THE BERLEKAMP-WELCH ALGORITHM

Suppose that we have a polynomial P of fairly low degree over a finite field $\mathbb{F}$. The data is corrupted, leaving a function $F : \mathbb{F} \to \mathbb{F}$, and we know that $F(x) = P(x)$ for a certain fraction of $x \in \mathbb{F}$. We want to understand whether we can recover P from F with an efficient algorithm. In particular, our main goal is to explain the Berlekamp-Welch algorithm. We present here the following case of the Berlekamp-Welch result. (We're not trying to be as general as possible, but just to get a flavor of the subject and see the key ideas in the proof.)

Theorem 0.1. *(Berlekamp-Welch, 1986) Suppose that*

- $\mathbb{F}$ *has* q *elements,*
- *the degree of* P *is* $< q/100$*, and*
- $F(x) = P(x)$ *for at least* $(51/100)q$ *values of* x*.*

Under these assumptions, there is an efficient algorithm to recover P *from* F*.*

The ideas depend on the following elementary but fundamental vanishing lemma for polynomials.

Lemma 0.2. *If* $P(x)$ *is a polynomial of degree* $\leq D$*, and* P *vanishes at* $D+1$ *distinct points, then* P *is the zero polynomial.*

We will recall the proof of this lemma later in the lecture.

One application is that a polynomial P of degree D can be recovered if we know its values at any $D+1$ points. Let S be a set of $D+1$ points, $S = \{x_1, ..., x_{D+1}\} \subset \mathbb{F}$. Let $V_1(D)$ be the vector space of polynomials in one variable of degree $\leq D$. Consider the evaluation map E_S which evaluates a given polynomial at the points of S:

$$E_S(P) := (P(x_1), ..., P(x_{D+1})) .$$

The evaluation map $E_S : V_1(D) \to \mathbb{F}^{D+1}$ is a linear map. By the vanishing lemma, the kernel of E_S is zero. Since the domain and range have the same dimension, E_S is an isomorphism. Therefore, P can be recovered from its values on the set S. Moreover, recovering P amounts to solving a system of linear equations, so it can be done efficiently.

There are at least $(51/100)q$ points x where $F(x) = P(x)$. If we knew which points they were, we could recover P by this interpolation procedure, because the degree of P is $< q/100$. The whole point is that we don't know where F is still correct and where F has been corrupted.

1. The algorithm

Now we turn to the algorithm. We will consider the graph of F. To set conventions, the graph of F is the set

$$\{(x, y) \in \mathbb{F}^2 | y = F(x)\}.$$

We are hoping to find some nice algebraic structure hidden in the graph of F. To do so, we will find a low degree polynomial $R(x, y)$ which vanishes on the graph of F. It turns out to be a good idea to consider polynomials of the form $R(x, y) = R_0(x) + R_1(x)y$. We may/will talk more about this choice later. Let's define $W(d)$ to be the vector space of polynomials of the form $R_0(x) + R_1(x)y$ where R_0 and R_1 have degree $\leq d$. The dimension of $W(d)$ is $2d + 2$. The graph of F has q elements. As long as $2d + 2 > q$, there is a non-zero polynomial in $W(d)$ which vanishes on the graph of F. In particular, there is such a polynomial of degree $d \leq q/2$. Finding such a polynomial just involves linear algebra, so we can find it in polynomial time. In fact, with a little more work we can find a polynomial that vanishes on the graph of F of minimal degree. Let us define $R(x, y)$ to be a lowest-degree polynomial of the form $R(x, y) = R_0(x) + R_1(x)y$ that vanishes on the graph of F. We know that the degree of R_0 and R_1 is $\leq q/2$.

The key observation in the whole argument is that R also vanishes on the graph of P.

Claim 1.1. *The polynomial R vanishes on the graph of P. In fact, $R(x, P(x))$ is the zero polynomial.*

Proof. We know that R vanishes on the graph of F.

Therefore, $R(x, F(x)) = 0$ for all x.

Since $F(x) = P(x)$ for most x, we see that $R(x, P(x)) = 0$ for at least $(51/100)q$ values of x.

Now $R(x, P(x)) = R_0(x) + R_1(x)P(x)$ is a polynomial in x of degree $< q/2 + q/100 = (51/100)q$. By the vanishing lemma, this polynomial is identically zero. □

We can now describe how to recover the polynomial P. We just proved that $R_0(x) + R_1(x)P(x)$ is the zero polynomial. In other words, $R_1(x)P(x) = -R_0(x)$. So R_0 is divisible by R_1, and P is equal to $-R_1/R_0$.

