BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 3

Sofia • 2013

Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2013-0024

Upper Bound on Correcting Partial Random Errors

Ankita Gaur, Bhu Dev Sharma

Department of Mathematics, JIIT, Noida, UP-201307, India,
Emails: ankitagaur50@yahoo.com bhudev_sharma@ yahoo.com bhudev.sharma@jiit.ac.in

Abstract: *Since coding has become a basic tool for practically all communication/electronic devices, it is important to carefully study the error patterns that actually occur. This allows correction of only partial errors rather than those which have been studied using Hamming distance, in non-binary cases.*

The paper considers a class of distances, SK-distances, in terms of which partial errors can be defined. Examining the sufficient condition for the existence of a parity check matrix for a given number of parity-checks, the paper contains an upper bound on the number of parity check digits for linear codes with property that corrects all partial random errors of an (n,k) *code with minimum SK-distance at least d. The result generalizes the rather widely used Varshamov-Gilbert bound, which follows from it as a particular case.*

Keywords: *Hamming distance, SK-metric, error patterns, error control in codes, bounds.*

1. Introduction

In the study of error correcting codes mainly the Hamming metric [2], being primarily developed for the binary case, is used even when the coding alphabet is *q*-nary, *q*>2. Another metric for *q*-nary is due to Lee. In case of Hamming metric, one digital change in one place is a single error, no matter what the magnitude is, whereas in case of Lee metric [2], a digital change of magnitude *i* is made in one place by one of the two $\pm i$ entries in one place. In the case of constant length codes, error patterns have places where errors occur, at different positions. To imbue the efficiency, one has to use tools that can handle these considerations

efficiently. Hamming weight-distance notion is not sensitive to them. A mathematically robust method of making the proper choices out of a class of distances was introduced by Sharma and Kaushik in 1977. This SK-class of metric provides freedom of choice that might logically correspond best to the error patterns that are encountered in different real communication systems.

In this paper, by the random-error-correction with SK-metric considerations, a class of errors is studied, which in some sense is a "part" of the class of errors that may arise from Hamming considerations. The paper contains the sufficient condition for correcting errors of a certain number of partial random errors. Results derived under Hamming considerations follow as particular cases from this study, and those for Lee metric can also be directly obtained.

2. Preliminaries

We shall consider n-vectors over $Z_q = \{0, 1, 2, ..., q-1\}$, q>3. Sharma and Kaushik consider partitions of Z_q into non-empty disjoint subsets $B_0, B_1, ..., B_{m-1}$ where m is a natural number greater than or equal to two for introducing a class of metrics and weights. For this to happen, the partition is such that

(i) $B_0 = \{0\}$,

(ii) for $i \in Z_q$, $i \in B_s \Leftrightarrow q - i \in B_s$,

(iii) if $i \in B_s$ and $j \in B_t$ and s>t (in the order of their natural occurrence in Z_q), then $\min\{i, q-i\} > \min\{j, q-j\}$.

(iv) if $s > .t$, $|B_s| \geq |B_t|$, except for $s = m-1$, in which case we may have

$$|B_{m-1}| \geq \frac{1}{2}|B_{m-2}|,$$

where $|B|$ is the number of elements in the set B [3].

Note. When m = 2, the partition is simply $\{B_0 = \{0\}, B_1 = \{1, 2, ..., q-1\}\}$, they called it Hamming partition of Z_q. This leads to Hamming distance and weights as considered in literature. Also, when each $|B_s| = 2$ for $0 < s < m-2$, the partition may be called Lee partition, since in this case it leads to Lee-metric, refer to B e r l e k e m p [1].

Definition. SK-Weights and distances [3].

SK-Weights: It may be mentioned that in SK-scheme of things, weights and distances are defined in reference to a SK-partition. Thus for different SK-partitions of Z_q, they, in general, will have different values for the same element of Z_q, or for a vector over Z_q.

