Newton Iteration and Algebraic Independence

1 Introduction

In this document, we look at a proof of part of the PSS criterion via Newton Iteration. This vastly simplifies the proof. It also gives an idea of why the random shift is necessary, and the role of the inseparability.

Let $\mathbb{F}$ denote the field $\overline{\mathbb{F}}_p$, the algebraic closure of the finite field of characterstic p. We use vector notation to represent sets of objects: for example $\mathbf{x}$ denotes the variables $x_1, \ldots, x_n$. We use the notation $\mathbf{x}^{\mathbf{e}}$ the monomial $x_1^{\mathbf{e}_1} \cdots x_n^{\mathbf{e}_n}$.

Assume that we are given polynomials $f_1, \ldots, f_n \in \mathbb{F}[\mathbf{x}]$. Assume that the $\mathbf{f}$ have transcendence degree n. Let g be any polynomial in $\mathbb{F}[\mathbf{x}]$. We know that the transcendence degree of $\{\mathbf{f}, g\}$ will also be n, and thus g depends algebraically on the $\mathbf{f}$. Let A be an annihilator of $\{\mathbf{f}, g\}$. By definition, $A(\mathbf{f}, g) = 0$. We will also look at A as a polynomial in one variable, say y, which is the variable in which we plug in g. We assume that A is an annihilator with minimum degree in y. We will show that after a random shift, we can write g as a power series in $\mathbf{f}$.

2 Proof

The key idea will be to use Newton iteration (NI). We use the following result, which is a slight modification of Theorem 2.31 from Bürgisser, Lickteig, Clausen, and Shokrollahi [1996]. This formulation is from Dutta, Saxena, and Sinhababu [2018]. For completeness, we provide a proof of this lemma in the last section.

Lemma 2.1 (Newton Iteration)**.** *Let $F(\mathbf{x}, y) \in \mathbb{F}[[\mathbf{x}]][y]$ be a polynomial in y with coefficients power series in $\mathbb{F}[[\mathbf{x}]]$. Suppose μ is such that $F(\mathbf{0}, \mu) = 0$ and also $F'(\mathbf{0}, \mu) \neq 0$, then there is a unique element $Y \in \mathbb{F}[[\mathbf{x}]]$ with constant term μ such that $F(\mathbf{x}, Y) = 0$. We also have*

$$y_{t+1} = y_t - \frac{F(\mathbf{x}, y_t)}{F'(\mathbf{x}, y_t)},$$

such that $Y \equiv y_t \pmod{\langle \mathbf{x} \rangle^{2^t}}$.

In the above lemma, it is essential that F is a polynomial in y. This allows us to evaluate it at power series with non-zero constant term. In general, if F were a power series in y, we would require $\mu = 0$.

In order to get g as a power series in $\mathbf{f}$, we will try and use NI to get a power series in $\mathbf{x}$, and then try and argue that what we get is actually a power series in $\mathbf{f}$. First note that if g depended inseparably on $\mathbf{f}$, then A' would be an identially zero polynomial. In this case, we will not be able to satisfy the conditions of the lemma. Thus we assume that g depends separably on $\mathbf{f}$. In general, this can be obtained by replacing g by a power g^{p^i}. Therefore we can assume now that g depends separably on $\mathbf{f}$.

