“ax + b” is dead, long live “a(x − b)”

Douglas Butler on transformations of polynomial and trigonometric functions

1. The pitfalls of starting with \( y = ax + b \) (or \( mx + c \))

The straight-line

Students learning about graphs start, naturally, with the straight-line. The traditional approach is to use the forms:

\[ y = mx + c, \text{ or } y = ax + b \]

and to discover, by exploration, that ‘m’ represents the gradient and ‘c’ the \( y \)-intercept. This is fine as far as it goes, but the skills learnt do not transfer to the transformation of polynomial or trigonometric functions, or even the circle.

One approach which prepares better for future transformations is to study the straight lines, initially in the forms:

\[ y = mx - c \]

for which ‘c’ is the transformation in the \( y \) direction, this the \( y \) intercept

Then try:

\[ y = m(x - c) \]

for which ‘c’ is now the transformation in the \( x \) direction, this the \( x \) intercept. The earlier you can instil the idea that \( (x - c) \) moves it to the RIGHT by ‘c’, the better.

Then

\[ y = (x - a)^2 \]

which is a translation to the right by \([0, a]\).

Moving on to:

\[ y = kx^2 \]

but, here you need to be careful with the interpretation: this is a stretch in the \( y \) direction, factor ‘k’, and really should be written as:

\[ y = k(x^2), \text{ or even } \sqrt{k} = x^2 \]

This form is of course very different from \( y = (mx)^2 \) which never goes negative, and is possibly best described as an ‘inverse stretch’ in the \( x \)-direction.

This is a concept that will surface again as ‘frequency’ in trig graphs, i.e. the higher the value of ‘m’, the more squashed up is the graph.

Putting these together, the most general quadratic is:

\[ y = k(m(x - a))^2 + b \]

which, relating to its ‘parent’ \( y = x^2 \), has the following interpretation for its parameters:

‘k’ is the vertical stretch (amplitude?), ‘m’ is the ‘inverse stretch’ in the \( x \)-direction (frequency?), ‘a’ is the translation by \([a, 0]\) in the \( x \)-direction (phase shift?), and … ‘b’ is the translation by \([0, b]\) in the \( y \)-direction

Consequently the vertex of the parent graph has been translated by the vector \([a, b]\) to \((a, b)\).

There is of course some redundancy here as it could be written as:

\[ y = (km^2)(x - a)^2 + b \]

but, it is nice to note that the interpretation of ‘k’ and ‘m’ are inter-related.

2. The Quadratic Function

If the straight line has been done properly, the transformations of:

\[ y = x^2 \]

can be explored though the progressive study of:

\[ y = x^2 + b \]

which should be written \( y - b = x^2 \), so that it represents a translation by \([0, b]\)

Then

\[ y = (x - a)^2 \]

which is a translation to the right by \([0, a]\).

Moving on to:

\[ y = kx^2 \]

but, here you need to be careful with the interpretation: this is a stretch in the \( y \) direction, factor ‘k’, and really should be written as:

\[ y = k(x^2), \text{ or even } \sqrt{k} = x^2 \]

This form is of course very different from \( y = (mx)^2 \) which never goes negative, and is possibly best described as an ‘inverse stretch’ in the \( x \)-direction.

This is a concept that will surface again as ‘frequency’ in trig graphs, i.e. the higher the value of ‘m’, the more squashed up is the graph.

Putting these together, the most general quadratic is:

\[ y = k(m(x - a))^2 + b \]

which, relating to its ‘parent’ \( y = x^2 \), has the following interpretation for its parameters:

‘k’ is the vertical stretch (amplitude?), ‘m’ is the ‘inverse stretch’ in the \( x \)-direction (frequency?), ‘a’ is the translation by \([a, 0]\) in the \( x \)-direction (phase shift?), and … ‘b’ is the translation by \([0, b]\) in the \( y \)-direction

Consequently the vertex of the parent graph has been translated by the vector \([a, b]\) to \((a, b)\).

There is of course some redundancy here as it could be written as:

\[ y = (km^2)(x - a)^2 + b \]

but, it is nice to note that the interpretation of ‘k’ and ‘m’ are inter-related.
The un-factored general form of the quadratic
\[ y = ax^2 + bx + c \]
is not at all friendly, yet it is frequently studied: there are simply no visual interpretations readily available for either ‘a’ or ‘b’, apart from the sign of ‘a’ leading to the ‘happy’ and ‘sad’ descriptors! However, in this form, ‘c’ does revisit the idea of a y-intercept.

On the other hand, the important ‘completing the square’ exercise does bring in the concepts above, for example:

\[ y = 2x^2 + 4x + 3 \Rightarrow y = 2(x + 1)^2 + 1 \]

so, vertex at \((-1, 1)\) with vertical stretch of 2.

3. Trigonometric plotting, and the dreaded \( \sin(bx + c) \)

A few thoughts about the meaning of “frequency”:

Frequency is usually denoted by ‘f’ measured in Hz, but it can be angular frequency, \( \omega \) rad/sec. These definitions only apply when the \( x \)-axis is ‘t (sec)’.

For example \( y = \sin(\omega t) \) or \( y = \sin(2\pi ft) \). However \( y = \sin(3x) \) moves three times as fast as \( y = \sin(x) \) and is often described as having a frequency 3.

All too often text books, and even examination questions, use the form:

\[ y = a \sin(bx + c) + d \]

From the above analysis we know that ‘a’ controls the vertical scaling (‘amplitude’), and ‘d’ is the vertical translation by \([0, d]\). But, what do ‘b’ and ‘c’ represent? If you rearrange as:

\[ y = a \sin(b(x + \tfrac{c}{b})) + d \]

it can be shown that ‘b’ represents an inverse stretch in the \( x \)-direction (‘frequency’), but ‘c’ does not represent anything useful. The translation in the \( x \)-direction (‘phase shift’) is by the vector \([-c/b, 0]\). Not nice.

Having learnt transformations of straight-lines and quadratics properly, it should be obvious that the correct form for the general sine wave is:

\[ y = a \sin(b(x - c)) + d \]

where ‘a’ is the vertical stretch (‘amplitude’), ‘b’ is the ‘inverse stretch’ in the \( x \)-direction (‘frequency’), ‘c’ is the translation by \([c, 0]\) (‘phase shift’) in the positive \( x \)-direction, and ‘d’ is the vertical translation by \([0, d]\). Furthermore, the origin of the graph has been translated by the vector \([c, d]\) to \((c, d)\), and, for differing values of ‘a’ and ‘b’, the graph always passes through \((c, d)\).

\( \sin(bx + c) \) is dead, long live \( \sin(b(x - c)) \) . . .