We first define the SK-weight of an element $j \in Z_q$ corresponding to a SK-partition $P=\{B_0, B_1,..., B_{m-1}\}$. Denoting it by $W_P(j)$, it is given by

$$W_P(j)=s \text{ if } j \in B_s,\ 0 \le s \le m-1.$$

Next, we define SK-weight $W_P(u)$ of $u=(a_1, a_2 ..., a_n),\ a_i \in Z_q$, corresponding to SK-partition $P=\{B_0, B_1,..., B_{m-1}\}$ as the sum of the class-weights of its components, i.e.,

$$W_P(u)=\sum_{i-1}^{n} W_p(a_i).$$

SK-Distance between two vectors: Given two *n*-vectors $u=(a_1, a_2,..., a_n)$ and $v=(b_1, b_2,..., b_n)$, where $a_i, b_i \in Z_q$, the SK-distance between vectors u and v associated with SK-partition P is defined as the sum of the SK-distances between their components, i.e.,

$$d_P(u,v)=\sum_{i=1}^{n} d_p(a_i,b_i)=\sum_{i=1}^{n} W_P(a_i-b_i),$$

$$d_P(u,v)=W_P(u-v)=\sum_{i=1}^{n} d_p(a_i-b_i).$$

In an earlier paper [6] we have introduced the idea of a partial error pattern and Limited Error pattern which is defined as follows.

Partial Error Pattern and Patterns of a Limited Error [6]

- **Partial sets:** It may be noted that in defining the SK-partition $P=\{B_0, B_1,..., B_{m-1}\}$, of Z_q arranged in a circular order, B_s is in fact the collection of all elements of Z_q at distance *s*, on either side of 0, each element of which is assigned an SK weight *s*. We can thus call B_s a subset or a "partial set" of Z_q of weight *s* , and at distance *s* from 0.

More generally, for an arbitrary element $j \in Z_q$ its "partial set at distance *s*" is given by $B_s(j)=\{B_s + j\}_q$, the addition being in each element of B_s mod *q*.

- **Patterns of limited errors:** We know that the error detection/correction studies are made taking into consideration the patterns of errors, which vary from a system to system. The study of block/linear codes is also limited to errors in which an entry in a code word is received as another symbol, the error is called a "substitution error". With SK-scheme of things, it is possible to consider various different limited kinds of substitution errors that were not possible under Hamming scheme of things. These can be in terms of

(i) number of places of the errors (random or bursts),
(ii) substitutions limited to one or more of B_i's,
(iii) maximum overall SK-weight of the error patterns,
(iv) combinations of any two or three of the above.

In obtaining a bound on the necessary number of parity-checks for an *e*-error correcting code, it is customary to define the volume of a sphere of radius *e* around every code word and consider their mutual disjointness, etc. In the SK-study that we undertake, this idea can be considered more closely. Given an *n*-vector *u*, we can find numbers of patterns which have specified SK-distance from vector *u* [6].

We will need a generalization, which we call "partial independence" of vectors.

Definition. Partial independence: Given a set of *n*-vectors *S* over the field GF(*q*), and a subset *B* of GF(*q*), the set of vectors *S* will be called partially independent in *B*, if all linear combinations of vectors in *S*, with coefficients from *B*, some non-zero, is not zero. In this situation *S* may be termed "*B*-independent".

In [6] we obtained the necessary condition giving the number of parity check digits for linear codes correcting an SK-distance limited partial random error on *e* or fewer positions.

3. Sufficient condition on the number of parity check digits

This section presents the sufficient condition on the number of parity check digits for linear codes with a property that corrects all random errors of an (n,k) code with minimum distance at least d, with entries limited to B_1, as also B_1 and B_2.

