LECTURE 9. DWORK'S PROOF OF RATIONALITY OF ZETA FUNCTIONS

In this lecture we present Dwork's proof [Dwo] for the first of the Weil conjectures, asserting the rationality of the Hasse-Weil zeta function for a variety over a finite field. We follow, with small modifications, the presentation in [Kob]. We freely make use of the basic facts about p-adic fields as covered in Appendix 2.

1. A formula for the number of $\mathbf{F}_q$-points on a hypersurface

Recall that our goal is to prove the rationality of the zeta function of an algebraic variety X over $\mathbf{F}_q$. As we have seen in Lecture 3, in order to prove this in general, it is enough to prove it in the case when X is a hypersurface in $\mathbf{A}^d_{\mathbf{F}_q}$, defined by some $f \in \mathbf{F}_q[x_1, \ldots, x_d]$. Furthermore, an easy argument based an induction and on the inclusion-exclusion principle, will allow us to reduce ourselves to proving the rationality of

$$\widetilde{Z}(X,t) := \exp\left(\sum_{n\geq 0} \frac{N'_n}{n} t^n\right),$$

where

$$N'_n = |\{u = (u_1, \ldots, u_d) \in \mathbf{F}^d_{q^n} \mid f(u) = 0, u_i \neq 0 \,\text{for all}\, i\}|.$$

Hence from now on we will focus on $\widetilde{Z}(X,t)$.

The starting point consists in a formula for N'_n in terms of an additive character of $\mathbf{F}_{q^n}$. By this we mean a group homomorphism $\chi\colon \mathbf{F}_{q^n} \to \overline{\mathbf{Q}_p}$. We say that such a character is trivial if $\chi(u) = 1$ for every $u \in \mathbf{F}_q$. The main example that we will need is the following,

Lemma 1.1. *If $\varepsilon \in \overline{\mathbf{Q}_p}$ is a primitive root of 1, then $\chi\colon \mathbf{F}_{q^n} \to \mathbf{Q}_p(\varepsilon)$ given by $\chi(u) = \varepsilon^{\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(u)}$ is a nontrivial additive character of $\mathbf{F}_{q^n}$.*

Proof. It is clear that $\psi\colon \mathbf{F}_p \to \mathbf{Q}_p(\varepsilon)$ given by $\psi(m \bmod p) = \varepsilon^m$ is an injective homomorphism. Since $\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_q}$ is additive, we deduce that χ is an additive character. If χ is trivial, then $\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(u) = 0$ for every $u \in \mathbf{F}_q$. This contradicts the fact that the bilinear pairing $(u,v) \to \mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(u,v)$ is nondegenerate (recall that $\mathbf{F}_{q^n}$ is separable over $\mathbf{F}_p$). □

Remark 1.2. Since $\mathbf{F}_{q^n}/\mathbf{F}_p$ is Galois, with Galois group cyclic and generated by the Frobenius morphism, it follows that for every $a \in \mathbf{F}_{q^n}$, we have $\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(a) = a + a^p + \ldots + a^{p^{ne-1}}$, where $q = p^e$.

Lemma 1.3. *If χ is a nontrivial additive character of $\mathbf{F}_{q^n}$, then $\sum_{u\in\mathbf{F}_{q^n}} \chi(u) = 0$.*

Proof. Let $v \in \mathbf{F}_{q^n}$ be such that $\chi(v) \neq 1$. We have

$$\sum_{u \in \mathbf{F}_{q^n}} \chi(u) = \sum_{u \in \mathbf{F}_{q^n}} \chi(u+v) = \chi(v) \cdot \sum_{u \in \mathbf{F}_{q^n}} \chi(u),$$

which implies the assertion in the lemma since $\chi(v) \neq 1$. □

Suppose now that $f \in \mathbf{F}_q[x_1, \ldots, x_n]$ is as above, and ψ_n is a nontrivial additive character of $\mathbf{F}_{q^n}$. It follows from Lemma 1.3 that for every $a \in \mathbf{F}_{q^n}$, we have $\sum_{v \in \mathbf{F}_{q^n}} \psi_n(va) = 0$, unless $a = 0$, in which case the sum is clearly equal to q^n. Therefore we have

$$\sum_{u \in (\mathbf{F}_{q^n}^*)^d} \sum_{v \in \mathbf{F}_{q^n}} \psi_n(vf(u)) = N'_n q^n.$$

Since the sum of the terms corresponding to $v = 0$ is $(q^n - 1)^d$, we conclude that

$$\sum_{u \in (\mathbf{F}_{q^n}^*)^d} \sum_{v \in \mathbf{F}_{q^n}^*} \psi_n(vf(u)) = N'_n q^n - (q^n - 1)^d. \tag{1}$$

The main result of this section will be a formula for the left-hand side of (1) by applying a suitable analytic function to the Teichmüller lifts of $u_1, \ldots, u_n, v$. Furthermore, the analytic functions corresponding to the various n will turn out to be related in a convenient way. Let us fix a primitive root ε of 1 of order p in $\overline{\mathbf{Q}_p}$. For every $a \in \mathbf{F}_{p^m}$, we denote by $\widetilde{a} \in \mathbf{Z}_p^{(m)}$ the Teichmüller lift of a (see Appendix, §2). The key ingredient is provided by a formal power series $\Theta \in \mathbf{Q}_p(\varepsilon)[\![t]\!]$, that satisfies the following two properties:

P1) The radius of convergence of Θ is > 1.
P2) For every $n \geq 1$ and every $a \in \mathbf{F}_{q^n}$, we have

$$\varepsilon^{\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(a)} = \Theta(\widetilde{a})\Theta(\widetilde{a}^q) \cdots \Theta(\widetilde{a}^{q^{n-1}}). \tag{2}$$

Note that by Lemma 1.1, the left-hand side of (2) is a nontrivial character of $\mathbf{F}_{q^n}$. Furthermore, note that if $a \in \mathbf{F}_{q^n}$, then $|\widetilde{a}|_p = 1$, hence $\Theta(\widetilde{a}^{q^i})$ is well-defined by P1).

Let us assume for the moment the existence of such Θ, and let us see how we can rewrite the left-hand side of (1). Suppose that $f = \sum_{m \in \mathbf{Z}^d_{\geq 0}} c_m x^m \in \mathbf{F}_q[x_1, \ldots, x_d]$, where for $m = (m_1, \ldots, m_d)$ we put $x^m = x_1^{m_1} \cdots x_d^{m_d}$. Note that only finitely many of the c_m are nonzero. It is clear that for $u = (u_1, \ldots, u_d) \in (\mathbf{F}_{q^n}^*)^d$ and $v \in \mathbf{F}_{q^n}^*$, we have

$$\psi_n(vf(u)) = \prod_{m \in \mathbf{Z}^d_{\geq 0}} \psi_n(c_m\, v u_1^{m_1} \cdots u_d^{m_d}). \tag{3}$$

We take $\psi_n(a) = \varepsilon^{\mathrm{Tr}_{\mathbf{F}_{q^n}/\mathbf{F}_p}(a)}$, and let

$$G(y, x) = \prod_{m \in \mathbf{Z}^d_{\geq 0}} \Theta(\widetilde{c_m}\, y x^m) \in \overline{\mathbf{Q}_p}[\![x_1, \ldots, x_d, y]\!]. \tag{4}$$

hence (2) and (3) imply

$$\sum_{u\in(\mathbf{F}_{q^n}^*)^d}\sum_{v\in\mathbf{F}_{q^n}^*}\psi_n(vf(u)) = \sum_{v,u_1,\dots,u_d\in\mathbf{F}_{q^n}^*}\left(\prod_{i=0}^{n-1}G(\widetilde{v}^{q^i},\widetilde{u_1}^{q^i},\dots,\widetilde{u_d}^{q^i})\right). \tag{5}$$

We will use this formula in §3 to prove that $\widetilde{Z}(X,t)$ can be written as the quotient of two formal power series in $\mathbf{C}_p[\![t]\!]$, both having infinite radius of convergence.

