

International Journal of Sciences: Basic and Applied Research (IJSBAR)



ISSN 2307-4531
(Print & Online)

<http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>



A Note on Exploration of Sequence Spaces and Function Spaces on Interval $[0,1]$ for DNA Sequencing

Shailendra K. Mishra ^{a*}, Vinod Parajuli ^b, Sergei Silvestrov ^c

^{a,b} *Department of Engineering Science and Humanities, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*

^c *Division of Applied Mathematics, The school of Education, Culture and Communication; Malardalen University, Vasteras, Sweden*

^a*Email: skmishra12@hotmail.com*

^b*Email: vparajuli@hotmail.com*

^c*Email: sergei.silvestrov@mdh.se*

Abstract

In [7] authors studied the sequence spaces and function spaces on interval $[0,1]$. Further they introduced new sequence spaces by using generalized p -summation method and proved these spaces of sequences and functions are Banach spaces. In this paper we extend the results of authors in [7] by introducing a new basis function and strongly p -summation method.

Keywords: DNA sequence; sequence space; function space; polynomial space; basis function; p -summation method.

* Corresponding author. Tel.: +977-1-5543072

E-mail address: skmishra12@hotmail.com

1. Introduction

John Maynard Smith in 1970 first introduced the notion of sequence space for protein evolution. He proposed a “sequence space” where all possible proteins are arranged in a protein space in which neighbors can be interconnected by single mutation [1]. These problems are not only unique to protein structures but relevant to many other areas such as DNA sequence, brain imaging, climate data, financial data and others. In these area of interest the data have common features that: data are enormous, information is multi dimensional and complex, the sample size is relevantly small, they posses finitely many non zero elements in the sequence and some elements in the sequence repeat many times. For instance, four types of nucleotide A, T, G and C are linked in different orders in extremely long DNA molecules. It now becomes a continuing challenge for scientists, engineers, mathematicians and others to record and preserve data in these endeavors.

When the data received from the reservoir to obtain some information have lower dimension and samples have larger size, the statistical methods such as that the covariance matrix [2,3], dot matrix[4] and position weight matrix [5,6] can deal with the cases promptly in a simplified way. However, when data have multidimensional character and the sample size is smaller, the statistical methods may lead to errors [7].

In this connection authors [7] have pointed out the necessity of the new definition of norm to fit a given data ‘a’ in a of set some class samples S .

Based on the sequence spaces and function spaces on interval [0,1], in the present paper we examine the behaviors of sequences generated by DNA nucleotides. It has been aimed to extend the results of authors [7] by: introducing new function space in [0,1], extending the basis function $\frac{x^n}{n!}$, introducing a new sequence $b = (b_n) = (\sum_{v=n}^{\infty} a_v)$ which can characterize DNA sequence and obtaining some new completion results among the existing spaces in [0,1].

2. Preliminaries and reviews

Before introducing new function spaces we begin with following definitions, notations and function spaces as defined in [7].

Let $a = (a_1, a_2, a_3, \dots, a_n, \dots)$ be a DNA sequence where $a_n \in \{A, C, T, G\}$ and

$$a(x) = A p_1(x) + C p_2(x) + T p_3(x) + G p_4(x) \quad (2.1)$$

Clearly, for different DNA sequence, we have different polynomials $p_j(x)$. It is a simpler reserve form. To extend it into a sequence of infinitely many non zero terms, we take $x \in [0,1]$. Here, $a(x)$ is called the generation function

in the classical queuing theory. We remark that the generation function is not continuous function defined in $[0, 1]$. Hence in order to find out a feasible form of $a(x)$ we integrate first and then differentiate.

Denoting by L the integral operation and performing it for constant 1 leads to,

$$L^1(1)(x) = \int_0^x 1 \, dx = x$$

$$L^2(1)(x) = \int_0^x L^1(1) \, x \, dx = \frac{x^2}{2}$$

.

.

.