Theorem 1. Given an SK-partition $P=\{B_0,B_1,...,B_{m-1}\}$ of Z_q, q prime, the sufficient condition for the existence of an (n,k) code over Z_q, with minimum SK-distance at least d, when entries in a position of any two code words differ only partially by the entry from B_1', is given by

$$\sum_{i=0}^{d-2}\binom{n}{i}\left|B_1\right|^i \geq q^{n-k}.$$

Proof: Obviously, the existence of such an (n,k) code is ensured by constructing a parity check matrix *H*, with *n* columns and $r=n-k$ rows. Here the condition for its existence is to be examined to suit the code in question. We proceed as follows.

First we select any nonzero *r*-tuple as the first column of the parity check matrix. Then select any non-zero *r*-tuple except those that are B_1 multiples of the first (or B_1 independent) as the second column in the parity check matrix. The third column then may be any *r*-tuple which is not a linear combination of the first and second column, with coefficients from B_1. In general, the *i*-th column is chosen as any *r*-tuple that is not a linear combination of any $d-2$ or less previous columns with coefficients from B_1. This construction ensures that no linear combination of $d-1$ or fewer columns with coefficients from B_1 will be zero, that is, $d-1$ columns will be "B_1-independent".

If there are $|B_1|$ possible coefficients at the time of finding *j*-th column, the number of *r*-tuples to be excluded is

$$\binom{j-1}{1}|B_1|+\binom{j-1}{2}|B_1|^2+\ldots+\binom{j-1}{d-2}|B_1|^{d-2},$$

the linear combination of $d-2$ or less columns out of total of $j-1$ columns. If this is less than the total number of non-zero *r*-tuples, then there is certainly one more column, the *j*-th that can be added to the matrix. That is, if

$$1+\binom{j-1}{1}|B_1|+\binom{j-1}{2}|B_1|^2+\ldots+\binom{j-1}{d-2}|B_1|^{d-2}\le q^r.$$

Now let *n* be the largest value of *j*, then an (n,k) code with minimum SK B_1-distance d is given by

$$1+\binom{n}{1}|B_1|+\binom{n}{2}|B_1|^2+\ldots+\binom{n}{d-2}|B_1|^{d-2}\ge q^r.$$

This proves the result.

Particular case:

(i) In case of Hamming metric, we have $m=2$, $B=\{|B_0|,|B_1|\}$, where $B_1 = \{1, 2, \ldots, q-1\}$, then the expression in Theorem 1 can be stated as

$$\sum_{i=0}^{d-2}\binom{n}{i}(q-1)^i\ge q^{n-k}.$$

(ii) In case of Lee metric, we have $m-1=\frac{q-1}{2}$, so $B=\left\{|B_0|,|B_1|,\ldots,\left|B_{\frac{q-1}{2}}\right|\right\}$, $B_1=\{1,q-1\}$ and then the expression in Theorem 1 takes the form

$$\sum_{i=0}^{d-2}\binom{n}{i}(2)^i\ge q^{n-k}.$$

Broadened partial case. The result of Theorem 1 can be further broadened from a single B_1 to that for entries from a group of B_i's.

Theorem 2. Given an SK-partition $P=\{B_0,B_1,\ldots,B_{m-1}\}$ of Z_q, q prime, the sufficient condition for the existence of an (n,k) code over Z_q with a minimum SK-distance at least d, and entries from B_1 and B_2, is given by

$$\sum_{s=1}^{d-2}\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{n}{s-2m,m,n-s+m}\left|B_1\right|^{s-2m}\left|B_2\right|^{m}\right)\geq q^{r}.$$

Proof: Let us first consider different linear combinations of length l with entries from B_1 and B_2 having SK-weight s. These are given by various columns of Table 1.

