Frequency Metric in Binary Spaces

N.Suresh Babu
Department of Mathematics
University College, Thiruvananthapuram
Kerala, India.

Abstract

This paper deals with a new metric named "Frequency Metric" defined in the space F_2^n of binary sequences of length n. We show that the frequency metric can be "generated" by a particular "basis" of F_2^n. Procedures to construct all "frequency classes" and an "orbit" (ie. the set of all binary sequences at a given frequency distance from any given fixed binary sequence) are obtained.

Key words: *Binary sequences, Norm, Metric, Binary code, Hamming metric, Transition number.*

2010 Mathematics Subject Classification: *94B25, 94B99.*

1 Introduction

Binary sequences have innumerable applications in diverse areas.They are used to represent different things in different applications. Depending on their specific representation, the binary sequences can be studied, compared and classified.

In general, consider a two-state system with states represented by **0** and **1**. For any positive integer $'n'$, an observation of the states of the system through n consecutive time points can be represented by the binary sequence $a_1a_2 \ldots a_n$ where a_i is the state of the system at the i^{th} time point. Thus the set of all possible observations of the system can be represented by F_2^n, the set of all binary sequences of length 'n'. We give F_2^n, the structure of a linear space over the binary Field F_2.

In many applications, it is required to find out how frequently the above system changes its states in a given observation. Based on this, it is also required to know how different is one observation from the other. Consequently, how to classify the set of all observations based on the frequency of the change of states of the system.

Equivalently, in terms of the binary

sequences, the above problems can be stated as:
(i).Define a *norm* in F_2^n(considering F_2^n as a linear space over F_2) which gives the number of transitions between 0 and 1 in any binary sequence of F_2^n.
(ii).Derive the *metric* induced by this norm.
(iii). Classify the binary sequences in the space F_2^n with respect to this norm.

A problem requiring such a norm is described in [1] where cellular activities are analysed using a binary mapping technique that detects difference in spike trains with respect to their frequency and/or pattern of firing discontinuities. Here a binary sequence composed of $\mathbf{1}'\mathbf{s}$ and $\mathbf{0}'\mathbf{s}$ indicates the the presence ($\mathbf{1}$) and absence ($\mathbf{0}$) of spikes in each bin. To compare two such binary sequences, the *transition number* (ie. the number of transitions from 1 to 0 and 0 to 1) is used. But the transition number does not define a *norm* on F_2^n as it does not satisfy the properties of a norm (it is infact a *semi-norm* [3]) and hence it will not induce a *metric* (distance function) in F_2^n.

The analysis of human heart beat is another area in which a metric based on the frequency of transitions is much needed. Refer to [4] in which a *distance* is defined between symbolic sequences which is actually not a metric [5].

In this work, a new norm called *frequency weight* in F_2^n is proposed. The metric induced by this norm (*frequency metric*) is derived. Frequency metric can be proved as a metric generated by a *basis* of F_2^n. Finally, procedures to construct all the *frequency classes* and *orbits* in F_2^n are described with example.

2 Definitions

Let F_2^n be the set of all binary sequences of length n. ie. $F_2^n = \{x = x_1x_2\ldots x_n : x_i = 0 \ or \ 1\}$. Clearly F_2^n can be considered as an n- dimensional linear space over the binary field $F_2 = \{0,1\}$ with respect to *modulo* 2 addition ($+_2$) and *modulo* 2 multiplication ($._2$).

Definition 2.1. *For each* $x = x_1x_2\ldots x_n \in F_2^n$, *the frequency weight or simply the frequency of* x *is defined as* $wt_f(x) = [\Sigma_{i=1}^{n-1}(x_i +_2 x_{i+1})] + x_n$.

Since $(x_i \ +_2 \ x_{i+1}) = 0 \ or \ 1$ respectively as $x_i = x_{i+1}$ or $x_i \neq x_{i+1}$, the sum $[\Sigma_{i=1}^{n-1}(x_i +_2 x_{i+1})]$ gives the number of of transitions between 0 and 1 (ie. transition number) in x.

