Enumerative Coding for Grassmannian Space

Natalia Silberstein and Tuvi Etzion, *Fellow, IEEE*

***Abstract*—The Grassmannian space $\mathcal{G}_q(n,k)$ is the set of all k-dimensional subspaces of the vector space $\mathbb{F}_q^n$. Recently, codes in the Grassmannian have found an application in network coding. The main goal of this paper is to present efficient enumerative encoding and decoding techniques for the Grassmannian. These coding techniques are based on two different orders for the Grassmannian induced by different representations of k-dimensional subspaces of $\mathbb{F}_q^n$. One enumerative coding method is based on a Ferrers diagram representation and on an order for $\mathcal{G}_q(n,k)$ based on this representation. The complexity of this enumerative coding is $O(k^{5/2}(n-k)^{5/2})$ digit operations. Another order of the Grassmannian is based on a combination of an identifying vector and a reduced row echelon form representation of subspaces. The complexity of the enumerative coding, based on this order, is $O(nk(n-k)\log n \log\log n)$ digit operations. A combination of the two methods reduces the complexity on average by a constant factor.**

***Index Terms*—Enumerative coding, Ferrers diagram, Grassmannian, identifying vector, partitions, reduced row echelon form.**

I. Introduction

Let $\mathbb{F}_q$ be a finite field of size q. The *Grassmannian space* (Grassmannian, in short), denoted by $\mathcal{G}_q(n,k)$, is the set of all k-dimensional subspaces of the vector space $\mathbb{F}_q^n$, for any given two integers k and n, $0 \leq k \leq n$. It is well known [1] that $|\mathcal{G}_q(n,k)| = \begin{bmatrix} n \\ k \end{bmatrix}_q$, where $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is a q-ary Gaussian coefficient, defined by

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} \frac{q^{n-i}-1}{q^{k-i}-1}, \tag{1}$$

where $\begin{bmatrix} n \\ 0 \end{bmatrix}_q = 1$, and $\begin{bmatrix} n \\ k \end{bmatrix}_q = 0$ if $k > n$ or $k < 0$.

Coding (and related designs) in the Grassmannian was considered in the last 40 years, e.g., [2]–[8]. Koetter and Kschischang [9] presented an application of error-correcting codes in $\mathcal{G}_q(n,k)$ to random network coding. This application has motivated extensive work in the area [10]–[20]. A natural question is how to encode/decode the subspaces in the Grassmannian in an efficient way. By encoding we mean a transformation of an information word into a k-dimensional subspace. Decoding is the inverse transformation of the k-dimensional subspace into the information word.

To solve this coding problem, we will use the general enumerative coding method which was presented by Cover [21]. Let $\{0,1\}^n$ denote the set of all binary vectors of length n. Let S be a subset of $\{0,1\}^n$. Denote by $n_S(x_1,x_2,\ldots,x_k)$ the number of elements of S for which the first k coordinates are given by $(x_1,x_2,\ldots,x_k)$, where x_1 is the most significant bit. A lexicographic order of S is defined as follows. We say that for $x,y \in \{0,1\}^n$, $x<y$, if $x_k < y_k$ for the least index k such that $x_k \neq y_k$. For example, $00101 < 00110$.

Theorem 1: [21] The lexicographic index (decoding) of $x \in S$ is given by

$$\mathrm{ind}_S(x) = \sum_{j=1}^{n} x_j \cdot n_S(x_1,x_2,\ldots,x_{j-1},0).$$

Let S be a given subset and let i be a given index. The following algorithm finds the unique element x of the subset S such that $\mathrm{ind}_S(x) = i$ (encoding).

Inverse Algorithm [21]: For $k = 1,\ldots,n$, if $i \geq n_S(x_1,x_2,\ldots,x_{k-1},0)$ then set $x_k = 1$ and $i = i - n_S(x_1,x_2,\ldots,x_{k-1},0)$; otherwise set $x_k = 0$.

Remark 1: The coding algorithms of Cover are efficient if $n_S(x_1,x_2,\ldots,x_{j-1},0)$ can be calculated efficiently.

Cover [21] also presented the extension of these results to arbitrary finite alphabets. For our purpose this extension is more relevant as we will see in the sequel. The formula for calculating the lexicographic index of $x \in S \subseteq \{1,2,3,\ldots,M\}^n$ is given as follows:

$$\mathrm{ind}_S(x) = \sum_{j=1}^{n} \sum_{m<x_j} n_S(x_1,x_2,\ldots,x_{j-1},m). \tag{2}$$

Enumerative coding has various applications and it was considered in many papers, e.g., [22]–[24]. Our goal in this paper is to apply this scheme to the set of all subspaces in a Grassmannian, using different lexicographic orders. These lexicographic orders are based on different representations of subspaces. Lexicographic orders also have other applications, e.g., in constructions of lexicographic codes (lexicodes) [25].

The rest of this paper is organized as follows. In Section II we discuss different representations of subspaces in the Grassmannian. We define the reduced row echelon form of a k-dimensional subspace and its Ferrers diagram. These two concepts combined with the identifying vector of a subspace [18] will be our main tools for the representation of subspaces. We also define and discuss some type of partitions which have an important role in our exposition. In Section III we present a new lexicographic order for the Grassmannian based on a representation of a subspace by its identifying vector and its reduced row echelon

Manuscript received November 17, 2009; revised June 03, 2010; accepted August 04, 2010. Date of current version December 27, 2010. This work was supported in part by the Israel Science Foundation (ISF), Jerusalem, under Grant 230/08. This work is part of N. Silberstein's Ph.D. thesis performed at the Technion—Israel Institute of Technology. The material in this paper was presented in part at the 2009 IEEE Information Theory Workshop, Taormina, Sicily, Italy, October 2009.

The authors are with the Department of Computer Science, Technion — Israel Institute of Technology, Haifa 32000, Israel. (email: natalys@cs.technion.ac.il; etzion@cs.technion.ac.il).

Communicated by N. Kashyap, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2010.2090252

form. For this order we describe an enumerative coding method, whose computation complexity is $O(nk(n-k)\log n \log\log n)$ digit operations per subspace. In Section IV we discuss the more intuitive order for the Grassmannian based on Ferrers diagram representation and present a second enumerative coding method for the Grassmannian. In Section V we show how we can combine the two coding methods mentioned above to find a more efficient enumerative coding for the Grassmannian. In Section VI we summarize our results and discuss some related problems.

II. Representation of Subspaces and Partitions

In this section we give the definitions for two concepts which are useful in describing a subspace in $\mathcal{G}_q(n,k)$: Ferrers diagram (which is defined in connection to a partition) and reduced row echelon form. Based on these concepts we present two representations for subspaces from which our enumerative coding techniques will be induced. Representation of subspaces is also important in other problems related to the Grassmannian. For example, in constructing error-correcting codes in the Grassmannian [18], [26].

A *partition* of a positive integer m is a representation of m as a sum of positive integers, not necessarily distinct. We order this collection of integers in a decreasing order. The partition function $p(m)$ is the number of different partitions of m [1], [27], [28].

A *Ferrers diagram* $\mathcal{F}$ represents a partition as a pattern of dots with the ith row having the same number of dots as the ith term in the partition [1], [27], [28] (In the sequel, a *dot* will be denoted by a "•"). A Ferrers diagram satisfies the following conditions.

- The number of dots in a row is at most the number of dots in the previous row.
- All the dots are shifted to the right of the diagram.

Remark 2: Our definition of Ferrers diagram (see [18]) is slightly different form the usual definition [1], [27], [28], where the dots in each row are shifted to the left of the diagram.

A k-dimensional subspace $X \in \mathcal{G}_q(n,k)$ can be represented by a $k \times n$ matrix, whose rows form a basis for X. Such a $k \times n$ matrix is in reduced row echelon form (RREF in short) if the following conditions are satisfied.

- The leading coefficient (pivot) of a row is always to the right of the leading coefficient of the previous row.
- All leading coefficients are *ones*.
- Every leading coefficient is the only nonzero entry in its column.

