JOURNAL OF COMBINATORIAL THEORY (A) 22, 107–109 (1977)

Note

On the Sequential and Random Selection of Subspaces over a Finite Field

EUGENIO CALABI* AND HERBERT S. WILF*

Department of Mathematics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Communicated by the Managing Editors

Received February 2, 1976

Let V_n be an n-dimensional vector space over $GF(q)$, and let $0 \leqslant k \leqslant n$. We consider the question of generating sets of basis vectors for k-dimensional subspaces V_k of V_n in two ways: (a) the sequential generation of a set of basis vectors for *every* such subspace and (b) the random selection of such a subspace V_k in such a way that all subspaces have equal a priori probabilities of being chosen. Such questions are of interest in coding theory.

The answer to the first question has been known for some time and comes from the row-echelon, or "Schubert," form of the $k \times n$ matrix B whose rows are the basis vectors: We fix a k-subset $(a_1, \ldots, a_k) = S$ of $\{1, 2, \ldots, n\}$. In the columns of B which correspond to S we enter a $k \times k$ identity matrix. In row i of B we enter 0 in all columns $j > a_i$, for $i = 1, \ldots, k$. Finally the

$$N = \sum_{i=1}^{k} a_i - \binom{k}{2} = N(S)$$

remaining entries of B may be assigned independently to elements of the field. Hence for the given subset S we obtain q^N sets of basis vectors, each describing a different subspace V_k. By varying S we obtain all of the desired V_k, and of course

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1) \cdots (q - 1)} = \sum_S q^{N(S)} \qquad (1)$$

is the total number of such V_k.

Before discussing our algorithm for uniformly random selection, we mention, briefly, two others which may come immediately to mind.

* Supported in part by the National Science Foundation.

First, we might imagine fixing a $k \times k$ identity matrix in the first k columns of a $k \times n$ matrix, and choosing the remaining $k(n-k)$ elements independently from the field. This proposal results in nonuniform distribution, and in an essential way. For example, with $k = 2$, $n = 3$ over $GF(2)$, we would choose from the four basis matrices

$$\begin{pmatrix}100\\010\end{pmatrix} \quad \begin{pmatrix}100\\011\end{pmatrix} \quad \begin{pmatrix}101\\010\end{pmatrix} \quad \begin{pmatrix}101\\011\end{pmatrix}$$

with equal probability. Of these four vector spaces, the last has all nonzero vectors of weight (number of 1's) at least 2, while the other three spaces have minimum weight 1. Actually, however, the probability that a random $V_2 \subseteq V_3$ over $GF(2)$ will have minimum weight 2 is $1/7$.

A second proposal might be to select the rows of the matrix independently, and check each one for independence from its predecessors. Here the distribution is uniform, but the independence check of the jth vector will require Cjn operations, for a total of Cnk^2 units of labor altogether. Our algorithm operates in time Cnk which is minimal for the problem since there are nk items of output: the matrix elements. Indeed, as will be seen, at each stage of the algorithm we either fill in a whole row and column with zeros and a single 1, or else we fill in a whole column with independently randomly chosen elements from the field. We, therefore, can output the basis matrix essentially just as fast as we can write it down!

We begin with the well known identity

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q + q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q \tag{2}$$

and we think of the two terms on the right as counting complementary subsets of the totality of basis matrices B in echelon form: the first term counts those $k \times n$ matrices $B = (b_{ij})_{i=1,k;j=1,n}$ for which $b_{11} = 1$, $b_{i1} = 0$ $(1 < i \leqslant k)$, $b_{1j} = 0$ $(1 < j \leqslant n)$ and for which necessarily the remaining $(k-1) \times (n-1)$ array is the basis matrix of a $V_{k-1} \subseteq V_{n-1}$ (i.e., has rank $k-1$). The second term counts the other basis matrices B, i.e., those whose first column is an arbitrary vector of V_k, and for which the remaining $k \times (n-1)$ array is the basis matrix of a $V_k \subseteq V_{n-1}$ (i.e., has rank k).

The fraction of all subspaces which are of the first kind is

$$p = p_{k,n} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \Big/ \begin{bmatrix} n \\ k \end{bmatrix}_q = (q^k - 1)/(q^n - 1), \tag{3}$$

and so if n, k, q are given, the full algorithm is as follows:

(A) $r \leftarrow 1$; $s \leftarrow 1$.

(B) Choose a random number ξ. If $\xi < p_{k+1-r,n+1-s}$, go to (C); Otherwise, choose independently at random from the field the elements b_{is} $(i = r, k)$, set $s \leftarrow s + 1$ and go to (D).

(C) Set $b_{rj} \leftarrow 0$ $(j = s + 1, n)$; $b_{is} \leftarrow 0$ $(i = r + 1, k)$; $b_{rs} \leftarrow 1$; $r \leftarrow r + 1$; $s \leftarrow s + 1$.

(D) If $r \leqslant k$, go to (B); exit. ∎

REFERENCES

1. W. V. D. HODGE AND D. PEDOE, "Methods of Algebraic Geometry," Vol. II, pp. 321–326, Cambridge, 1947.
2. J. W. MILNOR AND J. STASHEFF, Characteristic classes, *in* "Annals of Math., Studies #76," pp. 72–81, Princeton Univ. Press.