Vector space of codons sequence over galois field $\text{GF}(7^3)$

To cite this article: I Aisah et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 567 012019

View the article online for updates and enhancements.
Vector space of codons sequence over galois field GF(7³)

I Aisah¹, N U Sayyidatunnisa¹, B Subartini¹, A Kartiwa¹

¹Department of Mathematics, Faculty of Mathematics and Natural Science
Padjadjaran University, Bandung, Indonesia

Corresponding author: isah.aisah@unpad.ac.id

Abstract. DNA and RNA is genetic material that play an important role in living things metabolism process which called protein synthesis. DNA have four nucleic acid, they are adenine (A), guanine (G), cytosine (C), and thymine (T). Protein synthesis process closely related with standard genetic code. Standard genetic code is a set of rules that defines the order of nucleotide bases in DNA or RNA to determine the order of certain amino acids in protein synthesis. The standard genetic code is a combination of three nitrogen bases or triplet’s bases. This standard genetic code can mathematically represented by algebraic structure. In this paper we will give that representation using the extended set of 7 elements from the set of four DNA’s nucleic acid that are \{D, A, C, O, G, T, P\}. Then we construct new triplet set from \{N\}, called extended triplet set. In the end we analyze the vector space structure of it and find the significant field that correspond with that structure.

1. Introduction
The DNA is which one of nucleic acid that contain genetic information for humans and almost all other organisms. It contains four nucleic acid, that are adenine (A), guanine (G), cytosine (C), and thymine (T). Mostly, RNA have the same structure of DNA. One of the main differences is the RNA’s nucleic acid. They are adenine (A), guanine (G), cytosine (C), and uracil (U). RNA is the result of DNA’s transcription and RNA is translating into amino acid as protein constituent which our body needs. This whole process involve encoding the triplets or codons. This triplets called standard genetic code and this is our main object.

Recently, many research in analyzing the algebraic structure of DNA sequence was reported [1][2-7]. DNA sequence’s analyze over the Galois field of 64 codons [4] and over 125 codons with extended DNA bases into 5 elements [5] have been studied. The question arise when the DNA bases is extended into 7 elements. What kind of algebraic structure it will have and is that extended set can represent the standard genetic code like previous research? Based on that question, in this paper we will extend the DNA bases into 7 elements \{N\} = \{D, A, C, O, G, T, P\} and investigate the algebraic structure of it.

2. DNA, Group and Vector Space
Please In this section we give some definitions of DNA, group theory and vector space.

2.1 DNA (Deoxyribonucleic Acid)
DNA is the hereditary material in humans and almost all other organisms, which have four chemical bases that contain the genetic information. The four bases are adenine (A), guanine (G), cytosine (C),
and thymine (T), which called the nitrogen bases. Beside, RNA (Ribonucleic Acid) have same structure with DNA but with different nitrogen bases. In RNA we can’t find thymine (T) but uracil (U). DNA can replicate and be a source of protein synthesis. Protein synthesis is biological process to form protein particle which essential for living things metabolism. The short process is described by DNA-RNA-Protein which every step is represent by encoding triplet code.

2.2 Group Theory

**Definition 2.2.1** Let $G$ is a set that closed under binary operation $*$. $G$ is called group under that operation if satisfy:

1. Associative, that is for every $a, b, c$ in $G$, $(a * b) * c = a * (b * c)$ hold.
2. Have the identity element, in other words, there is $e$ in $G$ such that for every $a$ in $G$, $e * a = a * e = a$ hold.
3. Every element in $G$, have inverse, that is for every $a$ in $G$ there is $a'$ in $G$ such that $a * a' = a' * a = e$ hold [8].

The example of group is set $N = \{C, U, A, G\}$ with matching of the set $\mathbb{Z}_2 \times \mathbb{Z}_2$, that is $C = (0, 0), U = (0, 1), A = (1, 0)$, and $G = (1, 1)$. It can be show that $(N, +)$ form an Abelian group.

2.3 Vector Space

**Definition 2.3.1** Set $V$ is called vector space over field $F$ if $V$ is Abelian group under addition and for every $a \in F$ and $v \in V$ there is $av$ in $V$ such that this conditions hold for every $a, b$ in $F$ and $u, v$ in $V$.

i. $a(u + v) = au + av$
ii. $(a + b)v = av + bv$
iii. $a(bv) = (ab)v$
iv. $1v = v$ [9].