For a given subspace X, there is exactly one matrix in RREF and it will be denoted by $\mathrm{RE}(X)$. For simplicity, we will assume that the entries in $\mathrm{RE}(X)$ are taken from $\mathbb{Z}_q$ instead of $\mathbb{F}_q$, using an appropriate bijection.

The Ferrers tableaux form of a subspace X, denoted by $\mathcal{F}(X)$, is obtained by removing from each row of $\mathrm{RE}(X)$ the leading coefficient and the *zeroes* to the left of it. All the remaining entries are shifted to the right. $\mathcal{F}(X)$ defines a unique representation of X. The Ferrers diagram of X, denoted by $\mathcal{F}_X$, is obtained from $\mathcal{F}(X)$ by replacing the entries of $\mathcal{F}(X)$ with dots.

Example 1: We consider a three-dimensional subspace X of $\mathbb{F}_2^7$ with the following 3×7 matrix in RREF given by

$$\mathrm{RE}(X) = \begin{pmatrix} 1 & \mathbf{0} & 0 & 0 & \mathbf{1} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 1 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & 0 & 1 & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{pmatrix}.$$

Its Ferrers tableaux form and Ferrers diagram are given by

$$\mathcal{F}(X) = \begin{matrix} 0 & 1 & 1 & 0 \\ & 1 & 0 & 1 \\ & 0 & 1 & 1 \end{matrix} \text{ and } \mathcal{F}_X = \begin{matrix} \bullet & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \\ & & \bullet & \bullet & \bullet \end{matrix}, \text{ respectively.}$$

Let $|\mathcal{F}|$ denote the *size* of a Ferrers diagram $\mathcal{F}$, i.e., the number of dots in $\mathcal{F}$. A Ferrers diagram of a k-dimensional subspace has size at most $k \cdot (n-k)$. It can be embedded in a $k \times (n-k)$ box. Let $p(k,\eta,m)$ be the number of partitions of m whose Ferrers diagram can be embedded into a box of size $k \times \eta$. The following result was given in [27, pp. 33–34].

Lemma 1: $p(k,\eta,m)$ satisfies the following recurrence relation:

$$p(k,\eta,m) = p(k,\eta-1,m-k) + p(k-1,\eta,m) \quad (3)$$

with the initial conditions

$$p(k,\eta,m) = 0 \text{ if } m < 0 \text{ or } m > \eta \cdot k \text{ and } p(k,\eta,0) = 1.$$

Let $\mathcal{F}$ be a Ferrers diagram of size m embedded in a $k \times (n-k)$ box. The number of k-dimensional subspaces whose Ferrers diagram is $\mathcal{F}$, is equal to q^m. By (1) this implies the following theorem [1, p. 327] which shows the connection between the q-ary Gaussian coefficients and partitions.

Theorem 2: For any given integers k and n, $0 < k \le n$,

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \sum_{m=0}^{k(n-k)} \alpha_m q^m,$$

where $\alpha_m = p(k,n-k,m)$.

The order defined in Section IV is based on Theorem 2. We order the subspaces by the size of their Ferrers diagrams. The order of Ferrers diagrams with the same size is explained in Section IV. Two subspaces with the same Ferrers diagrams are ordered lexicographically by their Ferrers tableaux forms. This order seems to be the most natural order of $\mathcal{G}_q(n,k)$. But a less natural representation, which follows, and its related order, will lead to a more efficient enumerative coding.

Each k-dimensional subspace $X \in \mathcal{G}_q(n,k)$ has an *identifying vector* $v(X)$ [18]. $v(X)$ is a binary vector of length n and weight k, where the *ones* in $v(X)$ are exactly in the positions (columns) where $\mathrm{RE}(X)$ has the leading coefficients (of the rows).

Let $X \in \mathcal{G}_q(n,k)$ be a k-dimensional subspace. The *extended representation*, $\mathrm{EXT}(X)$, of X is a $(k+1) \times n$ matrix obtained by combining the identifying vector $v(X) = (v(X)_n, \ldots, v(X)_1)$ and the RREF $\mathrm{RE}(X) = (X_n, \ldots, X_1)$, as follows:

$$\mathrm{EXT}(X) = \begin{pmatrix} v(X)_n & \cdots & v(X)_2 & v(X)_1 \\ X_n & \cdots & X_2 & X_1 \end{pmatrix}.$$

Note that $v(X)_n$ is the most significant bit of $v(X)$. Also, X_i is a column vector and $v(X)_i$ is the most significant bit of the column vector $\begin{pmatrix} v(X)_i \\ X_i \end{pmatrix}$.

Example 2: Consider the three-dimensional subspace X of Example 1. Its identifying vector is $v(X) = 1011000$ and its extended representation is given by

$$\mathrm{EXT}(X) = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

The extended representation is redundant since the RREF define a unique subspace. Nevertheless, this representation will lead to more efficient enumerative coding. Some insight for this will be the following well known equality given in [1, p. 329].

Lemma 2: For all integers q, k, and n, such that $k \leq n$ we have

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q. \tag{4}$$

The order defined in Section III is based on Lemma 2 (applied recursively). Note that the number of subspaces in which $v(X)_1 = 1$ is $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$ and the number of subspaces in which $v(X)_1 = 0$ is $q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q$.

Remark 3: A simple connection between (3) and (4) was given in [29, p. 68].

III. Coding Based on Extended Representation

In this section we define a lexicographic order for the Grassmannian based on the extended representation. We present an enumerative coding technique for the Grassmannian using this order and discuss its complexity.

A. Order for $\mathcal{G}_q(n,k)$ Based on the Extended Representation

Let $\{x\}$ denote the value of $x = (x_1, x_2, \ldots, x_r) \in \mathbb{Z}_q^r$ (or $x = (x_1, x_2, \ldots, x_r)^T \in \mathbb{Z}_q^r$), where the vector x is viewed as a number in base-q notation. Let $\{i\}_q$ be the base-q representation of the nonnegative integer i. The resulting vector is either a row vector or a column vector depending on the context.

Let $X, Y \in \mathcal{G}_q(n,k)$ be two k-dimensional subspaces and $\mathrm{EXT}(X)$, $\mathrm{EXT}(Y)$ be the extended representations of X and Y, respectively. Let i be the least index such that $\mathrm{EXT}(X)$ and $\mathrm{EXT}(Y)$ have different columns. We say that $X < Y$ if $\begin{Bmatrix} v(X)_i \\ X_i \end{Bmatrix} < \begin{Bmatrix} v(Y)_i \\ Y_i \end{Bmatrix}$. Clearly, this definition induces an order for $\mathcal{G}_q(n,k)$.

Example 3: For $X, Y, Z \in \mathcal{G}_2(6,3)$ whose $\mathrm{EXT}(X)$, $\mathrm{EXT}(Y)$ and $\mathrm{EXT}(Z)$ are given by

$$\mathrm{EXT}(X) = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix},$$

$$\mathrm{EXT}(Y) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

$$\mathrm{EXT}(Z) = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

we have $Y < X < Z$.

B. Enumerative Coding Based on Extended Representation

Let $N\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix}$ be the number of elements in $\mathcal{G}_q(n,k)$ for which the first j columns in the extended representation are given by $\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix}$.

Remark 4: We view all the q-ary vectors of length $k+1$ as our finite alphabet. Let S be the set of all q-ary $(k+1) \times n$ matrices which form extended representations of some k-dimensional subspaces. Now, we can use Cover's method to encode/decode the Grassmannian. In this setting note that $N\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix}$ is equivalent to $n_S(x_1, x_2, \ldots, x_j)$, where $\begin{pmatrix} v_i \\ X_i \end{pmatrix}$ has the role of x_i.

Let w_j denotes the weight of the first j entries of $v(X)$, i.e., $w_j = \sum_{\ell=1}^{j} v_\ell$.

