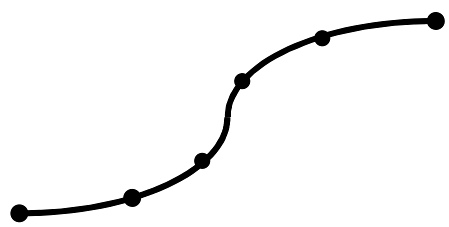
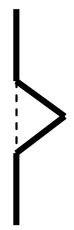
The Five Simple Facts About The Mandelbrot Set

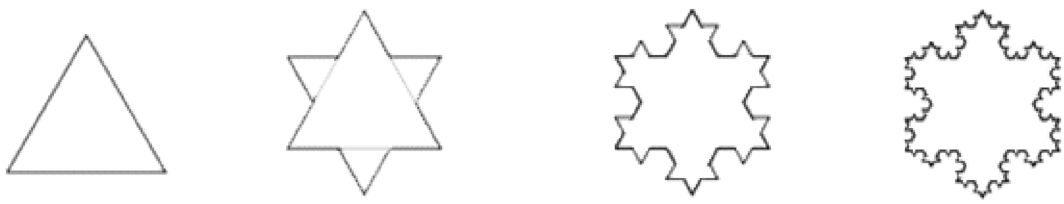
Imagine a square and in it a staircase going from the left bottom corner up to the right top one. The total length of the steps by counting both the horizontal and vertical lengths is obviously the double of the side. The same way, in a city with perpendicular streets only, it doesn't matter what way we go from A to B. But back to the square staircase, if we make smaller and smaller steps then the staircase approximates the diagonal. And so we approximated a segment yet we didn't approximate its length. This paradox is not mentioned before starting integration which is a didactical error. But our start is a reverse paradox.
If we don't use forced horizontal and vertical directions rather little connecting segments then we might think that the length approximation is always correct. And indeed in this next picture the curve segments could be approached by connecting strings:



Surprisingly, we can make a "curve" where this method definitely fails.
The idea is to start with a single interval and make a bump or sharp detour on the middle, say replace the middle third with two sides on a left or right triangle:



Then we can repeat this for the new 4 straight segments. And then, again and again.
We could even choose the left or right side of the detours by some rules.
The same idea can be applied to special closed curves or boundaries of a point set.
For example, we can start with a triangle and apply our previous line detours to all three sides.
If the three first detours are all done outward it will give a six-star or Star of David.
Continuing, we get the shape of a snow flake.



Now we again depart for a third question, whether such infinitely wiggly curve can be defined without such primitive repetitive rule, somehow "by itself".
The answer is yes but the "by itself" becomes tricky.
Namely, we will define the inner and outer points by an in itself very tricky method.
Then the "boundary", "the curve", can become wiggly beyond our imagination.

