Big Data Process Engineering under Manifold Coordinate Systems

VOLODYMYR RIZNYK
Lviv Polytechnic National University
Lviv, S. Bandery str. 12, 79013 UKRAINE

Abstract—This paper involves techniques for improving the quality indices of big data process engineering with respect to high-performance coded design, transmission speed, and reliability under manifold coordinate systems. The system formed with limited number of basis vectors. The set of modular sums of the vectors including themselves form t-dimensional toroidal coordinate grid over the toroid, and the basis is sub-set of general number of grid coordinate set. These design techniques make it possible to configure high performance information technology for big data coding design and vector signal processing. The underlying mathematical principles relate to the optimal placement of structural elements in spatially or temporally distributed systems by the appropriate algebraic constructions based on cyclic groups in extensions of Galois fields, and development of the scientific basis for optimal solutions for wide classes of technological problems in big data process engineering and computer science.

Keywords—Ideal Vector Ring, optimization, torus coordinate system, vector data coding design.

Received: May 11, 2020. Revised: December 27, 2020. Accepted: January 13, 2021. Published: February 2, 2021.

1. Introduction

The major goals of the modern information technology are the expansion of big data process engineering design and use concept of optimize teaching approach to the practical tasks and assessment methods. Another goal of the systems is creation of unified information space with intelligent components of upper management levels such as large amount of data, high computing amount, and data flow intensity. In this aspect of very profitable is the development of intelligent components for the practical tasks and lectures studies, using novel interpretations of mathematical principles relating to the optimal placement of structural elements in spatially or temporally distributed systems, including the appropriate mathematical models of the systems. The role of the models becomes evident if teacher selects methodology obviously to state the physical essence of a studied problem. The aim of the article involves techniques for improving the quality indices of integrating control functions of technological, business processes, creating unified big vector data information space. The main problem of designing big vector data coding systems is development of an approach to configure two- and multidimensional optimum model of the systems. The multidimensional coding systems, is known to be of very important in information technology, for improving the quality indices of the systems with optimum compressed structure (e.g. two-dimensional torus coordinate system). The paper regards innovative techniques for development of vector data coding design based on the idea of "perfect" spatially or temporally distributed systems, using the appropriate combinatorial configurations as a basis of expanded information field for big data process engineering.

2. Review of Literature

Geometric optimization as known can be performed as follows: optimize the geometry in internal coordinates, in redundant internal coordinates, and in Cartesian coordinates. Each step of the geometry optimization, Gaussian written to the output file the current structure of the system, the energy for this structure, the derivative of the energy with respect to the geometric variables, and a summary of the convergence criteria [1]. The vector data-sets often involve a number of attributes, such as name, type, length, content, and other indexes, which have led to difficulties in large-scale data processing. In recent times, a great number of new concepts, parallel algorithms, processing tools, platform, and applications are suggested and developed to improve the value of big vector data [2-10]. Developing a reversible rapid coordinate transformation big vector data model for the cylindrical projection given in paper [7]. The paper [8] contains fast multidimensional ensemble empirical mode decomposition for the analysis of big spatio-temporal datasets. The geometric computing algorithms are always very complex and time-consuming, which makes big spatial data processing very slow, even impossible [9]. A framework that couples

cloud and high-performance computing for the parallel map projection of vector-based big spatial data regarded in [10]. The projection provides large-scale spatial modeling of big vector data under a common coordinate system. High-dimensional datasets can be very difficult to visualize. While data in two or three dimensions can be plotted to show the inherent structure of the data, equivalent high-dimensional plots are much less intuitive. To aid geometric visualization and processing of a dataset, the processing must be optimized in some way with respect to the underlying criteria. The research in these brunches of science and technologies probes in the various aspects. There were many publications related to combinatorial vector field topology [11-15]. In the paper [11] a theoretical foundation of the combinatorial 2D vector field topology set forth. A discrete Morse theory for general vector fields Forman describes in [12, 13]. Morse-Smile complexes definition for piecewise linear data resulting in a combinatorial algorithm for its extraction considered in work [14]. A slightly different approach by developing a discrete Morse theory for scalar fields defined on cell complexes we see in [15]. This theory applied successfully to scalar fields on triangulated manifolds [16]. Classification of digital n-manifolds based on the notion of complexity and homotopic equivalence presents in paper [17]. The another theoretical approach founded on structural perfection of toroidal and multidimensional manifolds, namely the concept of "Glory to Ukraine Star" combinatorial configurations (GUS-configurations) stated in [18-19].

