

Binary Permutation Sequences as Subsets of Levenshtein Codes and Higher Order Spectral Nulls Codes

Khmaies Ouahada, Theo G. Swart and Hendrik C. Ferreira
 Department of Electrical and Electronic Engineering Science
 University of Johannesburg, P.O. Box 524,
 Auckland Park, 2006, South Africa
 Email: {ktw, ts, hcf}@ing.rau.ac.za

Abstract—We obtain long binary sequences by concatenating the columns of (0,1)-matrices derived from permutation sequences. We then prove that these binary sequences are subsets of the Levenshtein codes, capable of correcting insertion/deletion errors and subsets of the higher order spectral nulls codes, with spectral nulls at certain frequencies.

I. INTRODUCTION

Recent work in power-line communications [1], [2] has renewed interest in permutation codes, which led to several papers regarding permutation mappings [3], [4]. In this paper we investigate some interesting properties of binary sequences which are derived from permutation codes.

Related work was done by Ferreira *et al* [5], where it was proved that the higher order spectral null (HOSN) codes are subcodes of the balanced Levenshtein codes. In this paper we will show that the binary sequences obtained from permutation sequences are both subsets of the Levenshtein codes and the HOSN codes.

We consider permutation sequences written in the passive form, such as $12 \dots M$, where each of the symbols are written as a binary sequence of length M zeros, with the symbol value indicating where a 1 is to appear. As example for $M = 3$ we have

$$\begin{aligned} 1 &\rightarrow 100 \\ 2 &\rightarrow 010. \\ 3 &\rightarrow 001 \end{aligned} \tag{1}$$

The permutation sequences for $M = 3$ are thus changed to

$$\left\{ \begin{matrix} 123 \\ 132 \\ 213 \\ 231 \\ 312 \\ 321 \end{matrix} \right\} \rightarrow \left\{ \begin{matrix} 100010001 \\ 100001010 \\ 010100001 \\ 010001100 \\ 001100010 \\ 001010100 \end{matrix} \right\}. \tag{2}$$

Therefore, each of the $M!$ permutation sequences can be converted to binary sequences of length M^2 .

An alternative representation is that of (0,1)-matrices, where only one 1 is allowed in every column and every row.

As example, the permutation sequence 2431 will be

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and the binary sequence is then constructed by concatenating the columns to form 0100000100101000. We will use $\omega = 1$ to denote that only one 1 is allowed in each row and each column.

After converting all the permutation sequences of length M to binary sequences, we define the binary permutation code, $\mathcal{P}_1(M)$, as the code containing all these binary sequences of length M^2 . The cardinality of $\mathcal{P}_1(M)$ is $|\mathcal{P}_1(M)| = M!$.

Similarly, (0,1)-matrices with ω 1s in every column and every row can be regarded as an extension of permutations. For $\omega = 2$, we can have (12)(13)(24)(34), which is

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

and by permuting the columns we can get other possible matrices. More specifically, for $M = 4$ and $\omega = 2$ we have six unique “sequences”, namely

$$\begin{aligned} &(12)(13)(24)(34) \\ &(12)(14)(23)(34) \\ &(13)(14)(23)(24) \\ &(12)(12)(34)(34) \\ &(13)(13)(24)(24) \\ &(14)(14)(23)(23) \end{aligned}$$

and when these are permuted in all the possible ways, then all the possible matrices with two 1s in every column and every row are obtained.

In general, we will use $\mathcal{P}_\omega(M)$ to denote the code containing all the possible binary sequences that is obtained from (0,1)-matrices with ω ones in each row and each column.

