MATH 181 NOTES: FEBRUARY 14, 2024

JESSE LEO KASS

We are going to leave ancient Greece and return to nineteenth century America. In many ways, mathematics at a nineteenth century American university is recognizable although on a much more modest scale than we are use to. At a typical college, all the mathematics classes were taught by a single professor, and the topics covered were largely a mix of what is taught today in high school and in introductory college math classes. However, a close look at the algebra curriculum shows some topics are rarely taught today. Take a look at Figure 1 which shows part of the table of contexts from Ficklin's *Complete Algebra*. The last chapter of the cover "The Theory of Equations." The section includes a discussion of "Theorem of Descartes." This is Descartes' rule of signs which we saw earlier. Some of the other topics look unfamiliar. Both Sturm's Theorem and Horner's Method are still studied in computational math and parts of computer science. In this lecture and the next few ones, we will take a look at these topics and trace back their history.

Both Sturm's Theorem and Horner's Method are techniques for estimating the solutions to a polynomial equation $f(x) = 0$. Let's begin by describing the modern statement of Sturm's theorem.

Sturm's Theorem concerns a polynomial $f(x)$ that has no repeated roots (so there is no r such that $f(r) = f'(r) = 0$). Using long division, divide $f(x)$ by $f'(x)$ and let $f_1(x)$ be the *negative* of the remainder. Thus

$$f(x) = f'(x) \cdot q_1(x) - f_1(x).$$

CHAPTER XXIII.

THEORY OF EQUATIONS.

FIGURE 1. Part of the table of contents of Ficklin's textbook

Then define f_2 by dividing $f'(x)$ by $f_1(x)$ and so on until a remainder zero appears. Thus the polynomials satisfy

$$\begin{aligned}
f(x) =& f'(x)q_1(x) - f_1,\\
f'(x) =& f_1(x)q_2(x) - f_2,\\
f_1(x) =& f_2(x)q_3(x) - f_3,\\
f_2(x) =& f_3(x)q_4(x) - f_4,\\
&\cdots\cdots\\
f_{n-2}(x) =& f_{n-1}(x)q_n(x) - f_n,\\
f_{n-1}(x) =& f_n(x)q_n(x).
\end{aligned}$$

The polynomials $f(x), f'(x), f_1(x), \ldots, f_n(x)$ form the Sturm sequence associated to $f(x)$. For a number $r \in \mathbb{R}$, let $V_f(r) = V(r)$ denote the number of times the sign varies in the sequence $f(r), f'(r), f_1(r), \ldots, f_n(r)$. The significance of the variation is described by the following theorem of Sturm:

Theorem 1. *If* $r < s$, *then* $V(r) - V(s)$ *equal the number of roots which lie in the interval* (r, s).

Let's consider the simple example of $f(x) = (x-2)(x+1) = x^2+x-2$. Then $f'(x) = 2x+1$ and polynomial division shows

$$f(x) = f'(x) \cdot (1/2x - 1/4) - 7/4.$$

Thus the Sturm sequence is:

$$f(x) = x^2 + x - 2, f'(x) = 2x + 1, f_1(x) = 7/4.$$

Applying Sturm's theorem in this case is somewhat silly because we already know the roots (they are $x = 2, -1$), but let's illustrate how it used. Figure 1 is a table of the signs at various values.

x	$f(x)$	$f'(x)$	$f_1(x)$
-5	+	-	+
0	-	+	+
5	+	+	+

TABLE 1. Sturm's theorem for $x^2 + x - 2$

From the table, we get that

$$\begin{aligned}
V(-5) =& 2,\\
V(0) =& 1,\\
V(5) =& 0.
\end{aligned}$$

Because $V(-5) - V(5) = 2$, $f(x)$ must have two roots that lie between -5 and $+5$. In fact, one lies below 0 and one above since $V(-5) - V(0) = V(0) - V(5) = 1$. In this example, we could have verified this directly using the quadratic formula. The theorem is more interesting when the polynomial $f(x)$ is more complicated.