3. Optimized Torus Coordinate System

Let us regard an n-stage ring sequence of two-dimensional vectors $\{(k_{11}, k_{12}), (k_{21}, k_{22}),..., (k_{i1}, k_{i2}),...,(k_{n1}, k_{n2})\}$ as a basis for configure optimized toroidal coordinate system, where we require a set of all vector-sums combined on the basis, taking modulo m_1 and modulo m_2 accordingly, to be covered torus by coordinate grid of sizes $m_1 \times m_2$.

For, example, the two-dimensional cyclic sequence $\{(0,1), (2,2), (2,1), (1,1)\}$ containing four ($n=4$) 2D vectors allows an enumeration of all reference grid $m_1 \times m_2$ point coordinates, $m_1=3$, $m_2=4$, taking the (mod 3, mod 4) sums of the vectors as follows:

$(0,0) \equiv (1,1)+(0,1)+(2,2),$ $\quad$ $(0,2) \equiv (2,1)+(1,1),$
$(0,3) \equiv (2,1)+(1,1)+(0,1),$ $\quad$ $(1,0) \equiv (0,1)+(2,2),$
$(2,0) \equiv (2,2)+(2,1)+(1,1),$ $\quad$ $(1,2) \equiv (1,1)+(0,1),$
$(1,3) \equiv (2,2)+(2,1),$ $\quad\quad\quad$ $(2,3) \equiv (0,1)+(2,2),$

So long as the vectors (0,1), (2,1), (2,2), (1,1) of the ring sequence themselves are modular vector-sums too, the vector-sums configure the reference grid with sizes 3×4 over surface of torus (Table I).

TABLE I. REFERENCE GRID WITH SIZES 3×4 OVER SURFACE OF TORUS

(2,0)	(2,1)	(2,2)	(2,3)
(1,0)	(1,1)	(1,2)	(1,3)
(0,0)	(0,1)	(0,2)	(0,3)

Hence, the basic set of four ($n=4$) 2D vectors $\{(0,1), (2,1), (2,2), (1,1)\}$ forms optimized toroidal coordinate system with respect to minimizing basis as two-dimension reference grid of fixed sizes $m_1 \times m_2 = 3 \times 4$, which covers surface of usual torus about two ($t=2$) concurrent reference axes, $m_1=3$, and $m_2 = 4$.

It is now customary to represent a toroidal coordinate systems as two-dimension reference grid of sizes $m_1 \times m_2$ covered surface of usual torus about two ($t=2$) concurrent reference axes with (0,0) common point (Fig.1).
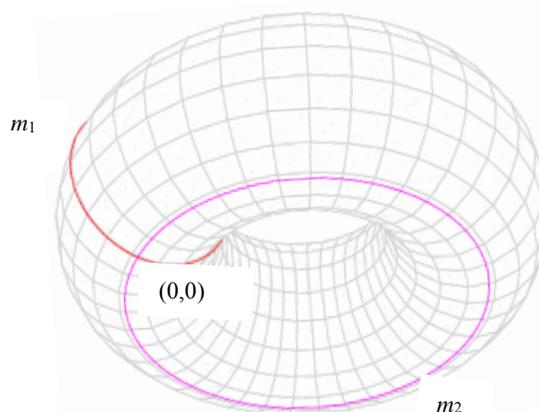
Fig.1 A torus coordinate systems as two-dimension reference grid of sizes $m_1 \times m_2$ covered surface of usual torus bout two ($t=2$) concurrent reference axes with common point (0,0)

Note, the optimization embedded in the underlying GUS-configuration forms binary two-dimensional code in the same basis $\{(0,1), (2,1), (2,2), (1,1)\}$, namely the "Star" code with parameters $n=4$, $m_1=3$, $m_2 = 4$, and code sizes $S = 3 \times 4$.

Example of binary 2D "Star" code with parameters $n=4$, $m_1=3$, $m_2 = 4$ under optimized torus coordinate system with sizes $S = 3 \times 4$ formed in the basis $\{(0,1), (2,2), (2,1), (1,1)\}$ given below (Table II).