II. LEVENSHTEIN CODES

Levenshtein [6] showed that the 2^n binary sequences of length n can be partitioned into codebooks, each capable of correcting a single insertion or deletion error. If $\mathbf{x} = x_1x_2 \cdots x_n$, then the binary sequences can be partitioned by using

$$\sum_{i=1}^n ix_i \equiv a \pmod{m},$$

for a fixed a , where $0 \leq a \leq m-1$. The sequences are thus partitioned into m distinct codebooks where each codebook is denoted by the integer a . Let $\mathcal{L}_a(n)$ denote the Levenshtein codebook in partition a . For Levenshtein's first class of codes, we require that $m \geq n+1$. To simplify the calculations to follow, we let $\sigma = \sum_{i=1}^n ix_i$.

A. Sequences with $\omega = 1$

As example, for the $\mathcal{P}_1(3)$ sequences 100010001 we get

$$\sigma = 1 + 6 + 9 = 15,$$

which can be verified for the other sequences in (2) as well. With $n = M^2 = 9$ and $m = n+1 = 10$, we have $\sigma \equiv a \pmod{10}$ which results in all the sequences being a subset of the Levenshtein code words in the $a = 5$ partition.

We will now prove for the general case that $\mathcal{P}_1(M) \subset \mathcal{L}_a(n)$, with $n = M^2$, $m = n+1$ and some a , $0 \leq a \leq n$.

Proposition 1 *The $\mathcal{P}_1(M)$ code is a subset of the Levenshtein code, $\mathcal{L}_a(n)$, as follows*

- 1) *If M is even, then $\mathcal{P}_1(M) \subset \mathcal{L}_0(n)$.*
- 2) *If M is odd, then $\mathcal{P}_1(M) \subset \mathcal{L}_{\frac{M^2+1}{2}}(n)$.*

Proof: As we saw in (1), each sequence of length M^2 consists of M subsequences of length M . Since it is a permutation, these subsequences will always be present, just in different positions. Therefore, each subsequence is shifted multiples of M relative to each other in the large sequence. If we use any arbitrary permutation sequence, $p_1p_2 \dots p_M$, then we have

$$\begin{aligned} \sigma &= p_1 + (p_2 + M) + (p_3 + 2M) + \\ &\quad (p_4 + 3M) + \cdots + (p_M + (M-1)M) \\ &= (p_1 + p_2 + p_3 + \cdots + p_M) + \\ &\quad (M + 2M + 3M + \cdots + (M-1)M), \end{aligned}$$

also showing that the actual position of the symbols in the permutation plays no role in the sum.

We have for

$$\begin{aligned} M = 2 &\rightarrow \sigma = (1 + 2) + 2 \\ M = 3 &\rightarrow \sigma = (1 + 2 + 3) + 3(1 + 2) \\ M = 4 &\rightarrow \sigma = (1 + 2 + 3 + 4) + 4(1 + 2 + 3) \\ &\vdots \end{aligned}$$

For M in general we have

$$\begin{aligned} \sigma &= \sum_{i=1}^M i + M \sum_{i=1}^M (i-1) \\ &= \frac{M(M+1)}{2} + \frac{M^2(M-1)}{2} \\ &= \frac{M}{2}(M^2+1). \end{aligned}$$

For Levenshtein codes, $\sigma \equiv a \pmod{n+1}$, then with $n = M^2$ we require that

$$\frac{M}{2}(M^2+1) \equiv a \pmod{M^2+1}.$$

- 1) If M is even, then $M/2$ will be some integer, say r , and σ will be divisible by M^2+1 such that

$$r(M^2+1) \equiv a \pmod{M^2+1} \Rightarrow a = 0.$$

Thus, for M even, $\mathcal{P}_1(M) \subset \mathcal{L}_a(n)$ with $a = 0$.

- 2) If M is odd, then $M/2$ will be some value, say $r + 1/2$ where r is some integer. Then,

$$\sigma = r(M^2+1) + \frac{M^2+1}{2}$$

and

$$(r + \frac{1}{2})(M^2+1) \equiv a \pmod{M^2+1} \Rightarrow a = \frac{M^2+1}{2}.$$

Thus, for M odd, $\mathcal{P}_1(M) \subset \mathcal{L}_a(n)$ with $a = \frac{M^2+1}{2}$. ■

B. Sequences with any ω

In a similar manner, we can show that $(0,1)$ -matrices with other ω -values are also subsets of the Levenshtein code.