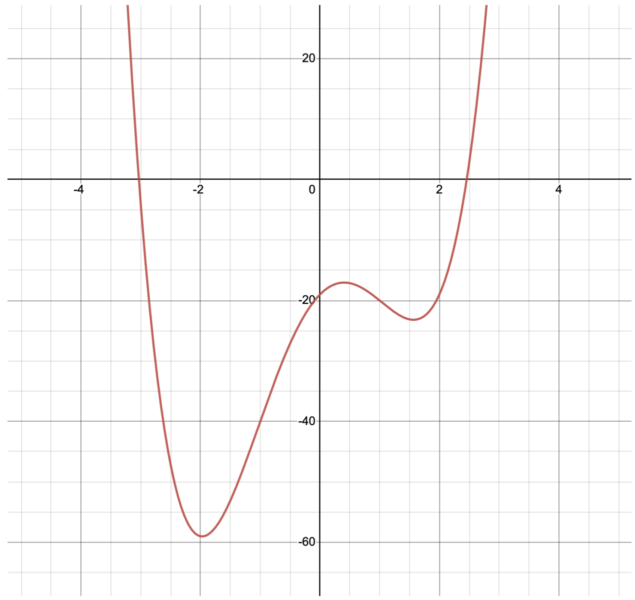


FIGURE 2. Graph of $2x^4 - 13x^2 + 10x - 19$

An example from the Charles Davies textbook that we looked at earlier is $f(x) = 2x^4 - 13x^2 + 10x - 19$. The graph is displayed in Figure 2.

It looks like the polynomial has two zeros, one which is positive and one which is negative. We can verify this using Sturm's Theorem. We compute

$$\begin{aligned}
f(x) =& 2x^4 - 13x^2 + 10x - 19, \\
f'(x) =& 8x^3 - 26x + 10, \\
f_1(x) =& (13x^2)/2 - (15x)/2 + 19, \\
f_2(x) =& (6546x)/169 + 2870/169 \\
f_3(x) =& -252148676/10712529.
\end{aligned}$$

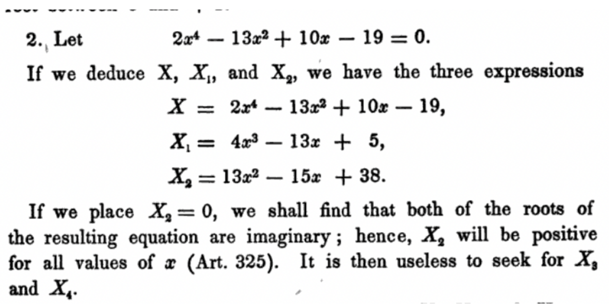
r	$f(r)$	$f'(r)$	$f_1(r)$	$f_2(r)$	$f_3(r)$
-100	+	-	+	-	-
0	-	+	+	+	-
+100	+	+	+	+	-

TABLE 2. Sturm's theorem for $2x^4 - 13x^2 + 10x - 19$

From this, we deduce that

$$\begin{aligned}
V(-100) =& 3, \\
V(0) =& 2, \\
V(+100) =& 1.
\end{aligned}$$

We have confirmed what we say from the graph. The polynomial has two real roots, one positive and one negative.



2. Let $2x^4 - 13x^2 + 10x - 19 = 0.$

If we deduce X, X_1, and X_2, we have the three expressions

$$X = 2x^4 - 13x^2 + 10x - 19,$$
$$X_1 = 4x^3 - 13x + 5,$$
$$X_2 = 13x^2 - 15x + 38.$$

If we place $X_2 = 0$, we shall find that both of the roots of the resulting equation are imaginary; hence, X_2 will be positive for all values of x (Art. 325). It is then useless to seek for X_3 and X_4.

FIGURE 3. Davies's computation

Davies's computation is displayed in Figure 3. He simplifies the computation in two ways. First, he simplifies some of the polynomials by multiplying by a positive constant: $X_1 = 1/2 \cdot f'$ and $X_2 = 2 \cdot f_1$. He also doesn't compute f_2. The reason for this is that polynomial is always nonnegative, as can be seen from the graph or the algebraic expression $f_1 = 13/2 \cdot ((x - 15/26)^2 + 23/26)$. It follows that the number of sign changes in $f_1(r), f_2(r), f_3(r)$ is always 1, so we can ignore f_2 and f_3 when computing $V(r) - V(s)$.