A possibly bigger issue is the fact that if the f_i have non-zero constant terms, then power series in f_i are not valid elements in $\mathbb{F}[[\mathbf{x}]]$. To fix this, we apply the shift operator, to remove the constant term: Define $\mathcal{H}f_i := f_i(\mathbf{x}+\mathbf{z}) - f_i(\mathbf{z})$, and similarly for $\mathcal{H}g = g(\mathbf{x}+\mathbf{z}) - g(\mathbf{z})$. For now we treat the $\mathbf{z}$ as part of the base field, that is, we switch from working with $\mathbb{F}$ to working with $\mathbb{F}(\mathbf{z})$. Eventually we show that we can replace $\mathbf{z}$ by an arbitrary element from $\mathbb{F}^n$, and the proof will continue to hold. We have $A(\mathcal{H}\mathbf{f} + \mathbf{f}(\mathbf{z}), \mathcal{H}g + g(\mathbf{z})) = A(\mathbf{f}(\mathbf{x}+\mathbf{z}), g(\mathbf{x}+\mathbf{z})) = 0$. We define $B(\mathbf{x}, y) = A(\mathcal{H}\mathbf{f} + f(\mathbf{z}), y + g(\mathbf{z})) = A(\mathbf{f}(\mathbf{x}+\mathbf{z}), y + g(\mathbf{z}))$. The polynomial B has root $y = \mathcal{H}g$. Now note that $B(\mathbf{0}, 0) = A(\mathbf{f}(\mathbf{z}), g(\mathbf{z})) = 0$, since A is an annihilator[1] . Further, consider $B'(\mathbf{0}, 0)$. We have

$$B = \sum_{i=0}^{d} c_i \left(y + g(\mathbf{z})\right)^i ,$$

where the c_i are polynomials in $\mathbf{x}$ and $\mathbf{z}$, and d is the degree with respect to y. Differentiating, we get

$$B' = \sum_{i=0}^{d} i c_i \left(y + g(\mathbf{z})\right)^{i-1} .$$

When evaluated at $(\mathbf{0}, 0)$, each of the c_i is a polynomial in $f_i(\mathbf{z})$. Therefore, $B'(\mathbf{0}, 0)$ is a polynomial in $f_i(\mathbf{z})$ and $g(\mathbf{z})$, of degree $d - 1$. As a polynomial in $\mathbf{z}$, this is non-zero: if it were not, we would have an annihilator for $\mathbf{f}, g$ of degree $d - 1$ in y, contradicting the assumption that A is the annihilator with minimum y degree. In general, when we replace $\mathbf{z}$ with a vector of random elements from $\mathbb{F}$, we can still say that $B'(\mathbf{0}, 0) \neq 0$ (for most choices), by using the Schwartz Zippel lemma.

We have now satisfied the conditions of the lemma. The lemma then gives us a root $Y \in \mathbb{F}[[\mathbf{x}]]$ such that $B(\mathbf{x}, Y) = 0$. Further, this is the unique root with constant term 0. But we know that $\mathcal{H}g$ is also a root of $B(\mathbf{x}, y)$ with constant term 0. Thus it must be that $Y = g$. All that is left to show is that we can actually get Y as a power series in $\mathcal{H}\mathbf{f}$, since

[1] The choice of setting $\mu = 0$ is motivated by the fact that we know that the root $\mathcal{H}g$ has no constant term. We also know that this is not a repeated root, due to minimality and separability assumption. The calculation of $B(\mathbf{0}, 0)$ and $B'(\mathbf{0}, 0)$ thus also act as a sort of sanity check.

the lemma only promises us a power series in $\mathbf{x}$. For this, we look at the series y_t whose limit is Y. We will inductively show that y_t can be written as a power series in $\mathcal{H}\mathbf{f}$ for all t.

The base case is $t = 0$. We have $y_0 = 0$, and thus vacuously y_0 is a power series in $\mathcal{H}\mathbf{f}$. Assume inductively that y_t is a power series in $\mathcal{H}\mathbf{f}$. First consider $B(\mathbf{x}, y_t) = A(\mathcal{H}\mathbf{f} + f(\mathbf{z}), y_t + g(\mathbf{z}))$. The first argument, $\mathcal{H}\mathbf{f} + f(\mathbf{z})$ is vacuously a power series in $\mathcal{H}\mathbf{f}$, and by the inductive hypothesis, so is the second argument $y_t + g(\mathbf{z})$. Thus $B(\mathbf{x}, y_t)$ is also a power series in $\mathcal{H}\mathbf{f}$. Now consider $(B'(\mathbf{x}, y_t))^{-1}$. The term $B'(\mathbf{x}, y_t)$ is a power series in $\mathcal{H}\mathbf{f}$ by an argument similar to the one above. In this form $B'(\mathbf{x}, y_t)$ must have a nonzero constant term, since the constant term will be exactly $B'(\mathbf{0}, 0)$, which is non-zero by assumption. Thus we have $B'(\mathbf{x}, y_t) = c_0 + D(\mathcal{H}\mathbf{f})$, where $c_0 \neq 0$, and D is a power series with no constant term. But then we have