Lemma 3: For $1 \leq j \leq n$ we have

$$N\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix} = \begin{bmatrix} n-j \\ k-w_j \end{bmatrix}_q.$$

Proof: Let X be a k-dimensional subspace in $\mathcal{G}_q(n,k)$ for which the first j columns in the extended representation are given by $\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix}$. Then in the last $n-j$ entries of $v(X)$ there are $k - w_j$ *ones*, and the w_j last rows of $n-j$ last columns of $\mathrm{EXT}(X)$ have only *zeroes*. Therefore, restriction of $\mathrm{EXT}(X)$ to the first $(k+1)-w_j$ rows of the last $n-j$ columns defines a subspace in $\mathcal{G}_q(n-j, k-w_j)$. Hence, we have

$$N\begin{pmatrix} v_j & \cdots & v_1 \\ X_j & \cdots & X_1 \end{pmatrix} = \begin{bmatrix} n-j \\ k-w_j \end{bmatrix}_q.$$

■

Theorem 3: Let $X \in \mathcal{G}_q(n,k)$ be a subspace, where

$$\mathrm{EXT}(X) = \begin{pmatrix} v_n & \cdots & v_2 & v_1 \\ X_n & \cdots & X_2 & X_1 \end{pmatrix}.$$

Then the lexicographic index (decoding) of X, $\mathrm{I}_{\mathrm{EXT}}(X)$, is given by

$$\mathrm{I}_{\mathrm{EXT}}(X)=\sum_{j=1}^{n}(v_j q^{k-w_{j-1}}+(1-v_j)\frac{\{X_j\}}{q^{w_{j-1}}})\begin{bmatrix} n-j \\ k-w_{j-1}\end{bmatrix}_q. \tag{5}$$

Proof: By (2) we have that $\mathrm{I}_{\mathrm{EXT}}(X)$ is equal to

$$\sum_{j=1}^{n}\sum_{\binom{u}{W}<\binom{v_j}{X_j}} N\begin{pmatrix} u & v_{j-1} & \cdots & v_1 \\ W & X_{j-1} & \cdots & X_1\end{pmatrix}. \tag{6}$$

To compute the jth summand of (6), we distinguish between two cases.

Case 1. $v_j=1$*:* It implies that X_j has weight one, and its bottom $w_{j-1}+1$ entries (as a column vector) are an *one* followed by w_{j-1} *zeroes*, i.e., $X_j=\{q^{w_j-1}\}_q$. Hence, $\mathrm{EXT}(X)$ has the form

$$\begin{pmatrix} v_n & \cdots & v_{j+1} & 1 & v_{j-1} & \cdots & v_1 \\ X_n & \cdots & X_{j+1} & \{q^{w_j-1}\}_q & X_{j-1} & \cdots & X_1\end{pmatrix}.$$

Therefore, a subspace $Y\in\mathcal{G}_q(n,k)$ is lexicographically preceding X, where $\mathrm{EXT}(Y)$ has the same first $j-1$ columns as $\mathrm{EXT}(X)$, if and only if $\mathrm{EXT}(Y)$ has the form

$$\begin{pmatrix} v'_n & \cdots & v'_{j+1} & 0 & v_{j-1} & \cdots & v_1 \\ Y_n & \cdots & Y_{j+1} & Y_j & X_{j-1} & \cdots & X_1\end{pmatrix}.$$

Note that Y_j has *zeroes* in the last w_{j-1} entries (since the leading coefficients of the last w_{j-1} rows are contained in $(X_{j-1}\ \cdots\ X_1)$). The first $k-w_{j-1}$ entries of Y_j can have any values.

Therefore, in this case the jth summand of (6) is equal to

$$\sum_{s=0}^{q^{k-w_j-1}-1} N\begin{pmatrix} 0 & v_{j-1} & \cdots & v_1 \\ \{s\cdot q^{w_j-1}\}_q & X_{j-1} & \cdots & X_1\end{pmatrix}$$

which is equal by Lemma 3 to

$$q^{k-w_{j-1}}\begin{bmatrix} n-j \\ k-w_{j-1}\end{bmatrix}_q. \tag{7}$$

Case 2. $v_j=0$*:* Since $w_{j-1}=\sum_{\ell=1}^{j-1}v_\ell$, it follows that the last w_{j-1} entries of X_j are *zeroes*, i.e., $\{X_j\}$ is a multiple of $q^{w_{j-1}}$. Hence, $\mathrm{EXT}(X)$ has the form

$$\begin{pmatrix} v_n & \cdots & v_{j+1} & 0 & v_{j-1} & \cdots & v_1 \\ X_n & \cdots & X_{j+1} & X_j & X_{j-1} & \cdots & X_1\end{pmatrix}.$$

Therefore, a subspace $Y\in\mathcal{G}_q(n,k)$ is lexicographically preceding X, where $\mathrm{EXT}(Y)$ has the same first $j-1$ columns as $\mathrm{EXT}(X)$, if and only if $\mathrm{EXT}(Y)$ has the form

$$\begin{pmatrix} v'_n & \cdots & v'_{j+1} & 0 & v_{j-1} & \cdots & v_1 \\ Y_n & \cdots & Y_{j+1} & \{s\cdot q^{w_j-1}\}_q & X_{j-1} & \cdots & X_1\end{pmatrix},$$

where $0\le s\le \frac{\{X_j\}}{q^{w_j-1}}-1$.

Thus, in this case the jth summand of (6) is equal to

$$\sum_{s=0}^{\frac{\{X_j\}}{q^{w_j-1}}-1} N\begin{pmatrix} 0 & v_{j-1} & \cdots & v_1 \\ \{s\cdot q^{w_{j-1}}\}_q & X_{j-1} & \cdots & X_1\end{pmatrix},$$

which is equal by Lemma 3 to

$$\frac{\{X_j\}}{q^{w_{j-1}}}\begin{bmatrix} n-j \\ k-w_{j-1}\end{bmatrix}_q. \tag{8}$$

Finally, combining (7) and (8) in Case 1 and Case 2 implies (5). ∎

Example 4: Let $X\in\mathcal{G}_2(6,3)$ be a subspace represented by

$$\mathrm{EXT}(X)=\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}.$$

By Theorem 3 we have that

$$\mathrm{I}_{\mathrm{EXT}}(X)=5\cdot\begin{bmatrix}5\\3\end{bmatrix}_2+2^3\cdot\begin{bmatrix}4\\3\end{bmatrix}_2+2^2\cdot\begin{bmatrix}3\\2\end{bmatrix}_2+1\cdot\begin{bmatrix}2\\1\end{bmatrix}_2 +2\cdot\begin{bmatrix}1\\1\end{bmatrix}_2+0\cdot\begin{bmatrix}0\\0\end{bmatrix}_2=928.$$

Now, suppose that an index $0\le i<\begin{bmatrix} n\\ k\end{bmatrix}_q$ is given. Encoding Algorithm A finds $X\in\mathcal{G}_q(n,k)$ such that $\mathrm{I}_{\mathrm{EXT}}(X)=i$.

Encoding Algorithm A:

Set $i_0=i$, $w_0=0$.

For $j=1,2,\ldots,n$ do

- if $w_{j-1}=k$ then set $v_j=v(X)_j=0$, $w_j=w_{j-1}$, $X_j=\{0\}_q$, and $i_j=i_{j-1}$;
- otherwise
  - if $i_{j-1}\ge q^{k-w_{j-1}}\begin{bmatrix} n-j\\ k-w_{j-1}\end{bmatrix}_q$ then set $v_j=v(X)_j=1$, $w_j=w_{j-1}+1$, $X_j=\{q^{w_{j-1}}\}_q$, and $i_j=i_{j-1}-q^{k-w_{j-1}}\begin{bmatrix} n-j\\ k-w_{j-1}\end{bmatrix}_q$;
  - otherwise let $val=\left\lfloor i_{j-1}/\begin{bmatrix} n-j\\ k-w_{j-1}\end{bmatrix}_q\right\rfloor$ and set $v_j=v(X)_j=0$, $w_j=w_{j-1}$, $X_j=\{val*q^{w_j-1}\}_q$, and $i_j=i_{j-1}-val*\begin{bmatrix} n-j\\ k-w_{j-1}\end{bmatrix}_q$.

Form the output

$$\mathrm{EXT}(X)=\begin{pmatrix} v_n & \cdots & v_2 & v_1 \\ X_n & \cdots & X_2 & X_1\end{pmatrix}.$$

Theorem 4: Encoding Algorithm A finds the subspace $X \in \mathcal{G}_q(n,k)$, such that $\mathrm{I}_{\mathrm{EXT}}(X) = i$.

