Characters, Jacobi Sums, and Gauss Sums: Counting solutions to diagonal equations in finite field towers

Alexander Russell Perlis

The following detailed account of Weil's counting technique is separate from the preceding expository account of Weil's paper. There is nothing here that does not appear in the first 9 pages of Weil's paper; however, my presentation includes more details in some places and my notation is slightly different. There is nothing here that does not appear in chapters 8–11 of the book by Ireland–Rosen; however, my presentation sticks to the point and thus is closer in length to Weil's own account. Corrections and comments are appreciated; send them to `aprl@math.arizona.edu`.

The equations to be considered here are those of the form

$$\langle 1\rangle \qquad a_0 x_0^{n_0} + a_1 x_1^{n_1} + \cdots + a_r x_r^{n_r} = b.$$

Such an equation might be termed *diagonal* to draw attention to the fact that there are no cross-terms (such as $x_1 x_3^2$) and that each variable appears once (as opposed to $x_1 + x_1^2$).

Fix a finite field $\mathbf{F}_q$ and fix an equation $\langle 1\rangle$. The coefficients a_i are nonzero elements of $\mathbf{F}_q$, the constant b is an arbitrary element of $\mathbf{F}_q$, and the exponents n_i are positive integers. The goal is to count solutions to $\langle 1\rangle$ simultaneously over each finite extension $\mathbf{F}_{q^\nu}$ of $\mathbf{F}_q$. The last point is important: although knowing how to count solutions over $\mathbf{F}_q$ certainly implies knowing how to do the same over an extension of $\mathbf{F}_q$—after all, q is arbitrary—there is no guarantee that the formulas involved should bear a useful resemblance to each other. The goal here, however, is to find a *single* formula, which depends on $\langle 1\rangle$, on q, and on ν (the degree of the extension $\mathbf{F}_{q^\nu}/\mathbf{F}_q$), that expresses the number of solutions to $\langle 1\rangle$ over $\mathbf{F}_{q^\nu}$.

Here is a brief description of the approach: The number of solutions to $\langle 1\rangle$ over $\mathbf{F}_q$ will be expressed in various ways, first in terms of character sums, then in terms of Jacobi sums, but ultimately in terms of Gauss sums. Since q was arbitrary, replacing q with q^ν gives us the number of solutions to $\langle 1\rangle$ over $\mathbf{F}_{q^\nu}$ in terms of Gauss sums over $\mathbf{F}_{q^\nu}$. Finally, the Hasse–Davenport theorem expresses Gauss sums over $\mathbf{F}_{q^\nu}$ in terms of Gauss sums over $\mathbf{F}_q$; thus we end up with a single formula.

Let us get started. For quite a while, our work will concentrate over $\mathbf{F}_q$. Recall that the multiplicative group $\mathbf{F}_q^\times$ is cyclic of order $q-1$; throughout, let w denote a generator of this group. (The notation remains consistent throughout; a reference table of notation appears at the end.) For $u \in \mathbf{F}_q$, let $\mathbf{N}(x^n = u)$ denote the number of solutions to $x^n = u$ over $\mathbf{F}_q$. If $u = 0$, then $x = 0$ is the unique solution. For $u \neq 0$, there are solutions if and only if $u \in \mathbf{F}_q^{\times n}$, i.e., if and only if u is an nth power; in this case, the number of solutions is the same as the number of distinct roots of unity of order dividing n. (Adjusting a solution by such a root of unity certainly gives another solution, and any two solutions must differ by such a root of unity.) Since the order of any root of unity in $\mathbf{F}_q^\times$ must divide $q-1$ (the order of the group), the roots of unity of order dividing n are precisely those of order dividing $d = \textsc{gcd}(n, q-1)$. In fact, $\mathbf{F}_q^\times$ contains precisely d such roots of unity, namely the first d powers of the element $w^{(q-1)/d}$, which compose the unique cyclic subgroup of $\mathbf{F}_q^\times$ of order d. As a consequence, one sees that $\mathbf{N}(x^n = u) = \mathbf{N}(x^d = u)$. This may be explained separately by observing that the map $x \mapsto x^{n/d}$, although not an automorphism of the group $\mathbf{F}_q^\times$ (since n/d is not necessarily coprime to $q-1$), is an automorphism of the group $\mathbf{F}_q^{\times d}$ onto the group $\mathbf{F}_q^{\times n}$. In summary,