Proposition 2 *The $\mathcal{P}_\omega(M)$ code is a subset of the Levenshtein code, $\mathcal{L}_a(n)$, as follows*

- 1) *If M is even and $\omega \in \{1, 2, \dots, M-1\}$, then $\mathcal{P}_\omega(M) \subset \mathcal{L}_0(n)$ with $a = 0$,*
- 2) *If M is odd and $\omega \in \{2, 4, \dots, M-1\}$, then $\mathcal{P}_\omega(M) \subset \mathcal{L}_0(n)$ with $a = 0$,*
- 3) *If M is odd and $\omega \in \{1, 3, \dots, M-2\}$, then $\mathcal{P}_\omega(M) \subset \mathcal{L}_{\frac{M^2+1}{2}}(n)$ with $a = \frac{M^2+1}{2}$.*

Proof: Any $\omega = 2$ $(0,1)$ -matrix can be constructed from two $\omega = 1$ $(0,1)$ -matrices by XORing them, as in

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

or equivalently 1000010000100001 \oplus 0001100001000010 = 100111000110001 for the binary sequences. For each $\omega = 1$ sequence we know that $\sigma = M(M^2+1)/2$ and thus for any $\omega = 2$ sequence we will have $\sigma = M(M^2+1)$. For ω in general, $\sigma = \omega M(M^2+1)/2$, therefore we require $\omega M(M^2+1)/2 \equiv a \pmod{M^2+1}$.

- 1) As before, with M even, $M/2$ will be some integer r , then

$$\omega r(M^2 + 1) \equiv a \pmod{M^2 + 1} \Rightarrow a = 0.$$

Thus, for M even and $\omega \in \{1, 2, \dots, M-1\}$, $\mathcal{P}_\omega(M) \subset \mathcal{L}_a(n)$ with $a = 0$.

- 2) Similarly, with M odd, $M/2$ will be some value $r + 1/2$ where r is some integer, then

$$\begin{aligned} \sigma &= \omega(r + \frac{1}{2})(M^2 + 1) \\ &= \omega r(M^2 + 1) + \frac{\omega}{2}(M^2 + 1) \\ &= \omega r(M^2 + 1) + q(M^2 + 1), \end{aligned}$$

with $\omega/2$ being some integer q if ω is even, resulting in

$$(\omega r + q)(M^2 + 1) \equiv a \pmod{M^2 + 1} \Rightarrow a = 0.$$

Thus, for M odd and ω even, $\mathcal{P}_\omega(M) \subset \mathcal{L}_a(n)$ with $a = 0$.

- 3) If ω is odd, $\omega/2$ will be some value $q + 1/2$ where q is some integer, then

$$\begin{aligned} \sigma &= \omega r(M^2 + 1) + (q + \frac{1}{2})(M^2 + 1) \\ &= (\omega r + q)(M^2 + 1) + \frac{1}{2}(M^2 + 1), \end{aligned}$$

with

$$(\omega r + \frac{q}{2})(M^2 + 1) \equiv a \pmod{M^2 + 1} \Rightarrow a = \frac{M^2 + 1}{2}.$$

Thus, for M odd and ω odd, $\mathcal{P}_\omega(M) \subset \mathcal{L}_a(n)$ with $a = \frac{M^2 + 1}{2}$. ■

It is interesting to note that for M even and ω any integer, $1 \leq \omega \leq M-1$, all the binary sequences are a subset of the Levenshtein code in partition $a = 0$. Thus, for M even, we have

$$\mathcal{P}_1(M) \cap \mathcal{P}_2(M) \cap \dots \cap \mathcal{P}_{M-1}(M) \subset \mathcal{L}_0(n).$$