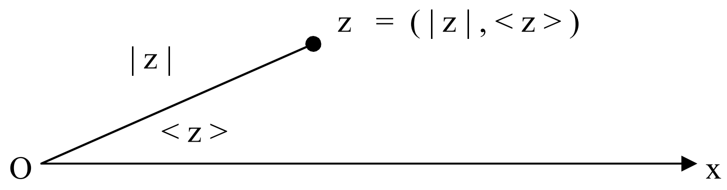
But this inner and outer intuitive picture must be handled first.
And so we start with an even simpler paradox.
Imagine a unit circle in a plane and its inside disc regarded as the B black points.
A black disk in a white plane, that is indeed simple.
But what about the circle itself? Is it black or white? Well, we can not see this from a picture!
If we defined our black points as those P that $|OP| \leq 1$ then the circle is black but if we defined them as $|OP| < 1$ then only the inside is black and the circle itself is white.
So only our algebraic definition tells the truth.
Now let's jump in abstraction and allow any B and W separation of the plane's points.
The most important points of the plane are where B and W mix and these are easy to define.
But first we can define if B approaches a P as P having in any surrounding infinite many points from B. Then the mixing points are those that are approached by both B and W.
Now observe that if B approaches P then trivially in any surrounding there has to be point of B beside P if it is in B at all. Not so trivially the reverse of this consequence is sufficient for approach already. That is, if in any surrounding of P there is point of B beside P then P is approached. Indeed, we can show this in the negative version. So if P is not approached then there is a surrounding where there is no point beside P. Indeed, not being approached means not having infinite many points of B in a surrounding. Thus either there is no point of B there at all or there is but only finite many. But then the finite many other than P has a closest to P and so a surrounding small enough would only have maybe P inside from B.
The proof provided also a characterization of a non approached P point as either having a surrounding where there is no B point at all, or having one where P is the only one from B.
In the first case P could be called as an outer point for B while in the second P is an isolated point of B. Now we just use the same logic for W.
Indeed, a non mixing point is simply where only one of B or W approaches P.
We just characterized the B not approaching situation where by the way W must approach.
So the other two cases of a non mixing P point is when it is outer point for W or isolated point of W. The more logical name of the first is that P is inner point of S, indeed having a surrounding fully inside B and of the second P being a hole in B. So to recap:
A P point of the space where B and W are not mixing can only be four kind.
1. outer point of B , that is inner point of W.
2. inner point of B , that is outer point of W.
3. isolated point of B , that is hole in W.
4. hole in B , that is isolated point of W.

The above black disk paradox hiding the circle's points whether they are black or white is now the question where the mixing points will belong. All to B, all to W or some here some there.
In our tricky definition of B and W they all will belong to B.
A totally hiding same answer by the conventional Formalist treatment is that Mandelbrot set is closed. Indeed, being closed means that all the points that B approaches are members of it.
Then of course the mixing points must also be all members in B too.

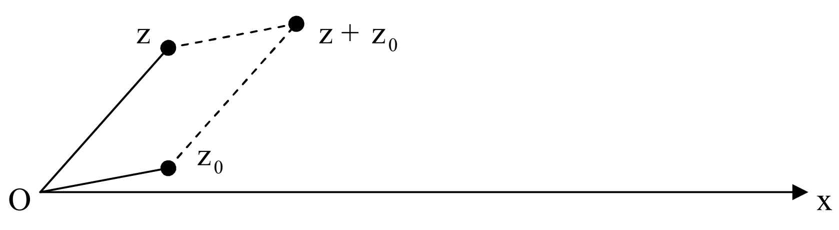
Now we come to the essence, the separation of the plane's points by a fairly primitive rule.
The trick is that this primitive rule does not directly give the answer if a point is in B or W.
Rather we apply the rule repeatedly, and then watch how the $z_1, z_2, z_3, \ldots$ sequence of points behave. Namely, whether they go to infinity or not, that is remain in a bounded region.
The false naive assumption was earlier that if a continuous transformation creates z_2 from z_1 then z_3 from z_2, and so on, then this would make the nearby points to z_1 go nearby to z_2 then nearby to z_3, and so on, and thus the divergence or boundedness would be similar too.
This however is not true! Continuity indeed implies that a fix finite many repetition creates points nearby, but this doesn't imply that the divergence can not happen for the full sequence.
So, closer and closer points can still have different end results in regard of divergence or boundedness. Only the actual examples make this affair clear.

A mystical spin was then attached to all this, as if the points had a secret unruled world.
The continuity of point sets is indeed mysterious, but these finer and finer wiggles on the borderline are not mysterious at all. The fact that they never repeat perfectly, that is don't become periodic, is perfectly natural. An analogue will explain this:
The rational numbers are the periodic decimals. An irrational, like $\sqrt{2}$ has totally "unpredictable digits" if it is meant by non repeatedness. Yet, these digits are very predictable. Here, in two dimension, the non periodicalness is simply more vivid visually.