Table 1

Number of entries from B_2	0	1	2	3	…	m	…	$\left\lfloor \frac{s}{2} \right\rfloor$
Number of entries from B_1	s	s–2	s–4	s–6	…	s–2m	…	$s-2\left\lfloor \frac{s}{2} \right\rfloor$
Number of 0's	l–s	l–s+1	l–s+2	l-s+3	…	l-s+m	…	$l-s+\left\lfloor \frac{s}{2} \right\rfloor$

Then the total number of such l-vectors is

$$\left(\binom{l}{s,0,(l-s)}\left|B_1\right|^{s}+\binom{l}{(s-2),1,(l-s+1)}\left|B_1\right|^{s-2}\left|B_2\right|+\binom{l}{(s-4),2,(l-s+2)}\times\right.$$

$$\times\left|B_1\right|^{s-4}\left|B_2\right|^{2}+\ldots+\binom{l}{(s-2m),m,(l-s+m)}\left|B_1\right|^{s-2m}\left|B_2\right|^{m}+\ldots+$$

$$\left.+\binom{l}{\left(s-2\left\lfloor \frac{s}{2} \right\rfloor\right),\left\lfloor \frac{s}{2} \right\rfloor,\left(l-s+\left\lfloor \frac{s}{2} \right\rfloor\right)}\left|B_1\right|^{s-2\left\lfloor \frac{s}{2} \right\rfloor}\left|B_2\right|^{\left\lfloor \frac{s}{2} \right\rfloor}\right)=$$

$$=\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{l}{s-2m,m,l-s+m}\left|B_1\right|^{s-2m}\left|B_2\right|^{m}\right).$$

Now we come to examining the existence of the parity-check matrix for the code, as in Theorem 1. We first select any nonzero r-tuple as the first column of the parity check matrix. Then we select any non-zero r-tuple except those that are B_1 and B_2 multiples of the first (or B_1 and B_2 independent) as the second column in the parity check matrix. The third column may then be any r-tuple which is not a linear combination of the first and second column, with coefficients from B_1 and B_2. In general, the i-th column is chosen as any r-tuple that is not a linear combination of any $d-2$ or less previous columns with coefficients from B_1 and B_2. This construction ensures that no linear combination of $d-1$ or fewer columns with coefficients from B_1 and B_2 will be zero, that is $d-1$ columns will

be "B_1 and B_2-independent".

If there are $|B_1|$ and $|B_2|$ possible coefficients at the time of finding *j*-th column, the number of *r*-tuples to be excluded is

$$\sum_{s=1}^{d-2}\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{j-1}{s-2m, m, j-s+m-1}|B_1|^{s-2m}|B_2|^{m}\right),$$

the linear combination of $d-2$ or less columns out of total of $j-1$ columns. If this is less than the total number of non-zero *r*-tuples, then there is certainly one more column, the *j*-th, column that can be added to the matrix. That is, if

$$\sum_{s=1}^{d-2}\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{j-1}{s-2m, m, j-s+m-1}|B_1|^{s-2m}|B_2|^{m}\right) \le q^r .$$

Now let *n* be the largest value of *j*, then an (n,k) code with minimum SK-distance d, with entries from B_1 and B_2 is given by

$$\sum_{s=1}^{d-2}\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{n}{s-2m, m, n-s+m}|B_1|^{s-2m}|B_2|^{m}\right) \ge q^r .$$

This proves the result.

Particular result. When *m* = 0, then the expression in Theorem 2 can be written as follows

$$\sum_{s=1}^{d-2}\binom{n}{s}|B_1|^{s} \ge q^r ,$$

which is a particular case when the entries are from $|B_1|$, the result considered in Theorem 1.

Corollary 1. Given an SK-partition *P* = $\{B_0, B_1, ..., B_{m-1}\}$ of Z_q, *q* prime, the sufficient condition for the existence of an (n,k) code over Z_q, at a minimum distance at least d**,** with coefficients from any two partial sets B_r and B_s, is given by

$$\sum_{s=1}^{d-2}\left(\sum_{m=0}^{\left\lfloor \frac{s}{2} \right\rfloor}\binom{n}{s-2m, m, n-s+m}|B_r|^{s-2m}|B_s|^{m}\right) \ge q^r .$$

It may be noted that the partial error correction results in reducing the redundancy. In the example below we demonstrate, by an example, that for the

same redundancy $r = 3$ and minimum distance at least $d = 5$, the existence of a much larger code word length is possible in comparison to Hamming case of all corrections.