This finishes the BW algorithm, but it's interesting to explore a little more about the minimal degree polynomial $R(x, y)$.

We let $E \subset \mathbb{F}$ be the set $\{e \in \mathbb{F} | F(x) \neq P(x)\}$. We call E the set of error locations. It turns out that the zero set of R is exactly the graph of P together with a vertical line $\{x = e\}$ for each error location e. (Picture?)

Claim 1.2. *For each $e \in E$, $R(x, y)$ vanishes on the line $x = e$.*

Proof. Fix $e \in E$. We consider $R(e, y) = R_0(e) + R_1(e)y$. We want to prove that $R(e, y)$ is the zero polynomial in y - in other words that $R_0(e) = R_1(e) = 0$. We know that $R(e, F(e)) = 0$ and $R(e, P(e)) = 0$. Since $F(e) \neq P(e)$, we see that the linear polynomial $R_0(e) + R_1(e)y$ vanishes at two different values of y. So the linear polynomial must vanish. □

In fact, we can say exactly what the minimal degree polynomial $R(x, y)$ is.

Claim 1.3. *$R(x, y) = c\,[y - P(x)] \prod_{e \in E}(x - e)$, for some non-zero constant $c \in \mathbb{F}$.*

From this last claim it follows that R vanishes exactly on the graph of P together with the vertical lines at the error locations. From this information, we can easily identify the set E, giving another way to recover the polynomial P.

The proof of this claim involves a fundamental idea/argument. The argument first appears in the proof of the vanishing lemma, so we begin by giving this proof. Then we develop the idea a bit further to prove a divisibility lemma which leads to the claim.

2. Vanishing and divisibility lemmas

Vanishing lemma. If $P(x)$ is a polynomial of degree $\leq D$, and P vanishes at $D+1$ distinct points, then P is the zero polynomial.

Proof of the vanishing lemma. We go by induction on D. The case $D = 0$ is trivial. The heart of the matter is in the following divisibility lemma.

Lemma 2.1. *If $P(x)$ is any polynomial and $P(x_1) = 0$ for some $x_1 \in \mathbb{F}$, then $P(x) = (x - x_1)P_1(x)$ for some polynomial P_1.*

Proof. Suppose $P(x) = \sum_{j=0}^{D} a_j x^j$. We can write any degree D polynomial P in the following form:

$$P(x) = (x - x_1)(b_{D-1}x^{D-1} + ... + b_0) + r.$$

To see this, first we choose the coefficient b_{D-1} in order to get the x^D term correct. None of the lower coefficients influence the x^D term, so we are still free to choose them. Next, we choose b_{D-2} to get the x^{D-1} term correct, etc. We choose b_0 to get the x term correct, and we choose r to get the units term correct.

But now, since $P(x_1) = 0$, we must have $r = 0$, and our factoring is done. □

We return to the vanishing lemma. Suppose that P vanishes at $x_1, ..., x_{D+1}$ distinct points. By the divisibility lemma, we see that $P(x) = (x - x_1)P_1(x)$, where $P_1(x)$ has degree $\leq D - 1$. But P_1 must vanish at $x_2, ..., x_{D+1}$. By induction, $P_1 = 0$, and we are done. This finishes the proof of the vanishing lemma.

With the same proof idea, we can prove a simple divisibility lemma for polynomials in two variables.

Lemma 2.2. *If $R(x, y)$ is a polynomial of two variables, and $P(x)$ is a polynomial in one variable, and $R(x, P(x))$ is the zero polynomial, then $R(x, y) = (y-P(x))R_1(x, y)$ for some polynomial R_1.*

Proof. Let $R(x, y) = \sum_{j=0}^{D} a_j(x)y^j$, where $a_j(x)$ is a polynomial in x. Now, we can write any polynomial $R(x, y)$ in the following form:

$$R(x, y) = (y - P(x))(b_{D-1}(x)y^{D-1} + ... + b_0(x)) + r(x),$$

where the $b_j(x)$ and $r(x)$ are polynomials in x. The proof is basically the same as above. First we choose the polynomial $b_{D-1}(x)$ in order to get the y^D term correct. None of the lower coefficients influence the y^D term, so we are still free to choose them. Next, we choose b_{D-2} to get the y^{D-1} term correct, etc. We choose b_0 to get the y term correct, and we choose r to get the units term correct.