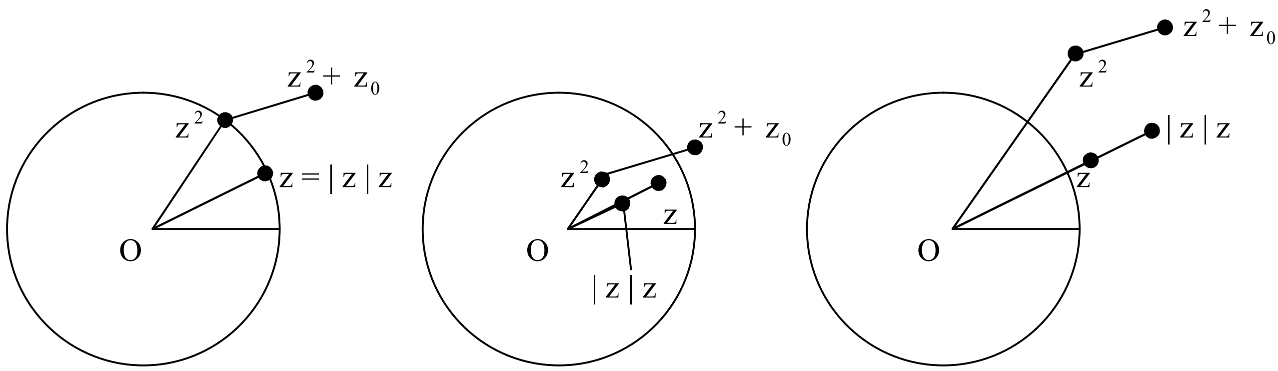
The mentioned primitive transformation from z_1 to z_2 then z_2 to z_3 and so on, can be regarded in three steps. The first is a stretch, that is the alteration of the distance from a fix O origin, but keeping the radial line from O. The second will be a rotation around O and the third is a shifting. To use our transformation formally, the best is to use not a Descartes, but rather a polar coordinate system.
Here, the z points have two components, their length from O denoted as $|z|$ the absolute value and the angle of z's radial from O to a fix horizontal x axis, denoted as $<z>$.
So, $z = (|z|, <z>)$



The multiplication of z with a c real number will be defined as $c\,z = (c\,|z|\,,\ <z>)$.
Thus, we'll keep the angle and only apply a c stretch.
Such fix c stretch is not enough to produce the interesting wiggles and rather we need a second order or squared stretch. So this actually means to form: $|z|\,z = (|z|^2\,, <z>)$
The beauty of this is that for $|z| = 1$, that is on the unit circle, this stretch does nothing, but inside we compress z toward O while outside, indeed it stretches more and more outward.
But in itself, this is still not enough. We need the further two steps.
First, a linear, that is first order turn, and then a fix, that is 0 order shift.
The linear turn would mean that $<z>$ is changed to $c<z>$.
We'll stick with the simplest $c = 2$ choice.
This also fits into our notations if we introduce the point multiplications in general as:
$u\,v = (|u|, <u>)(|v|, <v>) = (|u||v|\,,\ <u>+<v>)$
In short, we multiply the lengths and sum the angles.
Then, the $z = (|z|, <z>)$ point squared is exactly the one we'll use as second step, because:
$z\,z = (|z|, <z>)(|z|, <z>) = (|z|^2\,,\ 2<z>)$
Finally, the fix shift fits again into our notation as a $+\,z_0$ addition if we use vector addition.
This means that the $O\,z_0$ radial is moved parallel and same directionally to the point to which we want to add z_0 and then the end of this moved parallel version of z_0 will be the sum:



Thus, our transformation written in our introduced $z \rightarrow (\,z^2 + z_0)$ or so called complex notation, but broken into the three actual geometrical steps and shown on, in and outside the unit circle will look like this:



We apply this to a starting z_1 to form $z_2 = z_1^2 + z_0$ then we form $z_3 = z_2^2 + z_0$ then $z_4 = z_3^2 + z_0$ and so on. The turns and shifts can play against each other and this creates the crucial steppings in or out of the unit circle.

The Julia set collects all those z_1 points that if used as start will bring about a bounded sequence. So: Julia $(z_0) = \{ z_1 \;;\; z_2 , z_3 , \ldots$ are bounded $\}$

The Mandelbrot set collects all those z_0 shift values that if used with $z_1 = O$ as start value, will bring about a bounded sequence.
So: Mandelbrot $= \{ z_0 \;;\; O = z_1 , z_0 = z_2 , z_3 , z_4 , \ldots$ are bounded $\}$

In spite of the Julia set being the simpler logically, the Mandelbrot set became more famous.
The two inter-relate delicately.