$$\mathbf{N}(x^n = u) = \mathbf{N}(x^d = u) = \begin{cases} 1, & \text{if } u = 0; \\ d, & \text{if } u \neq 0 \text{ and } u \in \mathbf{F}_q^{\times n} = \mathbf{F}_q^{\times d}; \\ 0, & \text{if } u \neq 0 \text{ and } u \notin \mathbf{F}_q^{\times n} = \mathbf{F}_q^{\times d}. \end{cases}$$

From a computational standpoint, this result is useful only if one may easily determine whether a given nonzero u is a dth power. In terms of our generator w, this is easy: the dth powers are precisely those powers of w whose exponents are multiples of d. (Observe that this statement relies on the fact $d \mid q-1$; certainly the nth powers are in general not just those powers of w whose exponents are multiples of n.) Thus, if we write $u = w^k$, and then use the division algorithm to write $k = td + r$, where $0 \le r < d$, then the number $\mathbf{N}(x^d = w^k) = \mathbf{N}(x^d = w^{td+r})$ is d when $r = 0$, and 0 otherwise.

Now we will express $\mathbf{N}(x^n = u)$ in terms of character sums. A (multiplicative) character is a group homomorphism $\chi\colon \mathbf{F}_q^\times \to \mathbf{C}^\times$. Such a map χ is completely determined by its behavior on w (our generator of $\mathbf{F}_q^\times$), and the image of w must be some element of $\mathbf{C}^\times$ of order dividing $q-1$. Let $\zeta \in \mathbf{C}^\times$ denote a primitive root of unity, i.e., an element of order $q-1$; there are many choices for ζ, but for computation purposes one might as well set $\zeta = e^{2\pi i/(q-1)}$. There are precisely $q-1$ characters, namely the maps determined by $w \mapsto \zeta^k$, for $0 \le k < q-1$. The product of two characters is defined by $(\chi_1\chi_2)(x) = \chi_1(x)\chi_2(x)$, and under this operation the characters are seen to form a cyclic group, generated by $\omega\colon w \mapsto \zeta$. The identity element of this group is the trivial character $\varepsilon\colon w \mapsto 1$. Since $\chi^{-1}(u) = \overline{\chi(u)}$, the bar denoting complex conjugation, the inverse of a character χ is often denoted $\overline{\chi}$. As in any cyclic group, there is a unique subgroup of every order dividing the order of the group. Thus if $d \mid q-1$, then there are precisely d characters of order dividing d, and they are easy to write down: they are the first d powers of the character $\omega_d\colon w \mapsto \zeta^{(q-1)/d}$. In this notation, for $d \mid q-1$, subscripts may be converted to superscripts as follows: $\omega_d = \omega^{(q-1)/d}$; in particular, ω_1 is the same as ε, and ω_{q-1} is the same as ω.

The following simple fact is used repeatedly, and now is a good time to state it:

$$\langle 2\rangle \qquad u \neq 1 \Longrightarrow \sum_{\chi} \chi(u) = 0.$$

To see this, let T denote the sum; if you choose an arbitrary character χ and write out $\chi(u)T$, then you discover, after a change-of-variable (or change-of-character!), that you get back T; in short, $\chi(u)T = T$ for all χ. In particular, choosing $\chi = \omega$, we have $\chi(u) \neq 1$, forcing $T = 0$. A similar argument shows that $\chi \neq \varepsilon \Longrightarrow \sum_u \chi(u) = 0$; this time the character is held fixed while its argument ranges across $\mathbf{F}_q^\times$.