We haven't discussed why Sturm's Theorem is true. The key properties of the polynomials $f, f_1, \ldots, f_n$ are the following:

(1) if $f(r) = 0$, then $f'(r)$ and $f_1(r)$ have the same sign;
(2) if $f_i(r) = 0$, then $f_{i-1}(r)$ and $f_{i+1}(r)$ have opposite signs (and are nonzero);
(3) $f_n(x)$ is a nonzero constant.

Let's call any sequence of polynomials satisfying these conditions a **Sturm sequence**.

These properties hold by construction. For example, consider the second condition. Plugging r into the equations defining the polynomials $f_i(x)$, we have

$$\begin{aligned} f_{i-1}(r) =& f_i(r)q_i(r) - f_{i+1}(r) \\ =& - f_{i+1}(r) \end{aligned}$$

If $f_{i-1}(r) = f_{i+1}(r) = 0$, then we would have that

$$\begin{aligned} f_{i-2}(r) =& f_{i-1}(r)q_{i-1}(r) - f_i(r) \\ =& 0. \end{aligned}$$

Proceedings in this manner, we would get that $f(r) = f'(r) = 0$, contradicting the assumption that $f(r)$ has no multiple roots.

Sturm's theorem in fact holds for any Sturm sequence. This explains why Davies's computation is valid. If $f_i(x)$ is part of a Sturm sequence, then so is $c \cdot f_i(x)$ for some positive scalar c.

Now let's see why Sturm's theorem must hold for any Sturm sequence. Consider what happens if we take $x = r$ to be very negative and then slowly increase its value. Being polynomials, the signs of $f(x), f_1(x), \ddots, f_n(x)$ can only change when some x passes through a root. The content is that $V(x) - V(r)$ only changes when x passes through a

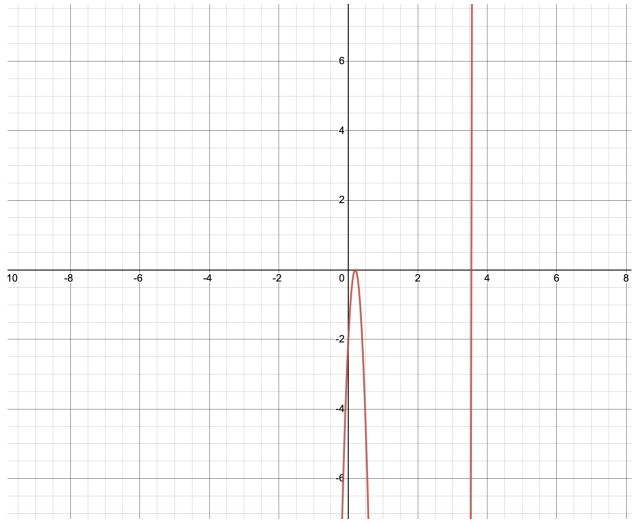


FIGURE 4. Graph of $x^5 - 50x^2 + 20x - 2$.

root of $f(x)$. When $f(s) = 0$, all the other polynomials be nonzero, and $f(s)$ must change sign (since the polynomial has only simple roots). Then other case to consider is where $f_i(s) = 0$. Without loss of generality, we can assume that $f_i(x)$ changes from positive to negative as x passes through s. The partial sign sequence $f_{i-1}(x), f_i(x), f_{i+1}(x)$ then either changes from $+, +, -$ to $+, -, +$ or from $-, +, +$ to $-, -, +$. In both cases, the total number of sign changes is zero. This is just a short sketch of an argument. Below I include the justifications offered by Davies and Ficklin in their textbooks. I encourage you to read what they wrote and see if it is a complete proof.