TABLE II. BINARY 2D "STAR" CODE UNDER OPTIMIZED TORUS COORDINATE SYSTEM WITH SIZES $S = 3 \times 4$ FORMED IN THE BASIS $\{(0,1), (2,2), (2,1), (1,1)\}$

No	Torus coordinates	Binary 2D vector "Star" code			
		(0,1)	(2,2)	(2,1)	(1,1)
1	(0,0)	1	1	0	1
2	(0,1)	1	0	0	0
3	(0,2)	0	0	1	1
4	(0,3)	1	0	1	1
5	(1,0)	1	1	1	0
6	(1,1)	0	0	0	1

7	(1,2)	1	0	0	1
8	(1,3)	0	1	1	0
9	(2,0)	0	1	1	0
10	(2,1)	0	0	1	0
11	(2,2)	0	1	0	0
12	(2,3)	1	1	0	0

We can see that four-stage (n=4) two-dimensional vectors $\{(0,1), (2,2), (2,1), (1,1)\}$ forms both optimized toroidal coordinate systems with reference grid of sizes $m_1 \times m_2 = 3 \times 4$, and the binary 2D vector code come as the complete systemic package.

4. Manifold Optimized Coordinate System

Let us regard an n-stage ring sequence of t-dimensional vectors $\{K_1, K_2, ..., K_i, ..., K_n\}$, $K_i = (k_{i1}, k_{i2}, ..., k_{it})$, each of them is an t-stage subsequence of the n-stage sequence, where we require a set of all t-modular vector-sums combined on this sequence cover t-dimensional manifold surface as a reference grid. Summation are modulo $m_1, m_2, ..., m_i, ..., m_t$ accordingly about t of spatially disjointed concurrent reference axes with common ground point forming a grid of sizes $m_1 \times m_2 \times ... \times m_i \times ... \times m_t$ over a manifold, $i=1,2,...,t$. We define, the t-dimensional vectors $\{K_1, K_2, ..., K_i, ..., K_n\}$ is a basis for configure optimized t-dimensional manifold coordinate system, generated by GUS –configuration. Graphic presentation of ring ordered vectors $\{K_1, K_2, ..., K_i, ..., K_n\}$ is in Fig.2.
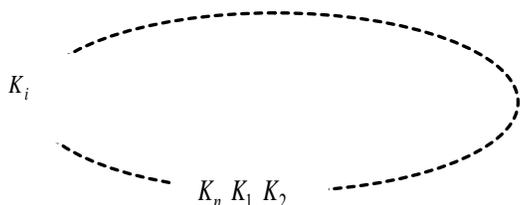
Fig.2 Graphic presentation of ring ordered vectors $\{K_1, K_2, ..., K_i, ..., K_n\}$

Example of binary 3D "Star" code with parameters n=6, m_1=2, m_2=3, m_3=5, and optimized manifold coordinate system with sizes S =2×3×5 formed by GUS –configuration $\{(0,2,3), (1,1,2), (0,2,2), (1,0,3), (1,1,1), (0,1,0)\}$:

1. **(0,0,0)** ≡ (0,2,3)+(1,1,2)+(0,2,2)+(1,0,3)+(0,1,0)

2. **(0,0,1)** ≡ (0,2,2)+(1,0,3)+(0,1,0)+ (1,1,1)

3. **(0,0,2)** ≡ (1,1,2)+(0,2,2)+(1,0,3)

4. **(0,0,3)** ≡ (0,2,3)+(0,1,0)

5. **(0,0,4)** ≡ (0,2,3)+ (0,2,2)+(1,0,3)+(0,1,0)+ (0,1,0)

…………………………………………………………

30. **(1,2,4)** ≡ (0,2,3)+(1,1,2)+(1,0,3)+(1,1,1)+(0,1,0)

As the vector weights of binary 3D "Star" code {(0,2,3), (1,1,2), (0,2,2), (1,0,3), (1,1,1), (0,1,0)} take on binary values, vector **(0,0,0)** is code word 111101, vector **(0,0,1)** is 001110, vector **(0,0,2)** is 011100, vector **(0,0,3)** – 100001, vector **(0,0,4)** – 101111, etc….
Finally, **(1,2,4)** is code word 110111.

We can doing computation these design, using remarkable properties multidimensional GUS-configurations of arbitrary higher dimensionality with parameters n, S, according to selected sizes and spatial configurations of t-dimensional manifold reference grid, essentially, without theoretical bounds of the underlying parameters [18].