2. The construction of Θ

We now explain how to construct the formal power series Θ whose existence was assumed in the previous section. Note first that it is enough to do the construction when $q=p$: indeed, if $\Theta_1\in\mathbf{Q}_p(\varepsilon)[\![t]\!]$ satisfies P1) and P2) for $q=p$, and for $q=p^e$ we take $\Theta(t)=\Theta_1(t)\Theta_1(t^p)\cdots\Theta_1(t^{p^{e-1}})$, then Θ satisfies P1) and P2) for q. Indeed, if $R>1$ is the radius of convergence of Θ_1, then the radius of convergence of Θ is at least $R^{1/p^{e-1}}>1$. Furthermore,

$$\varepsilon^{\mathrm{Tr}_{\mathbf{F}_{p^{ne}}/\mathbf{F}_p}(a)} = \prod_{i=0}^{ne-1}\Theta_1(\widetilde{a}^{p^i}) = \prod_{j=0}^{e-1}\Theta(\widetilde{a}^{q^j}).$$

Therefore, in the rest of this section we assume $q=p$.

We begin by considering the formal power series in two variables given by the following infinite product

$$F(x,y) = (1+y)^x(1+y^p)^{\frac{x^p-x}{p}}\cdots(1+y^{p^n})^{\frac{x^{p^n}-x^{p^{n-1}}}{p^n}}\cdots\in\mathbf{Q}[\![x,y]\!].$$

Note that if $1+h_i$ is the i^{th} factor in the above product, then $h_i\in(y^{p^{i-1}})$, hence the above product gives, indeed, a formal power series[1].

Proposition 2.1. *We have $F(x,y)\in\mathbf{Z}_p[\![x,y]\!]$*[2].

The following lemma gives a general criterion for proving an assertion as in the proposition.

Lemma 2.2. *If $f\in\mathbf{Q}_p[\![x,y]\!]$ is such that $f(0,0)=1$, then $f\in\mathbf{Z}_p[\![x,y]\!]$ if and only if*

$$\frac{f(x^p,y^p)}{f(x,y)^p}\in 1+p(x,y)\mathbf{Z}_p[\![x,y]\!]. \tag{6}$$

Proof. Suppose first that $f\in\mathbf{Z}_p[\![x,y]\!]$. Since $f(0,0)=1$, it follows that f is invertible and $\frac{1}{f}\in 1+(x,y)\mathbf{Z}_p[\![x,y]\!]$. We deduce that $\frac{1}{f^p}$, hence also $\frac{f(x^p,y^p)}{f(x,y)^p}$ lies in $1+(x,y)\mathbf{Z}_p[\![x,y]\!]$. If $\overline{f}\in\mathbf{F}_p[\![x,y]\!]$ is the reduction of f mod $p\mathbf{Z}_p[\![x,y]\!]$, we clearly have $\overline{f}(x^p,y^p)=\overline{f}(x,y)^p$. This implies that $\frac{f(x^p,y^p)}{f(x,y)^p}$ lies in $1+p(x,y)\mathbf{Z}_p[\![x,y]\!]$, as required.

[1]The general assertion is that if $h_i\in(x,y)^{N_i}$ are such that $\lim_{i\to\infty}N_i=\infty$, then $\prod_i(1+h_i)$ is a formal power series, as the coefficient of each monomial x^my^n comes from only finitely many factors in the product.

[2]Of course, since F has coefficients in $\mathbf{Q}$, this is equivalent to saying that F has coefficients in $\mathbf{Z}_{(p\mathbf{Z})}$.

Conversely, suppose that (6) holds, and let us write $f = \sum_{i,j\geq 0} a_{i,j}x^iy^j$, with $a_{i,j} \in \mathbf{Q}_p$ and $a_{0,0} = 1$. By hypothesis, we may write

$$\sum_{i,j\geq 0} a_{i,j}x^{pi}y^{pj} = \left(\sum_{i,j\geq 0} a_{i,j}x^iy^j\right)^p \cdot \sum_{i,j\geq 0} b_{i,j}x^iy^j, \tag{7}$$

where $b_{0,0} = 1$, and all other $b_{i,j}$ lie in $p\mathbf{Z}_p$. Arguing by induction, we see that it is enough to show the following: if $\alpha, \beta \in \mathbf{Z}_{\geq 0}$, not both zero, are such that $a_{k,\ell} \in \mathbf{Z}_p$ for all (k, ℓ) with $k \leq \alpha$ and $\ell \leq \beta$ such that one of the inequalities is strict, then $a_{\alpha,\beta} \in \mathbf{Z}_p$. Let us consider the coefficient $c_{\alpha,\beta}$ of $x^\alpha y^\beta$ in the power series in (7). By considering the left-hand side of (7), we see that $c_{\alpha,\beta} = 0$ unless p divides both α and β, in which case it is equal to $a_{\alpha/p,\beta/p}$. By considering the right-hand side of (7), we see that $c_{\alpha,\beta} = pa_{\alpha,\beta} + Q_1 + \ldots + Q_r$, where each Q_j is a product of the form $Nb_{k,\ell}a_{k_1,\ell_1}\cdots a_{k_s,\ell_s}$, for some multinomial coefficient $N \in \mathbf{Z}$, and with all (k_i, ℓ_i) having the property that $k_i \leq \alpha$ and $\ell_i \leq \beta$, with one of the inequalities being strict. It follows that every Q_j lies in $\mathbf{Z}_p$, and if Q_j is not in $p\mathbf{Z}_p$, then $k = \ell = 0$, and $c_{\alpha,\beta}x^\alpha y^\beta = (a_{k,\ell}x^k y^\ell)^p$ for some k and ℓ. This can happen only when both α and β are divisible by p, and in this case Q_j is equal to $a^p_{\alpha/p,\beta/p}$. Furthermore, since in this case we have $a^p_{\alpha/p,\beta/p} \equiv a_{\alpha/p,\beta/p} \pmod p$, we deduce that $a_{\alpha,\beta} \in \mathbf{Z}_p$, and this completes the proof of the proposition. □

Remark 2.3. It should be clear from the proof of the lemma that a similar statement holds for formal power series in any number of variables. We restricted to the case of two variables, which is the one we will need, in order to avoid complicating too much the notation.

Proof of Proposition 2.1. Since we clearly have $F(0, 0) = 1$, we may apply Lemma 2.2, so it is enough to show that $\frac{F(x^p,y^p)}{F(x,y)^p}$ lies in $1 + p(x, y)\mathbf{Z}_p$. By definition, we have

$$\frac{F(x^p, y^p)}{F(x, y)^p} = \frac{(1 + y^p)^{x^p} \cdot (1 + y^{p^2})^{\frac{x^{p^2}-x^p}{p}} \cdot (1 + y^{p^3})^{\frac{x^{p^3}-x^{p^2}}{p^2}} \cdots}{(1 + y)^{px} \cdot (1 + y^p)^{x^p-x} \cdot (1 + y^{p^2})^{\frac{x^{p^2}-x^p}{p}} \cdots} = \frac{(1 + y^p)^x}{(1 + y)^{px}} = \left(\frac{(1 + y^p)}{(1 + y)^p}\right)^x.$$

In order to see that this lies in $1 + p(x, y)\mathbf{Z}_p[\![x, y]\!]$, we apply Lemma 2.2 in the other direction: since $g = 1 + y \in \mathbf{Z}_p[\![y]\!]$, and $g(0) = 1$, we deduce that $\frac{1+y^p}{(1+y)^p} = 1 + pw$, for some $w \in y\mathbf{Z}_p[\![y]\!]$. It follows from definition that

$$(1 + pw)^x = 1 + \sum_{m\geq 1} \frac{x(x - 1)\ldots(x - m + 1)}{m!} p^m w^m,$$

and $\frac{p^m}{m!} \in p\mathbf{Z}_p$ for every $m \geq 1$. Indeed, we have

$$\mathrm{ord}_p(m!) = \sum_{i\geq 1} \lfloor m/p^i \rfloor < \frac{m}{p} \sum_{i\geq 0} \frac{1}{p^i} = \frac{m}{p - 1} \leq m.$$

We conclude that $\left(\frac{(1+y^p)}{(1+y)^p}\right)^x \in 1 + p(x, y)\mathbf{Z}_p[\![x, y]\!]$, which completes the proof. □

Recall that $\varepsilon \in \overline{\mathbf{Q}_p}$ is our fixed primitive root of order p of 1. Let $\lambda = \varepsilon - 1$. The following estimate for $|\lambda|_p$ is well-known, but we include a proof for completeness.