Generalizing we get,

$$L^n(1)(x) = \frac{x^n}{n!} \text{ for all } n \in \mathbb{N}. \tag{2.2}$$

For any polynomial $p_n(x)$ of order n , it can be written as

$$\begin{aligned} p_n(x) &= a_0 \cdot 1 + a_1 x + a_2 \frac{x^2}{2!} + \dots + a_n \frac{x^n}{n!} \\ &= \left[\sum_{k=0}^n a_k L^k \right] (1)(x) \end{aligned} \tag{2.3}$$

Next , we consider the differential operator D for the for the function $\frac{x^n}{n!}$ which yields,

$$D^1 \left(\frac{x^n}{n!} \right) = \frac{x^{n-1}}{(n-1)!} , D^2 \left(\frac{x^n}{n!} \right) = \frac{x^{n-2}}{(n-2)!} , \dots$$

In general for $1 \leq k \leq n$, it holds that

$$D^k \left(\frac{x^n}{n!} \right) = \frac{x^{n-k}}{(n-k)!} \tag{2.4}$$

Therefore the coefficient sequence is given by

$$(a_0, a_1, a_2, \dots, a_n) = (D^0, D^1, D^2, \dots, D^n) p_n |_{x=0} \tag{2.5}$$

and $\frac{x^n}{n!}$ is defined to be the basis function.

Moreover, the polynomial space over $[0, 1]$, denoted by $P[0,1]$, is a Banach space with the norm

$$\|p\|_{\phi} = \sup_{n \geq 0} \left\{ \|D^n p\|_{\infty} \right\} \tag{2.6}$$

where $\|f\|_{\infty} = \max_{0 \leq x \leq 1} |f(x)|$.

In this space, the integral and differential operations are bounded linear operators. To extend to an infinite sequence, we take a subset $C_M^{\infty}[0,1]$ of $C^{\infty}[0,1]$ defined by

$$C_M^{\infty} = \left\{ f \in C^{\infty}[0,1]: \sup_{n \geq 0} \|D^n p\|_{\infty} < \infty \right\} \tag{2.7}$$

$C_M^{\infty}[0,1]$ is a Banach space. Now for the function space on interval $[0,1]$, there exist the following set inclusion relations

$$P[0,1] \subset C_M^{\infty}[0,1] \subset C^{\infty}[0,1] \subset C^k[0,1] \subset C[0,1] \subset L^{\infty}[0,1] \subset L^p[0,1] \subset L^1[0,1] \tag{2.8}$$

But the completion of $(P[0,1], \|\cdot\|_{\phi})$ is not the space $(C_M^{\infty}[0,1], \|\cdot\|_{\phi})$. For the completion of the space

$$(P[0,1], \|\cdot\|_{\phi}) \text{ authors have defined the following spaces on } [0,1]: C_{\phi,0}[0,1] = \left\{ f(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} : \lim_{n \rightarrow \infty} a_n = 0 \right\},$$

$$C_{\phi,p}[0,1] = \left\{ f(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} : \sum_{n=0}^{\infty} |a_n|^p < \infty \right\} \text{ for } p \geq 1 \text{ and}$$

$$C_{\phi,\infty}[0,1] = \left\{ f(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} : \sup_{n \geq 0} |a_n| < \infty \right\}$$

These spaces are isomorphic to C_0, l_p and l_{∞} respectively [7].

Obviously $P[0,1] \subset C_{\phi,0}[0,1] \subset C_M^{\infty}[0,1]$ and authors have shown the following set inclusion relations:

$$P[0,1] \subset C_{\phi,1}[0,1] \subset C_{\phi,p}[0,1] \subset C_{\phi,0}[0,1] \subset C_{\phi,\infty}[0,1] = C_M^{\infty}[0,1], \quad 1 \leq p < \infty \tag{2.9}$$