To date, there have been no methods to optimize multidimensional manifold coordinate systems. A chart of manifold coordinate system with t-dimensional reference grid $m_1 \times m_2 \times ... \times m_t$ about concurrent axes m_1, m_2, ..., m_t with common ground " +" shows in Fig.3.
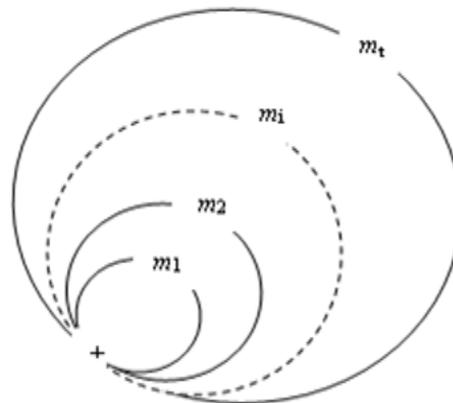
Fig.3 A chart of manifold coordinate system based on t-dimensional coordinate grid $m_1 \times m_2 \times ... \times m_t$ about concurrent axes m_1, m_2, ..., m_t with common ground coordinate " +"

The remarkable property of two- and multidimensional binary "Star" codes are that all modular sums of t-dimensional vectors in the code enumerate a set of distinct vectors of a finite space, which allows encoding and data processing the vectors as sequences of attributes, characters or their packages in a binary vector "star" codes combination simultaneously. This makes it possible to use numerous combinatorial varieties of methods for big vector data coded design under t-dimensional manifold coordinate system, based on suitable GUS-combinatorial configurations to be applicable in needed situations.

The researches [18, 19] probe in various aspects of the subject cover a number of vital points concerning its theoretical and applied problems in the area.

The "star" codes may be separating as self-correcting (redundant vector codes), and the non-redundant (optimum codes) with merits and limitations for each of them [18]. The rate of number growth of 2D and 3D "star" codes with rising their sizes given in Table III.

TABLE III. RATE OF NUMBER GROWTH OF 2D AND 3D "STAR" CODES WITH RISING THEIR SIZES

Code length, n	Number of code variants		Code size variants, S	
	2D codes	3D codes	2D codes	3D codes
2	1	-	1×2	-
3	4	-	2×3	-
4	24	-	3×4	-
5	272	-	4×5, 3×7	-
6	256	128	5×6, 3×10	2×3×5
7	360	180	6×7, 3×14	2×3×6

We can see fast growing varieties of 2D and 3D self-correcting "star" codes under manifold reference systems with rising their length n and code size S. This type of the "stars" belong to the self-correcting codes due to monolithic arranged both symbols "1" and "0" for allowed code words as being ring n-sequences of unified blocks each of the same symbols. This characteristic property offers at once detect or correct code words by its monolithic fashion. The second class of "star" codes which provide vector data coding or processing with the smallest possible information loss under optimized manifold coordinate system of fixed values n, S, R and dimensionality t of reference grid [18]:

$$S = R \prod_1^t m_i < 2^n, \ (m_1, m_2, \ldots, m_t) = 1 \qquad (1)$$

From (1) follows, there are theoretically infinite set of alternate design for configure optimized manifold coordinate systems based on GUS combinatorial configurations, and the more fixed values n, S, R and dimensionality t the more combinatorial varieties creates manifold topological space.

The discovery unknown remarkable properties of the combinatorial varieties with transforming their spatial structures and dimensionality as multidimensional manifolds provides forming optimized vector codes, signal and data processing, vector computing, and systemic researches based on the underlying configurations. There are existing whole classes of GUS-configurations [19].

4.1 Big Vector Data Processing under t-Dimensional Coordinate System

The basic ideas of BVD processing under torus (t=2) or manifold ($t \geq 3$) coordinate system are as following:
- determine sizes of manifold coordinate system and its dimensionality accordingly to entity- attribute-value of a BVD list;
- make indexing entity- attribute- value list under the manifold coordinate system;
- fetch from an information base applicable vector star-code with respect to computer power and processing program;
- make Big Vector Data processing under the manifold coordinate system.

The underlying process engineering provides opportunities to apply them to configure t-dimensional vector data set indexed numerically with t attributes and m_i categories ($i = 1, 2, \ldots, t$) accordingly under the t- dimensional coordinate system. The efficiency is that the less number of encoded signals use for given vector data coding and processing set than at present.

5. Conclusion and Outlook

The optimized manifold coordinate systems based on the "Gloria to Ukraine Stars" (GUS)s combinatorial configurations provide, essentially, a new concept for vector data coding, processing, computing, and systematic researches originated on idea of minimizing vector basis of the coordinate system, using vector combining summation over the basis to cover the coordinate system reference grid. Moreover, the minization has been embedded in the underlying configurations. The remarkable properties and structural perfection of two- and multidimensional GUSs provide an ability to reproduce the maximum number of combinatorial varieties in the system with limited number of vectors. These researches involves techniques for improving the quality indices of engineering devices or systems with respect to big vector data processing and computing speed, data transfer rate, information redundancy and code security communication. Vector data processing under the optimized manifold cordinate systems and "star" codes provide competitive advantages of the vector information technologies with respect to processing speed and storage capacity due to data coding of compound attributes for needed their number and categories simultaneously. The creative qualities of the combinatorial structures allows classify them among intelligent information systems [18, 19]. Study the properties allows a better understanding of the role of geometric structure in the behaviour of artificial and natural objects in different dimensionalities. The underlying skills are useful at high schools and universities for in-depth training of students, which study computer sciences and information technologies, involving contemporary combinatorial and algebraic theory for increasing interest to scientific researches.