Lemma 2.4. *With the above notation, we have* $|\lambda|_p = \left(\frac{1}{p}\right)^{1/(p-1)}$.

Proof. Since $(1+\lambda)^p = 1$, it follows that λ is a root of the polynomial $h(x) = x^{p-1} + \sum_{i=1}^{p-1} \binom{p}{i} x^{p-1-i}$. Since all coefficients of f but the leading one are divisible by p, and $h(0)$ is not divisible by p^2, it follows from Eisenstein's criterion that $h \in \mathbf{Q}_p[x]$ is an irreducible polynomial. This shows that $\mathbf{Q}_p(\varepsilon) = \mathbf{Q}_p(\lambda)$ has degree $(p-1)$ over $\mathbf{Q}_p$.

Every $\sigma\colon \mathbf{Q}_p(\varepsilon) \to \overline{\mathbf{Q}_p}$ must satisfy $\sigma(\varepsilon) = \varepsilon^i$ for some $1 \leq i \leq p-1$, and σ is uniquely determined by i. This shows that $\mathbf{Q}_p(\varepsilon)$ is a Galois extension of $\mathbf{Q}_p$, and since $[\mathbf{Q}_p(\varepsilon) : \mathbf{Q}_p] = n-1$, we conclude that the Galois conjugates of ε are precisely the ε^i, with $1 \leq i \leq p-1$. In particular, we have $|1-\varepsilon|_p = |1-\varepsilon^i|_p$ for every $1 \leq i \leq p$. On the other hand, we have

$$1 + x + \ldots + x^{p-1} = \prod_{i=1}^{p-1} (x - \varepsilon^i),$$

hence $\prod_{i=1}^{p-1}(1-\varepsilon^i) = p$. We thus deduce

$$|\varepsilon - 1|_p = |p|_p^{1/(p-1)} = \left(\frac{1}{p}\right)^{1/(p-1)}.$$

□

We put $\Theta(t) = F(t, \lambda)$. We first show that this is well-defined and has radius of convergence > 1.

Lemma 2.5. *We have* $\Theta \in \mathbf{Q}_p(\varepsilon)[\![t]\!]$, *and its radius of convergence is at least* $p^{1/(p-1)} > 1$.

Proof. Let us write $F(x,y) = \sum_{m \geq 0} \left(\sum_{n \geq 0} a_{m,n} y^n\right) x^m$. By Proposition 2.1, we have $a_{m,n} \in \mathbf{Z}_p$ for every m and n. We claim that $a_{m,n} = 0$ whenever $m > n$. Indeed, note that in

$$(1+y)^x = \sum_{n \geq 0} \frac{x(x-1)\ldots(x-n+1)}{n!} y^n,$$

every monomial $x^i y^j$ that appears with nonzero coefficient, has $i \leq j$. The same then holds for each $(1+y^{p^i})^{\frac{x^{p^i} - x^{p^{i-1}}}{p^i}}$, for $i \geq 1$. Since this property holds for each of the factors in the definition of $F(x,y)$, it also holds for F, as claimed.

Since $|a_{m,n}|_p \leq 1$ for every m and n, each series $\sum_{n \geq 0} a_{m,n} y^n$ has radius of convergence at least $1 > |\lambda|_p$, hence $F(t,\lambda)$ is a well-defined series in $\mathbf{Q}_p(\varepsilon)[\![t]\!]$. Furthermore, for every m we have

$$|\sum_{n \geq 0} a_{m,n} \lambda^n|_p = |\sum_{n \geq m} a_{m,n} \lambda^n|_p \leq |\lambda|_p^m.$$

This implies that the radius of convergence of $F(t,\lambda)$ is at least $|\lambda|_p^{-1} = p^{1/(p-1)} > 1$. □

We now show that Θ also satisfies the property P2) from §1 and thus complete the proof of the existence of Θ with the required properties.

Lemma 2.6. *For every $n \geq 1$, and every $a \in \mathbf{F}_{p^n}$, we have*

$$\varepsilon^{\mathrm{Tr}_{\mathbf{F}_{p^n}/\mathbf{F}_p}(a)} = \Theta(\widetilde{a})\Theta(\widetilde{a}^p)\cdots\Theta(\widetilde{a}^{p^{n-1}}).$$

Proof. Note first that since Θ has radius of convergence larger than 1, and $|\widetilde{a}^{p^i}|_p$ is either 1 or 0, we may apply Θ to the $\widetilde{a}^{p^i}$. Let us compute, more generally,

$$\prod_{i=0}^{n-1} F(\widetilde{a}^{p^i}, y) = \prod_{i=0}^{n-1}(1+y)^{\widetilde{a}^{p^i}} \cdot \prod_{m\geq 1}(1+y^{p^m})^{\sum_{i\geq 0}^{n-1} \frac{\widetilde{a}^{p^{m+i}} - \widetilde{a}^{p^{m+i-1}}}{p^m}}$$

$$= (1+y)^{\widetilde{a}+\widetilde{a}^p+\ldots+\widetilde{a}^{p^{n-1}}} \cdot \prod_{m\geq 1}(1+y^{p^m})^{\frac{\widetilde{a}^{p^{m+n-1}} - \widetilde{a}^{p^{m-1}}}{p^m}} = (1+y)^{\widetilde{a}+\widetilde{a}^p+\ldots+\widetilde{a}^{p^{n-1}}},$$

where the last equality follows from the fact that $\widetilde{a}^{p^n} = \widetilde{a}$. Since $\lambda = \varepsilon - 1$, in order to complete the proof of the lemma it is enough to show that

$$\varepsilon^{\widetilde{a}+\widetilde{a}^p+\ldots+\widetilde{a}^{p^{n-1}}} = \varepsilon^{\mathrm{Tr}_{\mathbf{F}_{p^n}/\mathbf{F}_p}(a)}. \tag{8}$$

Recall that $\mathbf{F}_{p^n}$ is a Galois extension of $\mathbf{F}_p$ with Galois group isomorphic to $\mathbf{Z}/n\mathbf{Z}$, and generated by σ, where $\sigma(u) = u^p$. By Theorem 2.1 in Appendix 2, we have an isomorphism $G(\mathbf{Q}_p^{(n)}/\mathbf{Q}_p) \simeq G(\mathbf{F}_{p^n}/\mathbf{F}_p)$, and let $\widetilde{\sigma}$ be the automorphism of $\mathbf{Q}_p^{(n)}$ corresponding to σ. Since $\widetilde{\sigma}(\widetilde{a})^{p^n} = \widetilde{\sigma}(\widetilde{a})$, it follows that $\widetilde{\sigma}(\widetilde{a})$ is the Teichmüller lift of its residue class, which is $\sigma(a) = a^p$. Therefore $\widetilde{\sigma}(\widetilde{a}) = \widetilde{a}^p$. We conclude that $\sum_{i=0}^{n-1} \widetilde{a}^{p^i} \in \mathbf{Z}_p$, and it clearly lies over $\sum_{i=0}^{n-1} a^{p^i} = \mathrm{Tr}_{\mathbf{F}_{p^n}/\mathbf{F}_p}(a)$. Therefore in order to show (8), we see that it suffices to show that if $w \in \mathbf{Z}_p$ lies over $b \in \mathbf{F}_p$, then $\varepsilon^w = \varepsilon^b$, where the left-hand side is defined as $(1+\lambda)^w$. We may write $w = pw_0 + \ell$ for some $\ell \in \mathbf{Z}$, and using Proposition 5.3 in Appendix 2, we obtain

$$(1+\lambda)^w = ((1+\lambda)^p)^{w_0} \cdot (1+\lambda)^\ell = 1 \cdot \varepsilon^\ell = \varepsilon^b.$$

This completes the proof of the lemma. □

3. Traces of certain linear maps on rings of formal power series

Our goal in this section is to establish the following intermediary step towards the proof of the rationality of the zeta function.