For M odd, the binary sequences are split between partitions $a = 0$ and $a = \frac{M^2 + 1}{2}$, depending on ω being even or odd. For M odd we have

$$\mathcal{P}_2(M) \cap \mathcal{P}_4(M) \cap \dots \cap \mathcal{P}_{M-1}(M) \subset \mathcal{L}_0(n),$$

$$\mathcal{P}_1(M) \cap \mathcal{P}_3(M) \cap \dots \cap \mathcal{P}_{M-2}(M) \subset \mathcal{L}_{\frac{M^2 + 1}{2}}(n).$$

III. HIGHER ORDER SPECTRAL NULL CODES

The technique of designing a baseband data stream to have a spectrum with nulls occurring at certain frequencies [5], is the same as having the power spectral density function (PSD) equal to zero at those frequencies [7]. Usually for simplification we choose the codeword length n as an integer multiple of N , where $f = r/N$ represents the spectral nulls at rational submultiples r/N of the symbol frequency. The

parameter N can be chosen either prime or not prime and divides n [8], i.e.

$$n = Nz.$$

In the case where N is a prime number, we have to satisfy

$$A_0 = A_1 = \dots = A_N, \quad (4)$$

where

$$A_i = \sum_{\lambda=0}^{z-1} x_{i+\lambda N}, \quad i = 1, 2, \dots, N. \quad (5)$$

In the case where N is not prime we we have to suppose that $N = cd$, where c and d are integer factors of N . The equation, which leads to nulls, is

$$A_u = A_{u+vc},$$

where $u = 1, 2, \dots, c$, $v = 1, 2, \dots, d-1$ and A_u is the same as in (5).

A. Sequences with $\omega = 1$

Proposition 3 The $\mathcal{P}_1(M)$ code is a subset of the HOSN codes with nulls at the frequency $f = r/M$ and they are not dc-free, except for $M = 2$ which is dc-free.

Proof: For $\mathcal{P}_1(M)$ we have $n = M^2$, $N = M$ and $z = M$.

Using an arbitrary permutation, $p_1 p_2 \dots p_M$, we convert each symbol to a binary sequences (as in (1)) such that $p_i \rightarrow b_{i1} b_{i2} \dots b_{iM}$, $1 \leq i \leq M$ with

$$b_{ij} = \begin{cases} 1, & j = p_i \\ 0, & \text{otherwise.} \end{cases}$$

This results in the matrix

$$\begin{bmatrix} b_{11} & b_{21} & \dots & b_{M1} \\ b_{12} & b_{22} & \dots & b_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1M} & b_{2M} & \dots & b_{MM} \end{bmatrix}.$$

When we concatenate the columns we obtain a $\mathcal{P}_1(M)$ sequence, x , as in

$$\begin{aligned} x &= x_1 x_2 \dots x_{M^2} \\ &= \underbrace{b_{11} \dots b_{1M}}_{p_1} \underbrace{b_{21} \dots b_{2M}}_{p_2} \dots \underbrace{b_{M1} \dots b_{MM}}_{p_M}, \end{aligned} \quad (6)$$

where

$$x_{j+(i-1)M} = b_{ij}. \quad (7)$$

We know that for each matrix the sum of each row and column are the same, thus the sum R_j for the j -th row is

$$R_j = \sum_{i=1}^M b_{ij}, \quad 1 \leq j \leq M.$$

Using (7), this sum becomes

$$R_j = \sum_{i=1}^M x_{j+(i-1)M}$$

and this is the same as the sum in (5). Therefore,

$$A_1 = A_2 = \dots = A_M,$$

irrespective of M being prime or not, proving that $\mathcal{P}_1(M)$ is a HOSN code.

