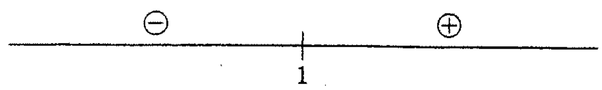
Signs of a Polynomial

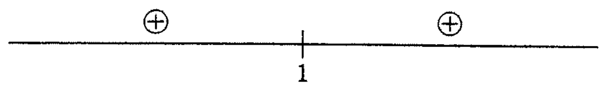
Suppose P is a polynomial. We wish to find the intervals on which the value of P is positive, and the intervals on which it is negative; briefly, we wish to find the *signs* of P.

EXAMPLE 1. Let $P_a(x) = x - 1$. It's easy to see that $P_a(x) > 0$ if $x > 1$, and $P_a(x) < 0$ if $x < 1$. So P_a is positive on the interval $(1, \infty)$, negative on $(-\infty, 1)$. We may indicate this by drawing what we call a *sign line* for P_a:



Notice that 1 is a root of P_a (a root of a polynomial P is a solution of $P = 0$). In this case, we may say that the polynomial $P_a(x) = x - 1$ "changes sign at its root" $x = 1$.

EXAMPLE 2. Next let $P_b(x) = (x-1)^2$. Again 1 is a root, since $P_b(1) = 0$; but now the polynomial does not change sign at its root, in fact $P_b(x) \geq 0$ for all x, since P_b is a square. So the sign line in this case is just



In each of these examples the polynomial had only one factor, and it was easy to determine the sign line by direct inspection; usually things will be a bit more complicated. In general the first step towards finding the signs is:

- Begin by factoring the polynomial and finding its roots. Mark the roots on the x-axis; this will partition the axis into intervals between adjacent roots.

EXAMPLE 3. Let us do this for the polynomial

$$P_c(x) = x^3(x+1)^2(x-1).$$

The factors are x, $x+1$, and $x-1$. $P_c = 0$ if and only if a factor $= 0$. So the roots of this polynomial are, respectively, 0, -1 and 1. These roots partition the x-axis into the 4 intervals

$$(-\infty, -1), \qquad (-1, 0), \qquad (0, 1), \qquad (1, \infty).$$

Next we shall use the fact (to be proved below) that the sign of a polynomial is constant throughout each interval between adjacent roots; that is, "a polynomial can change sign only at a root." There is no need to look elsewhere for sign changes. Thus in the previous example, the sign of $P_c(x) = x^3(x+1)^2(x-1)$ has to be constant in each of the 4 intervals determined by its roots.

Armed with this fact we can find the signs of a polynomial P by either of the following 2 methods, once we have marked the intervals between adjacent roots.

- Find the sign of $P(x)$ where x is large positive; this gives the sign in the right hand interval. Then work down the line from right to left, checking each root in turn to see whether P changes sign at it. As in the first example, if the factor corresponding to a root has *odd* exponent, that factor will change sign at the root; the other factors won't change sign, so P will change sign there, too. But if the exponent is *even*, as in the second example, no factor will change sign, so P will not change sign there. And in any case, no signs will change in between adjacent roots.

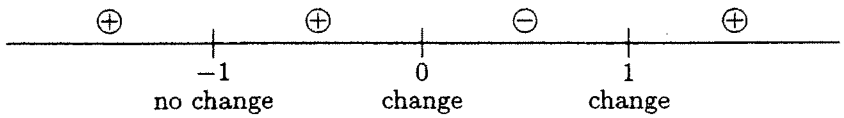
- Or, inside each of the intervals between adjacent roots, evaluate P at any test point. The sign of the value at the test point is the sign of all the values of P in that interval, so mark the interval with that sign.

EXAMPLE 4. Let us again take $P_c(x) = x^3(x+1)^2(x-1)$, and apply each method to find its signs.

First method. The factors of P_c are x, $x+1$ and $x-1$, as we said, and the corresponding roots are 0, -1 and 1. The sign of P_c changes at the roots $x=0$ and $x=1$, because the factors x and $x-1$ each occur with *odd* exponent; the sign does not change at $x=-1$ because the factor $x+1$ occurs with *even* exponent. Furthermore, when x is large — for instance, larger than 1 — all 3 factors are positive, so then $P_c(x)$ is positive.

Therefore, starting at the right and proceeding towards the left, we first put $\oplus$ in the interval $(1,+\infty)$; then we put $\ominus$ in the interval $(0,1)$ because the sign does change as we pass $x=1$; then we put $\oplus$ in the interval $(-1,0)$ because the sign changes again at $x=0$; finally we put another $\oplus$ in the interval $(-\infty,-1)$ because the sign does not change at $x=-1$.

So the sign line of $x^3(x+1)^2(x-1)$ is as shown.



Second method. You can get the same result by popping in test points, one from each interval. For instance, we find $P(2) = 72$, $P(1/2) = -9/64$, $P(-1/2) = 3/64$, and $P(-2) = 24$. The signs of these values agree with what we found above.

Of course, a root can't be a test point. So if consecutive roots are also consecutive integers, as in this example, a test point between them has to be a fraction.

Now we prove that a polynomial can change sign only at a root. Since polynomials are continuous, this is implied by the following result.

PROPOSITION. *A function can change sign only at a root or discontinuity.*

Proof. It suffices to show: If f is continuous in (a,b) and has no root there, then sign f is constant in (a,b). Otherwise, we would have $f(x_1) < 0 < f(x_2)$ for some x_1, x_2 in (a,b). According to the Intermediate Value Theorem, this would imply that $f(c) = 0$ for some c between x_1 and x_2, hence in (a,b). But that would contradict the assumption that f has no root in (a,b). So sign f is constant as asserted.

EXERCISES.

1. Find the sign lines of the polynomials.

   a) $x^3 - x$ *b)* $x^4 - x^2$ *c)* $4x - x^3$ *d)* $x^2(x^2-4)$ *e)* $x^3(x^2-4)$.

2. *a)* Use your answer to 1*a*) to solve the inequality $x^3 - x > 0$.
   b) Now use your answer to 2*a*) to solve the inequality $x^3 > x$.
   c) In the same manner, using your answer to 1*c*), solve the inequality $4x \le x^3$.