3. New function space on [0,1] and new set inclusion relations

We define for any $x \in [0,1]$, a polynomial function of order n

$$\begin{aligned}
 p_n(x) &= \sum_{\nu=1}^n a_{\nu} \left(\sum_{k=1}^{\nu} L^k(1)(x) \right), \quad a_0=0. \\
 &= \sum_{\nu=1}^n a_{\nu} \left(\sum_{k=1}^{\nu} \frac{x^k}{k!} \right)
 \end{aligned} \tag{3.1}$$

where L is integral operator and

$$\sum_{k=1}^{\nu} \frac{x^k}{k!} \text{ for } \nu = 1, 2, 3, \dots, n \tag{3.2}$$

is new basis function defined in the polynomial function $P[0,1]$ which illustrates better approximation to the problem. Further by using differential operator for the basis function

$$\sum_{k=1}^{\nu} \frac{x^k}{k!} \text{ for } \nu = 1, 2, 3, \dots, n \text{ we find that ,}$$

$$D^k \left[\sum_{i=1}^{\nu} \frac{x^i}{i!} \right] = \frac{x^{\nu-k}}{(\nu-k)!}, \quad 1 \leq k \leq \nu$$

Obviously ,

$$D^1 p_n(0) = a_1 + a_2 + \dots + a_n$$

$$D^2 p_n(0) = a_2 + a_3 + \dots + a_n$$

.

.

$$D^n p_n(0) = a_n$$

Therefore the coefficient sequence $b = (b_n)$ is given by $(a_1 + a_2 + \dots + a_n, a_2 + a_3 + \dots + a_n, \dots, a_n) = (D^1, D^2, \dots, D^n) p_n(x) \Big|_{x=0}$ (3.3)

Thus we obtained new coefficient sequence to characterize DNA sequence. With the coefficient sequence $b = (b_k)$ defined by $b_k = \sum_{v=k}^n a_v$, for all k ; we can characterize DNA sequence and the result is helpful to explore for the possible application in DNA sequencing. The following table shows the distribution of the coefficient sequence $b = (b_k)$ with all possible alignments of DNA nucleotides.

Table 1. Distribution of the coefficient sequence $b = (b_k)$

$b_1 =$	$a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 \dots a_n$
$b_2 =$	$a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 \dots a_n$
$b_3 =$	$a_3 a_4 a_5 a_6 a_7 a_8 a_9 \dots a_n$
$b_4 =$	$a_4 a_5 a_6 a_7 a_8 a_9 \dots a_n$
.....
.....
.....
$b_n =$	a_n

where $a_n \in \{A, C, T, G\}$. In computational process, if we input a DNA sequence, BLAST (Basic Local Alignment Search Tool) will display all possible gene matches with closure similarities between the existing DNA sequence in Gene Bank and the input sequence. The most likely matches will be displayed from top to bottom sequence alignments.

As in [7] polynomial function $P[0,1]$ is now a finite dimensional normed space normed by,

$$\|p\|_{\psi} = \sup_{n \geq 0} \left\{ \sum_{k=1}^n \|D^k p - D^{k+1} p\|_{\infty} \right\}.$$

To extend the case to an infinite dimension, consider a subset of function space $C^{\infty}[0,1]$ defined by

$$C_M^{\infty}[0,1] = \left\{ f \in C^{\infty}[0,1] : \sup_{n \geq 0} \|D^n f\|_{\infty} < \infty \right\}$$

which is a linear space.

The authors [7] have shown the sets inclusion relations as

$$P[0,1] \subset C_{\phi,1}[0,1] \subset C_{\phi,p}[0,1] \subset C_{\phi,0}[0,1] \subset C_{\phi,\infty}[0,1] = C_M^{\infty}[0,1], \quad 1 \leq p < \infty.$$

Let the completion of the space $C_{\phi,0}[0,1]$ be $C_{\psi,0}[0,1]$. Then we have the following representation theorem.

Theorem 3.1: $C_{\psi,0}[0,1] = \left\{ g(x) = \sum_{k=1}^{\infty} a_k \left(\sum_{\nu=1}^k \frac{x^{\nu}}{\nu!} \right) : \lim_{n \rightarrow \infty} b_n = 0 \right\}$ is isomorphic to the space $C_{\phi,0}[0,1]$, where

$$b = (b_n) = \left(\sum_{\nu=n}^{\infty} a_{\nu} \right).$$

Proof: We define an operator

$$T : C_{\psi,0}[0,1] \rightarrow C_{\phi,0}[0,1] \text{ by}$$

$$(b_n) \mapsto (a_n) = T((b_n)).$$

The linearity of T is obvious. Now,

$$T((b_n)) = g(x)$$

$$= \sum_{n=1}^{\infty} b_n \frac{x^n}{n!} = b_1 \frac{x}{1!} + b_2 \frac{x^2}{2!} + b_3 \frac{x^3}{3!} + \dots$$

