equation: $\sqrt{-2+6} \neq -2$. When the equations are evaluated at x=-2, we started with $\sqrt{-2+6} \neq -2$ or $2 \neq -2$ and applied $f(x) = x^2$ to both -2 and to 2 and obtained 4=4.

Extraneous roots then can appear when a function is applied to both sides of the equation for which it is possible that $a \neq b$ but $f(a) = f(b)$. This is another way of saying that two different input values can give the same result when the function f is applied, which means that *f* is not a one-to-one function. In the last example, the function that was applied to both sides is the squaring function, $f(x) = x^2$, which is a standard example of a function that is not one-to-one.

Thus the application of functions that aren't one-to-one can produce extraneous roots. The six basic trigonometric functions fail to be one-to-one, and so applying them to both sides of an equation can give rise to extraneous roots. For example, in solving the equation $\arcsin x = \pi$, which has no roots, if the sine function is applied to both sides of the equation, then new equation is $\sin(\arcsin x) = \sin(\pi)$, or x=0, which provides an extraneous root to the original equation.

Functions that are not one-to-one abound, although functions that are commonly applied to both sides of an equation in solving an equation are less in abundance. Common examples include even power functions (functions of the form $f(x)=x^n$, where n is even) and more generally functions of the form $f(x)=x^{m/n}$ where m is an even integer and n is an odd integer.

It is also possible to arrive at extraneous roots when multiplying both sides of an equation by a function. If the original equation is $a(x)=b(x)$ and both sides are multiplied by $h(x)$, then the resulting equation $a(x)h(x)=b(x)h(x)$ can also have roots at zeroes of $h(x)$. For example, in solving the equation $\frac{1}{x}=\frac{4}{x^2}$, if both sides of the equation are multiplied by the function $h(x)=x^3$ (achieving the same effect as cross multiplying), then the result is the equation $x^2=4x$, which has the root of the original equation $x=4$ as well as an extraneous root $x=0$, acquired because it was a zero of $h(x)$.

Extraneous roots can also appear as a result of applying a function that is one-to-one to both sides of an equation, due to a subtlety in the apparent domain of the functions that form the sides of the equation.

The general understanding for an equation $a(x)=b(x)$ is that, unless otherwise noted, the values of x that are to be considered possible roots of the equation comprise the intersection of the domains of $a(x)$ and of $b(x)$, which are understood to be the largest collection of (typically real) values of x at which the functions are defined. Thus for example, in solving the equation $\frac{3}{x}=x-2$ for x, the understanding is that roots are to be selected from all non-zero numbers, because 0 is outside the implied domain of one of the sides.

This can cause confusion, because when a function $f(x)$ is applied to both sides of $a(x)=b(x)$, the equation becomes $f(a(x))=f(b(x))$ and the domain should now be a subset of the domain of $a(x)=b(x)$ under the function *f*, but because of the standard understanding that the domain is the largest possible domain implied by the sides of the equation, the domain can be mistaken to be the largest possible domain of $f(a(x))$ and $f(b(x))$. For example, consider the initial equation $\frac{3}{x}=\frac{4}{x}$, which has no roots and whose domain is understood to be all non-zero numbers. If the function $f(x)=\frac{1}{x}$ is applied to both sides of the equation, the domain should still be all non-zero numbers. But the resulting equation is $\frac{x}{3}=\frac{x}{4}$, whose implied domain is all numbers, and now $x=0$ appears as an extraneous root.

Lost Roots

If applying a function to both sides of an equation produces an equation that is implied by that original equation, how is it possible to lose roots when solving an equation?

For example, in solving $x^2 = 5$, the root $x = -\sqrt{5}$ is easily lost, typically because of a forgotten $\pm$ symbol. The equation $x^2 = 5$ is understood to imply the equation $x = \pm\sqrt{5}$, but what function was applied to both sides of the equation? If the function that is applied to both sides is *f*, then this would say that $f(5) = \pm\sqrt{5}$, meaning that *f* is not single-valued and hence not a function, so this step apparently involves something other than just applying a function to both sides.

Careful application of the function $f(x) = \sqrt{x}$ to both sides of the equation $x^2 = 5$ produces an equation that is implied by that equation: $\sqrt{x^2} = \sqrt{5}$. No roots were lost in this step. Now recognize that for real numbers x, $\sqrt{x^2} = |x|$ and so the equation $\sqrt{x^2} = \sqrt{5}$ is equivalent to the equation $|x| = \sqrt{5}$. The final step is to use knowledge about the absolute value function to conclude that $x = \pm\sqrt{5}$. Note that this last step is not the same thing as applying a function to both sides of an equation.

Another common way to lose a root is to multiply both sides of the equation by a function $h(x)$ whose domain does not include the root. For example, in solving the equation $x^2 = x$, if both sides are multiplied by $h(x) = \frac{1}{x}$, then the resulting equation is $x = 1$ and the root $x = 0$ is lost, because it is not in the domain of $h(x)$. The apparent domain of $x^2 = x$ is all numbers, but when both sides were multiplied by $h(x) = \frac{1}{x}$, the domain was decreased by $x = 0$, because that is not in the domain of $h(x)$. In this way the root $x = 0$ was lost.

Roots can also be lost by applying functions to both sides of an equation if the function that is applied has a limited domain. If the function $f(x)$ is applied to both sides of $a(x) = b(x)$, the equation becomes $f(a(x)) = f(b(x))$. The domain of $a(x) = b(x)$ is not necessarily the same as the domain for $f(a(x)) = f(b(x))$, as the latter might be restricted by the domain of *f*. For example, if we start with the equation $(x+2)^2 = 1$, which has roots $x = -1$ and $x = -3$, and apply a logarithmic function to both sides, we have imposed a restriction on the equation that whatever is inside the log function must be positive. This subtle restriction can easily go unnoticed:

$$(x+2)^2 = 1$$

$$\log(x+2)^2 = \log(1)$$

$$2\log(x+2)=0$$
$$x+2=1$$
$$x=-1$$

The root $x=-3$ was lost when the equation $2\log(x+2)=0$ arrived, as $x+2$ is required to be positive because of the domain of the logarithmic function.