Proposition 3.1. *With the notation introduced in §1, for every $X = V(f)$, where $f \in \mathbf{F}_q[x_1, \ldots, x_n]$, the formal power series $\widetilde{Z}(X,t)$ can be written as a quotient $\frac{g(t)}{h(t)}$, where $g, h \in \mathbf{C}_p[\![t]\!]$ have infinite radii of convergence.*

The proof of the proposition will rely on the formula for the numbers N'_n coming out of (1) and (5) in §1, and on a formalism for treating certain linear maps on a formal power series ring, that we develop in this section.

For $N \geq 1$, we consider the formal power series ring $R = \mathbf{C}_p[\![x_1, \ldots, x_N]\!]$, and we denote by $\mathfrak{m}$ the maximal ideal in R. We will apply this with $N = d+1$, where d is as in the previous sections. As usual, for $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbf{Z}_{\geq 0}^N$, we put $x^\alpha = x_1^{\alpha_1} \cdots x_N^{\alpha_N}$

and $|\alpha| = \sum_{i=1}^N \alpha_i$. The order $\operatorname{ord}(h)$ of $h \in R$ is the largest $r \geq 0$ such that $h \in \mathfrak{m}^r$ (we make the convention that $\operatorname{ord}(0) = \infty$). On R we consider the $\mathfrak{m}$-adic topology. Recall that this is invariant under translations, and a basis of open neighborhoods of the origin is given by $\{\mathfrak{m}^r \mid r \geq 0\}$. Therefore we have $h_m \to h$ when m goes to infinity if and only if $\lim_{m\to\infty} \operatorname{ord}(f_m - f) = \infty$. As in the case of a DVR, one shows that one can put a metric on R that induces the $\mathfrak{m}$-adic topology.

We will consider $\mathbf{C}_p$-linear maps $A\colon R \to R$ that are continuous with respect to the $\mathfrak{m}$-adic topology. Such a map is determined by its values on the monomials in R. More precisely, such a map must satisfy

$$\lim A(x^\alpha) = 0 \text{ when } |\alpha| \to \infty, \tag{9}$$

and for $f = \sum_\alpha c_\alpha x^\alpha$, we have $A(f) = \sum_\alpha c_\alpha A(x^\alpha)$. Conversely, given a set of elements $(A(x^\alpha))_{\alpha \in \mathbf{Z}^N_{\geq 0}}$ that satisfies (9), we obtain a continuous linear map A given by the above formula. If we write $A(x^\beta) = \sum_\alpha a_{\alpha\beta} x^\alpha$, with $a_{\alpha\beta} \in \mathbf{C}_p$, then we can represent A by the "matrix" $(a_{\alpha\beta})_{\alpha,\beta\in\mathbf{Z}^N_{\geq 0}}$. Note that condition (9) translates as follows: for every α, we have $a_{\alpha\beta} = 0$ for $|\beta| \gg 0$.

We say that A has finite support if the corresponding "matrix" $(a_{\alpha\beta})$ has only finitely many nonzero entries. In this case A can be identified to an endomorphism of a finite-dimensional subspace of $\mathbf{C}_p[x_1, \ldots, x_N] \subseteq \mathbf{C}_p[\![x_1, \ldots, x_n]\!]$, and $(a_{\alpha\beta})$ can be identified to the corresponding matrix.

The usual rules for dealing with matrices apply in this setting. If A is described by the "matrix" $(a_{\alpha\beta})$, then

$$A(\sum_\beta c_\beta x^\beta) = \sum_\alpha \left(\sum_\beta a_{\alpha\beta} c_\beta \right) x^\alpha$$

(note that by hypothesis, the sum $\sum_\beta a_{\alpha\beta} c_\beta$ has only finitely many nonzero terms). If A and B are linear, continuous maps as above, described by the "matrices" $(a_{\alpha\beta})$ and $(b_{\alpha\beta})$, then the composition $A \circ B$ is again linear and continuous, and it is represented by the product $(c_{\alpha\beta})$ of the two "matrices": $c_{\alpha\beta} = \sum_\gamma a_{\alpha\gamma} b_{\gamma\beta}$.

We now introduce the two main examples of such maps that we will consider. Given $H \in R$, we define $\Psi_H\colon R \to R$ to be given by multiplication by H: $\Psi_H(f) = fH$. This is clearly $\mathbf{C}_p$-linear and continuous. If $H = \sum_\alpha h_\alpha x^\alpha$, then Ψ_H is represented by the "matrix" $(h_{\alpha-\beta})_{\alpha,\beta}$, where we put $h_{\alpha-\beta} = 0$ if $\alpha - \beta \notin \mathbf{Z}^N_{\geq 0}$. Note that $\Psi_{H_1} \circ \Psi_{H_2} = \Psi_{H_1 H_2}$.

For another example, if q is any positive integer, let $T_q\colon R \to R$ be given by $T_q(\sum_{\alpha\in\mathbf{Z}^N_{\geq 0}} a_\alpha x^\alpha) = \sum_{\alpha\in\mathbf{Z}^N_{\geq 0}} a_{q\alpha} x^\alpha$. It is clear that T_q is $\mathbf{C}_p$-linear and continuous. If $H = \sum_\alpha h_\alpha x^\alpha \in R$, let $\Psi_{q,H} = T_q \circ \Psi_H$. We have

$$\Psi_{q,H}(x^\beta) = T_q(\sum_\alpha h_\alpha x^{\alpha+\beta}) = T_q(\sum_\alpha h_{\alpha-\beta} x^\alpha) = \sum_\alpha h_{q\alpha-\beta} x^\beta.$$

Therefore $\Psi_{q,H}$ is represented by the "matrix" $(h_{q\alpha-\beta})_{\alpha,\beta}$.

Lemma 3.2. *We have $\Psi_H \circ T_q = \Psi_{q,H_q}$, where $H_q(x_1, \ldots, x_N) = H(x_1^q, \ldots, x_N^q)$.*

Proof. Let $H = \sum_{\alpha \in \mathbf{Z}^d_{\geq 0}} h_\alpha x^\alpha$, and we put $h_\alpha = 0$ if $\alpha \notin \mathbf{Z}^N_{\geq 0}$. We have

$$\Psi_H \circ T_q(\sum_\beta b_\beta x^\beta) = H \cdot \sum_\beta b_{q\beta} x^\beta = \sum_\gamma \left(\sum_{\alpha+\beta=\gamma} h_\alpha b_{q\beta} \right) x^\gamma. \tag{10}$$

On the other hand,

$$T_q \circ H_q \left(\sum_\beta b_\beta x^\beta \right) = T_q \left(\sum_\gamma \left(\sum_{q\alpha+\beta=\gamma} h_\alpha b_\beta \right) x^\gamma \right) = \sum_\gamma \left(\sum_{q\alpha+\beta=q\gamma} h_\alpha b_\beta \right) x^\gamma \tag{11}$$

In the last sum in (11) we see that β has to be divisible by q, and we deduce that the two expressions in (10) and (11) are equal. □

We now discuss the trace of a continuous linear map as above. Given such a map $A\colon R \to R$ described by the "matrix" $(a_{\alpha\beta})$, we consider the series $\sum_{\alpha \in \mathbf{Z}^N_{\geq 0}} a_{\alpha\alpha}$. If this is convergent in $\mathbf{C}_p$, we denote its sum by $\mathrm{Trace}(A)$. Note that if A has finite support, then $\mathrm{Trace}(A)$ is equal to the trace of any corresponding endomorphism of a finite-dimensional vector space of polynomials.