Proof: First we will show that the output of the algorithm is a k-dimensional subspace. In other words, we will prove that the weight w_n of identifying vector of the resulting subspace X is equal to k. We observe that the first "if" of the algorithm implies that $w_n \leq k$. Note also that $i_j \geq 0$ for all $1 \leq j \leq n$. Suppose that $w_n = k - t$ for some $t > 0$. Let $n-k+t \leq j' \leq n$ be the last index where $v(X)_{j'} = 0$. Then $w_{j'} = k - t - n + j' = w_{j'-1}$. According to the algorithm, $i_{j'-1} < q^{k-w_{j'-1}} \begin{bmatrix} n-j' \\ k-w_{j'-1} \end{bmatrix}_q = q^{t+n-j'} \begin{bmatrix} n-j' \\ t+n-j' \end{bmatrix}_q = 0$ (since $t > 0$), which contradicts the observation that $i_j \geq 0$ for each $1 \leq j \leq n$.

Let S_j be the jth summand of $\mathrm{I}_{\mathrm{EXT}}(X)$, given in (5), i.e., $\mathrm{I}_{\mathrm{EXT}}(X) = \sum_{t=1}^{n} S_t$. To prove the theorem it is sufficient to show that $i_j = i - \sum_{t=1}^{j} S_t$ for all $1 \leq j \leq n$ and $i_n = 0$. The proof will be inductive.

By the algorithm, for each coordinate $1 \leq j \leq n-k$,

$$i_j = \begin{cases} i_{j-1} - q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q, & \text{if } v(X)_j = 1 \\ i_{j-1} - \frac{\{X_j\}}{q^{w_{j-1}}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q, & \text{if } v(X)_j = 0 \end{cases}.$$

Thus,

$$i_j = i_{j-1} - v(X)_j q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q - (1-v(X)_j) \frac{\{X_j\}}{q^{w_{j-1}}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q = i_{j-1} - S_j \quad (9)$$

for all $1 \leq j \leq n-k$. Thus, for $j = 1$ we have $i_1 = i - S_1$. We assume that $i_j = i - \sum_{t=1}^{j} S_t$, for $j \geq 1$. By (9), $i_{j+1} = i_j - S_{j+1}$, therefore, $i_{j+1} = i - \sum_{t=1}^{j} S_t - S_{j+1} = i - \sum_{t=1}^{j+1} S_t$.

Now, we will show that for all $0 \leq j \leq n$, i_j is the lexicographic index of a subspace in $\mathcal{G}_q(n-j, k-w_j)$ with given j first columns of its representation matrix. It will complete the proof since i_n is the index of subspace in $\mathcal{G}_q(0,0)$ and thus it is equal to 0.

It is sufficient to prove that $i_j < \begin{bmatrix} n-j \\ k-w_j \end{bmatrix}_q$ for all $0 \leq j \leq n$. The proof will be inductive. For $j = 0$ we observe that $i_0 = i < \begin{bmatrix} n \\ k \end{bmatrix}_q$ is given. Assume that $i_{j-1} < \begin{bmatrix} n-j+1 \\ k-w_{j-1} \end{bmatrix}_q$. We will show that $i_j < \begin{bmatrix} n-j \\ k-w_j \end{bmatrix}_q$. We distinguish between two cases.

Case 1. $i_{j-1} \geq q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q$*:* Then, by the algorithm, $v_j = 1$, $w_j = w_{j-1} + 1$, and $i_j = i_{j-1} - q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q$. By the assumption, $i_j < \begin{bmatrix} n-j+1 \\ k-w_{j-1} \end{bmatrix}_q - q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q$ and thus by Lemma 2, $i_j \leq \begin{bmatrix} n-j \\ k-w_{j-1}-1 \end{bmatrix}_q = \begin{bmatrix} n-j \\ k-w_j \end{bmatrix}_q$.

Case 2. $i_{j-1} < q^{k-w_{j-1}} \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q$*:* Then, by the algorithm, $v_j = 0$, $w_j = w_{j-1}$, and

$$\begin{aligned} i_j &= i_{j-1} - \left\lfloor i_{j-1} / \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q \right\rfloor \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q \\ &< \left(\left\lfloor i_{j-1} / \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q \right\rfloor + 1 \right) \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q \\ &\quad - \left\lfloor i_{j-1} / \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q \right\rfloor \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q = \begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q, \end{aligned}$$

since we can write $\lfloor \frac{a}{b} \rfloor \leq a < (\lfloor \frac{a}{b} \rfloor + 1)b$ for all positive integers a and b. ∎

Example 5: Let $q = 2$, $n = 6$, $k = 3$, and $i = 928$. By using the Encoding Algorithm A we will find the subspace $X \in \mathcal{G}_2(6,3)$ such that $\mathrm{I}_{\mathrm{EXT}}(X) = i$. We apply the following steps of the algorithm.

$j = 1$: $i_0 = 928 < 2^3 \begin{bmatrix} 5 \\ 3 \end{bmatrix}_2 = 1240$ and hence $v_1 = v(X)_1 = 0$, $val = \lfloor 928/155 \rfloor = 5$, $X_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, and $i_1 = 928 - 5 \cdot 155 = 153$.

$j = 2$: $i_1 = 153 \geq 2^3 \begin{bmatrix} 4 \\ 3 \end{bmatrix}_2 = 120$ and hence $v_2 = v(X)_2 = 1$, $X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, and $i_2 = 153 - 120 = 33$.

$j = 3$: $i_2 = 33 \geq 2^2 \begin{bmatrix} 3 \\ 2 \end{bmatrix}_2 = 28$ and hence $v_3 = v(X)_3 = 1$, $X_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $i_3 = 33 - 28 = 5$.

$j = 4$: $i_3 = 5 < 2^1 \begin{bmatrix} 2 \\ 1 \end{bmatrix}_2 = 6$ and hence $v_4 = v(X)_4 = 0$, $val = \lfloor 5/3 \rfloor = 1$, $X_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, and $i_4 = 5 - 3 = 2$.

$j = 5$: $i_4 = 2 \geq 2^1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}_2 = 2$ and hence $v_5 = v(X)_5 = 1$, $X_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, and $i_5 = 2 - 2 = 0$.

$j = 6$: $w_5 = 3 = k$ and hence $v_6 = v(X)_6 = 0$, $X_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, and $i_6 = i_5 = 0$.

Therefore, we obtain a subspace $X \in \mathcal{G}_2(6,3)$ whose extended representation is given by

$$\mathrm{EXT}(X) = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

C. Complexity

We consider the complexity of computation of lexicographic index $\mathrm{I}_{\mathrm{EXT}}(\cdot)$ in (5). Note that all the integers that we use in the calculations are q-ary integers. Let $M[a,b]$ denotes the number of operations for the multiplication of two q-ary integers of length a and b. It is known [30, p. 634], that for $a > b$, $M[a,b] = a \log b \log\log b$.

First, we calculate the length of the q-ary integer which represents the largest Gaussian coefficient in (5). This Gaussian coefficient is

$$\begin{bmatrix} n-1 \\ k \end{bmatrix}_q = \frac{(q^{n-1}-1)\cdots(q^{n-k}-1)}{(q^k-1)\cdots(q-1)},$$

and hence this length is less than $k(n-k)$.

If $w_j = w_{j-1}$ then

$$\begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q = \begin{bmatrix} n-(j+1) \\ k-w_j \end{bmatrix}_q \cdot \frac{q^{n-j}-1}{q^{n-k-j+w_j}-1}. \quad (10)$$

If $w_j = w_{j-1}+1$ then

$$\begin{bmatrix} n-j \\ k-w_{j-1} \end{bmatrix}_q = \begin{bmatrix} n-(j+1) \\ k-w_j \end{bmatrix}_q \cdot \frac{q^{n-j}-1}{q^{k-w_j+1}-1}. \quad (11)$$

The Gaussian coefficients in (5) can be derived from the identifying vector. Their computation is done by (10) and (11). Hence, the complexity for computation of all the Gaussian coefficients that we need in (5) is $O(nM[k(n-k),n])$.