Based on the example of definition 2.2.1, the set $NNN$ form a vector space over $\mathbb{Z}_2$.

3. Vector Space Construction over Galois Field $GF(7^3)$

In this section, in order to construct the vector space structure of codons sequence, we analyze first the extended triplet set then construct a mapping based on it algebraic structure.

3.1 Extended Triplet Set

The algebra structure of DNA sequence over Galois field with 64 and 125 elements have been studied [8]. In this paper, the DNA nitrogen bases is extend to be $N = \{D, A, C, O, G, T, P\}$. Then we formed $C_{343} = \{X_1, X_2, X_3 | X_i \in N, i = 1, 2, 3\}$ as extended triplet set with cardinality 343 and showed in Figure 1 on the first and third columns.
### Column I is triplet index number, column II the polynomial coefficient, column III the extended triplet

#### 3.2 Mapping The Galois Field with $C_{343}$

To construct a mapping from Galois field $GF(7^3)$ to $C_{343}$, defined $f$ as follows.

| Column I | Column II | Column III |
|----------|-----------|------------|
| D | 000 DDD | 49 001 DAD |
| | 010 DDA | 50 101 DAA |
| | 200 DDC | 51 201 DAC |
| | 300 DDO | 52 301 DAO |
| | 500 DOT | 54 501 DAT |
| | 600 DPP | 55 601 DAP |
| E | 701 ADD | 56 011 AAD |
| | 810 ADA | 57 111 AAA |
| | 910 ADE | 58 211 AAC |
| | 1010 ADO | 59 311 AAO |
| | 1110 AGD | 60 411 AAG |
| | 1210 ADT | 61 511 AAT |
| | 1310 ADD | 62 611 AAP |
| | 1410 CDD | 65 021 CAD |
| | 1510 CDA | 66 121 CAA |
| | 1610 CDC | 67 221 CAC |
| | 1710 CDO | 68 321 CAO |
| | 1810 CCD | 69 421 CAP |
| | 2110 ODD | 70 031 OAD |
| | 2210 ODA | 71 131 OAA |
| | 2310 ODC | 72 231 OAC |
| | 2410 ODO | 73 331 OAO |
| | 2510 ODG | 74 431 OAG |
| | 2610 ODT | 75 531 OAT |
| | 2710 ODP | 76 631 OAP |
| | 2810 ODD | 77 041 OAD |
| | 2910 ODA | 78 141 OAA |
| | 3010 ODC | 79 241 DAC |
| | 3110 ODO | 80 341 DAO |
| | 3210 ODG | 81 441 DAD |
| | 3310 ODT | 82 541 DAT |
| | 3410 OSP | 83 641 GAP |
| | 3510 TDD | 84 051 TAD |
| | 3610 TDA | 85 151 TAA |
| | 3710 TDC | 86 251 TAC |
| | 3810 TDO | 87 351 TAO |
| | 3910 TGD | 88 451 TAG |
| | 4010 TGT | 89 551 TAT |
| | 4110 TDP | 90 651 TAP |
| | 4210 PDD | 91 061 PAD |
| | 4310 PDA | 92 161 PAA |
| | 4410 PDC | 93 261 PAC |
| | 4510 PDO | 94 361 PAC |
| | 4610 PDD | 95 461 PAC |
| | 4710 PDR | 96 561 PAT |
| | 4810 PDP | 97 661 PAP |

![](https://via.placeholder.com/150)

**Figure 1.** Column I is triplet index number, column II the polynomial coefficient, column III the extended triplet.
\[ f : GF(7) \rightarrow \{D, A, C, O, G, T, P\} \]

\[
\begin{align*}
0 & \mapsto D \\
1 & \mapsto A \\
2 & \mapsto C \\
3 & \mapsto O \\
4 & \mapsto G \\
5 & \mapsto T \\
6 & \mapsto P 
\end{align*}
\]

To construct Galois field \( GF(p^n) \) we can use \( \mathbb{Z}_p[x] \) ring and irreducible polynomial \( p(x) \) \( n \)-degrees such that \( GF(p^n) = \mathbb{Z}_p[x] / (p(x)) \). Then we have \( GF(7^3) = \mathbb{Z}_7[x] / (p(x)) = \{a_0 + a_1x + a_2x^2 | a_0, a_1, a_2 \in \mathbb{Z}_7\} \) for an irreducible polynomial \( p(x) \).