Hence the *frequency* of x can be defined as:

$$wt_f(x) = \begin{cases} \text{transition number of x} & \text{if} x_n = 0 \\ (\text{transition number of x}) + 1 & \text{if} x_n = 1. \end{cases}$$

Result 1. *Frequency weight* $wt_f(x)$ *defines a norm on* F_2^n.

Proof: We have the following from the definition of $wt_f(x)$.
(i). For all $x \in F_2^n$, $wt_f(x) \geq 0$ and $wt_f(x) = 0$ if and only if x = 0 .
(ii). $wt_f(1.x) = wt_f(x)$ and $wt_f(0.x) = wt_f(0) = 0$ for all $x \in F_2^n$.

2

Now, let $x = x_1x_2x_3...x_n$, $y = y_1y_2y_3...y_n$ and $z = x + y = z_1z_2z_3...z_n$. Then $z_i = x_i +_2 y_i$for i = 1 to n.

For i = 1 to n-1, if $z_i +_2 z_{i+1} = 1$ then $(x_i +_2 y_i) +_2 (x_{i+1} +_2 y_{i+1}) = 1$

$\Rightarrow (x_i +_2 y_i) \neq (x_{i+1} +_2 y_{i+1})$

$\Rightarrow (x_i +_2 x_{i+1}) \neq (y_i +_2 y_{i+1})$

$\Rightarrow [(x_i +_2 x_{i+1}) = 1 \;\; and \;\; (y_i +_2 y_{i+1}) = 0]$

or $[(x_i +_2 x_{i+1}) = 0 \;\; and \;\; (y_i +_2 y_{i+1}) = 1]$.

Also $z_n = 1 \Rightarrow x_n +_2 y_n = 1 \Rightarrow x_n = 1 \;\; and \;\; y_n = 0 \;\; or \;\; x_n = 0 \;\; and \;\; y_n = 1$.

This implies that (iii) $wt_f(x + y) \leq wt_f(x) + wt_f(y)$

By (i), (ii) and (iii) it is clear that $wt_f(x)$ defines a norm on F_2^n.

Definition 2.2. *Let $x, y \in F_2^n$. The frequency distance $d_f(x, y)$ between x and y is the frequency weight of $x - y$. ie. $d_f(x, y) = wt_f(x - y) = wt_f(x +_2 y)$.*

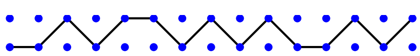
By Result 1, the frequency distance d_f is a proper metric on F_2^n [3].

Example

Space : $F_2{}^{15}$

1. Binary Sequence (x) : 001011010100101

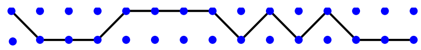
   Binary Curve :

   Frequency Weight $(wt_f(x))$: 12

2. Binary Sequence (y) : 100011110101000

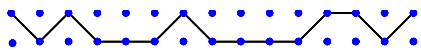
   Binary Curve :

   Frequency Weight $(wt_f(y))$: 7

3. Binary Sequence (x-y) : 101000100001101

   Binary Curve :

   Frequency distance between x and $y = d_f(x, y) = (wt_f(x - y)) = 9$

3 Generator of the frequency norm

In this section we construct a basis for the linear space F_2^n which *generates* the frequency norm on F_2^n. In [2], a new family of metrics is introduced in F_2^n. Each such metric is defined (or generated) by a spanning set B of the linear space F_2^n. The norm of any vector in the space is defined as the size of the minimal subset of B whose span contains this vector. We can prove that the frequency norm can be represented in a similar way. The spanning set corresponding to the frequency norm can be constructed as follows.

For $1 \leq i \leq n$, we denote by $[i]$, the binary sequence $x_1x_2x_3...x_n$ whose first i terms are 1 and the remaining $n-i$ terms are 0. Consider $B = \{[1], [2], [3], ..., [n]\}$. Clearly B is a spanning set for F_2^n. In fact, B is a basis for F_2^n.