Why was Sturm's theorem included in the algebra textbooks? It is an important algorithm or method for **real root isolation**. The general idea is that, given a polynomial $f(x)$, we want to find intervals $\{(r_1, s_1), (r_2, s_2), \ldots, (r_n, s_n)\}$ such that each real root lies in exactly one interval. In other words, the intervals isolate the roots. Finding these intervals is the first step when applying many algorithms for computing real roots. In the examples we have seen so far, we can find suitable intervals by examining the graph of $f(x)$. For example, when $f(x) = 2x^4 - 13x^2 + 10x - 19$, it is clear from the graph that the intervals $(-4, -2)$ and $(2, 4)$ work. A more realistic example is the Mignotte polynomial $x^5 - 50x^2 + 20x - 2$.

The graph of the Mignotte polynomial is shown in Figure 4. It looks like the polynomial has two real roots: a root lying between 3 and 4 and a positive double root that is close to zero. In fact, the polynomial does not have a double root. Rather, it has two roots near zero that are very close together. Separating the two roots is exactly what Sturm's theorem is designed to do. Of course, there are example that are much more complicated than $x^5 - 50x^2 + 20x - 2$. Examples that occur in engineering and scientific applications can exhibit similar behavior but have degree 100 or larger.

In a later lecture, we'll see how Sturm's theorem can be combined with Horner's method to approximate the roots of a polynomial like $x^5 - 50x^2 + 20x - 2$.

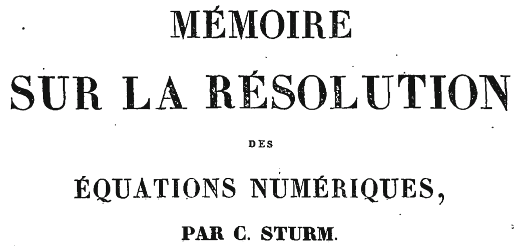

MÉMOIRE
SUR LA RÉSOLUTION
DES
ÉQUATIONS NUMÉRIQUES,
PAR C. STURM.

FIGURE 5. The title of Sturm's paper.

So much for the mathematics behind Sturm's theorem. What about the history? I'm actually not entirely clear on why Davies and Ficklin present Sturm's theorem. Few, if any, of the American college students who used their textbooks went on to do the sort of difficult numerical computations that would require an application of Sturm's theorem. In general, US higher education at the time focused on rote learning, memorization, and accurately following instructions. When it was taught, I'm guessing Sturm's theorem was viewed as simply a complicated procedure to challenge students, not that different from, say, conjugating verbs in ancient Greek. Its inclusion in American algebra textbooks is likely an artifact of textbook authors following the model set by the French mathematics textbook author M. Bourdon. (Recall looked at his textbook when we studied Descartes' Rule of Signs.)

In France, and Europe more generally, the situation with Sturm's theorem was very different. Mathematical and scientific research was being done a very high level, and there certainly was a need to provide students with deep training in solving polynomial equations numerically. At the time, Sturm's theorem was a relatively recent technique. The theorem was indeed proved by the French mathematician C. Sturm in 1829. The title of the paper is displayed as Figure 5.

Sturm's theorem can still be found in textbooks today, but whole topic of numerically computing roots to polynomial equations has been transformed by the widespread use of computers. Today, no serious computation of the roots of a polynomial equation is done by hand, and as anyone who has taken an algorithms class can tell you, algorithms that run efficiently on a computer are often very different from those that can easily be run "by hand." Sturm's theorem is relatively inefficient for the purposes of computer computation, so while it is mentioned in textbooks (proving it is an exercise in Donald Knuth's *Art of Computer Programming*), it does not play a central role.

A worthwhile activity would be to closely read Sturm's paper, compare it with the treatment of his theorem in American textbooks, and see how the treatment evolved during the twentieth century as computer use became widespread. However, Sturm's paper is tough going: it was written in French for research mathematicians. Rather than doing this, we are going to take a look at the last topic in the textbooks by Davies and Ficklin: Horner's method. For us, Horner's work has one major advantage over Sturm's: it is written in English. We'll see what he said in the next lecture.

DEPT. OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT SANTA CRUZ