References

[1] Geometry Optimization - Basic Considerations, https://www.cup.uni-muenchen.de/ch/compchem/geom/basic.html

[2] Y. Xiaochuang, and Li Guoqing, "Big spatial vector data management: a review," in Big Earth Data, vol.2, No1, 108-129, DOI: 10.1080/20964471.2018.1432115.

[3] M. Chi, A.Plaza, J.A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu," Big data for remote sensing: Challenges and opportunities," IEEE, 104 (11), 2207–2219, DOI:10.1109/Jproc. 2016.598228.

[4] Y.Ma, H. Wu, L. Wang, B.Huang, R. Ranjan, A. Zomaya, and W. Jie, "Remote sensing big data computing: Challenges and opportunities," Future Generation Computer Systems , 51 , 47–60,2015 doi:10.1016/j.future.2014.10.029.

[5] M. K. Pekturk, and M. Unal, "A review on real-time big data analysis in remote sensing applications," 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, May 15–18, 2017.

[6] A. Eldawy, M. F. Mokbel, S. Alharthi, A.Alzaidy, K. Tarek,S. Ghani. SHAHED,"A Map Reduce-based system for querying and visualizing spatio-temporal satellite data, "IEEE International Conference on Data Engineering, Seoul, South Korea, April 13–17, 2015.

[7] S.Ye, T. Yan, Y. Yue, W. Lin, L. Li, X.Yao, D. Zhu, "Developing a reversible rapid coordinate transformation model for the cylindrical projection,"Computers & Geosciences , 89 , 44–56. doi:10.1016/j.cageo.2016.01.007.

[8] Z.Wu, J.Feng, F. Qiao, and Z. Tan - M. (2016),"Fast multidimensional ensemble empirical mode decomposition for the analysis of big spatio-temporal datasets,"Philos Trans A Math Phys Eng Sci, 374(2065), 2015.01.97.

[9] S. Ray, B. Simion, A.D. Brown, and R. Johnson, "A parallel spatial data analysis infrastructure for the cloud,"ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 2013, November 5–8.

[10] W.Tang, and W. Feng,"Parallel map projection of vector-based big spatial data: Coupling cloud computing with graphics processing units," Computers, Environment and Urban Systems , 61 , 187–197. doi:10.1016/j.compenvurbsys.2014.01.001

[11] Jan Reininghaus, and Ingrid Hotz, Combinatorial 2D vector Field Topology Extraction and Simplification. In book: Topological methods in data analysis and visualization. Theory, algorithms, and applications. Based on the 3rd workshop on topological methods in data analysis and visualization, Snowbird, UT, USA, February 23–24, 2009, pp.103-114.

[12] Robin Forman. Combinatorial vector fields and dynamical systems. Mathematische Zeitschrift, 228: 629-681, 1998.

[13] Robin Forman. Morse theory for cell complexes. Advances in Mathematics, 134: 90-145, 1998.

[14] Herbert Edelsbrunner, John Harer, Vijay Natarajan, and Valerio Pascucci. Morse-smale complexes for picewice linear 3-manifolds. In SCG '03: Proceedings of the nineteenth annual symposium on Computational geometry, p.p. 361-370, New York, NY, USA, 2003. ACM.

[15] Robin Forman. Morse theoryial vector fields and dynamical systems. Mathematische Zeitschrift, 228: 629-681, 1998.

[16] Thomas Lewiner, Helio Lopes, and Geovan Tavares. Applications of forman's discrete morse theory to topology visualization and mesh compression. IEEE Transactions on visualization and Computer Graphics, 10 (5): 499-508, 2004.

[17] Alexander V. Evako. Classification of digital n-manifolds. Discrete Applied Mathematics, Volume 181, 30 January 2015, Pages 289-296.

[18] V.Riznyk, "Multi-modular Optimum Coding Systems Based on Remarkable Geometric Properties of Space", Advances in Intelligent Systems and Computing. Springer Int. Publ.- Vol. 512, 2017, pp. 129-148.

[19] V. Riznyk, "Multidimensional Systems Optimization Developed from Perfect Torus Groups", International Journal of Applied Mathematics and Informatics, Vol. 9, 2015.