$$= (a_1 + a_2 + a_3 + \dots) \frac{x}{1!} + (a_2 + a_3 + a_4 + \dots) \frac{x^2}{2!} + \dots$$

$$= a_1 \frac{x}{1!} + a_2 \left(\frac{x}{1!} + \frac{x^2}{2!} \right) + a_3 \left(\frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} \right) + \dots$$

$$= \sum_{n=1}^{\infty} a_n \left(\sum_{k=1}^n \frac{x^k}{k!} \right)$$

Hence T is bijective. Thus T is isomorphism mapping and $C_{\psi,0}[0,1]$ is isomorphic to $C_{\phi,0}[0,1]$.

Now for $p \geq 1$ we define new norm on the space $P[0,1]$ by

$$\|g\|_{\psi,p} = \left\{ \sum_{k=1}^{\infty} \left\| D^k f - D^{k+1} f \right\|_{\infty}^p \right\}^{\frac{1}{p}}.$$

Let $C_{\psi,p}[0,1]$ be the completion of the space $C_{\phi,p}[0,1]$. Then we have the following representation theorem.

Theorem 3.2: The space $C_{\psi,p}[0,1] = \left\{ g(x) = \sum_{k=1}^{\infty} a_k \left(\sum_{\nu=1}^k \frac{x^{\nu}}{\nu!} \right) : \sum_{n=0}^{\infty} |b_n|^p < \infty \right\}$ is isomorphic to the space $C_{\phi,p}[0,1]$.

The proof of the theorem follows immediately by using isomorphism operator defined as in the proof of theorem 3.1.

Further, letting $p \rightarrow \infty$ we define new norm on the space $P[0,1]$ by

$$\|g\|_{\psi,\infty} = \sup_{n \geq 0} \left\{ \sum_{k=1}^n \left\| D^k f - D^{k+1} f \right\|_{\infty} < \infty \right\}. \text{ Then we have the following theorem.}$$

Theorem 3.3: The space $C_{\psi,\infty}[0,1] = \left\{ g(x) = \sum_{k=1}^{\infty} a_k \left(\sum_{\nu=1}^k \frac{x^{\nu}}{\nu!} \right) : \sup_{n \geq 0} |b_n| < \infty \right\}$ is isomorphic to the space $C_{\phi,\infty}[0,1]$.

The proof is similar to the proof of theorem 3.1 .

We, therefore, observe the following sets inclusion relations:

$$P[0,1] \subset C_{\phi,p}[0,1] \subset C_{\psi,p}[0,1] \subset C_{\phi,0}[0,1] \subset C_{\psi,0}[0,1] \subset C_{\phi,\infty}[0,1] \subset C_{\psi,\infty}[0,1] = C_M^\infty [0,1], 1 \leq p < \infty$$

(3.4)

Moreover the spaces $C_{\psi,0}[0,1]$, $C_{\psi,p}[0,1]$ and $C_{\psi,\infty}[0,1]$ are respectively equivalent to $C_{\phi,0}[0,1]$, $C_{\phi,p}[0,1]$ and $C_{\phi,\infty}[0,1]$. Hence $C_{\psi,0}[0,1]$, $C_{\psi,p}[0,1]$ and $C_{\psi,\infty}[0,1]$ are Banach spaces with their natural norms.

4. Strongly Summation Method

Let (b_n) be a sequence of real or complex numbers and satisfy $\lim_{n \rightarrow \infty} b_n = 0$. We define a new strongly p- summation method for the sequence (b_n) as

$$s_{0,p} = |b_n| = \left(|b_n|^p \right)^{\frac{1}{p}}$$

$$s_{1,p} = \left(|b_{n-1}|^p + |b_{n-2}|^p \right)^{\frac{1}{p}}$$

$$s_{2,p} = \left(|b_{n-2}|^p + |b_{n-3}|^p + |b_{n-4}|^p \right)^{\frac{1}{p}}$$

.