Since multiplication or division by q^i is done by a shift of i digits, there are $n-k$ indices where $v_j = 0$, and the length of $\{X_j\}$ is k, it follows that the complexity of these operations is $O((n-k)M[k(n-k),k])$. Finally, in (5) there are at most n additions of integers whose length is at most $k(n-k+1)$, and therefore the complexity of these operations can be omitted.

Hence, the complexity of computation of $\mathrm{I}_{\mathrm{EXT}}(\cdot)$ in (5) is $O(nM[k(n-k),n])$, i.e., $O(nk(n-k)\log n \log\log n)$.

Therefore, we have proved the following theorem:

Theorem 5: The computation complexity of the lexicographic index (decoding) in (5) is $O(nk(n-k)\log n\log\log n)$ digit operations.

If $k < \log n \log\log n$ then the Gaussian coefficients in (5) can be computed more efficiently. For their computation we can use Lemma 2. To compute $\begin{bmatrix} n \\ k \end{bmatrix}_q$ we need to compute $\begin{bmatrix} \eta \\ \kappa \end{bmatrix}_q$ for all η and κ such that $0 \le \kappa \le k$ and $0 \le \eta - \kappa \le n-k$. It requires at most $k(n-k)$ additions of integers whose length is at most $k(n-k)$, and a total of at most $k(n-k)$ shifts. All other computations do not change and can be omitted from the total complexity. Thus, we have

Theorem 6: If $\min\{k, n-k\} < \log n \log\log n$, then the computation complexity of the lexicographic index in (5) is $O(n^2 \min\{k, n-k\}^2)$ digit operations.

Finally, in a similar way we can show that the computation complexity of Encoding Algorithm A is the same as the computation complexity given for the decoding in Theorem 5 and in Theorem 6.

IV. Coding Based on Ferrers Tableaux Form

In this section we present an enumerative coding for the Grassmannian based on the Ferrers tableaux form representation of k-dimensional subspaces. Note that even so this enumerative coding is less efficient, it is more intuitive and might have its own applications. Lexicodes based on the related order, were found to be larger than the known codes [26].

A. Enumerative Coding for Ferrers Diagrams of the Same Size

Let $\mathcal{F}$ be a Ferrers diagram of size m embedded in a $k \times (n-k)$ box. We represent $\mathcal{F}$ by an integer vector of length $n-k$, $(\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1)$, where $\mathcal{F}_i$ is equal to the number of dots in the ith column of $\mathcal{F}$, $1 \le i \le n-k$. Note that the columns are numbered from right to left and that $0 \le \mathcal{F}_{i+1} \le \mathcal{F}_i \le k$ for all $1 \le i \le n-k-1$. Let $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ be two Ferrers diagrams of the same size. We say that $\mathcal{F} < \widetilde{\mathcal{F}}$ if $\mathcal{F}_i > \widetilde{\mathcal{F}}_i$ for the least index i such that $\mathcal{F}_i \ne \widetilde{\mathcal{F}}_i$, i.e., in the least column where they have a different number of dots, $\mathcal{F}$ has more dots than $\widetilde{\mathcal{F}}$. This is similar to the lexicographic order defined in the literature for unrestricted partitions, e.g., [31], [32, pp. 93–98].

Let $N_m(\mathcal{F}_j, \ldots, \mathcal{F}_2, \mathcal{F}_1)$ be the number of Ferrers diagrams of size m embedded in a $k \times (n-k)$ box, for which the first j columns are given by $(\mathcal{F}_j, \ldots, \mathcal{F}_2, \mathcal{F}_1)$.

Lemma 4: If $1 \le j \le n-k$ and $0 < m \le k(n-k)$ then

$$N_m(\mathcal{F}_j, \ldots, \mathcal{F}_2, \mathcal{F}_1) = p(\mathcal{F}_j, n-k-j, m - \sum_{i=1}^{j} \mathcal{F}_i).$$

Proof: The lemma is an immediate consequence from the fact that $\mathcal{F} = (\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1)$ is a Ferrers diagram with m dots embedded in a $k \times (n-k)$ box if and only if $(\mathcal{F}_{n-k}, \ldots, \mathcal{F}_{j+1})$ is a Ferrers diagram with $m - \sum_{i=1}^{j} \mathcal{F}_i$ dots embedded in an $\mathcal{F}_j \times (n-k-j)$ box. ∎

Remark 5: We view the set $\mathbb{Z}_{k+1} = \{0, 1, \ldots, k\}$ as our finite alphabet since $0 \le \mathcal{F}_i \le k$. Let S be the set of all $(n-k)$-tuples over $\mathbb{Z}_{k+1}$ which represent Ferrers diagrams embedded in a $k \times (n-k)$ box. In other words, $(\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1) \in S$ if and only if $0 \le \mathcal{F}_i \le \mathcal{F}_{i-1} \le k$ for each $2 \le i \le n-k$. Now, we can use Cover's method to encode/decode the set of Ferrers diagrams with m dots embedded in a $k \times (n-k)$ box. In this setting note that $N_m(\mathcal{F}_j, \ldots, \mathcal{F}_2, \mathcal{F}_1)$ is equivalent to $n_S(x_1, x_2, \ldots, x_j)$, where $\mathcal{F}_i$ has the role of x_i.

Theorem 7: Let $\mathcal{F} = (\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1)$ be a Ferrers diagram of size m embedded in a $k \times (n-k)$ box. Then the lexicographic index (decoding), ind_m, of $\mathcal{F}$ among all the Ferrers diagrams with the same size m is given by

$$\mathrm{ind}_m(\mathcal{F}) = \sum_{j=1}^{n-k} \sum_{a=\mathcal{F}_j+1}^{\mathcal{F}_{j-1}} p(a, n-k-j, m - \sum_{i=1}^{j-1} \mathcal{F}_i - a), \quad (12)$$

where we define $\mathcal{F}_0 = k$.

Proof: By (2) we have that

$$\mathrm{ind}_m(\mathcal{F}) = \sum_{j=1}^{n-k} \sum_{a=\mathcal{F}_j+1}^{\mathcal{F}_{j-1}} N_m(a, \mathcal{F}_{j-1}, \ldots, \mathcal{F}_2, \mathcal{F}_1).$$

The theorem follows now from Lemma 4. ∎

Remark 6: The summation in Theorem 7 is over larger values, while the summation in (2) is over smaller values, due to the defined order ($\mathcal{F} < \widetilde{\mathcal{F}}$ if $\mathcal{F}_i > \widetilde{\mathcal{F}}_i$ for the least index i).

Theorem 7 implies that if we can calculate $p(k, \eta, m)$ efficiently then we can calculate $\mathrm{ind}_m(\mathcal{F})$ efficiently for a Ferrers diagram of size m embedded in a $k \times (n-k)$ box.

Now suppose that an index $0 \leq i < p(k, n-k, m)$ is given. Encoding Algorithm B finds a Ferrers diagram $\mathcal{F}$ of size m embedded in a $k \times (n-k)$ box, such that $\mathrm{ind}_m(\mathcal{F}) = i$.

Encoding Algorithm B:

Step 1: Set $\mathcal{F}_0 = k,\ \ell_1 = 0,\ h = i,\ i_0 = i$;

- while $h \geq N_m(\mathcal{F}_0 - \ell_1)$ set $h = h - N_m(\mathcal{F}_0 - \ell_1)$, $\ell_1 = \ell_1 + 1$;
- set $\mathcal{F}_1 = \mathcal{F}_0 - \ell_1$, and $i_1 = h$;

Step 2: For $j = 2, \ldots, n-k$ do

- if $\sum_{i=1}^{j-1} \mathcal{F}_i = m$ then set $\mathcal{F}_j = 0$;
- otherwise do
  begin
  — set $\ell_j = 0, h = i_{j-1}$;
  — while $h \geq N_m(\mathcal{F}_{j-1} - \ell_j, \mathcal{F}_{j-1}, \ldots, \mathcal{F}_1)$ set $h = h - N_m(\mathcal{F}_{j-1} - \ell_j, \mathcal{F}_{j-1}, \ldots, \mathcal{F}_1), \ell_j = \ell_j + 1$;
  — set $\mathcal{F}_j = \mathcal{F}_{j-1} - \ell_j$, and $i_j = h$;

  end {begin}

Step 3: Form the output $\mathcal{F} = (\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1)$.