After that, we define a mapping named \( \varphi \), from \( GF(7^3) \) to \( C_{343} \)

\[ \varphi : GF(7^3) \rightarrow C_{343} \]

with \( \varphi(a_0 + a_1x + a_2x^2) = (f(a_1) f(a_2) f(a_0)) = (X_1 X_2 X_3) \). Consider that the polynomial coefficient \( a_2 \) with maximal degree \( a_2x^2 \) correspond with bases on second codon/triplet position and polynomial coefficient \( a_1 \) correspond with bases on first codon/triplet position and polynomial coefficient \( a_0 \) correspond with bases on third codon/triplet position. This order is correspond with error frequency on that three bases from a triplet which analogous as degree of variable on \( GF(7^3) \) polynom.

For example, if we have ADA triplet. Based on \( \varphi \) and \( f \) that we have defined first, then

\[ ADA = (f(a_1) f(a_2) f(a_0)) = (f(1) f(0) f(1)) = \varphi(1 + 1 \cdot x + 0 \cdot x^2) \]

So we obtained \( a_0 = 1, a_1 = 1 \) and \( a_2 = 0 \), the polynomial coefficient which correspond to ADA and can be written

\[ 110 \leftrightarrow ADA \ldots (\ast) \]

Therefore we obtained figure 1 for second column using above mapping which is the polynomial coefficient that correspond with every codons on \( C_{343} \).

According to figure 1, we can see that \( \varphi \) mapping is a bijection and so we can obtained a mapping between the triplet’s index number and the triplets as follows.

For ADA, the corresponding index number in figure 1 is 8.

Consider that \( 8 = 0 \cdot 7^0 + 1 \cdot 7^1 + 1 \cdot 7^0 \), so with this result and \( (\ast) \) we can write as follows :

\[ 8 \leftrightarrow 011 \leftrightarrow 110 \leftrightarrow 1 + x \leftrightarrow ADA \]

According to above result, we can defined a bijection

\[ f : S \rightarrow GF(7^3) \]

with \( S \in \{0,1,2,\ldots,342\} \) and \( f[\alpha] = a_0 + a_1x + a_2x^2 \) where \( a_0, a_1, a_2 \) is a form of basis 7’s number from \( \alpha \). So we have \( f[8] = 1 + x \).

### 3.3 Algebra Structure of \( C_{343} \)

Based on definition 2.3.1, in order to analyze the vector space structure of \( C_{343} \) we must investigate the group structure first.

#### 3.3.1 Sum operation and Group Structure on \( C_{343} \) over Addition

To investigate the group structure of \( C_{343} \), first we defined sum operation on \( C_{343} \) as follows. For every \( X_1Y_1Z_1 \) and \( X_2Y_2Z_2 \in C_{343} \), define

\[ X_1Y_1Z_1 + X_2Y_2Z_2 = \varphi[(\varphi^{-1}(X_1Y_1Z_1) + \varphi^{-1}(X_2Y_2Z_2)) \mod 7] \]
For example, we would find the sum between $ADA$ and $OAA$. According to figure 1:

$ADA \leftrightarrow f[8] = 1 + x$

$OAA \leftrightarrow f[71] = 1 + 3x + x^2$

then

$ADA + OAA = \varphi(\varphi^{-1}(ADA) + \varphi^{-1}(OAA) \mod 7)$

$= \varphi((1 + x) + (1 + 3x + x^2) \mod 7)$

$= \varphi(2 + 4x + x^3)$

$= GAC$

So, $ADA + OAA = GAC$. Then, we can showed that $(C_{343}, +)$ is Abelian group.

### 3.3.2 Product Operation and Group Structure on $C_{343} - \{000\}$ over Multiplication

Let $C_{343} - \{000\} = C_{343}^*$, and we defined product operation as follows. For all $X_1Y_1Z_1$ and $X_2Y_2Z_2 \in C_{343}^*$, define

$X_1Y_1Z_1 \cdot X_2Y_2Z_2 = \varphi(\varphi^{-1}(X_1Y_1Z_1)\varphi^{-1}(X_2Y_2Z_2) \mod g(x))$

where $g(x)$ is irreducible polynomial degree 3 with coefficient on $GF(7)$. Based on that definition, we can showed that $(C_{343}, \cdot)$ is Abelian group. For example, we would find the product between $ADA$ and $OAA$ with irreducible polynomial $g(x) = x^3 + 3x^2 + 5x + 4$. Consider that

