ON THE NUMBER OF NILPOTENT MATRICES WITH COEFFICIENTS IN A FINITE FIELD

BY

MURRAY GERSTENHABER

Fine and Herstein have demonstrated [1] that the number of nilpotent $n \times n$ matices with coefficients in the finite field of q elements, $GF(q)$, is q^{n^2-n}. The present (self-contained) note gives an alternate proof, suggested by algebraic geometry, and not involving sums over partitions of n. Using a lemma of [1], Reiner [2] has determined the number of matrices over $GF(q)$ having a given characteristic polynomial. This result is here obtained directly from the Fine-Herstein theorem.

1. Proof of the Fine-Herstein theorem

Throughout, N_k will denote the $k \times k$ matrix having zeros everywhere but on the first diagonal above the principal one and unity everywhere there. If $k = n$, we write simply N. Given a nilpotent $n \times n$ matrix A, we shall denote by $L(A)$ the linear space of all matrices Y such that $NY = YA$, and by $\mathfrak{A}$ the union of the spaces $L(A)$ for all nilpotent A. The matrices in $\mathfrak{A}$ will be called admissible. One sees that $L(T^{-1}AT) = L(A)T$, whence Y is admissible if and only if YT is admissible for any nonsingular T.

We determine now a necessary and sufficient condition that $Y \in \mathfrak{A}$. Given Y, let v be a row vector such that $vY = 0$. If $Y \in \mathfrak{A}$, then $vNY = vYA$ for some A, so $(vN)Y = 0$, i.e., the null space of Y is preserved by N. Let $v = (v_1, \cdots, v_n)$. Then $vN = (0, v_1, \cdots, v_{n-1})$. This implies that if the rank of Y is r, then the last $n - r$ rows of Y are zero (and the first r, therefore, are independent). Conversely, suppose Y has this property. Then for some nonsingular T, $YT = E$ is the direct sum of the $r \times r$ identity matrix I_r and the $(n - r) \times (n - r)$ zero matrix O_{n-r}. Now E is admissible, for $NE = N_r \oplus O_{n-r}$ is nilpotent and $NE = E(NE)$. Therefore $Y = ET^{-1}$ is admissible. The necessary and sufficient condition that $Y \in \mathfrak{A}$ is therefore

If rank $Y = r$, then the last $n - r$ rows of Y vanish.

We see next that $\dim L(A) = n$ for all nilpotent A. Observe that $NY = YA$ implies $N^kY = YA^k$ for all k. Let e_i denote the row vector having one in the i^{th} place and zeros elsewhere. Then $e_1 Y$ (i.e., the first row of Y) may be prescribed arbitrarily, but $e_k Y$ is then determined for all k by the relation $e_k Y = e_1 N^{k-1}Y = e_1 YA^{k-1}$.

Given $Y \in \mathfrak{A}$, let there be assigned to it a multiplicity $m(Y)$ equal to the number of distinct nilpotent matrices A such that $NY = YA$. If O is the zero matrix, then $m(O)$ is just the number of all nilpotent matrices, which we

Received May 18, 1960; received in revised form October 6, 1960.

wish to determine. We shall denote this quantity by u. Consider the correspondence f which assigns to a nilpotent A the set of all elements of $L(A)$. This f is, by our previous result, exactly q^n-valued for every nilpotent A, and if $Y \in \mathfrak{A}$, then $m(Y)$ is just the number of elements in $f^{-1}(Y)$. Therefore $\sum m(Y) = q^n u$, the sum being taken over all $Y \in \mathfrak{A}$. Since $m(O) = u$, we may write $\sum' m(Y) = (q^n - 1)u$, where in $\sum'$ the zero matrix is omitted. Now $m(Y)$ and $m(YT)$ are identical for any nonsingular T. If Y has rank r, then we have seen that the last $n - r$ rows of Y vanish, and for suitable T we have $YT = E = I_r \oplus O_{n-r}$. Therefore m depends only on the rank of Y, so we may define $m(r) = m(E) = m(Y)$ for any $Y \in \mathfrak{A}$ with rank $Y = r$. Let $\alpha(r)$ denote the number of elements of $\mathfrak{A}$ with rank r; this is just the number of matrices whose first r rows are independent and last $n - r$ rows vanish. Then