The mentioned five simple facts in the title give only a rough knowledge of the Mandelbrot set, but they are so easy to obtain, that form a definite start of the subject.
Yet these are never shown in the popularizing exposes.

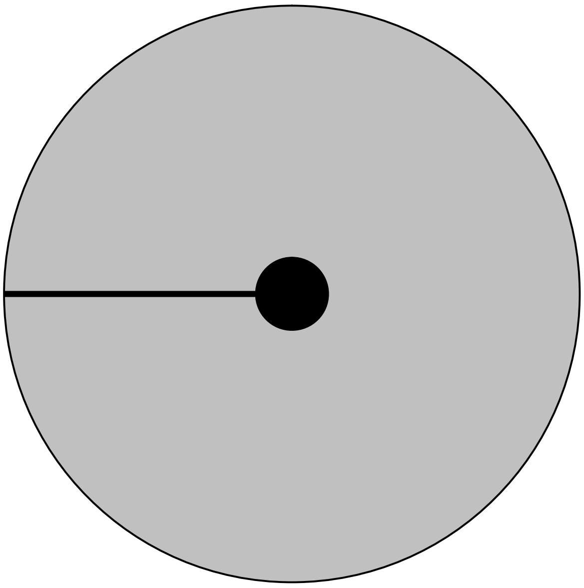
The first easy fact is that the Mandelbrot set is symmetrical to the x line.

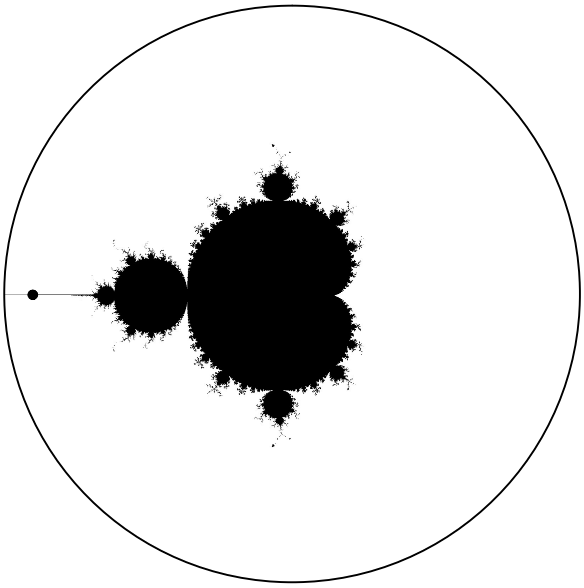
The second fact says that if $|z_0| > 2$ then we must have divergence.
So, the plane outside of the 2 radius circle is a definite "outer zone", it is white.

The third fact says quite oppositely that if $|z_0| \leq \frac{1}{4}$ then we must have boundedness.

So, the disc of the $\frac{1}{4}$ radius circle is a definite "inner zone", it is black.
These are the trivial points of the Mandelbrot set.
Between these two circles lies a ring that is not obviously white nor black.
So, it is the ring where all the non trivial points of the Mandelbrot set must be.
Luckily, on the x axis, the points of the ring are trivial too. Namely, on the right of the trivial $\frac{1}{4}$ disc all points cause divergence, so are white, while to the left, all points cause boundedness, so are black and belong to the Mandelbrot set.

This will be the left picture, and the much wilder reality of the ring is on the right:





Now we'll come to proving the five facts.

1. Symmetry.
Indeed, firstly, angles measured downward from x count as minus and so their doubling will be doubled downward too. Secondly, the shift is collected and it keeps symmetry, that is if a z_0' is symmetrical with z_0 then adding it to symmetrical z , z' will stay symmetrical.

2. Monotony of the outer zone.

If $2 < |z_0| \leq |z|$ then $|z^2 + z_0| > |z|$.