$ADA \cdot OAA = \varphi(\varphi^{-1}(ADA)\varphi^{-1}(OAA) \mod x^3 + 3x^2 + 5x + 4)$

$= \varphi((1 + x)(1 + 3x + x^2) \mod x^3 + 3x^2 + 5x + 4)$

$= \varphi((1 + 4x + 4x^2 + x^3) \mod x^3 + 3x^2 + 5x + 4)$

$= \varphi(1 + 4x + x^3)$

$= GAC$

So, we have $ADA \cdot OAA = GAA$.

### 3.4 Codon Sequence's Vector Space Construction

After we know that $(C_{343}, +)$ is Abelian group and $(C_{343}, \cdot)$ is isomorphic to $GF(7^3)$, we give scalar product construction in order to investigate the vector space structure of $C_{343}$ as follows.

For all $XYZ \in C_{343}$ and $\alpha \in GF(7^3)$, defined:

$x: GF(7^3) \times C_{343} \rightarrow C_{343}$

$\alpha(XYZ) \leftrightarrow \alpha(XYZ)$

with

$\alpha(XYZ) = \varphi(\alpha \varphi^{-1}(XYZ) \mod 7)$

**Proof**

i) Let $\alpha \in GF(5^3)$ and $X_1Y_1Z_1, X_2Y_2Z_2 \in C_{343}$, we will show that $\alpha(X_1Y_1Z_1) \in C_{343}$

Because $X_1Y_1Z_1 \in C_{343}$ then $\varphi^{-1}(X_1Y_1Z_1) \in GF(7^3)$. Because $GF(7^3)$ is field then $GF(7^3)$ closed under multiplication then $\alpha \varphi^{-1}(X_1Y_1Z_1) \in GF(7^3)$.

So $\alpha(X_1Y_1Z_1) = \varphi(\alpha \varphi^{-1}(X_1Y_1Z_1) \mod 7) \in C_{343}$

ii) Let $\alpha \in GF(7^3)$ and $X_1Y_1Z_1, X_2Y_2Z_2 \in C_{343}$, we will show that $\alpha(X_1Y_1Z_1 + X_2Y_2Z_2) = \alpha(X_1Y_1Z_1) + \alpha(X_2Y_2Z_2)$.

According to scalar multiplication definition, we have

$\alpha(X_1Y_1Z_1 + X_2Y_2Z_2) = \varphi(\alpha \varphi^{-1}(X_1Y_1Z_1 + X_2Y_2Z_2))$

$= \varphi(\alpha \varphi^{-1}(X_1Y_1Z_1) + \varphi^{-1}(X_2Y_2Z_2))$

$= \varphi(\alpha \varphi^{-1}(X_1Y_1Z_1) + \alpha \varphi^{-1}(X_2Y_2Z_2))$

$= \varphi(\alpha \varphi^{-1}(X_1Y_1Z_1)) + \varphi(\alpha \varphi^{-1}(X_2Y_2Z_2))$

$= \alpha(X_1Y_1Z_1) + \alpha(X_2Y_2Z_2)$
iii) Let \( \alpha_1, \alpha_2 \in GF(7^3) \) and \( X_1, Y_1, Z_1, X_2, Y_2, Z_2 \in C_{343} \), we will show that \( (\alpha_1 + \alpha_2)(X_1Y_1Z_1) = \alpha_1(X_1Y_1Z_1) + \alpha_2(X_1Y_1Z_1) \).

According to scalar multiplication definition, we have
\[
(\alpha_1 + \alpha_2)(X_1Y_1Z_1) = \varphi\left((\alpha_1 + \alpha_2)\varphi^{-1}(X_1Y_1Z_1)\right)
= \varphi(\alpha_1\varphi^{-1}(X_1Y_1Z_1) + \alpha_2\varphi^{-1}(X_1Y_1Z_1))
= \varphi(\alpha_1(X_1Y_1Z_1) + \alpha_2(X_1Y_1Z_1))
= \alpha_1(X_1Y_1Z_1) + \alpha_2(X_1Y_1Z_1)
\]

iv) Let \( \alpha_1, \alpha_2 \in GF(7^3) \) and \( X, Y, Z \in C_{343} \), we will show that \( \alpha_1(\alpha_2(XYZ)) = (\alpha_1\alpha_2)(XYZ) \).