Definition 3.1. *For each binary sequence $x = x_1x_2x_3...x_n \in F_2^n$, the unique representation of x as a linear combination of the elements of B over F_2 ie. $a_1.[1]+_2a_2.[2]+_2...+_2a_n.[n]$ where $a_i \in F_2$ is called the 'B- expansion' of x.*

Definition 3.2. *The 'B- weight' of $x \in F_2^n$ is the size of the set $I_x = \{i : a_i \neq 0 \ \ in\ the\ \ B-expansion\ of\ x\ \}$.*

Thus B - weight of $x \in F_2^n$ is the size of the smallest subset $I_x \subseteq \{1, 2, , n\}$ such that x is the linear combination of $\{[i] : i \in I_x\}$ over F_2. If $I_x = \{i_1, i_2, ..., i_r\}$, then the $B-weight(x) = r$ and $x = [i_1] +_2 [i_2] +_2 ... +_2 [i_r]$. Note that this representation is unique. It can be proved that B-weight defines a norm in F_2^n

Theorem 3.3. *Let $x = x_1x_2...x_n \in F_2^n$. Then the B- weight(x) $= wt_f(x)$. ie. B-weight of x is equal to the frequency weight of x.*

Proof: Let $x = x_1x_2...x_n \in F_2^n$. If $a_1.[1] +_2 a_2.[2] +_2 ... +_2 a_n.[n]$ is the B-expansion of x, then $x_i = a_i +_2 a_{i+1} +_2 ... +_2 a_n$ for $1 \leq i \leq n$. Let $I = \{i \leq n-1 : a_i \neq 0\}$. Suppose $i \in I$. Then $a_i = 1$ and $1 \leq i \leq n-1$. Then $x_i = a_i +_2 a_{i+1} +_2 ... +_2 a_n$ and $x_{i+1} = a_{i+1} +_2 a_{i+2} +_2 ... +_2 a_n$. This implies that $x_i \neq x_{i+1}$.

Also if $a_n = 1$, then $x_n = 1$. Hence each $i \in I$ $(1 \leq i \leq n-1)$ contributes a weight of 1 to $wt_f(x)$ and if $a_n = 1$, then it contributes a weight of 1 to $wt_f(x)$. Thus,

$$wt_f(x) = \begin{cases} sizeof\ I & if a_n = 0 \\ (sizeof\ I) + 1 & if a_n = 1. \end{cases}$$

Hence,

$$\begin{aligned} wt_f(x) &= size\ of\ I\ +\ a_n \\ &= size\ of\ I_x. \\ &\quad (Note\ that\ n\ \in\ I_x\ if\ and \\ &\quad\ only\ if\ a_n\ =\ 1) \\ &= B-weight(x). \end{aligned}$$

Thus, if $i \in I_x$ then the basis element $[i]$ is 'present' in the B-expansion of x contributing a weight of '1' to $wt_f(x)$. In other words, $wt_f(x)$ is the number of elements of B present in the B-expansion of 'x'. Hence we say that the set B generates $wt_f(x)$.

4 Frequency Classes and Orbits

In this section, two practically relevant problems with respect to the frequency metric are discussed. One is to classify the set of all binary sequences in F_2^n into different frequency classes B_i where $B_i = \{x \in F_2^n : wt_f(x) = i\}$ for $1 \leq i \leq n$.We call B_i as the i^{th} frequency class.

The second problem is to construct the set of all binary sequences at a required (frequency) distance r from a fixed sequence x in F_2^n.

Definition 4.1. *For $1 \leq i \leq n$, the i^{th} frequency class is the set $B_i = \{x \in F_2^n : wt_f(x) = i\}$.*

Construction of Frequency classes

As $B = \{[1], [2], ..., [n]\}$ is the generator set of the frequency weight,

$wt_f([i]) = 1$ for each i. Hence $B_1 = B$ $= \{[1], [2], ..., [n]\}$.

Now, $wt_f([i] +_2 [j]) = 2$ for all $[i], [j] \in B, (i \neq j)$. Hence $B_2 = \{[i] +_2 [j] : 1 \leq i, j \leq n, i \neq j\}$.

Similarly, $B_3 = \{[i_1] +_2 [i_2] +_2 [i_3] : 1 \leq i_1, i_2, i_3 \leq n\}$ and so on.