$$(q^n - 1)u = \sum' m(Y) = \sum_{r=1}^{n} \alpha(r)m(r).$$

We compute $\alpha(r)$ by observing that the first row of any $Y \in \mathfrak{A}$ of rank r ($\geqq 1$) is anything but zero, for which vector there are $q^n - 1$ possibilities, and that for $k \leqq r$ the k^{th} row may be any vector not in the space spanned by the first $k - 1$ rows, for which vector there are therefore $q^n - q^{k-1}$ possibilities. The last $n - r$ rows are determined, being all zero. It follows that

$$\alpha(r) = (q^n - 1)(q^n - q) \cdots (q^n - q^{r-1}).$$

To compute $m(r) = m(E)$, observe that $NE = EA$ for some A if and only if A is of the form

$$\begin{pmatrix} N_r & O \\ P & Q \end{pmatrix}$$

where P and Q are matrices of dimensions $r \times (n - r)$ and $(n - r) \times (n - r)$, respectively. Therefore $m(r)$ is the number of nilpotent matrices of this form. Such a matrix is nilpotent if and only if Q is nilpotent. Denoting by $u(k)$ the number of nilpotent $k \times k$ matrices with coefficients in $GF(q)$, let us make the inductive assumption that $u(k) = q^{k^2-k}$ for $k < n$. It follows that

$$m(r) = q^{r(n-r)} q^{(n-r)^2-(n-r)} = q^{(n-r)(n-1)} \quad \text{for} \quad 1 \leqq r \leqq n.$$

Writing the terms in $\sum_{r=1}^{n} \alpha(r)m(r)$ in reverse order, starting with the term for $r = n$, we now have

$$\begin{aligned} \sum_{r=1}^{n} \alpha(r)m(r) = {} & (q^n - 1)(q^n - q) \cdots (q^n - q^{n-3})(q^n - q^{n-2})(q^n - q^{n-1}) \\ & + (q^n - 1)(q^n - q) \cdots (q^n - q^{n-3})(q^n - q^{n-2})q^{n-1} \\ & + (q^n - 1)(q^n - q) \cdots (q^n - q^{n-3})q^{2(n-1)} \\ & + \cdots \\ & + (q^n - 1)q^{(n-1)^2}. \end{aligned}$$

The sum of the first term and the second is

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-2})q^n,$$

the sum of this and the third is

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-3})q^{2n},$$

and continuing so, one finds the sum of all to be $(q^n - 1)q^{(n-1)n}$. Therefore

$$(q^n - 1)q^{(n-1)n} = (q^n - 1)u,$$

i.e., $u = q^{n^2-n}$, completing the induction and the proof.

2. Determination of the number of matrices over $GF(q)$ with given characteristic polynomial

We determine first the number of $md \times md$ matrices over $GF(q)$ satisfying an equation $f(x)^m = 0$, where f is an irreducible polynomial of degree d. The set of all $r \times r$ matrices with coefficients in $GF(q)$ will be denoted by $GF(q)_r$, and the number of nonsingular ones by