Remark 7: We did not join Step 1 and Step 2, since $N_m(\mathcal{F}_{j-1} - \ell_j, \mathcal{F}_{j-1}, \ldots, \mathcal{F}_1)$ is not defined for $j = 1$.

B. Order for $\mathcal{G}_q(n,k)$ Based on Ferrers Tableaux Form

Let $X,\ Y \in \mathcal{G}_q(n,k)$ be two k-dimensional subspaces and let $\mathcal{F}_X,\ \mathcal{F}_Y$ be the related Ferrers diagrams. Let $x = (x_1, x_2, \ldots, x_{|\mathcal{F}_X|})$ and $y = (y_1, y_2, \ldots, y_{|\mathcal{F}_Y|})$ be the entries vectors of $\mathcal{F}(X)$ and $\mathcal{F}(Y)$, respectively. These entries are numbered from right to left, and from top to bottom.

We say that $X < Y$ if one of the following conditions holds.

- $|\mathcal{F}_X| > |\mathcal{F}_Y|$;
- $|\mathcal{F}_X| = |\mathcal{F}_Y|$ and $\mathcal{F}_X < \mathcal{F}_Y$;
- $\mathcal{F}_X = \mathcal{F}_Y$ and $\{x\} < \{y\}$.

Clearly, this definition induces an order for $\mathcal{G}_q(n,k)$.

Example 6: Let $X, Y, Z, W \in \mathcal{G}_2(6,3)$ be given by

$$\mathcal{F}(X) = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ & & 1 \end{matrix}, \quad \mathcal{F}(Y) = \begin{matrix} 1 & 0 & 1 \\ & 0 & 0 \\ & 1 & 1 \end{matrix},$$

$$\mathcal{F}(Z) = \begin{matrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 0 \end{matrix}, \quad \mathcal{F}(W) = \begin{matrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{matrix}.$$

$\mathcal{F}_Z = \mathcal{F}_W$ and by definition $Z < W$. Clearly, $|\mathcal{F}_X| = |\mathcal{F}_Y| > |\mathcal{F}_Z|$ and $\mathcal{F}_Y < \mathcal{F}_X$. Thus, $Y < X < Z < W$.

C. Enumerative Coding Based on Ferrers Tableaux Form

In this subsection, we use the given order of Ferrers tableaux forms and Theorem 2 for enumerative coding for $\mathcal{G}_q(n,k)$.

Theorem 8: Let $X \in \mathcal{G}_q(n,k)$, $\mathcal{F}_X$ be the Ferrers diagram of X, and let $x = (x_1, x_2, \ldots, x_{|\mathcal{F}_X|})$ be the entries vector of $\mathcal{F}(X)$. Then the lexicographic index (decoding) of X, $\mathrm{Ind}_{\mathcal{F}}(X)$, defined by the order based on Ferrers tableaux form, is given by

$$\mathrm{Ind}_{\mathcal{F}}(X) = \sum_{i=|\mathcal{F}_X|+1}^{k(n-k)} \alpha_i q^i + \mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X) q^{|\mathcal{F}_X|} + \{x\}, \quad (13)$$

where α_i, $|\mathcal{F}_X| + 1 \leq i \leq k(n-k)$, is defined in Theorem 2.

Proof: To find $\mathrm{Ind}_{\mathcal{F}}(X)$ we have to calculate the number of k-dimensional subspaces which are preceding X according to the order defined above.

1) All the k-dimensional subspaces with Ferrers diagrams which have more dots than $\mathcal{F}_X$ are preceding X. Their number is $\sum_{i=|\mathcal{F}_X|+1}^{k(n-k)} \alpha_i q^i$.
2) There are $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X)$ Ferrers diagrams with $|\mathcal{F}_X|$ dots which are preceding X. Hence, there are $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X) q^{|\mathcal{F}_X|}$ k-dimensional subspaces whose Ferrers diagrams have $|\mathcal{F}_X|$ dots and preceding X.
3) Finally, the number of k-dimensional subspaces whose Ferrers diagram is $\mathcal{F}_X$ which are preceding X is $\{x\}$.

∎

Example 7: Let $X \in \mathcal{G}_2(6,3)$ be the subspace of Example 4, whose Ferrers tableaux form and Ferrers diagram are

$$\mathcal{F}(X) = \begin{matrix} 1 & 1 \\ & 0 \\ & 1 \end{matrix} \quad \text{and} \quad \mathcal{F}_X = \begin{matrix} \bullet & \bullet \\ & \bullet \\ & \bullet \end{matrix}.$$

By Theorem 8 we have that

$$\mathrm{Ind}_{\mathcal{F}}(X) = \sum_{i=5}^{9} \alpha_i 2^i + \mathrm{ind}_4(\mathcal{F}_X) 2^4 + \{(1011)\}.$$

Since $\alpha_5 = 3$, $\alpha_6 = 3$, $\alpha_7 = 2$, $\alpha_8 = 1$, $\alpha_9 = 1$ (see [1, pp. 326–328]), $\mathrm{ind}_4(\mathcal{F}_X) = 0$, and $\{(1011)\} = 11$, it follows that $\mathrm{Ind}_{\mathcal{F}}(X) = 1323$.

Now suppose that an index $0 \leq i < \begin{bmatrix} n \\ k \end{bmatrix}_q$ is given. Encoding Algorithm C finds a subspace $X \in \mathcal{G}_q(n,k)$ such that $\mathrm{Ind}_{\mathcal{F}}(X) = i$.

Encoding Algorithm C:

Set $i_0 = i$.

For $j = 0, \ldots, k(n-k)$ do

- if $i_j < \alpha_{k(n-k)-j} q^{k(n-k)-j}$ then set $|\mathcal{F}_X| = k(n-k) - j$, $\mathcal{F}_X = \mathrm{ind}^{-1}_{|\mathcal{F}_X|}(\lfloor \frac{i_j}{q^{k(n-k)-j}} \rfloor)$; $\{i_j - \lfloor \frac{i_j}{q^{k(n-k)-j}} \rfloor q^{k(n-k)-j}\}_q$ is assigned to x (the entries vector of $\mathcal{F}(X)$) and stop;
- otherwise set $i_{j+1} = i_j - \alpha_{k(n-k)-j} q^{k(n-k)-j}$.

D. Complexity

We consider the complexity of the calculation of the lexicographic index $\mathrm{Ind}_{\mathcal{F}}(X)$, for $X \in \mathcal{G}_q(n,k)$, whose Ferrers diagram is $\mathcal{F}_X = (\mathcal{F}_{n-k}, \ldots, \mathcal{F}_2, \mathcal{F}_1)$. We will use the following lemma concerning partitions to find a bound on the length of q-ary integers which represent the value of $p(k, n-k, i)$.

Lemma 5: For any given n, k, and i, we have $p(k, n-k, i) < e^{\pi\sqrt{\frac{2}{3}i}}$.

Proof: Clearly, $p(k, n-k, i) \leq p(i)$, where $p(i)$ is the number of unrestricted partitions of i. It is known [1, p. 160] that $p(i) < e^{\pi\sqrt{\frac{2}{3}i}}$ and the lemma follows. ∎

Theorem 9: The computation complexity of the lexicographic index (decoding) in (14) is $O(k^{5/2}(n-k)^{5/2})$ digit operations.