For any multiplicative character χ, extend its domain of definition to all of $\mathbf{F}_q$ by defining $\chi(0) = 0$ for nontrivial characters, and $\varepsilon(0) = 1$ for the trivial character. Warning: if the product of two nontrivial characters is trivial, say $\chi_1\chi_2 = \varepsilon$, then $(\chi_1\chi_2)(0) \neq \chi_1(0)\chi_2(0)$. Here is the first interesting result:

$$\langle 3\rangle \qquad \mathbf{N}(x^n = u) = \mathbf{N}(x^d = u) = \sum_{0 \le k < d} \omega_d^k(u).$$

The sum is taken across all characters of order dividing d, i.e., across the character subgroup of order d. The result clearly holds for $u = 0$. For $u \neq 0$, we must consider, as before, whether u is a dth power. We saw earlier that by writing $u = w^{td+r}$, with $0 \le r < d$, then u is a dth power precisely when $r = 0$. In terms of characters, this may be stated as follows: u is a dth power precisely when each character of order dividing d assigns 1 to u, or more simply, when $\omega_d(u) = 1$. After all, if $r = 0$, then $\omega_d(w^{td}) = \omega_d(w^t)^d = 1$, while if $r \neq 0$, then $\omega_d(w^{td+r}) = \omega_d(w^r) = \zeta^{r(q-1)/d}$, which is not 1 since the exponent is not a multiple of $q-1$. Now back to the displayed result: if u is a dth power, then each term is 1, giving the desired sum of d. For the remaining case, letting T denote the sum in $\langle 3\rangle$, an argument similar to that used for $\langle 2\rangle$ shows $\omega_d(u)T = T$, and since u is not a dth power, $\omega_d(u) \neq 1$, forcing $T = 0$.

From what we know thus far, we can count the solutions to the original equation $\langle 1\rangle$ over $\mathbf{F}_q$ as follows. Let N denote the number of solutions. Then, putting $d_i = \mathrm{GCD}(n_i, q-1)$, we have

$$\begin{aligned} N &= \mathbf{N}(a_0x_0^{n_0} + a_1x_0^{n_1} + \cdots + a_rx_r^{n_r} = b) \\ &= \sum_{\Sigma a_iu_i = b} \mathbf{N}(x^{n_0} = u_0)\mathbf{N}(x^{n_1} = u_1)\cdots\mathbf{N}(x^{n_r} = u_r) \\ &= \sum_{\Sigma a_iu_i = b} \sum_{0 \le k_0 < d_0} \omega_{d_0}^{k_0}(u_0) \sum_{0 \le k_1 < d_1} \omega_{d_1}^{k_1}(u_1) \cdots \sum_{0 \le k_r < d_r} \omega_{d_r}^{k_r}(u_r) \\ &= \sum_{0 \le k_i < d_i} \sum_{\Sigma a_iu_i = b} \omega_{d_0}^{k_0}(u_0)\omega_{d_1}^{k_1}(u_1)\cdots\omega_{d_r}^{k_r}(u_r). \end{aligned}$$

When each index k_i is 0, the rightmost character product is trivially equal to 1. The inner sum runs over q^r distinct tuples of the u_i: having assigned arbitrary values to $u_1, \ldots, u_r$, the value of u_0 is uniquely determined. Thus, in the case that each index k_i is 0, the sum is q^r.

I now claim that when some indices are 0, but other indices are not 0, then the sum is 0. In fact, consider the case that $k_0 = 0$ but $k_1 \neq 0$; I wish to show $\sum_{\Sigma a_iu_i = b} \varepsilon(u_0)\omega_{d_1}^{k_1}(u_1)\cdots\omega_{d_r}^{k_r}(u_r) = 0$. Arbitrary assignment of $u_1, \ldots, u_r$ forces the value of u_0, but each term of the sum is independent of u_0, and thus our sum is simply

$\sum_{u_1,\dots,u_r} \omega_{d_1}^{k_1}(u_1)\cdots\omega_{d_r}^{k_r}(u_r)$, where each u_i freely runs over $\mathbf{F}_q$. This sum is now seen to be 0, as we can pull out the factor $\sum_{u_1}\omega_{d_1}^{k_1}(u_1)$, which is 0 by ⟨2⟩. Summary: the presence of at least one trivial character allowed us to free the sum, and the presence of at least one nontrivial character forced the resulting free sum to be 0.