$$\begin{aligned}
\frac{1}{B'(\mathbf{x}, y_t)} &= \frac{1}{c_0 + D(\mathcal{H}\mathbf{f})} \\
&= \frac{1}{c_0}\frac{1}{1 - D_1(\mathcal{H}\mathbf{f})} && \text{(Setting } D_1 = -D/c_0\text{)} \\
&= \frac{1}{c_0}\left(1 + D_1(\mathcal{H}\mathbf{f}) + D_2(\mathcal{H}\mathbf{f})^2 + \ldots\right)
\end{aligned}$$

This converges since each $D_1(\mathcal{H}\mathbf{f})^i$ has x-adic valuation atleast i. It is also a power series in $\mathcal{H}\mathbf{f}$. The product $B(\mathbf{x}, y_t)\,(B'(\mathbf{x}, y_t))^{-1}$ is thus also a power series in $\mathcal{H}\mathbf{f}$, and so is y_{t+1}. Note that c_0 is a non-zero element in $\mathbb{F}(\mathbf{z})$, and by Schwartz Zippel, it continues to remain non-zero after we replace $\mathbf{z}$ by random field elements. It is crucial that the term c_0 is independent of t, since otherwise the random choice of $\mathbf{z}$ would have had to be such that a countable number of equations are non-zero. This completes the proof.

3 Characterisation of truncated functional dependence

Assume that the transcendence degree of the extension $\mathbb{F}(\mathbf{x})$ over $\mathbb{F}(\mathbf{f})$ is p^i. Without loss of generality, assume that x_1 is a witness for this inseparable degree. This implies that x_1 has inseparable degree exactly p^i, and no other x_j has inseparable degree greater than p^i.

The first thing we show is that if we truncate our computation and to terms of x-adic valuation smaller than or equal to $p^i - 1$, we will be able to write some $\mathcal{H}f_i$ as a polynomial function of the other $\mathcal{H}f_j$, despite the fact that the $\mathbf{f}$ are independent. By definition, $x_1^{p^i}$ depends separably and algebraically on $\mathbf{f}$. The claim now is that some f_j depends spearably and algebraically on $x_1^{p^i}, f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_n$. If this is true, then we can write $\mathcal{H}f_j$ as a power series in $\mathcal{H}x_1^{p^i}, \mathcal{H}f_1, \ldots, \mathcal{H}f_{j-1}, \mathcal{H}f_{j+1}, \ldots, \mathcal{H}f_n$. Note that $\mathcal{H}x_1^{p^i} = (x_1 + z_1)^{p^i} - z_1^{p^i} = x_1^{p^i}$. Thus if we truncated our computation to terms of valuation smaller than or equal to $p^i - 1$, the occurance of $\mathcal{H}x_1^{p^i}$ would vanish, and we would have written one of the $\mathcal{H}f_j$ as a polynomial in the others. Suppose first that $i > 0$. If all the f_i depend inseparably, then the annihilator of x_1 and the $\mathbf{f}$ will be such that all exponents are multiples of p. This will let us factor it, which is a contradiction. Suppose now that $i = 0$. In this case, the extension $\mathbb{F}(\mathbf{z})(\mathbf{x})$ is algebraic over $\mathbb{F}(\mathbf{z})(\mathbf{f})$, and thus it must be that any f_j that occurs non-trivially in the annihilator of x_1 and $\mathbf{f}$ depends separably on x_1 and the remaining $f_{j'}$.