Proof: First, we combine the expressions in (12) and (13) to obtain

$$\mathrm{Ind}_{\mathcal{F}}(X) = \sum_{i=|\mathcal{F}_X|+1}^{k(n-k)} p(k, n-k, i) q^i + \{x\}$$
$$+ q^{|\mathcal{F}_X|} \sum_{j=1}^{n-k} \sum_{a=\mathcal{F}_j+1}^{\mathcal{F}_{j-1}} p(a, n-k-j, |\mathcal{F}_X| - \sum_{i=1}^{j-1} \mathcal{F}_i - a). \quad (14)$$

By the recurrence relation of Lemma 1, we can compute the table of $p(j, \ell, i)$ for $j \leq k$, $\ell \leq \eta$, and $i \leq m$ with no more than $mk\eta$ additions. By Lemma 5 each integer in such addition has $O(\sqrt{k(n-k)})$ digits. Therefore, the computation of all the values which are needed from the table takes $O(k^{5/2}(n-k)^{5/2})$ digit operations.

The number of additions in (14) is $O(k(n-k))$. Each integer in this addition has $O(k(n-k))$ digits (as a consequence of Lemma 5 and the powers of q in (14)). The multiplication by q^i is a shift by i symbols. Hence, these additions and shifts do not increase the complexity. ∎

Similarly, we can prove the following theorem.

Theorem 10: The computation complexity of Encoding Algorithm C is $O(k^{5/2}(n-k)^{5/2})$ digit operations.

Remark 8: If $k(n-k) - |\mathcal{F}_X|$ is a small integer then the complexity of the computation becomes much smaller than the complexity given in Theorems 9 and 10. For example, if $|\mathcal{F}_X| = k(n-k)$ then the complexity of the enumerative decoding is $O(k(n-k))$ since $\mathrm{Ind}_{\mathcal{F}}(X) = \{x\}$ in (14).

It is worth to mention in this context that the number of operations in the algorithms can be made smaller if we will consider the following two observations [27, p. 47]:

- If $m_1 < m_2 \leq \frac{k\eta}{2}$ then $p(k, \eta, m_1) \leq p(k, \eta, m_2)$.
- $p(k, \eta, m) = p(k, \eta, k\eta - m)$ and hence we can assume that $m \leq \frac{k\eta}{2}$.

V. Combination of the Coding Techniques

By Theorems 5, 6, and 9, it is clear that the enumerative coding based on the extended representation is more efficient than the one based on Ferrers tableaux form. But for some of k-dimensional subspaces of $\mathbb{F}_q^n$ the enumerative coding based on Ferrers tableaux form is more efficient than the one based on the extended representation (see Remark 8). This is the motivation for combining the two methods.

The only disadvantage of the Ferrers tableaux form coding is the computation of the α_i's and $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X)$ in Theorem 8. This is the reason for its relatively higher complexity. The advantage of this coding is that once the values of the α_i's and the value of $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X)$ are known, the computation of $\mathrm{Ind}_{\mathcal{F}}(X)$, for $X \in \mathcal{G}_q(n,k)$, is immediate. Our solutions for the computation of the α_i's and $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X)$ are relatively not efficient and this is the main reason why we suggested to use the enumerative coding based of the RREF and the identifying vector of a subspace. The only disadvantage of this enumerative coding is the computation of the Gaussian coefficients in (5). It appears that a combination of the two methods is more efficient than the efficiency of each one separately. The complexity will remain $O(nk(n-k)\log n \log\log n)$, but the constant will be considerably reduced on the average. This can be done if there won't be any need for the computation of the α_i's and the computation of $\mathrm{ind}_{|\mathcal{F}_X|}(\mathcal{F}_X)$ will be efficient.

It was proved in [9] that $q^{k(n-k)} < \begin{bmatrix} n \\ k \end{bmatrix}_q < 4q^{k(n-k)}$ for $0 < k < n$. Thus, more than $\frac{1}{4}$ of the k-dimensional subspaces in $\mathcal{G}_q(n,k)$ have the unique Ferrers diagram with $k(n-k)$ dots, where the identifying vector consists of k *ones* followed by $n-k$ *zeroes*. All the codewords of the Reed–Solomon-like code in [9] have this Ferrers diagram. Note that most of the k-dimensional subspaces have Ferrers diagrams with a large number of dots. We will encode/decode these subspaces by the Ferrers tableaux form coding and the other subspaces by the extended representation coding. We will choose a set $S_{\mathcal{F}}$ with a small number of Ferrers diagrams. $S_{\mathcal{F}}$ will contain the largest Ferrers diagrams. The Ferrers tableaux form coding will be applied on these diagrams.

We say that a subspace $X \in \mathcal{G}_q(n,k)$ is of Type $S_{\mathcal{F}}$ if $\mathcal{F}_X \in S_{\mathcal{F}}$. In the new order these subspaces are ordered first, and their internal order is defined as the order of the Ferrers tableaux forms in Section IV. The order of the other subspaces is defined by the order of the extended representation in Section III. We define a new index function $\mathrm{I}_{\mathrm{comb}}$ as follows:

$$\mathrm{I}_{\mathrm{comb}}(X) = \begin{cases} \mathrm{Ind}_{\mathcal{F}}(X) & \mathcal{F}_X \in S_{\mathcal{F}} \\ \mathrm{I}_{\mathrm{EXT}}(X) + \Delta_X(S_{\mathcal{F}}) & \text{otherwise,} \end{cases} \quad (15)$$

where $\Delta_X(S_{\mathcal{F}})$ is the number of subspaces of Type $S_{\mathcal{F}}$, which are lexicographically succeeding X by the extended representation ordering. These $\Delta_X(S_{\mathcal{F}})$ subspaces are preceding X in the ordering induced by combining the two coding methods.

We demonstrate the method for the simple case where $S_{\mathcal{F}}$ consists of the unique Ferrers diagram with $k(n-k)$ dots.

Lemma 6: Let $S_{\mathcal{F}}$ be a set of Ferrers diagrams, embedded in a $k \times (n-k)$ box, which contains only one Ferrers diagram, the unique one with $k(n-k)$ dots. Let $X \in \mathcal{G}_q(n,k)$, $X \notin S_{\mathcal{F}}$, $\mathrm{RE}(X) = (X_n, \ldots, X_1)$, and let ℓ, $0 \leq \ell \leq n-k-1$, be the number of consecutive *zeroes* before the first *one* (from the right) in the identifying vector $v(X)$. Then $\Delta_X(S_{\mathcal{F}}) = \sum_{i=1}^{\ell} (q^k - 1 - \{X_i\}) q^{k(n-k-i)}$.

Proof: If $\ell = 0$ then $v(X)_1 = 1$ and hence there are no subspaces of Type $S_{\mathcal{F}}$ which are lexicographically succeeding

X and hence $\Delta_X(S_{\mathcal{F}}) = 0$. For $1 \leq \ell \leq n-k-1$, let $X_1, \ldots, X_\ell$ be the first ℓ columns of $\mathrm{RE}(X)$. All the subspaces of Type $S_{\mathcal{F}}$ in which the value of the first column is greater than $\{X_1\}$, are lexicographically succeeding X. There are $(q^k - 1 - \{X_1\})q^{k(n-k-1)}$ such subspaces. All the subspaces of Type $S_{\mathcal{F}}$ in which the first $i-1$ columns, $2 \leq i \leq n-k-1$, are equal to the first $i-1$ columns of $\mathrm{RE}(X)$, and the value of the ith column is greater than $\{X_i\}$, are lexicographically succeeding X. There are $(q^k - 1 - \{X_i\})q^{k(n-k-i)}$ such subspaces. Therefore, there are $\sum_{i=1}^{\ell}(q^k - 1 - \{X_i\})q^{k(n-k-i)}$ subspaces of Type $S_{\mathcal{F}}$ which are lexicographically succeeding X by the extended representation ordering. ∎

Example 8: Let X be the subspace of Example 4. By Example 4 we have $\mathrm{I}_{\mathrm{EXT}}(X) = 928$, and by Lemma 6 we have $\Delta_X(S_{\mathcal{F}}) = (2^3 - 1 - 5)2^{3\cdot 2} = 2^7$. Hence, $\mathrm{I}_{\mathrm{comb}}(X) = \mathrm{I}_{\mathrm{EXT}}(X) + \Delta_X(S_{\mathcal{F}}) = 928 + 128 = 1056$.