Example. Let $Z_q = \{0, 1, 2, 3, 4, 5, 6\}$, $q = 7$ and $r = 3$. Then we consider the following two cases:

(i) Partial case. Illustrate the results of Theorem 1, considering the SK-partitions of Z_q, given by

$$P_1 = \{B_0, B_1, B_2\},$$

where $B_0 = \{0\}$, $B_1 = \{1, 2, 5, 6\}$ and $B_2 = \{3,4\}$.

The existence of an (n,k) code over Z_q, at a minimum distance at least 5, from Theorem 1 with coefficients from B_1, is given by

$$\binom{n}{1}4 + \binom{n}{2}16 + \binom{n}{3}64 \geq 342,$$

$$4n + \frac{n(n-1)}{2}16 + \binom{n}{3}\frac{n(n-1)(n-2)}{6}64 \geq 342,$$

$$32n^3 - 72n^2 + 52n \geq 1026.$$

The minimum value of *n* for which it holds is 4. The theorem guarantees the existence of a code of minimum length 4, correcting 2 entries obtained by addition of entries in $B_1 = \{1, 2, 5, 6\}$. In practice it is possible to try a greater length. In the case of the present example, *n* = 6 has been quite possible as shown by constructing the parity-check matrix

$$\begin{bmatrix} 1\,0\,0\,3\,0\,0 \\ 0\,1\,0\,0\,3\,0 \\ 0\,0\,1\,0\,0\,3 \end{bmatrix}.$$

(ii) **Hamming case.** For this we have to consider the SK-partition

$$P = \{B_0, B_1\},$$

where $B_0 = \{0\}$ and $B_1 = \{1, 2, 3, 4, 5, 6\}$.

For the existence of an (n,k) code over Z_q, the correcting single error in an (*n*, *k*) code with a minimum distance d = 5, is given by

$$\binom{n}{1}6 + \binom{n}{2}36 + \binom{n}{3}216 \geq q^r - 1,$$

$$6n + 18n(n-1) + 36n(n-1)(n-2) \geq 342,$$

$$36n^3 - 90n^2 + 60n \geq 342.$$

The minimum value of n for which it holds is 3, and by trial for $n \geq 3$ we get the check matrix for $n = 4$

$$\begin{bmatrix} 1\,0\,0\,1 \\ 0\,1\,0\,1 \\ 0\,0\,1\,1 \end{bmatrix}.$$

Comparing the two results, it is clear that for $r = 3$, the partial case has $n = 6$, and the total case has $n = 3$. This should be considered as a rather significant advantage.

4. Concluding remarks

The study of the partial error corrections is guided by practical considerations. Here we have considered only the existence problem. This is an area of research on the actual construction of more efficient codes with designed partial errors. We propose to undertake further investigations in this direction.

References

1. Berlekamp, E. R. Algebraic Coding Theory. New York, McGraw-Hill, 1968.
2. Peterson, W. W., E. J. Weldon. Error-Correcting Codes. 2nd Ed. Cambridge, Mass., MIT Press, 1972.
3. Sharma, B. D., M. L. Kaushik. Algebra of Sharma-Kaushik's Metric Inducing Partitions of Z_q. − J. Combin. Information System Science, Vol. **11**, 1986, 19-32.
4. Sharma, B. D., G. Dial. Some Tighter Bounds on Code Size with Sharma-Kaushik Metrics. – In: Presented at the Intern. Conf. on Math., Mao, Menorca, June 1987, 15-17.
5. Sharma, B. D., M. L. Kaushik. Limited Intensity Random and Burst Error Correcting Codes with Class Weight Consideration. − Elektronische Informationsverarbeitung und Kybernetik, Vol. **15**, 1979, 315-321.
6. Sharma, B. D., A. Gaur. Codes Correcting Limited Patterns of Random Errors Using SK- METRIC. – Cybernetics and Information Technologies, Vol. **13**, 2013, No 1, 34-45.