Let R_0 be the set of those $H = \sum_\alpha h_\alpha x^\alpha \in R$ with the property that there is $M > 0$ such that $|h_\alpha|_p \leq \left(\frac{1}{p}\right)^{M|\alpha|}$ for every $\alpha \in \mathbf{Z}^N_{\geq 0}$.

Remark 3.3. If $H \in R_0$, then there is $\rho > 1$ such that $H(u_1, \ldots, u_N)$ is convergent whenever $u_i \in \mathbf{C}_p$ are such that $|u_i| \leq \rho$ for all i. Indeed, with M as above, if $\rho = p^a$, where $0 < a < M$, then

$$|h_\alpha u_1^{\alpha_1} \ldots u_N^{\alpha_N}|_p \leq |h_\alpha|_p \cdot \rho^{|\alpha|} \leq \left(\frac{1}{p}\right)^{(M-a)|\alpha|},$$

which converges to zero when $|\alpha|$ goes to infinity.

Lemma 3.4. *R_0 is a subring of R. Furthermore, if $j_1, \ldots, j_N$ are positive integers, and if $H \in R_0$, then $H(x_1^{j_1}, \ldots, x_N^{j_N}) \in R_0$.*

Proof. The first assertion follows from the fact that if $M > 0$ works for both H_1 and H_2, then it also works for $H_1 - H_2$ and $H_1 H_2$. The second assertion follows from the fact that if M works for H, and if $j = \max\{j_1, \ldots, j_N\}$, then M/j works for $H(x_1^{j_1}, \ldots, x_N^{j_N})$. □

Proposition 3.5. *Let $H \in R_0$ and $\Psi = \Psi_{q,H}$ for some integer $q \geq 2$. For every $s \geq 1$ the trace of Ψ^s is well-defined, and*

$$(q^s - 1)^N \mathrm{Trace}(\Psi^s) = \sum_u H(u) H(u^q) \ldots H(u^{q^{s-1}}),$$

where the sum is over all $u = (u_1, \ldots, u_N) \in \mathbf{C}_p^N$ such that $u_i^{q^s - 1} = 1$ for all i.

Proof. Let us first consider the case $s = 1$. Recall that if $H = \sum_\alpha h_\alpha x^\alpha$, then Ψ is described by the matrix $(h_{q\alpha-\beta})_{\alpha,\beta}$. Therefore $\mathrm{Trace}(\Psi) = \sum_{\alpha \in \mathbf{Z}^N_{\geq 0}} h_{(q-1)\alpha}$. By assumption, there is $M > 0$ such that $|h_\alpha|_p \leq \left(\frac{1}{p}\right)^{M|\alpha|}$ for every α. In particular, $\lim_{|\alpha| \to \infty} h_{(q-1)\alpha} = 0$.

Furthermore, we have seen in Remark 3.3 that $H(u_1, \dots, u_N)$ is well-defined when $|u_i| \leq 1$ for all i. The subset $U = \{\lambda \in \mathbf{C}_p \mid \lambda^{q-1} = 1\}$ is a cyclic subgroup of $\mathbf{C}_p$. If $\lambda_0 \in U$ is a generator, then

$$\sum_{\lambda \in U} \lambda^i = \sum_{j=0}^{q-2} \lambda_0^{ij} = \begin{cases} q-1, & \text{if } (q-1)|i; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore

$$\sum_{u \in U^N} H(u) = \sum_{u \in U^N} \sum_{\alpha \in \mathbf{Z}_{\geq 0}^N} h_\alpha u^\alpha = \sum_{\alpha \in \mathbf{Z}_{\geq 0}^N} h_\alpha \cdot \prod_{i=1}^N \left(\sum_{u_i \in U} u_i^{\alpha_i} \right)$$
$$= (q-1)^N \sum_{\alpha \in (q-1)\mathbf{Z}_{\geq 0}^N} h_\alpha = (q-1)^N \mathrm{Trace}(\Psi).$$

This completes the proof when $s = 1$. Suppose now that $s \geq 2$. Using repeatedly Lemma 3.2, we obtain

$$\Psi^s = (T_q \circ \Psi_H)^s = (T_q^2 \circ \Psi_{H_q} \circ \Psi_H) \circ (T_q \circ \Psi_H)^{s-2} = (T_q^2 \circ \Psi_{H_q H}) \circ (T_q \circ \Psi_H)^{s-2} = \dots$$
$$= T_q^s \circ \Psi_{H_{q^{s-1}} \dots H_q H} = \Psi_{q^s, H_{q^{s-1}} \dots H_q H}.$$

It follows from Lemma 3.4 that since H lies in R_0, we also have $H_{q^{s-1}} \dots H_q H \in R_0$. Therefore we may apply the case $s = 1$ to deduce that $\mathrm{Trace}(\Psi^s)$ is well-defined, and that

$$(q^s - 1)^N \mathrm{Trace}(\Psi^s) = \sum_{u \in U^N} H(u) H(u^q) \dots H(u^{q^{s-1}}).$$

□

Suppose now that $A \colon R \to R$ is a $\mathbf{C}_p$-linear continuous map, described by the "matrix" $(a_{\alpha\beta})_{\alpha,\beta}$. We define the characteristic power series of A by

$$\det(\mathrm{Id} - tA) := \sum_{m \geq 0} (-1)^m \left(\sum_\sigma \varepsilon(\sigma) a_{\alpha_1 \sigma(\alpha_1)} \cdots a_{\alpha_m \sigma(\alpha_m)} \right) t^m, \tag{12}$$

where the second sum is over all subsets with m elements $\{\alpha_1, \dots, \alpha_m\}$ of $\mathbf{Z}_{\geq 0}^N$, and over all permutations σ of such a set. Of course, the definition makes sense if the series that appears as the coefficient of t^m is convergent in $\mathbf{C}_p$ for every m. It is clear that if A has finite support, then $\det(\mathrm{Id} - tA)$ is equal to the characteristic polynomial of a corresponding endomorphism of a finite-dimensional vector space of polynomials.

Lemma 3.6. *If $H \in R_0$, then for every integer $q \geq 2$ the characteristic power series of $\Psi = \Psi_{q,H}$ is well-defined, and it has infinite radius of convergence.*

Proof. Let us write $H = \sum_\alpha h_\alpha x^\alpha$, and let $M > 0$ be such that $|h_\alpha|_p \leq \left(\frac{1}{p}\right)^{M|\alpha|}$ for every α. We have seen that Ψ is described by the "matrix" $(a_{\alpha\beta})$, where $a_{\alpha\beta} = h_{q\alpha - \beta}$. Given $\{u_1, \dots, u_m\} \subseteq \mathbf{Z}_{\geq 0}^N$, and a permutation σ of this set, we have

$$|a_{\alpha_1 \sigma(\alpha_1)} \cdots a_{\alpha_m \sigma(\alpha_m)}|_p \leq \left(\frac{1}{p}\right)^{M \sum_{i=1}^m |q\alpha_i - \sigma(\alpha_i)|}.$$

Note that $|q\alpha_i - \sigma(\alpha_i)| = q|\alpha_i| - |\sigma(\alpha_i)|$ if $q\alpha_i - \sigma(\alpha_i)$ is in $\mathbf{Z}^N_{\geq 0}$, and $|q\alpha_i - \sigma(\alpha_i)| = 0$, otherwise. Furthermore, in the latter case we also have $a_{\alpha_i\sigma(\alpha_i)} = 0$. We thus conclude that

$$|a_{\alpha_1\sigma(\alpha_1)} \cdots a_{\alpha_m\sigma(\alpha_m)}|_p \leq \left(\frac{1}{p}\right)^{M(q-1)(|\alpha_1|+\ldots+|\alpha_m|)}.$$

Since the right-hand side tends to zero when $\max\{|\alpha_i|\}$ goes to infinity, it follows that $\det(\mathrm{Id} - tA)$ is well-defined.