Going back to the displayed expression for N, we can separate the case when each index k_i is 0, and ignore the case when some indices are 0 while others are not, to obtain

$$N = q^r + \sum_{0<k_i<d_i} \sum_{\Sigma a_i u_i = b} \omega_{d_0}^{k_0}(u_0)\omega_{d_1}^{k_1}(u_1)\cdots\omega_{d_r}^{k_r}(u_r).$$

If $b=0$, introduce the change-of-variables $v_i = a_i u_i$, while if $b\neq 0$, introduce the change-of-variables $v_i = a_i u_i/b$, to obtain

$$N_{b=0} = q^r + \sum_{0<k_i<d_i} \omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\overbrace{\sum_{\Sigma v_i=0} \omega_{d_0}^{k_0}(v_0)\omega_{d_1}^{k_1}(v_1)\cdots\omega_{d_r}^{k_r}(v_r)}^{J_0},$$
$$N_{b\neq 0} = q^r + \sum_{0<k_i<d_i} \omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\omega_{d_0}^{k_0}\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r}(b)\underbrace{\sum_{\Sigma v_i=1} \omega_{d_0}^{k_0}(v_0)\omega_{d_1}^{k_1}(v_1)\cdots\omega_{d_r}^{k_r}(v_r)}_{J}.$$

Here I have denoted certain sums using J_0 and J. In general, for a collection of characters $\chi_0, \dots, \chi_r$, define the sums

$$J_0(\chi_0,\dots,\chi_r) = \sum_{\Sigma v_i=0} \chi_0(v_0)\chi_1(v_1)\cdots\chi_r(v_r),$$
$$J(\chi_0,\dots,\chi_r) = \sum_{\Sigma v_i=1} \chi_0(v_0)\chi_1(v_1)\cdots\chi_r(v_r).$$

The latter sum is known as a Jacobi sum. I now show, in the case that all characters are nontrivial, how to rewrite J_0 in terms of J; thus, even when $b=0$, the number of solutions to our original equation ⟨1⟩ can always be written in terms of Jacobi sums. The crucial fact is

$$\langle 4\rangle \qquad \chi_i \neq \varepsilon, \chi_0\chi_1\cdots\chi_r \neq \varepsilon \Longrightarrow J_0(\chi_0,\dots,\chi_r) = 0,$$
$$\langle 5\rangle \qquad \chi_i \neq \varepsilon, \chi_0\chi_1\cdots\chi_r = \varepsilon \Longrightarrow J_0(\chi_0,\dots,\chi_r) = (q-1)\overbrace{\chi_1\cdots\chi_r(-1)}^{\chi_0(-1)}J(\chi_1,\dots,\chi_r).$$

Observe that in the interesting case $\chi_0\chi_1\cdots\chi_r = \varepsilon$, in the process of rewriting J_0 in terms of a Jacobi sum, one character (it doesn't matter which) disappears. Let me now establish the claim. Start with

$$J_0 = \sum_{v_0} \chi_0(v_0) \sum_{v_1+\cdots+v_r=-v_0} \chi_1(v_1)\cdots\chi_r(v_r).$$

Since $\chi_0\neq\varepsilon$, we have $\chi_0(0)=0$. Introducing the change-of-variables $u_i = -v_i/v_0$, we obtain

$$\begin{aligned} J_0 &= \sum_{v_0\neq 0} \chi_0(v_0) \sum_{u_1+\cdots+u_r=1} \chi_1(-u_1v_0)\cdots\chi_r(-u_rv_0) \\ &= \sum_{v_0\neq 0} \chi_1\cdots\chi_r(-1)\chi_0\chi_1\cdots\chi_r(v_0) \overbrace{\sum_{u_1+\cdots+u_r=1}\chi_1(u_1)\cdots\chi_r(u_r)}^{J(\chi_1,\dots,\chi_r)} \\ &= \chi_1\cdots\chi_r(-1)J(\chi_1,\dots,\chi_r)\sum_{v_0\neq 0}\chi_0\chi_1\cdots\chi_r(v_0). \end{aligned}$$

The inner sum is either $q-1$ or 0; the former occurs when $\chi_0\chi_1\cdots\chi_r$ is trivial. This establishes ⟨4⟩ and ⟨5⟩.