Indeed, firstly $|z^2 + z_0| \geq |z|^2 - |z_0| \geq |z|^2 - |z|$.

But also, $|z| > 2 \rightarrow |z| - 1 > 1 \rightarrow (|z| - 1)^2 > 1 \rightarrow (|z| - 1)^2 - 1 > 0$

$\rightarrow |z|^2 - 2|z| > 0 \rightarrow |z|^2 - |z| > z$.

Since $z_2 = z_0$ thus using z_2 as z we get that $|z_3| > |z_2|$.

Then using z_3 as z we get that $|z_4| > |z_3|$ and so on.

3. Boundedness in the inner zone.

If $|z_0| \leq \frac{1}{4}$ then the sequence cannot diverge.

Namely, if $|z| < \frac{1}{2}$ then $|z^2 + z_0| < \frac{1}{2}$ too.

Indeed, $|z^2 + z_0| \leq |z|^2 + |z_0| < \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

So, since $z_2 = z_0$ thus $|z_2| \leq \frac{1}{4} < \frac{1}{2}$ indeed and so, $|z_3|, |z_4|, \ldots < \frac{1}{2}$ too.

4. Monotony on x.

If $x_0 > \frac{1}{4}$ then $x < |x^2 + x_0|$

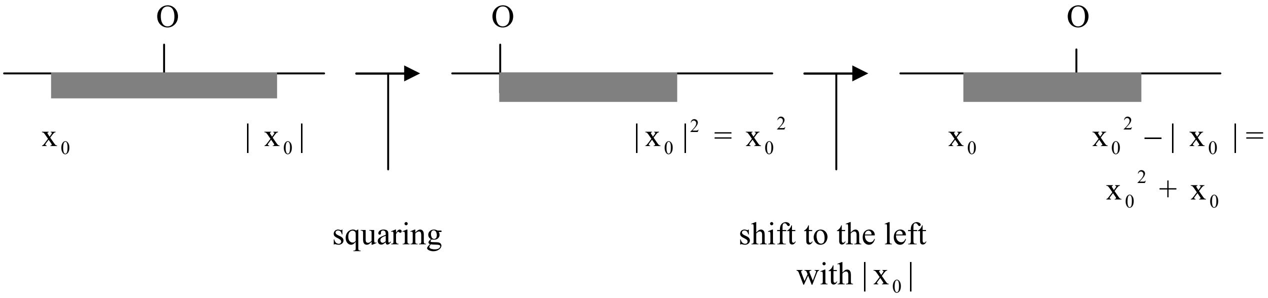
Indeed, firstly $|x^2 + x_0| = x^2 + x_0$.

But also, $x_0 > \frac{1}{4} \rightarrow x_0 - \frac{1}{4} > 0 \rightarrow (x - \frac{1}{2})^2 + x_0 - \frac{1}{4} > 0$

$\rightarrow x^2 - x + x_0 > 0 \rightarrow x^2 + x_0 > x$.

So, actually $x_1 < x_2 < x_3 < \ldots$

5. Boundedness from O to -2 on the x axis.
If x_0 is in $[-2, O]$ and x is in $[x_0, |x_0|]$ then $x^2 + x_0$ is there too.
Indeed, the $[x_0, |x_0|]$ interval is squared and shifted left to x_0:



Here $x_0^2 + x_0 \le |x_0|$ if $|x_0| < 2$ because:

$$0 \le |x_0| \le 2 \rightarrow ||x_0|-1| \le 1 \rightarrow -1 \le |x_0|-1 \le 1 \rightarrow (|x_0|-1)^2 \le 1$$
$$\rightarrow x_0^2 - 2|x_0| = x_0^2 + 2x_0 \le 0 \rightarrow x_0^2 + x_0 \le |x_0|$$

A final exactification is needed for the above proofs in 2. and 4.
We indeed proved the monotony but, this does not instantly prove divergence. Indeed, it could be possible that the absolute values are always increasing, but never go beyond a certain value. This however would also mean that they approach a v value. This also implies that the actual points also approach a z point. The transformed of this z would have to be the limit of the limits and thus z itself, contradicting monotony.