According to scalar multiplication definition, we have
\[
\alpha_1(\alpha_2(XYZ)) = \alpha_1\left(\varphi(\alpha_2\varphi^{-1}(XYZ))\right)
= \varphi\left(\alpha_1\varphi^{-1}\left(\varphi(\alpha_2\varphi^{-1}(XYZ))\right)\right)
= \varphi(\alpha_1\alpha_2\varphi^{-1}(XYZ))
= (\alpha_1\alpha_2)(XYZ)
\]

v) Let \( 1 \in GF(7^3) \) and \( X, Y, Z \in C_{343} \), we will show that \( 1(XYZ) = XYZ \).

According to scalar multiplication definition, we have
\[
1(XYZ) = \varphi(1\varphi^{-1}(XYZ))
= \varphi(\varphi^{-1}(XYZ))
= XYZ
\]

So we have that \( C_{343} \) is one dimensional vector space a vector space over Galois field \( GF(7^3) \).

Let \( S = (C_{343})^N = C_{343} \oplus C_{343} \oplus \ldots \oplus C_{343} \) (N-factor) and \( s + s' = (s_1, \ldots, s_N) + (s'_1, \ldots, s'_N) = (s_1 + s'_1, \ldots, s_N + s'_N) \)
\[
s = a(s_1, \ldots, s_N) = (a_1s_1, \ldots, aNs_N)
\]

It can be showed that \( (S,+) \) is Abelian group with \( (DDD, DDD, \ldots, DDD) \) as its identity.

Therefore, \((C_{343})^N\) forms \( N \)-dimensional vector space over \( GF(7^3) \) with \( e_i = (DA, DDD, \ldots, DDD), e_N = (DDD, DDD, \ldots, DDA) \) as their canonic bases. So, we can conclude that for every \( s \in S \) there is unique representation as follows
\[
s = a_1e_1 + \cdots + a_N e_N
\]
or, \( (a_1, \ldots, a_N) \) is representation coordinate from \( S \) over canonical bases \( e_i \).

4. Conclusion

Based on our explanation before, here are some conclusions that we have:

1. The standard genetic code can represented as the elements of extended triplet set (notated by \( C_{343} \)) with \( X, Y, Z \in \{D, A, C, O, G, T, P\} \) as its elements.
2. The \( C_{343} \) set form commutative group structure over addition and the set \( C_{343} - \{OOO\} \) form commutative group under multiplication with operation that have been defined. Then with scalar product that have been defined, \( C_{343} \) form one dimensional vector space structure over Galois field \( GF(7^3) \) and the codons sequence is represented as \( (\alpha_1, \alpha_2, \ldots, a_N) \in (C_{343})^N \).

Acknowledgments

The authors thank to the University of Padjadjaran, due to the funding of this work through to the scheme of Riset Fundamental Unpad (RFU) with contract number 2400/UN6.D/KS/2018.
References
[1] Hornos J E and Hornos Y M 2006 Algebraic Model for The Evolution of The Genetic Code Phys. Rev. Lett. 71 4401
[2] Sanchez R, Perfetti L A, Grau R and Morgado E 2005 A New DNA Sequences Vector Space on a Genetic Code Galois Field MATCH Commun. Math. Comput. Chem. 54 3
[3] Sanchez R and Grau R 2009 An algebraic hypothesis about the primeval genetic code Math. Biosci. 221 60
[4] Sanchez R, Grau R and Morgado E 2006 A Novel Lie Algebra of the Genetic Code over the Galois Field of Four DNA Bases Math. Biosci. 202 156
[5] Sanchez R and Grau R 2006 A Novel Algebraic Structure of The Genetic Code Over The Galois Field of Four DNA Bases Acta Biotheor. 54 27
[6] Sanchez R, Morgado E and Grau R 2005 Gene Algebra From a Genetic Code Algebraic Structure J. Math. Biol 51 431
[7] Sanchez R, Grau R and Morgado E 2006 A Novel DNA Sequence Vector Space over an extended Genetic Code Galois Field MATCH Commun. Math. Comput. Chem. 56 5
[8] Herstein, I. N. 1996. Abstract Algebra. New Jersey: Prentice-Hall
[9] Gallian, Joshep A. 2010. Contemporary Abstract Algebra. Belmont: Richard Stratton