To show that the codes are not dc-free, we use the mapping $\{0, 1\} \rightarrow \{-1, +1\}$ and calculate the running digital sum (RDS) as

$$\begin{aligned} RDS &= \sum_{i=1}^{M^2} x_i \\ &= R_1 + R_2 + \dots + R_M \\ &= MR_1 \\ &= M[1 - (M - 1)] \\ &= M(2 - M), \end{aligned}$$

For $M = 2$ we have $RDS = 0$ and $M > 2$ we have $RDS \neq 0$, proving the proposition. ■

As example, for $\mathcal{P}_1(3)$ with 123 written as 100010001 and mapping $\{0, 1\} \rightarrow \{-1, +1\}$ we find

$$\begin{aligned} A_1 &= 1 + (-1) + (-1) = -1 \\ A_2 &= (-1) + 1 + (-1) = -1 \\ A_3 &= (-1) + (-1) + 1 = -1, \end{aligned}$$

and checking all sequences in $\mathcal{P}_1(3)$, we get

$$A_1 = A_2 = A_3.$$

In the case of M not a prime number, we have as example $M = 4$, $c = 2$ and $d = 2$, where for 1234 written as 10000100000100001 and mapping $\{0, 1\} \rightarrow \{-1, +1\}$ we find

$$\begin{aligned} A_1 &= 1 + (-1) + (-1) + (-1) = -2 \\ A_2 &= (-1) + 1 + (-1) + (-1) = -2 \\ A_3 &= (-1) + (-1) + 1 + (-1) = -2 \\ A_4 &= (-1) + (-1) + (-1) + 1 = -2, \end{aligned}$$

and for all sequences in $\mathcal{P}_1(4)$, we have

$$\left. \begin{aligned} A_1 &= A_3 \\ A_2 &= A_4 \end{aligned} \right\} \Rightarrow A_1 = A_2 = A_3 = A_4.$$

It is clear that for $\mathcal{P}_1(3)$ and $\mathcal{P}_1(4)$, we have HOSN codes with nulls at frequency multiples of $1/3$ and $1/4$ respectively, as depicted in Fig. 1 and 2, as well as not being dc-free.

B. Sequences with any ω

Proposition 4 The $\mathcal{P}_\omega(M)$ code, with $\omega \in \{1, 2, \dots, M-1\}$, is a subset of the HOSN codes with nulls at the frequency $f = r/M$ and they are not dc-free codes, except when M is even and $\omega = M/2$, then $\mathcal{P}_{M/2}(M)$ is dc-free as well.

Proof: As before, the sum of the rows and columns of the matrices are going to be the same, however this time it is

$$R_j = \omega - (M - \omega) = 2\omega - M, \quad 1 \leq j \leq M. \quad (8)$$

Using the same approach to the previous proof, we find $A_1 = A_2 = \dots = A_M$, proving that $\mathcal{P}_\omega(M)$ is a HOSN code.

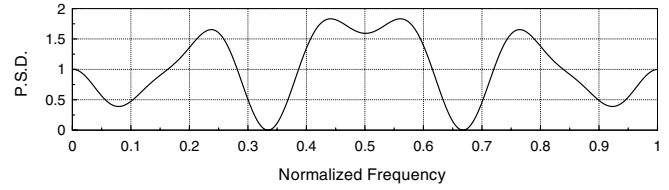


Fig. 1. Power spectral density of $\mathcal{P}_1(3)$

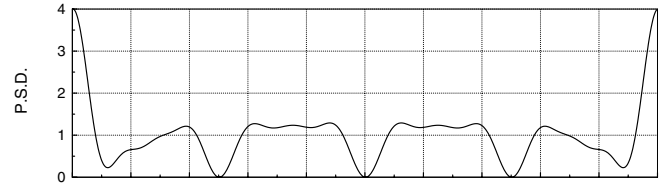


Fig. 2. Power spectral density of $\mathcal{P}_1(4)$

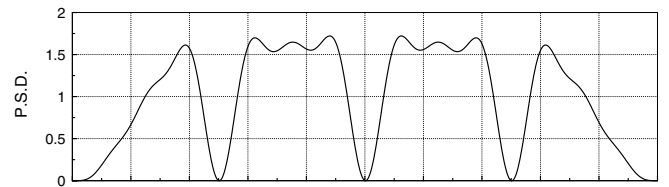


Fig. 3. Power spectral density of $\mathcal{P}_2(4)$

Using (8), the RDS for this case is derived as

$$RDS = M(2\omega - M).$$

Clearly when M is even and $\omega = M/2$ we have $RDS = 0$, proving that $\mathcal{P}_{M/2}(M)$ are dc-free codes. ■