For the second part, we will show that if we compute to precision greater than or equal to p^i, then none of the $\mathcal{H}f_j$ can be written as a polynomial function of the other $\mathcal{H}f_{j'}$. For this, assume by contradiction that at precision p^i, we can write $\mathcal{H}f_1$ as polynomial in $\mathcal{H}f_2, \ldots, \mathcal{H}f_n$. We know that for each x_j, we can write $x_j^{p^i}$ as a power series in $\mathcal{H}\mathbf{f}$, and upon truncating our calculations, as a polynomial in $\mathcal{H}\mathbf{f}$ upto precision p^i. Plugging in the polynomial for $\mathcal{H}f_1$ that exists by assumption, we are able to write each $x_j^{p^i}$ as a polynomial in $\mathcal{H}f_2, \ldots, \mathcal{H}f_n$ when truncating calculations at precision p^i. In the following paragraph, we will show that this leads to a contradiction. This will complete the proof.

For the rest of this paragraph, we will use the vector notation $\mathcal{H}\mathbf{f}$ to denote $\mathcal{H}f_2, \ldots \mathcal{H}f_n$, that is, we only consider the last $n-1$ polynomials. We have written each $x_j^{p^i}$ as a polynomial in $\mathcal{H}\mathbf{f}$ when computing with precision p^i. Note that we have also replaced f_i with the part of f_i of degree atmost p^i, since the higher degree part does not affect the equation. Suppose $x_j^{p^i} = F_j(\mathcal{H}\mathbf{f})$ when computed with precision p^i. Again note that F_j is a polynomial, not a power series. Then we have $x_j^{p^i} - F_j(\mathcal{H}\mathbf{f}) \in \langle \mathbf{x} \rangle^{p^i+1}$. For each j, let $-\alpha_j = x_j^{p^i} - F_j(\mathcal{H}\mathbf{f})$. Then we have the exact equation $x_j^{p^i} + \alpha_j = F_j(\mathcal{H}\mathbf{f})$ in the ring $\mathbb{F}(\mathbf{z})[\mathbf{x}]$. In the reverse graded lexicographic monomial ordering, the set of polynomials $\mathbf{h} := \left\{ x_j^{p^i} + \alpha_j \,|\, j \in [n] \right\}$ have leading monomials $x_1^{p^i}, x_2^{p^i}, \ldots, x_n^{p^i}$, since each α_j belongs to $\langle \mathbf{x} \rangle^{p^i+1}$. Since the leading monomials are independent, these polynomials are independent. The transcendence degree of the extension $\mathbb{F}(\mathbf{z})(\mathbf{h})$ over $\mathbb{F}(\mathbf{z})$ is thus n. The set of polynomials $F_1, \ldots, F_n$, by virtue of depending only $n-1$ polynomials are such that the degree of the extension $\mathbb{F}(\mathbf{z})(\mathbf{F})$ over $\mathbb{F}(\mathbf{z})$ is atmost $n-1$. But since each $x_j^{p^i} = F_j$, it follows that $\mathbb{F}(\mathbf{z})(\mathbf{h}) \subseteq \mathbb{F}(\mathbf{z})(\mathbf{F})$. This is a contradiction.