But $R(x, P(x))$ is $r(x)$, so $r(x)$ is the zero polynomial. This gives the required factoring of $R(x, y)$. □

As a corollary, we can quickly prove the last claim about the polynomial $R(x, y)$ in the Berlekamp-Welch algorithm. We know that $R(x, P(x))$ is the zero polynomial, so $R(x, y) = (y - P(x))R_1(x, y)$. Because $R(x, y)$ has degree 1 in y, it follows that R_1 must have degree 0 in y: in other words, $R_1 = R_1(x)$ is a polynomial in x only. So $R(x, y) = (y - P(x))R_1(x)$. At each $e \in E$, $R(x, F(x))$ vanishes, but $F(x) - P(x)$ doesn't, and so $R_1(e) = 0$. Using the divisibility lemma in one variable, we see that $R(x, y) = (y - P(x)) \prod_{e\in E}(x - e)R_2(x)$. Any polynomial of this form vanishes on the graph of F. Since R is a polynomial of minimal degree, it follows that $R_2(x)$ is just a constant $c \neq 0$.

Our last divisibility lemma is closely related to Bezout's theorem. The formulation of the last lemma depended on the special form of the polynomial $y - P(x)$, but Bezout's theorem says something similar about two arbitrary polynomials. Here is one formulation of Bezout's theorem.

Theorem 2.3. *Suppose that $P(x, y)$ and $Q(x, y)$ are polynomials. Let $Z(P, Q)$ be the set of common zeroes of P and Q. In other words,*

$$Z(P, Q) := \{(x, y) \in \mathbb{F}^2 | P(x, y) = Q(x, y) = 0\}.$$

Then either

1. *$Z(P, Q)$ has at most $(degP)(degQ)$ points, or*
2. *P and Q have a non-trivial common factor. In other words, $P = R(x, y)P_1(x, y)$ and $Q = R(x, y)Q_1(x, y)$ for some polynomial $R(x, y)$ with degree ≥ 1.*

This is an important theorem that we will prove and discuss more during the course. With a little extra trick, it recovers the last divisibility theorem as a special case. It's an interesting problem to try to prove the Bezout theorem by generalizing the last proof.

The arguments above are also related to the proof that there is unique factorization in the ring of polynomials $\mathbb{F}[x_1, ..., x_n]$ for any number of variables.

3. Correcting polynomials from badly corrupted data

In the Berlekamp-Welch algorithm, we considered corrupted data F which was correct a little more than half the time. If F is correct only half the time, then it's impossible to recover the polynomial P even in theory. For example, start with two low degree polynomials P_1 and P_2, and arrange for F to agree with P_1 half the time and with P_2 half the time. There is no way to tell if the original polynomial was P_1 or P_2. Following this observation, it may seem that data F which is correct only 1 % of the time would not be very useful. Surprisingly, it turns out that a great deal of information can be recovered from such data. In the mid 90's, Sudan generalized the algorithm of Berlekamp-Welch to deal with highly corrupted data. For example, he proved the following result.

Theorem 3.1. *(Sudan, 1997) Suppose that $\mathbb{F}$ is a field with q elements, and that $F : \mathbb{F} \to \mathbb{F}$ is any function. There is an efficient algorithm that lists all the polynomials of degree $\leq (1/200)q^{1/2}$ that agree with F at least 1 % of the time.*

We have the tools to follow most of the steps of Sudan's argument. We again consider the graph of F in $\mathbb{F}^2$. We find a low-degree polynomial $Q(x, y)$ that vanishes on the graph. This time we consider all the polynomials of two variables. If we let $V(d)$ be the space of polynomials in two variables of degree $\leq d$, then the dimension of $V(d)$ is $\binom{d+2}{2}$. The graph of F has q elements. As long as $\binom{d+2}{2} > q$, we can find a polynomial $Q(x, y)$ of degree $\leq d$ that vanishes on the graph. So we can find a non-zero Q with degree $d \leq 2q^{1/2}$.

Suppose that P has degree $\leq (1/200)q^{1/2}$, and that $P(x) = F(x)$ for at least $q/100$ values of x. We claim that $Q(x, P(x))$ is the zero polynomial. This follows for the same reason as above. We know that $Q(x, F(x))$ is zero for every x. So $Q(x, P(x))$ has at least $q/100$ zeroes. But $Q(x, P(x))$ is a polynomial of degree at most $(degQ)(degP) < 2q^{1/2}(1/200)q^{1/2} = q/100$. Therefore $Q(x, P(x))$ is identically zero.

By the divisibility lemma in the last section, we see that $y - P(x)$ divides $Q(x, y)$. There is a polynomial time algorithm that factors Q into irreducible factors. This step is not at all obvious, and it requires different ideas. Now we can recover all the good polynomials P by examining the factors of Q for factors of the form $(y - P(x))$.