$$\beta(q, r) = (q^r - 1)(q^r - q) \cdots (q^r - q^{r-1}).$$

Let σ be a fixed representation of $GF(q^d)$ in $GF(q)_d$, $\sigma^{(m)}$ the naturally induced representation of $GF(q^d)_m$ in $GF(q)_{md}$, and λ a fixed root of $f(x)$. Set $\sigma(\lambda) = M$. An element of $GF(q)_d$ is in the image of σ if and only if it commutes with M; likewise, an element of $GF(q^d)_m$ is in the image of $\sigma^{(m)}$ if and only if it commutes with $\sigma^{(m)}(\lambda I_m) = M \otimes I_m$. If $R \in GF(q^d)_m$ has coefficients in $GF(q)$, then $\sigma^{(m)}(R) = I_d \otimes R$. In particular, this is the case if R is a nilpotent matrix in Jordan normal form, i.e., a direct sum of matrices of the form N_k. For such an R, the Jordan normal form of $\sigma^{(m)}(\lambda I_m + R)$ is, writing λ for λI_m, $(\lambda_1 + R) \oplus \cdots \oplus (\lambda_d + R)$, where $\lambda_1, \cdots, \lambda_d$ are the zeros of $f(x)$ (distinct since $GF(q)$ is perfect). On the other hand, if A is any matrix in $GF(q)_{md}$ satisfying $f(x)^m = 0$, then the Jordan normal form of A is also of the form $(\lambda_1 + R) \oplus \cdots \oplus (\lambda_d + R)$, where R is some nilpotent matrix in Jordan normal form, the nilpotent part associated with each proper value λ_i being the same since $\lambda_i \to \lambda_j$ induces an automorphism of $GF(q^d)$ over $GF(q)$. A fortiori, every A in $GF(q)_{md}$ satisfying $f(x)^m = 0$ is similar to $\sigma^{(m)}(\lambda + P)$ for some nilpotent P in $GF(q^d)_m$, and on the other hand it is clear that matrices of the latter form all satisfy $f(x)^m = 0$. If C is a nonsingular matrix in $GF(q)_{md}$ and P, P' nilpotent matrices in $GF(q^d)_m$ such that

$$C^{-1}\sigma^{(m)}(\lambda + P)C = \sigma^{(m)}(\lambda + P'),$$

then in fact C commutes with $\sigma^{(m)}(\lambda)$, for

$$(\lambda + P)^{q^{md}} = (\lambda + P')^{q^{md}} = \lambda.$$

Therefore C is in the image of $\sigma^{(m)}$, and conjugation by C carries the set of all matrices $\sigma^{(m)}(\lambda + P)$, P nilpotent, onto itself. The number of such C is the

number of nonsingular matrices in $GF(q^d)_m$, namely, $\beta(q^d, m)$. There being $\beta(q, md)$ nonsingular matrices in $GF(q)_{md}$, and, by the Fine-Herstein theorem, $(q^d)^{m^2-m}$ nilpotent matrices P in $GF(q^d)_m$, it follows that the number of solutions of $f(x)^m = 0$ in $GF(q)_{md}$ is $(q^d)^{m^2-m}\beta(q, md)/\beta(q^d, m)$.

Finally, let A be an element of $GF(q)_n$ whose characteristic polynomial is $f = f_1^{m_1} \cdots f_k^{m_k}$, where the f_i are distinct irreducible polynomials of degree d_i over $GF(q)$, and necessarily $\sum m_i d_i = n$. Then A is similar to a direct sum $A_1 \oplus \cdots \oplus A_k$, where the characteristic polynomial of A_i is $f_i^{m_i}$. Every such direct sum has characteristic polynomial f, and if two such are similar, then the matrix C effecting the similarity must itself be a direct sum $C = C_1 \oplus \cdots \oplus C_k$, where the dimensions of C_i are $m_i d_i \times m_i d_i$. Letting t_i denote the number of elements of $GF(q)_{m_i d_i}$ satisfying $f_i^{m_i} = 0$, it follows that the number t of elements of $GF(q)_n$ whose characteristic polynomial is f must be $\prod t_i \cdot \beta(q, n)/\prod \beta(q, m_i d_i)$. Substituting for t_i its value from the preceding paragraph, one has (after cancellations)

$$t = \prod (q^{d_i})^{m_i^2-m} \cdot \beta(q, n)/\prod \beta(q^{d_i}, m_i).$$

If one sets $F(q, r) = q^{-r^2}\beta(q, r) = (1 - q^{-1})(1 - q^{-2}) \cdots (1 - q^{-r})$, then one may also write, observing that $\sum m_i d_i = n$,

$$t = q^{n^2-n}F(q, n)/\prod F(q^{d_i}, m_i).$$

References

1. N. J. Fine and I. N. Herstein, *The probability that a matrix be nilpotent*, Illinois J. Math., vol. 2 (1958), pp. 499–504.
2. I. Reiner, *On the number of matrices with given characteristic polynomial*, Illinois J. Math., vol. 5 (1961), pp. 324–329.

University of Pennsylvania
Philadelphia, Pennsylvania