Going back to our expression for $N_{b=0}$, in which J_0 occurred, we now ask for which tuples of indices k_i do we have $\omega_{d_0}^{k_0}\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r} = \varepsilon$? The character in question is $\omega^{k_0(q-1)/d_0 + k_1(q-1)/d_1+\cdots+k_r(q-1)/d_r}$, and this is trivial precisely when the exponent is a multiple of $q-1$, which happens precisely when $\sum k_i/d_i$ is an integer. Hence we obtain

$$\langle 6\rangle \qquad N_{b=0} = q^r + (q-1)\sum_{\substack{0<k_i<d_i \\ \Sigma k_i/d_i\in\mathbf{Z}}} \omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r}(-1)J(\omega_{d_1}^{k_1},\dots,\omega_{d_r}^{k_r}),$$

$$\langle 7\rangle \qquad N_{b\neq 0} = q^r + \sum_{0<k_i<d_i} \omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\omega_{d_0}^{k_0}\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r}(b)J(\omega_{d_0}^{k_0},\omega_{d_1}^{k_1},\dots,\omega_{d_r}^{k_r}).$$

There is a different Jacobi sum for each collection of characters—computing N using the formulas above can be quite tedious! On the other hand, associated to each single character is a Gauss sum, to be defined below. Thus there are precisely $q-1$ Gauss sums. Even better, Jacobi sums can be expressed in terms of Gauss sums. Thus, having computed the Gauss sums once and for all, Jacobi sums are easily computed as they arise.

Choose, once and for all, a nontrivial additive character $\psi\colon \mathbf{F}_q^+ \to \mathbf{C}^\times$; for example, fix $\xi = e^{2\pi i/p}$, and set $\psi(u) = \xi^{\mathrm{TRACE}(u)}$, where $\mathrm{TRACE}\colon \mathbf{F}_q^+ \to \mathbf{F}_p^+ = (\mathbf{Z}/p\mathbf{Z})^+$ is the trace map, defined by $\mathrm{TRACE}(u) = u^p + u^{p^2} + \cdots + u^q$. For any multiplicative character χ, define the Gauss sum

$$g(\chi) = \sum_{u\in\mathbf{F}_q} \chi(u)\psi(u).$$

The result $g(\varepsilon) = 0$ is immediately established using the argument from $\langle 2\rangle$. For nontrivial χ, we have:

$$\langle 8\rangle \qquad \chi \neq \varepsilon \Longrightarrow g(\chi)\overline{g(\chi)} = |g(\chi)|^2 = q,$$

$$\langle 9\rangle \qquad \chi \neq \varepsilon \Longrightarrow g(\chi)g(\overline{\chi}) = q\chi(-1).$$

These results are established by calculation:

$$\begin{aligned} g(\chi)\overline{g(\chi)} &= \sum_{u\neq 0}\chi(u)\psi(u)\sum_{v\neq 0}\overline{\chi}(v)\overline{\psi}(v) = \sum_{u\neq 0}\chi(u)\psi(u)\sum_{v\neq 0}\overline{\chi}(uv)\overline{\psi}(uv) = \sum_{u\neq 0}\sum_{v\neq 0}\overline{\chi}(v)\psi(u-uv)\\ &= \underbrace{\sum_{v\neq 0}\overline{\chi}(v)\sum_{u\neq 0}\psi(u(1-v))}_{q-1 \text{ when } v=1} = q-1+\sum_{v\neq 0,1}\overline{\chi}(v)\underbrace{\sum_{u\neq 0}\psi(u)}_{-1} = q-1-\underbrace{\sum_{v\neq 0,1}\overline{\chi}(v)}_{-1} = q. \end{aligned}$$

$$g(\chi)\overline{g(\chi)} = g(\chi)\sum_{u}\overline{\chi}(u)\overline{\psi}(u) = g(\chi)\sum_{u}\overline{\chi}(-u)\overline{\psi}(-u) = g(\chi)\sum_{u}\overline{\chi}(u)\overline{\chi}(-1)\psi(u) = g(\chi)\overline{\chi}(-1)\overbrace{\sum_{u}\overline{\chi}(u)\psi(u)}^{g(\overline{\chi})}.$$