The x dimensional result in two dimension

The $[-2, \frac{1}{4}]$ boundedness result on the x axis has a very nice visual meaning.
On the x axis of course the transformation is simply $x^2 + x_0$.
But if we regard the $y = x^2$ parabola then this x_0 is actually a y value shift.

The surprisingly simple method of the iteration is then regarding the 45 degree $y = x$ line through the origin. Any point of this line can be regarded as a starting x value.
A vertical trip to the parabola gives the first function point.
From here a horizontal trip back to the $y = x$ line gives the next x value.
Continuing this again and again an infinite staircase or spiral will show where x is going.
In truth, a third option can be a finite loop and a fourth option can be not even moving if we start from where the parabola crosses $y = x$.

A beautiful example of an upward staircase shows visually why $\frac{1}{4}$ is the end of the boundedness section. Indeed, the $y = x^2 + \frac{1}{4}$ parabola touches the $y = x$ line at $(\frac{1}{2}, \frac{1}{2})$.
Starting from the origin with a vertical line we get to $(0, \frac{1}{4})$, the minimum of the parabola.
The horizontal from here takes us to $(\frac{1}{4}, \frac{1}{4})$ on the $y = x$ line.
From here a vertical goes up to the parabola and then we repeat and walk the upward staircase.
Quite obviously we will approach the touching point $(\frac{1}{2}, \frac{1}{2})$.

If the parabola is shifted up more than $\frac{1}{4}$ then the $y = x$ line will not touch it.
So the staircase might narrow first but then go to infinity.

The other end is not as simple:
At the $y = x^2 - 2$ parabola we go to $(0, -2)$ vertically from the origin, again to the minimum of the parabola. From here horizontally to $(-2, -2)$ on the $y = x$ line.
Then we get vertically to $(-2, 2)$ on the parabola and then $(2, 2)$ on the line.
But this is actually a crossing point so a fix point or end in our method.
By the way, the other crossing is at $(-1, -1)$ which suggests that it has to be a limit value for some other starts than our conventional from the origin.
That with shifts above -2 we approach the $(2, 2)$ crossing, while with shifts under -2 the trips are unbounded, is not as trivial as above but visually can be checked perfectly.

A new layer of the "fractal behavior" is appearing by observing the sequence values with non origin starts, as also suggested by the previously mentioned second crossing at $(-1, -1)$.
The -2 lowest edge of the boundedness section is for example a spiral that will have values arbitrary close to a range of values and this stays the same for all starting values.
So this lowest edge is actually a most complex point for the sequences.
Moving the parabola up we see cycles of less and less complexity.
The Mandelbrot Set was blind to all this! It only cared about boundedness or not.

An amazing meaning behind the previous one dimensional result

To see this new depth even better, we can shift our normal parabola to the right and then apply a negative y directional stretch.
This way we can put an upside down parabola symmetrically onto the $[0, 1]$ interval.
So then the new $y = r(x - x^2)$ equation describes a parabola mountain on $[0, 1]$.
Its height is at the middle, that is at $\frac{1}{2}$ and with value $r(\frac{1}{2} - \frac{1}{4}) = \frac{r}{4}$.

The $r = 4$ case with height 1 corresponds to the earlier -2 edge and here both the possible initial values and the sequence values will be between 0 and 1.
So now the mentioned total complexity means that the sequence values seem like randomly generated real numbers between 0 and 1. Von Neumann observed this already in the forties.
Later computer calculations showed the cyclic intricacies for lower r values.

But a whole new meaning emerged by regarding our equation as $y = r x (1 - x)$.
It means that y is proportional with both x and $1 - x$.

Now if an animal population has 1 as limit and x is a portion of this as actual population, then $1 - x$ is the still potential population portion.
Then r is the reproduction rate and then a next y population is indeed proportional not only with $r x$ but also with $1 - x$ due to the food supply.
Amazingly, then this primitive model foretells that slight variations in the reproduction constant and initial population can cause dramatically different limit populations.
And this is observed in nature.