4 Proof of NI

Proof of lemma 2.1. In order to see the existance of Y, we plug in a power series with unknown coefficients, equate with zero, and compare coefficients on both sides. This gives us a system of linear equations, with unknowns corresponding to monomials, and equations also correspoding to monomials. In particular, let $Y = \sum c_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}$ where the sum runs over all $\mathbb{N}$ valued vectors of length n. We will first show that $c_{\mathbf{0}} = \mu$ satisfies the equation corresponding to the constant term. Then we will use the y_t described in the statement of the lemma, to get coefficients $c_{\mathbf{e}}$ in the following way: we will look at some y_t, and use the coefficients of monomials upto degree 2^t as the values for the corresponding variables in our system. We will show that these satisfy the equations corresponding to the monomials of degree atmost 2^t. Note that these equations do not have any other variables. This is equivalent to showing that $F(\mathbf{x}, y_t) \equiv 0 \pmod{\langle \mathbf{x} \rangle^{2^t}}$. When showing that the y_t satisfy these equations, we will additionally show that the values for the variables that we already had from y_{t-1}, namely those for the coefficients of degree atmost 2^{t-1}, are the same as those in y_{t-1}. More succinctly, we will show that $y_t \equiv y_{t-1} \pmod{\langle \mathbf{x} \rangle^{2^{t-1}}}$. As hinted, the proof will proceed by induction on t.

First we show the base case, namely $t = 0$. Consider the equation $F(\mathbf{x}, Y) = 0$. The constant term in this expression is $F(\mathbf{0}, c_{\mathbf{0}})$. By assumption, since $F(\mathbf{0}, \mu) = 0$, we can set $c_{\mathbf{0}} = \mu$. This also ensures we satisfy the requirement of our Y having constant term μ. In the notation of the question, we also get $y_0 = \mu$. The statement about equality of coefficients holds vacuously.

Assume now that the statement holds for t. First note that $F(\mathbf{x}, y_t) \equiv F(\mathbf{x}, y_0) \pmod{\langle \mathbf{x} \rangle}$, since $y_t \equiv y_0 \pmod{\langle \mathbf{x} \rangle}$ by the induction hypothesis. This implies that $F'(\mathbf{x}, y_t)$ has constant term $F'(\mathbf{0}, \mu)$, which is non-zero by assumption. This implies that $F'(\mathbf{x}, y_t)$ is invertible in the power series ring, and that the expression for y_{t+1} is well defined. Further, by induction, we have that $F(\mathbf{x}, y_t) \equiv 0 \pmod{\langle \mathbf{x} \rangle^{2^t}}$. This implies that $y_{t+1} - y_t \equiv 0 \pmod{\langle \mathbf{x} \rangle^{2^t}}$, proving the consistency requirement. Now we compute $P(\mathbf{x}, y_{t+1})$. For this, we will use the Taylor expansion. We have

$$\begin{aligned} F(\mathbf{x}, y_{t+1}) &= F\left(\mathbf{x}, y_t - \frac{F(\mathbf{x}, y_t)}{F'(\mathbf{x}, y_t)}\right) \\ &= F(\mathbf{x}, y_t) + \frac{F'(\mathbf{x}, y_t)}{1!}\left(-\frac{F(\mathbf{x}, y_t)}{F'(\mathbf{x}, y_t)}\right) + \frac{F''(\mathbf{x}, y_t)}{2!}\left(-\frac{F(\mathbf{x}, y_t)}{F'(\mathbf{x}, y_t)}\right)^2 + \ldots \end{aligned}$$

On the right hand side, the first two summands cancel. All other summands, and hence the entire right hand side, are $0 \pmod{\langle \mathbf{x} \rangle^{2^{t+1}}}$. This shows that y_{t+1} has the required property.

Finally we must show that Y is unique. This follows from the fact that μ is not a repeated root of $F(\mathbf{0}, y)$.

□

5 Bibliography

P. Bürgisser, T. Lickteig, M. Clausen, and A. Shokrollahi. *Algebraic Complexity Theory*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1996. ISBN 9783540605829. URL https://books.google.co.in/books?id=dYcgjfXsYk8C.

Pranjal Dutta, Nitin Saxena, and Amit Sinhababu. Discovering the roots: Uniform closure results for algebraic classes under factoring. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 1152–1165, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5559-9. doi: 10.1145/3188745.3188760. URL `http://doi.acm.org/10.1145/3188745.3188760`.