$B_n = \{[1] +_2 [2] +_2 +... + [n]\}$.

Note that the size of the i^{th} frequency class B_i in F_2^n is nC_i. ie. $|B_i| = nC_i$.

Definition 4.2. *Let $x \in F_2^n$ be fixed. The i^{th} **orbit** around x denoted by $B_i(x)$ is the set of all binary sequences in F_2^n at a frequency distance 'i' from x.*

ie. $B_i(x) = \{y \in F_2^n : d_f(x, y) = i\}$.

Construction of Orbits

Given $x \in F_2^n$, we have

$$\begin{aligned} B_i(x) &= \{y \in F_2^n : d_f(x, y) = i\} \\ &= \{x +_2 z \in F_2^n : d_f(0, z) = i\} \\ &= \{x +_2 z \in F_2^n : wt_f(z) = i\} \\ &= \{x +_2 z \in F_2^n : z \in B_i\} \\ &= x +_2 B_i. \end{aligned}$$

Hence $B_i(x) = x +_2 B_i$ for $1 \leq i \leq n$.

5 Example

Consider the binary space,

$$\begin{aligned} F_2^5 &= \{a_1a_2a_3a_4a_5 : a_i \in F_2\} \\ &= \{00000, 10000, 01000, \ldots, 11111\}. \end{aligned}$$

The generator is given by:

$B = \{10000, 11000, 11100, 11110, 11111\}$.

The Frequency classes are given by:

$B_1 = \{10000, 11000, 11100, 11110, 11111\}$.

$B_2 = \{01000, 01100, 01110, 01111, 00100, 00110, 00111, 00010, 00011, 00001\}$.

$B_3 = \{10100, 10110, 10111, 10010, 10011, 10001, 11010, 11011, 11101, 11001\}$.

$B_4 = \{01010, 01011, 01001, 01101, 00101\}$.

$B_5 = \{10101\}$.

If we fix $x = 11001$, then the orbits of x are given by:

$B_1(x) = x + B_1 = \{01001, 00001, 00100, 00111, 00110\}$.

$B_2(x) = x + B_2 = \{10001, 10101, 10111, 10110, 11101, 11111, 11110, 11011, 11010, 11000\}$

$B_3(x) = x + B_3 = \{010101, 01111, 01110, 01011, 01010, 01000, 00011, 00010, 00100, 00000\}$

$B_4(x) = x + B_4 = \{10011, 10010, 10000, 10100, 11100\}$

$B_5(x) = x + B_5 = \{01100\}$.

6 Conclusion

We have introduced a new norm and the corresponding metric in the linear space F_2^n of binary sequences of length n. We call these as the frequency norm and frequency metric respectively. Frequency norm is proved to be generated by a basis of F_2^n. Two practically relevant problems of constructing all the frequency classes and orbits in this space were discussed.

5

References

[1] **M.Bodner,Y.D.Zhou,J.M.Fuster**, "High-Frequency Transitions in Cortical Spike Trains Related to Short-term memory ", Neuroscience, vol.86, no.4, pp.1083 - 1087, 1998.

[2] **Ernst M.Gabidulin, Jurian Simonis**, "Metrics Generated by Families of Subspaces " IEEE Trans.on Information theory, vol.44, no.3, May 1998.

[3] **Erwin Kreyszig**, "Introductory Functional Analysis with Applications ", John Wiley and Sons.

[4] **A.C.C. Yang., etal.**, "Linguistic Analysis of the Human heart rate using Frequency and Rank order statistics " Physical Review Letters, vol.90, 108103 (2003).

[5] **Alexander Kraskov., etal.**, "Comment on "Linguistic Analysis of the Human heart rate using Frequency and Rank order statistics "" Physical Review Letters, vol.92, no.10 (2004).

[6] **Monika, Dr.Ranjeet Singh**, "Primitive Idempotents of Abelian Codes of length $4p^nq^m$ "" International Journal of Mathematics Trends and Technology, vol.35 no.3 (July, 2016).