Furthermore, the above computation shows that if we write $\det(\mathrm{Id}-tA) = \sum_{m\geq 0} b_m t^m$, then

$$|b_m|_p^{1/m} \leq \max_{\alpha_1,\ldots,\alpha_m} \left(\frac{1}{p}\right)^{\frac{M(q-1)(|\alpha_1|+\ldots+|\alpha_m|)}{m}},$$

where the maximum is over *distinct* $\alpha_1, \ldots, \alpha_m \in \mathbf{Z}^N_{\geq 0}$. When m goes to infinity, we have

$$\min_{\alpha_1,\ldots,\alpha_m} \frac{M(q-1)(|\alpha_1| + \ldots + |\alpha_m|)}{m} \to \infty.$$

The above estimate therefore implies that $\lim_{m\to\infty} |b_m|_p^{1/m} = 0$, hence $\det(\mathrm{Id} - tA)$ has infinite radius of convergence. □

Proposition 3.7. *If $A\colon R \to R$ is a continuous $\mathbf{C}_p$-linear map such that $\det(\mathrm{Id} - tA)$ and $\mathrm{Trace}(A^s)$ are well-defined for all $s \geq 1$, then*

$$\det(\mathrm{Id} - tA) = \exp\left(-\sum_{s\geq 1} \frac{\mathrm{Trace}(A^s)}{s} t^s\right).$$

Proof. If A has finite support, then the assertion follows from Lemma 2.2 in Lecture 5. Our goal is to use this special case to deduce the general one.

Let us consider a sequence $(A^{(m)})_{m\geq 1}$ of maps with finite support, each described by the matrix $(a^{(m)}_{\alpha\beta})_{\alpha,\beta\in\mathbf{Z}^N_{\geq 0}}$, that satisfies the following condition. For every α and β, we have $a^{(m)}_{\alpha\beta} = a_{\alpha\beta}$ or $a^{(m)}_{\alpha\beta} = 0$, and the former condition holds for all $m \gg 0$. It is clear that we can find a sequence $(A^{(m)})_{m\geq 1}$ with this property.

It is convenient to consider on $\mathbf{C}_p[\![t]\!]$ (identified to a countable product of copies of $\mathbf{C}_p$) the product topology, where each $\mathbf{C}_p$ has the usual p-adic topology. Explicitly, a sequence of formal power series $(f_m)_{m\geq 1}$, with $f_m = \sum_{i\geq 0} b_{m,i} t^i$, converges to $f = \sum_{i\geq 0} b_i t^i$ if and only if $\lim_{m\to\infty} b_{m,i} = b_i$ for every i. Note that if this is the case, and all $f_m(0)$ are zero, then $\exp(f_m)$ converges to $\exp(f)$ when m goes to infinity (this is the case if we replace exp by any other element of $\mathbf{C}_p[\![t]\!]$). Since each $A^{(m)}$ satisfies the conclusion of the proposition, in order to complete the proof it is enough to show that

i) $\lim_{m\to\infty} \det(\mathrm{Id} - tA^{(m)}) = \det(\mathrm{Id} - tA)$.
ii) $\lim_{m\to\infty} \mathrm{Trace}((A^{(m)})^s) = \mathrm{Trace}(A^s)$ for every $s \geq 1$.

Let us first check i). We consider the coefficients $b_\ell^{(m)}$ and b_ℓ of t^ℓ in $\det(\mathrm{Id} - tA^{(m)})$ and $\det(\mathrm{Id} - tA)$, respectively. By definition, we have

$$b_\ell^{(m)} = (-1)^\ell \sum_\sigma \varepsilon(\sigma) a^{(m)}_{\alpha_1\sigma(\alpha_1)} \cdots a^{(m)}_{\alpha_\ell\sigma(\alpha_\ell)}. \tag{13}$$

By our choice of $A^{(m)}$, every product in the sum above is either zero, or it shows up in the corresponding expression for b_ℓ. Furthermore, given any $\{\alpha_1, \ldots, \alpha_\ell\}$ and any permutation σ of this set, the product $\varepsilon(\sigma) a_{\alpha_1\sigma(\alpha_1)} \cdots a_{\alpha_\ell\sigma(\alpha_\ell)}$ appears in (13) for $m \gg 0$. Since we know that $\det(\mathrm{Id} - tA)$ exists, the assertion in i) follows.

The proof of ii) is similar. By definition, we have

$$\mathrm{Trace}((A^{(m)})^s) = \sum_{\alpha_1, \ldots, \alpha_s} a^{(m)}_{\alpha_1\alpha_2} \cdots a^{(m)}_{\alpha_{s-1}\alpha_s} a^{(m)}_{\alpha_s\alpha_1}. \tag{14}$$

By hypothesis, each product $a^{(m)}_{\alpha_1\alpha_2} \cdots a^{(m)}_{\alpha_s\alpha_1}$ is either zero, or it is equal to $a_{\alpha_1\alpha_2} \cdots a_{\alpha_s\alpha_s}$. Moreover, by hypothesis every product $a_{\alpha_1\alpha_2} \cdots a_{\alpha_s\alpha_{s+1}}$ appears in (14) if $m \gg 0$. Since $\mathrm{Trace}(A^s)$ exists, we deduce the assertion in ii). This completes the proof of the proposition. □

By Lemmas 3.5 and 3.6, we may apply the above proposition, to get the following

Corollary 3.8. *If $H \in R_0$ and $\Psi = \Psi_{q,H}$ for an integer $q \geq 2$, then*

$$\det(\mathrm{Id} - t\Psi) = \exp\left(- \sum_{s \geq 1} \frac{\mathrm{Trace}(\Psi^s)}{s} t^s \right).$$

We now apply the above framework to give a proof of Proposition 3.1. Given $f \in \mathbf{F}_q[x_1, \ldots, x_d]$, we let $N = d + 1$. We begin with the following lemma.

Lemma 3.9. *For every $n \geq 1$, the formal power series $G \in R = \mathbf{C}_p[\![y, x_1, \ldots, x_d]\!]$ defined in (4) lies in R_0.*

Proof. Since G is a product of factors of the form $\Theta(\widetilde{c}yx_1^{m_1} \cdots x_d^{m_d})$, it follows from Lemma 3.4 that it is enough to see that $\Theta(ayx_1^{m_1} \cdots y_d^{m_d})$ lies in R_0 whenever $|a|_p = 1$ and $m_1, \ldots, m_d \in \mathbf{Z}_{\geq 0}$. Furthermore, if $q = p^e$, then we have taken $\Theta(t) = \prod_{i=0}^{e-1} \Theta_0(t^{p^i})$, where Θ_0 is constructed for $q = p$. A second application of Lemma 3.4 allows us to reduce to the case when $q = p$.