Gauss sums have been fun thus far by themselves, but we are really interested in writing our Jacobi sums in terms of Gauss sums. We only need to do this when all characters are nontrivial, and the result is

$$\langle 10\rangle \qquad \chi_i \neq \varepsilon, \chi_0\chi_1\cdots\chi_r \neq \varepsilon \Longrightarrow J(\chi_0,\chi_1,\ldots,\chi_r) = \frac{g(\chi_0)g(\chi_1)\cdots g(\chi_r)}{g(\chi_0\chi_1\cdots\chi_r)},$$

$$\begin{aligned} \langle 11\rangle \qquad \chi_i \neq \varepsilon, \chi_0\chi_1\cdots\chi_r = \varepsilon \Longrightarrow J(\chi_0,\chi_1,\ldots,\chi_r) &= -\chi_0(-1)J(\chi_1,\ldots,\chi_r)\\ &= -\chi_0(-1)\frac{g(\chi_1)\cdots g(\chi_r)}{g(\chi_1\cdots\chi_r)}. \end{aligned}$$

To see this, start with

$$\begin{aligned} g(\chi_0)g(\chi_1)\cdots g(\chi_r) &= \sum_{u_0}\chi_0(u_0)\psi(u_0)\sum_{u_1}\chi_1(u_1)\psi(u_1)\cdots\sum_{u_r}\chi_r(u_r)\psi(u_r)\\ &= \sum_{u_i}\chi_0(u_0)\chi_1(u_1)\cdots\chi_r(u_r)\psi(\underbrace{u_0+u_1+\cdots+u_r}_{t}) = \sum_{t}\psi(t)\underbrace{\sum_{\Sigma u_i=t}\chi_0(u_0)\chi_1(u_1)\cdots\chi_r(u_r)}_{J_0 \text{ when } t=0}\\ &= J_0(\chi_0,\ldots,\chi_r) + \sum_{t\neq 0}\psi(t)\sum_{\Sigma u_i=t}\chi_0(u_0)\chi_1(u_1)\cdots\chi_r(u_r)\\ &= J_0(\chi_0,\ldots,\chi_r) + \sum_{t\neq 0}\psi(t)\sum_{\Sigma v_i=1}\chi_0(tv_0)\chi_1(tv_1)\cdots\chi_r(tv_r) \qquad (\text{here } v_i = u_i/t)\\ &= J_0(\chi_0,\ldots,\chi_r) + \underbrace{\sum_{t\neq 0}\psi(t)\chi_0(t)\chi_1(t)\cdots\chi_r(t)}_{T}\underbrace{\sum_{\Sigma v_i=1}\chi_0(v_0)\chi_1(v_1)\cdots\chi_r(v_r)}_{J(\chi_0,\ldots,\chi_r)}. \end{aligned}$$

In the case $\chi_0\chi_1\cdots\chi_r \neq \varepsilon$, the earlier result $\langle 4\rangle$ gives $J_0 = 0$, while the sum T above is just $g(\chi_0,\chi_1,\ldots,\chi_r)$, thus establishing $\langle 10\rangle$. Consider now the case $\chi_0\chi_1\cdots\chi_r = \varepsilon$. Result $\langle 5\rangle$ gives $J_0 = (q-1)\chi_0(-1)J(\chi_1,\ldots,\chi_r)$, while the sum T, being one trivial term short of a trivial Gauss sum, is -1. Now we work on the product of Gauss sums on the left side. Temporary removal of one character allows application of $\langle 10\rangle$ to obtain

$$\begin{aligned} g(\chi_1)\cdots g(\chi_r) &= g(\chi_1\cdots\chi_r)J(\chi_1,\ldots,\chi_r),\\ g(\chi_0)g(\chi_1)\cdots g(\chi_r) &= \underbrace{g(\chi_0)g(\overline{\chi}_0)}_{q\chi_0(-1)}J(\chi_1,\ldots,\chi_r).\end{aligned}$$

Substituting the pieces into our work above gives

$$q\chi_0(-1)J(\chi_1,\ldots,\chi_r) = (q-1)\chi_0(-1)J(\chi_1,\ldots,\chi_r) - J(\chi_0,\chi_1,\ldots,\chi_r),$$

which is the same as $\langle 11\rangle$.