Now, suppose that an index $0 \leq i < \begin{bmatrix} n \\ k \end{bmatrix}_q$ is given. Based on (15) and Lemma 6 we can find the subspace X such that $\mathrm{I}_{\mathrm{comb}}(X) = i$, where $S_{\mathcal{F}}$ consists of the unique Ferrers diagram with $k(n-k)$ dots. We omit the details of the encoding algorithm.

VI. Conclusion

Three methods of enumerative coding for the Grassmannian are presented. The first is based on the representation of subspaces by their identifying vector and their reduced row echelon form. The second is based on the Ferrers tableaux form representation of subspaces. The complexity of the first method is superior on the complexity of the second one. The third method is a combination of the first two. On average it reduces the constant in the first term of the complexity compared to the complexity of the first method. Improving on these methods is a problem for future research.

The enumerative coding is based on an order for the Grassmannian related to a specific representation. This order can be used to form lexicographic codes [25] in the Grassmannian. To our surprise some of these lexicographic codes form the best known error-correcting codes in the Grassmannian. For example, a lexicode of size 4605 in $\mathcal{G}_2(8,4)$ with minimum subspace distance 4 (see [9] for the distance definition) was generated based on Ferrers tableaux form order (compared to the largest previously known code of size 4573 generated by a multilevel construction [18]). These codes also revealed a new method to form error-correcting codes in the Grassmannian. This topic is considered in [26].

Construction of a lexicode might require to generate all subspaces of $\mathcal{G}_q(n,k)$ by the given lexicographic order. Usually, this does not require to use the enumerative coding since the subspaces are generated one after another. By using one of our orders it is not difficult to prove that given a subspace $X \in \mathcal{G}_q(n,k)$, it takes no more than $O(kn)$ digit operations to generate the next subspace.

Acknowledgment

The authors thank the anonymous reviewers whose comments have helped to improve the presentation of this paper.

References

[1] J. H. van Lint and R. M. Wilson, *A course in Combinatorics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2001.
[2] D. E. Knuth, "Subspaces, subsets, and partitions," *J. Combin. Theory*, vol. 10, pp. 178–180, 1971.
[3] S. Thomas, "Designs over finite fields," *Geometriae Dedicata*, vol. 21, pp. 237–242, 1987.
[4] W. J. Martin and X. J. Zhu, "Anticodes for the Grassman and bilinear forms graphs," *Des. Codes, Crypt.*, vol. 6, pp. 73–79, 1995.
[5] S. Thomas, "Designs and partial geometries over finite fields," *Geometriae Dedicata*, vol. 63, pp. 247–253, 1996.
[6] R. Ahlswede, H. K. Aydinian, and L. H. Khachatrian, "On perfect codes and related concepts," *Des. Codes, Crypt.*, vol. 22, pp. 221–237, 2001.
[7] M. Schwartz and T. Etzion, "Codes and anticodes in the Grassman graph," *J. Combin. Theory, Ser. A*, vol. 97, pp. 27–42, 2002.
[8] M. Braun, A. Kerber, and R. Laue, "Systematic construction of q-analogs of $t-(v,k,\lambda)$-designs," *Des. Codes, Crypt.*, vol. 34, pp. 55–70, 2005.
[9] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3579–3591, Aug. 2008.
[10] S. T. Xia and F. W. Fu, "Johnson type bounds on constant dimension codes," *Des. Codes, Crypt.*, vol. 50, pp. 163–172, 2009.
[11] T. Etzion and A. Vardy, "Error-correcting codes in projective space," in *Proc. Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 871–875.
[12] F. Manganiello, E. Gorla, and J. Rosenthal, "Spread codes and spread decoding in network coding," in *Proc. Int. Symp. Inf. Theory*, Jul. 2008, pp. 881–885.
[13] D. Silva, F. R. Kschischang, and R. Koetter, "A rank-metric approach to error control in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3951–3967, Sep. 2008.
[14] D. Silva and F. R. Kschischang, "On metric for error correction in network coding," *IEEE Trans. Inf. Theory*, vol. 55, pp. 5479–5490, Dec. 2009.
[15] M. Gadouleau and Z. Yan, "Constant-rank codes and their connection to constant-dimension codes," *IEEE Trans. Inf. Theory*, vol. 56, pp. 3207–3216, Jul. 2010.
[16] M. Gadouleau and Z. Yan, "On the decoder error probability of bounded rank distance decoders for maximum rank distance codes," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3202–3206, Jul. 2008.
[17] M. Gadouleau and Z. Yan, Construction and covering properties of constant-dimension codes [Online]. Available: arxiv.org/abs/0903.2675
[18] T. Etzion and N. Silberstein, "Error-correcting codes in projective space via rank-metric codes and Ferrers diagrams," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2909–2919, Jul. 2009.
[19] A. Kohnert and S. Kurz, "Construction of large constant dimension codes with a prescribed minimum distance," *Lecture Notes on Computer Science*, vol. 5393, pp. 31–42, 2008.
[20] V. Skachek, "Recursive code construction for random networks," *IEEE Trans. Inf. Theory*, vol. IT-56, pp. 1378–1382, Mar. 2010.
[21] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 73–77, Jan. 1973.
[22] V. Braun and K. A. S. Immink, "An Enumerative coding technique for DC-free runlength-limited sequences," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 2024–2031, Dec. 2000.
[23] O. f. Kurmaev, "Enumerative coding for constant-weight binary sequences with constrained run-length of zeros," *Probl. Inf. Transm.*, vol. 38, no. 1, pp. 249–254, 2002.
[24] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Eindhoven, The Netherlands: Shannon Found. Publ., 1999.
[25] J. H. Conway and N. J. A. Sloane, "Lexicographic codes: Error-correcting codes from game theory," *IEEE Trans. Inf. Theory*, vol. IT-32, pp. 337–348, May 1986.
[26] N. Silberstein and T. Etzion, Large constant dimension codes and lexicodes [Online]. Available: arxiv.org/abs/1003.4879.2010
[27] G. E. Andrews, *The Theory of Partitions*. Cambridge, U.K.: Cambridge Univ. Press, 1984.
[28] R. P. Stanley, *Enumerative Combinatorics*. Monterey, CA: Wadsworth, 1986, vol. 1.
[29] G. E. Andrews and K. Eriksson, *Integer Partitions*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[30] D. E. Knuth, *The Art of Computer Programming*, 3rd ed. Reading, MA: Addison-Wesley, 1997, vol. 2, Seminumerical Algorithms.

[31] T. V. Narayana, R. M. Mathsen, and J. Sarangi, "An algorithm for generating partitions and its applications," *J. Combin. Theory*, vol. 11, pp. 54–61, 1971.

[32] F. Ruskey, *Combinatorial Generation*, Working Version ed. Victoria, BC, Canada: Univ. Victoria, 2001.

Natalia Silberstein was born in Novosibirsk, Russia, in 1977. She received the B.A. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, in 2004 and 2007, respectively, from the Computer Science Department and the Applied Mathematics Department, respectively. She is currently working toward the Ph.D. degree in the department of Computer Science at the Technion.

Her research interests include algebraic error-correction coding, coding theory, and combinatorial designs.

Tuvi Etzion (M'89–SM'94–F'04) was born in Tel Aviv, Israel, in 1956. He received the B.A., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1982, and 1984, respectively.

Since 1984 he has held a position in the Department of Computer Science at the Technion, where he is currently a Professor. During 1986–1987 he was also Visiting Research Professor with the Department of Electrical Engineering—Systems at the University of Southern California, Los Angeles. During the summers of 1990 and 1991 he was visiting Bellcore in Morristown, NJ. During 1994–1996 he was a Visiting Research Fellow in the Computer Science Department at Royal Holloway College, Egham, U.K. He also had several visits to the Coordinated Science Laboratory at the University of Illinois in Urbana-Champaign during 1995–1998, two visits to HP Bristol during the summers of 1996 and 2000, a few visits to the department of Electrical Engineering, University of California at San Diego during the years 2000–2010, and several visits to the Mathematics department at Royal Holloway College during 2007–2009. His research interests include applications of discrete mathematics to problems in computer science and information theory, coding theory, and combinatorial designs.

Dr Etzion was an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2006 to 2009.