Recall that we have seen in the proof of Lemma 2.5 that if $\Theta = \sum_{i \geq 0} b_i t^i$, then $|b_i|_p \leq |\lambda|_p^i = \left(\frac{1}{p}\right)^{i/(p-1)}$. If a and $m_1, \ldots, m_d$ are as above, then

$$\Theta(ayx_1^{m_1} \cdots x_d^{m_d}) = \sum_i b_i a^i y^i x_1^{im_1} \cdot x_d^{im_d}.$$

Note that

$$|b_i a^i|_p = |b_i|_p \leq \left(\frac{1}{p}\right)^{i/(p-1)} = \left(\frac{1}{p}\right)^{M|(i, im_1, \ldots, im_d)|},$$

where $M = \frac{1}{(p-1)(1+m_1+\ldots+m_d)}$. Therefore $\Theta(ayx_1^{m_1} \cdots x_d^{m_d})$ lies in R_0. □

We can now prove the result stated at the beginning of this section.

Proof of Proposition 3.1. Since $G \in R_0$, we may apply Proposition 3.5 in order to compute $\mathrm{Trace}(\Psi_{q,G})$. Note that $\{w \in \mathbf{C}_p \mid w^{q^n-1} = 1\} = \{\widetilde{u} \mid u \in \mathbf{F}_{q^n}^*\}$. We deduce using (1) and (5) that

$$N'_n q^n - (q^n-1)^d = \sum_{v,u_1,\dots,u_d \in \mathbf{F}_{q^n}^*} \left(\prod_{i=0}^{n-1} G(\widetilde{v}^{q^i}, \widetilde{u_1}^{q^i}, \dots, \widetilde{u_d}^{q^i}) \right) = (q^n-1)^{d+1} \mathrm{Trace}(\Psi_{q,G}^n). \tag{15}$$

Let us compute

$$\exp\left(\sum_{n\geq 1} \frac{N'_n q^n - (q^n-1)^d}{n} t^n \right) = \widetilde{Z}(X,qt) \cdot \exp\left(-\sum_{i=0}^{d} (-1)^{d-i} \binom{d}{i} \frac{q^{ni}}{n} t^n \right) \tag{16}$$

$$= \widetilde{Z}(X,qt) \cdot \prod_{i=0}^{d} \exp\left((-1)^{d-i} \binom{d}{i} \log(1-q^i t) \right) = \widetilde{Z}(X,qt) \cdot \prod_{i=0}^{d} (1-q^i t)^{(-1)^{d-i}\binom{d}{i}}.$$

On the other hand, using Corollary 3.8 and Lemma 3.9 we get

$$\exp\left(\sum_{n\geq 1} (q^n-1)^{d+1} \mathrm{Trace}(\Psi_{q,G}^n) \frac{t^n}{n} \right) = \exp\left(\sum_{i=0}^{d+1} (-1)^{d+1-i} \binom{d+1}{i} \mathrm{Trace}(\Psi_{q,G}^n) \frac{q^{ni} t^n}{n} \right) \tag{17}$$

$$= \prod_{i=0}^{d+1} \det(\mathrm{Id} - q^i t \Psi_{q,G})^{(-1)^{d-i}\binom{d+1}{i}}.$$

It follows from Lemma 3.6 that each $\det(\mathrm{Id} - q^i t \Psi_{q,G})$ has infinite radius of convergence. Since the expressions in (16) and (17) are equal, we conclude that $\widetilde{Z}(X,qt)$ is the quotient of two formal power series in $\mathbf{C}_p[\![t]\!]$ with infinite radius of convergence, hence $\widetilde{Z}(X,t)$ has the same property. □

4. The rationality of the zeta function

The last ingredient in Dwork's proof for the rationality of the zeta function is the following proposition. In order to avoid confusion, we denote by $|m|_\infty$ the usual (Archimedean) absolute value of an integer m.

Proposition 4.1. *Let $Z(t) = \sum_{n\geq 0} a_n t^n$ be a formal power series in $\mathbf{Z}[\![t]\!]$, that satisfies the following two properties:*

1) *There are $C, s > 0$ such that $|a_n|_\infty \leq C s^n$ for all $n \geq 0$.*
2) *The image of Z in $\mathbf{C}_p[\![t]\!]$ can be written as a quotient $\frac{g(t)}{h(t)}$, where $g, h \in \mathbf{C}_p[\![t]\!]$ have infinite radii of convergence.*

Then $Z(t)$ lies in $\mathbf{Q}(t)$.

We first need a lemma that gives a sharper version of the rationality criterion in Proposition 2.3 in Lecture 5. We will consider a formal power series $f = \sum_{n\geq 0} a_n t^n$ with coefficients in a field K. For every $i, N \geq 0$, we consider the matrix $A_{i,N} = (a_{i+\alpha+\beta})_{0\leq\alpha,\beta\leq N}$.

Lemma 4.2. *With the above notation, the power series f is rational if and only if there is N such that $\det(A_{i,N}) = 0$ for all $i \gg 0$.*

Proof. We have $f \in K(t)$ if and only if there is a nonzero polynomial $Q(t)$ such that Qf is a polynomial. If we write $Q = b_0 + b_1 t + \ldots + b_N t^N$, then the condition we need is that

$$b_N a_i + b_{N-1} a_{i+1} + \ldots + b_0 a_N = 0 \tag{18}$$

for all $i \gg 0$. The existence of $b_0, \ldots, b_N$, not all zero, that satisfy these conditions clearly implies that $\det(A_{i,N}) = 0$ for $i \gg 0$.

Conversely, suppose that we have N such that $\det(A_{i,N}) = 0$ for $i \gg 0$ (say, for $i \geq i_0$), and that N is minimal with this property. For every i, we put

$$L_i = (a_i, \ldots, a_{i+N}) \in K^{N+1} \text{ and } L'_i = (a_i, \ldots, a_{i+N-1}) \in K^N.$$

Claim. We have $\det(A_{i,N-1}) \neq 0$ for every $i \geq i_0$. If this is the case, since $\det(A_{i,N}) = 0$, it follows that for every $i \geq i_0 + N$, we have $L_i \in \sum_{j=1}^{N} L_{i-j}$, so that $\sum_{i\geq i_0} K \cdot L_i$ is spanned by $L_{i_0}, \ldots, L_{i_0+N-1}$. In this case, it is clear that we can find $b_0, \ldots, b_N$ not all zero such that (18) holds for all $i \geq i_0$. Therefore, in order to complete the proof it is enough to show the claim.

By the minimality assumption in the definition of N, it is enough to show that if $i \geq i_0$ and $\det(A_{i,N-1}) = 0$, then $\det(A_{i+1,N-1}) = 0$. Since $\det(A_{i,N-1}) = 0$, we have $L'_i, \ldots, L'_{i+N-1}$ linearly dependent. We have two cases to consider. If $L'_{i+1}, \ldots, L'_{i+N-1}$ are linearly dependent, then it is clear that $\det(A_{i+1,N-1}) = 0$. On the other hand, if this is not the case, then we can write $L'_i = \sum_{j=1}^{N-1} c_j L'_{i+j}$. Let us replace in the first row of $A_{i,N}$ each $a_{i+\ell}$ by $a_{i+\ell} - \sum_{j=1}^{N-1} c_j a_{i+\ell+j}$. We thus obtain $0 = \det(A_{i,N}) = \det(A_{i+1,N-1}) \cdot \delta$, where $\delta = a_{i+N} - \sum_{j=1}^{N-1} c_\ell a_{i+N+j}$. If $\delta \neq 0$, we clearly get $\det(A_{i+1,N-1}) = 0$. On the other hand, if $\delta = 0$, then it follows that L_i lies in the linear span of $L_{i+1}, \ldots, L_{i+N-1}$. Hence the top-right N-minor of $A_{i,N}$ vanishes, but this is precisely $\det(A_{i+1,N-1})$. This completes the proof of the claim, hence that of the proposition. □