Looking back at $\langle 6\rangle$ and $\langle 7\rangle$, the Jacobi sums are now easily computed in terms of Gauss sums, and so $\langle 6\rangle$ and $\langle 7\rangle$ provide useful formulas for computing $N_{b=0}$ and $N_{b\neq 0}$. Since q was arbitrary, the same results apply in $\mathbf{F}_{q^\nu}$, keeping in mind that the Gauss sums and Jacobi sums must be defined using characters on $\mathbf{F}_{q^\nu}$. It turns out that Gauss sums and Jacobi sums defined over $\mathbf{F}_{q^\nu}$ can be calculated in terms of Gauss sums and Jacobi sums defined over $\mathbf{F}_q$. I shall now describe how this goes.

As usual, the trace and norm maps compute the sum and product of all conjugates (with multiplicities):

$$\begin{aligned} \text{TRACE}\colon \mathbf{F}_{q^\nu}^+ \to \mathbf{F}_q^+, &\qquad u\mapsto u^p + u^{p^2} + \cdots + u^q,\\ \text{NORM}\colon \mathbf{F}_{q^\nu}^\times \to \mathbf{F}_q^\times, &\qquad u\mapsto u^{p+p^2+\cdots+q}.\end{aligned}$$

The important fact in finite fields is that each map is surjective. Therefore, if ψ is a nontrivial additive character on $\mathbf{F}_q$, then $\psi'\colon u\mapsto \psi(\text{TRACE}(u))$ is a nontrivial additive character on $\mathbf{F}_{q^\nu}$. Similarly, given a multiplicative character χ on $\mathbf{F}_q$, one defines a multiplicative character $\chi'\colon u\mapsto\chi(\text{NORM}(u))$. This sets up a one-to-one correspondence between characters of a given order on $\mathbf{F}_q$ and characters of the same order on $\mathbf{F}_{q^\nu}$. Let g' and J' denote the Gauss and Jacobi sums over $\mathbf{F}_{q^\nu}$. The Hasse–Davenport theorem asserts

$$\langle 12\rangle \qquad -g'(\chi') = (-g(\chi))^\nu.$$

An elegant proof of this result appears in the book by Ireland–Rosen. Since the proof gives no insight into counting solutions to equations over finite fields, I have not included it here. As a consequence of $\langle 12\rangle$, a careful counting of signs shows that the separate results $\langle 10\rangle$ and $\langle 11\rangle$ become the single result†

$$\langle 13\rangle \qquad \chi_i\neq\varepsilon \Longrightarrow J'(\chi_0',\chi_1',\ldots,\chi_r') = (-1)^{r(\nu+1)}J(\chi_0,\chi_1,\ldots,\chi_r)^\nu.$$

We should now be in a position to modify formulas $\langle 6\rangle$ and $\langle 7\rangle$ to hold over $\mathbf{F}_{q^\nu}$, but there is one remaining obstacle. Those formulas involve $d_i = \text{GCD}(n_i, q-1)$, and it is not immediately clear how to calculate $\text{GCD}(n_i, q^\nu - 1)$ in terms of d_i. So we cheat: let us restrict attention to the case that each $n_i \mid q-1$. Then $d_i = n_i$, and our problem is solved. Letting N_ν denote the number of solutions to $\langle 1\rangle$ over $\mathbf{F}_{q^\nu}$, applying $\langle 13\rangle$ to results $\langle 6\rangle$ and $\langle 7\rangle$ gives:

$\langle 14\rangle \quad b=0, n_i\mid q-1 \Longrightarrow$

$$N_\nu = q^{\nu r} + (q^\nu-1)(-1)^{(r-1)(\nu+1)}\sum_{\substack{0<k_i<n_i\\ \Sigma k_i/n_i\in\mathbf{Z}}}\Bigl(\omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r}(-1)J(\omega_{d_1}^{k_1},\ldots,\omega_{d_1}^{k_r})\Bigr)^\nu,$$