.

.

$$s_{k,p} = \left(\sum_{j=0}^k |b_{n-k-j}|^p \right)^{\frac{1}{p}}$$

We, therefore, obtained a new non negative sequence $(s_{k,p}; k \geq 0)$; where $s_{k,p}^m \leq s_{k,p} \leq s_{k,p}^M$ and $s_{k,p}^m$ and $s_{k,p}^M$ are the values in decreasing and increasing queuing. Then it is a normed space normed by

$$\|(b_n)\|_{H,p} = \sup_{k \geq 0} s_{k,p} \tag{4.1}$$

where H is the generalized strongly summation and p is the p-norm in finite dimensional space .

In particular when $\{b_n\} \in l^p$, $s_{k,p} \rightarrow 0$, as $k \rightarrow \infty$ hence $\|(b_n)\|_{H,p} < \left(\sum_{n \geq 1} |b_n|^p\right)^{\frac{1}{p}} < \infty$.

Now we define the sequence spaces by

$$c_{H,p,M} = \left\{ (b_n) : \sup_{k \geq 0} s_{k,p} < \infty \right\},$$

$$c_{H,p} = \left\{ (b_n) : \lim_{k \rightarrow \infty} s_{k,p} = 0 \right\}$$

These spaces are evidently Banach spaces with their norm as defined in (4.1).

5. Conclusions

We have extended the basis function $\frac{x^n}{n!}$ ($n \in \mathbb{N}$) in [7] into $\sum_{k=1}^n \frac{x^k}{k!}$ ($n \in \mathbb{N}$) as a new basis function and introduced a new sequence $b = (b_n) = \left(\sum_{v=n}^{\infty} a_v\right)$ where $a_n \in \{A, C, T, G\}$ to characterize DNA sequence. We have also established some isomorphism theorems on newly introduced function spaces and obtained some new completion results between the existing spaces in [7]. In future one can find the dual spaces for newly defined function spaces.

Acknowledgements

This work was supported by the Erasmus-Mundus project IDEAS. Shailendra Kumar Mishra and Vinod Parajuli are grateful to the research environment in Mathematics and Applied Mathematics at the Division of Applied Mathematics of the School of Education, Culture and Communication (UKK) at Mälardalen University for hospitality and creating excellent conditions for research and cooperation as well as to the Erasmus-Mundus project IDEAS for support. Sergei Silvestrov is grateful also for the support at different times by The Swedish Research Council, The Royal Swedish Academy of Sciences, The Crafoord Foundation, The Royal Physiographic Society in Lund and The Swedish Foundation for International Cooperation in Research and Higher Education (STINT).

Our thanks are also due to the referee for valuable comments and suggestions.

References

- [1] C. Gustafsson, S. Govindarajan and R. Emig. "Exploration of sequence space for protein engineering". *Journal of Molecular Recognition*, vol.14, pp.308-314, 2001
- [2] N. El Karoui. "Operator norm consistent estimation of large dimensional sparse covariance matrices". *The Annals of Statistics*, vol. 36, pp. 2717-2756, 2008
- [3] A. J. Rothman, E. Levina, and J. Zhu. "A new approach to Cholesky-based covariance regularization in high dimensions" *Biometrika*, vol.97, pp. 539-550, 2010
- [4] J. Jun Shu and L. S. Ouw. "Pairwise Alignment of the DNA sequence Using Hypercomplex Number Representation". *Bulletin of Mathematical Biology*, vol.66, pp.1423-1438, 2004
- [5] S. Sinha. "On counting position weight matrix matches in a sequence, with application to discriminative motif finding". *Bioinformatics*, vol. 22, pp. e454-463, 2006
- [6] X. Xia. "Position Weight Matrix, Gibbs Sampler, and the Associated Significance Tests in Motif Characterization and Prediction". *Scientifica* , vol.2012, pp .(Article ID 917540, 15 pages) 1-15, 2012
- [7] C. Zhong Xu and G. Qi Xu. "New sequence spaces and Function spaces on Interval [0,1]". *Journal of function spaces and Applications* , vol.2013, pp.(Article ID 601490,10 pages) 1-10, 2013