Proof of Proposition 4.1. We begin by choosing $\alpha > 0$ such that $\alpha > \frac{\log(s)}{\log p}$. We then apply Proposition 4.4 in Appendix 2 to h and $R > p^\alpha$, to write $h = Pu$, where $P \in \mathbf{C}_p[t]$ and $u \in \mathbf{C}_p[\![t]\!]$ is invertible, and u and u^{-1} have radius of convergence $> p^\alpha$. We may clearly assume that $P(0) = 1$. We thus can write $f = \frac{gu^{-1}}{P}$, and the radius of convergence of gu^{-1} is $> p^\alpha$. If we write $gu^{-1} = \sum_{n\geq 0} b_n t^n$, then by Proposition 4.1 in Appendix 2 we have $\limsup_m |b_m|_p^{1/m} < p^{-\alpha}$. Therefore there is m_0 such that

$$|b_m|_p \leq p^{-m\alpha} \text{ for all } m \geq m_0. \tag{19}$$

Let us write $f = \sum_{n\geq 0} a_n t^n$. Using the notation in Lemma 4.2, we need to show that we can choose N such that $\det(A_{i,N})$ for all $i \gg 0$. The key is to compare $|\det(A_{i,N})|_p$

and $|\det(A_{i,N})|_\infty$. Using condition 1) is the proposition, we get

$$|\det(A_{i,n})|_\infty \leq \sum_{\sigma\in S_{n+1}} |\prod_{\alpha=0}^{N} |a_{i+\alpha+\sigma(\alpha)}|_\infty \leq C^{N+1}(N+1)!\cdot s^{2\sum_{j=0}^{N}(i+j)} = C^{N+1}(N+1)!\cdot s^{(N+1)(2i+N)}.$$

On the other hand, let us write $P = 1+\lambda_1 t+\ldots+\lambda_r t^r$, so that $b_i = a_i + c_1 a_{i-1} + \ldots + c_r a_{i-r}$ for every $i \geq r$. Suppose that $N+1 = r+\ell$, and let $T_0,\ldots,T_N$ denote the columns of the matrix $A_{i,N}$. Starting with $j = N$ and going down up to $j = r$, we may replace T_j by $T_j+\lambda_1 T_{j-1}+\ldots+\lambda_r T_{j-r}$, without changing $\det(A_{i,N})$. In this way, we have replaced in the last ℓ columns each a_j by b_j. Since all a_m are in $\mathbf{Z}$, we have $|a_m|_p \leq 1$, and is we assume $i \geq m_0$, we deduce using (19) that

$$|\det(A_{i,N})|_p \leq p^{-2\alpha\sum_{j=0}^{\ell-1}(i+r+j)} = p^{-\alpha\ell(2i+2r+\ell-1)}.$$

It follows from definition that if m is any nonzero integer, then $|m|_\infty \geq |m|_p^{-1}$. We conclude from the above that if $\det(A_{i,N})$ is nonzero, then

$$p^{\alpha\ell(2i+2r+\ell-1)} \leq |\det(A_{i,N})|_p^{-1} \leq |\det(A_{i,N})|_\infty \leq C^{N+1}(N+1)!s^{(N+1)(2i+N)}.$$

By taking log, we get

$$\alpha\ell(2i+2r+\ell-1)\log(p) \leq (r+\ell)(i+r+\ell)\log(s) + \log(C^{\ell+r}(\ell+r)!).$$

If ℓ is fixed and $i \gg 0$, this can only happen if $\alpha\ell\cdot\log(p) \leq (r+\ell)\log(s)$. However, by assumption we have $\alpha\cdot\log(p) > \log(s)$, hence if $\ell \gg 0$ we have $\alpha\ell\cdot\log(p) > (r+\ell)\log(s)$, and therefore $\det(A_{i,N}) = 0$ for all $i \gg 0$. This completes the proof of the proposition. □

We can now complete Dwork's proof of the rationality of the zeta function.

Theorem 4.3. *If X is a variety defined over a finite field $\mathbf{F}_q$, then the zeta function $Z(X,t)$ is rational.*

Proof. We have seen in Remark 2.3 in Lecture 3 that, arguing by induction on $\dim(X)$, it is enough to show that $Z(X,t)$ is a rational function when X is a hypersurface in $\mathbf{A}^d_{\mathbf{F}_q}$, defined by some nonzero $f \in \mathbf{F}_q[x_1,\ldots,x_d]$. We denote by H_i the hyperplane $(x_i = 0)$, where $1 \leq i \leq d$. For every $I \subseteq \{1,\ldots,d\}$ (including $I = \emptyset$), we put

$$X_I = X \cap \left(\bigcap_{i\in I} H_i\right) \text{ and } X_I^\circ = X_I \smallsetminus \left(\bigcup_{i\notin I} H_i\right).$$

We have a disjoint decomposition into locally closed subsets $X = \bigsqcup_I X_I^\circ$, hence Proposition 3.7 in Lecture 2 implies

$$Z(X,t) = \prod_{I\subseteq\{1,\ldots,d\}} Z(X_I^\circ). \tag{20}$$

Note that X_I is isomorphic to a hypersurface in $\mathbf{A}^{d-\#I}_{\mathbf{F}_q}$, and using the notation introduced in §1, we have $Z(X_I^\circ,t) = \widetilde{Z}(X_I,t)$. By Proposition 3.1, we can write $Z(X_I^\circ,t)$ as the quotient of two formal power series in $\mathbf{C}_p[\![t]\!]$, having infinite radii of convergence. Formula (20), implies that $Z(X,t)$ has the same property.

Recall that $Z(X,t)$ has nonnegative integer coefficients. Furthermore, if we write $Z(X,t) = \sum_{n\geq 0} a_n t^n$, then $a_n \leq q^{dn}$ for every n. Indeed, we have $|X(\mathbf{F}_{q^n})| \leq q^{dn}$ for every $n \geq 1$. Since the exponential function has non-negative coefficients, we deduce that $a_n \leq b_n$, where

$$\sum_{n\geq 0} b_n t^n = \exp\left(\sum_{n\geq 1} \frac{q^{dn}t^n}{n}\right) = \exp(-\log(1-q^d t)) = \frac{1}{1-q^d t} = \sum_{n\geq 0} q^{dn}t^n.$$

Therefore $a_n \leq q^{nd}$ for all $n \geq 0$, and we can apply Proposition 4.1 to conclude that $Z(X,t)$ is a rational function. □

Note the unlike the proof of the rationality of the zeta function described in Lecture 5 (using ℓ-adic cohomology), the above proof is much more elementary, as it only uses some basic facts about p-adic fields. At the same time, its meaning is rather mysterious. A lot of activity has been devoted to giving a cohomological version; in other words, to constructing a p-adic cohomology theory, and a corresponding trace formula, that would "explain" Dwork's proof. Such cohomology theories are the Monsky-Washnitzer cohomology (which behaves well for smooth affine varieties, see [vdP]) and the crystalline cohomology of Berthelot and Grothendieck (which behaves well for smooth projective varieties, see [Ber]). More recently, Berthelot introduced the *rigid cohomology* [LeS] that does not require smoothness, and which extends the Monsky-Washnitzer and the crystalline cohomolgy theories, when these are well-behaved.

References

[Ber] P. Berthelot, *Cohomologie cristalline des schémas de caractristique $p > 0$*, Lecture Notes in Mathematics, Vol. 407, Springer-Verlag, Berlin-New York, 1974. 15

[Dwo] B. Dwork, On the rationality of the zeta function of an algebraic variety, Amer. J. Math. **82** (1960), 631–648. 1

[Kob] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions*, Second edition, Graduate Texts in Mathematics, 58, Springer-Verlag, New York, 1984. 1

[LeS] B. Le Stum, *Rigid cohomology*, Cambridge Tracts in Mathematics, 172, Cambridge University Press, Cambridge, 2007. 15

[vdP] M. van der Put, The cohomology of Monsky and Washnitzer, in *Introductions aux cohomologies p-adiques (Luminy, 1984)*, Mém. Soc. Math. France (N.S.) No. **23** (1986), 33–59. 15