As example, for $\mathcal{P}_2(4)$ with the sequence (12)(23)(34)(14), (which is the matrix in (3)) and mapping $\{0, 1\} \rightarrow \{-1, +1\}$, we have

$$\begin{aligned} A_1 &= 1 + 1 + (-1) + (-1) = 0 \\ A_2 &= (-1) + 1 + 1 + (-1) = 0 \\ A_3 &= (-1) + (-1) + 1 + 1 = 0 \\ A_4 &= 1 + (-1) + (-1) + 1 = 0, \end{aligned}$$

and checking all sequences in $\mathcal{P}_2(4)$, we get

$$A_1 = A_2 = A_3 = A_4.$$

Fig. 3 shows that the $\mathcal{P}_2(4)$ code is dc-free, in addition to having nulls at multiples of $f = 1/4$.

IV. CONCLUSION

We showed how (0,1)-matrices of permutation sequences of length M can be represented as binary sequences of length M^2 and proved that these binary sequences are subsets of the Levenshtein and HOSN codes, as well as generalizing the (0,1)-matrices to include those containing more than one 1 in each row and column. Although we did not include $\mathcal{P}_0(M)$ and $\mathcal{P}_M(M)$ (all zeros and all ones sequences, respectively), both are also subsets of the Levenshtein and HOSN codes.

Though trivial, these sequences are also subsets of constant weight codes and run-length limited codes.

We conclude with a summary of the properties of $\mathcal{P}_\omega(M)$:

- minimum Hamming distance of 4,
- constant weight ωM sequences,
- subset of the Levenshtein codes, with minimum Levenshtein distance of 4,
- HOSN codes, with nulls at multiples of $f = 1/M$.

REFERENCES

- [1] A. J. H. Vinck, "Coded modulation for powerline communications," *AEÜ Int. J. Elec. Comm.*, vol. 54, no. 1, pp. 45–49, Jan. 2000.
- [2] H. C. Ferreira and A. J. H. Vinck, "Interference cancellation with permutation trellis codes," in *Proc. IEEE Vehic. Tech. Conf. (Fall)*, (Boston, MA, U.S.A., Sep. 2000), pp. 2401–2407.
- [3] J.-C. Chang, R.-J. Chen, T. Kløve and S.-C. Tsai, "Distance preserving mappings from binary vectors to permutations," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 1054–1059, Apr. 2003.
- [4] T. G. Swart, I. de Beer, and H. C. Ferreira, "On the distance optimality of permutation mappings," in *Proc. IEEE Int. Symp. Inform. Theory*, (Adelaide, Australia, Sept. 4–9, 2005), pp. 1068–1072.
- [5] H. C. Ferreira, W. A. Clarke, A. S. J. Helberg, K. A. S. Abdel-Ghaffar, and A. J. H. Vinck, "Insertion/deletion correction with spectral nulls," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 722–732, Mar. 1997.
- [6] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," *Sov. Phys. Dokl.*, vol. 10, no. 8, pp. 707–710, Feb. 1966, in Russian.
- [7] K. A. S. Immink, *Codes for Mass Data Storage Systems*, Shannon Foundation Publishers, The Netherlands, 1999.
- [8] E. Gorog, "Alphabets with Desirable Frequency Spectrum Properties," *IBM J. Res. Develop.*, vol. 12, pp. 234–241, May 1968.