$\langle 15\rangle \quad b\neq 0, n_i\mid q-1 \Longrightarrow$

$$N_\nu = q^{\nu r} + (-1)^{r(\nu+1)}\sum_{0<k_i<n_i}\Bigl(\omega_{d_0}^{-k_0}(a_0)\omega_{d_1}^{-k_1}(a_1)\cdots\omega_{d_r}^{-k_r}(a_r)\omega_{d_0}^{k_0}\omega_{d_1}^{k_1}\cdots\omega_{d_r}^{k_r}(b)J(\omega_{d_0}^{k_0},\omega_{d_1}^{k_1},\ldots,\omega_{d_r}^{k_r})\Bigr)^\nu.$$

Finally, if instead of counting the precise number of solutions one merely wishes a good estimate, then this can be obtained as follows. Applying the measurement in $\langle 8\rangle$ to results $\langle 10\rangle$ and $\langle 11\rangle$ gives

$$\begin{aligned} \chi_i\neq\varepsilon, \chi_0\chi_1\cdots\chi_r\neq\varepsilon &\Longrightarrow |J(\chi_0,\chi_1,\ldots,\chi_r)|^2 = q^r,\\ \chi_i\neq\varepsilon, \chi_0\chi_1\cdots\chi_r=\varepsilon &\Longrightarrow |J(\chi_0,\chi_1,\ldots,\chi_r)|^2 = q^{r-1}.\end{aligned}$$

† I have not seen this elegant fact stated elsewhere.

Applying this measurement to $\langle 14 \rangle$ gives

$$|N_{\nu,b=0} - q^{\nu r}| \leq M q^{\nu(r-1)/2},$$

where M is the number of tuples of k_i satisfying $0 < k_i < d_i$, $\sum k_i/d_i \in \mathbf{Z}$, and where $d_i = \operatorname{GCD}(n_i, q^\nu - 1)$. Since M is bounded (by $\prod n_i$, for example), this result shows that N_ν doesn't deviate from $q^{\nu r}$ too much as $\nu \to \infty$; in particular, there are always solutions if ν is taken sufficiently large. (Observe also that the restriction $n_i \mid q-1$, important in $\langle 14 \rangle$, is not required here.)

Applying the Jacobi sum measurement to $\langle 15 \rangle$ gives

$$|N_{\nu,b\neq 0} - q^{\nu r}| \leq M_\varepsilon q^{\nu(r-1)/2} + (M - M_\varepsilon) q^{\nu r/2},$$

where M is the total number of tuples of k_i satisfying $0 < k_i < d_i$, and M_ε is the number of those tuples that furthermore satisfy $\sum k_i/d_i \in \mathbf{Z}$. As before, these quantities are bounded, and we always get solutions if ν is taken sufficiently large.

TABLE OF NOTATION

Symbol	Explanation
p	Fixed prime number
q	Some power of p
$\mathbf{F}_q$	Finite field of q elements
$\mathbf{F}_{q^\nu}$	Degree ν extension of $\mathbf{F}_q$
$\mathbf{N}(\text{eqn})$	Number of solutions over $\mathbf{F}_q$ of the given equation
χ	Arbitrary multiplicative character $\mathbf{F}_q^\times \to \mathbf{C}^\times$
w	Fixed generator of multiplicative group $\mathbf{F}_q^\times$
ζ	Fixed primitive root of unity of order $q-1$
ω	Multiplicative character $w \mapsto \zeta$ of order $q-1$
d	Some divisor of $q-1$
ω_d	Multiplicative character $w \mapsto \zeta^{(q-1)/d}$ of order d
ε	Trivial multiplicative character $w \mapsto 1$
ξ	Primitive root of unity of order p
ψ	Fixed nontrivial additive character $\mathbf{F}_q^+ \to \mathbf{C}^\times$
$g(\chi)$	Gauss sum associated to a multiplicative character
$J(\chi,\ldots)$	Jacobi sum associated to multiplicative characters
$J_0(\chi,\ldots)$	Variant of Jacobi sum