International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-4, Issue-6, June 2019
ISSN: 2395-3470
www.ijseas.com

A Survey on Algorithms for Consecutive Ones Block Minimization Problem

Zoubir Layouni
Badji Mokhtar University, Algeria
zoubir.layouni@univ-annaba.dz

Abstract:
This paper survey highlight on algorithms for the well-known problem called consecutive ones block minimization. This is a NP-Hard problem which tries to reach to a feasible solution in polynomial time. All approaches seem to be based on two subsidiaries which start by searching how to obtain the consecutive ones propriety, second explore the approximation theory. This can be done by the introduction of notion of approximability preserving reduction. The aim of this survey is to present and discusses the last development of the structural aspects of the polynomial time approximation theory for this problem, it's includes several recent results about completeness in approximability classes and interesting applications.

Keywords:
Consecutive Ones Propriety, Complexity, Approximation algorithms, Consecutive block minimization, CBM.

1. Introduction

Let $A=(a_{ij})$ be a $m\ by\ n$ matrix whose entries a_{ij} are called either 0 or 1. The matrix A may be regarded as the incidence matrix of elements: $a_1, a_2, ..., a_m$ vs. sets $M_1, M_2, ...M_n$; that is, a_{ij} =0 or according as a_1 is not or is a member of M_j. For several applications it is of interest to know whether or not one can order the elements in such a way that each set M_j consists of elements that appear consecutively in the ordering. In the other way it is important to know whether there is a m by m permutation matrix P such that the 1's in each column of PA occurs in consecutive positions.

2. The Consecutive ones problem

A binary matrix has the consecutive ones property (C1P) for columns if there exist a permutation of its rows that places the 1's consecutive in every column. One can symmetrically define the equivalent property for rows. The C1P has been widely studies. The first mention of this property according to Kandall [1]. Fulkerson and Gross [2], presented the first polynomial complexity solution.
This is also mentioned by an archeologist named Petrie in 1899 [1]. Some heuristic methods were proposed for the problem (Robinson)[3]. In 1972 Tucker [4], presented a characterization of the problem based on forbidden configurations for matrices. In the last three decades, the study of the C1P property for a given 0-1 matrix has found applications in graph theory, computer sciences has also a relationsheep with an interval graph, it is the intersection graph of intervals on a line. That is, there in one vertex for each interval and two vertices are adjacent if the corresponding intervals intersect.
The problem of recognizing interval graphs has come up in molecular biology. Seymour Benzer [5], was able to show that the intersection graph of a large number of fragments of genetic material was an interval graph. This was regarded as compelling evidence that genetic information was somhow arranged inside a structure that had a linear topology.
Benzer's work of interval graphs motivated a search for efficient algorithms for recognizing interval graphs, and for constructing a set of intervals to represent the graph when it is one [2]. It is well known that the recognizing of the interval graphs is the recognizing of the Consecutive ones property.
Fulkerson and Gross[2], introduced consecutive one property and gave the first polynomial time sequential algorithm for it, the classic algorithm for solving consecutive one property is a linear

time sequential algorithm of Booth and Lueker[6], based on PQ-trees which is quite complicated.

Booth and Lueker are the first who define this notion and determined the way to obtain this propriety. They invented the PQ-trees as a compact way of storing and manipulating all the permutations on n elements that keep consecutive the elements in certain given sets. They also gave a linear algorithm for building PQ-trees, themselves are useful in solving other problem related to the consecutive ones problem.

This structure, called PQ-tree, has a complicated implementation. Hsu [7] also presented a linear complexity solution for the problem without using PQ-tree and avoiding its complexity of implementation, but the details of his method are difficult to GRASP.

The new trees, called PQR-trees invented by Meidanis and Munuera [8] exist for any instance and carry extra information that can be useful in several applications.

3. Applications Domain

Essentially, the C1P finds applications in any problem in which we are required to linearly arrange a set of objects subjected to restrictions of the form. In molecular biology, one such problem arose when Benzer [5], in 1959, performed a series of experiments aimed at verifying whether a chromosome was a linear arrangement of genes. In combinatorial terms, he had the adjacency relations defining a graph and wanted to know whether it was an interval graph.

Motivated by problems of comparative genomics and paleogenomics, C.Chauve, J.Manuch and M.Patterson [9], introduce the gapped consecutives-ones property.

The approach consists of transforming of M into a matrix that has the C1P, while minimizing the modifications to M; such modifications can involve either in removing rows, or columns, or both , or in flipping some entries from 0 to 1 (1 to 0), the corresponding optimization problems have been proven NP-hard. On the other hand relaxing the condition of consecutivity of the ones of each row by allowing gaps. [10]. Also finding an ordering of the columns that minimizes the number of gaps is NP-complete even if each row of M has at most two ones [Cited in.11].

Many interested applications may be introduced here like the physical Mapping problem that can be solved using fingerprints of the clone, for a given set of small synthetic DNA molecules called probes. This data can be stored in a Boolean probe-clone incidence matrix. In the paper of Jonathan Atkins and Martin Middendorf [12] they gives a simplified model for physical Mapping with probes that tend to occur very rarely along the DNA and show that the problem is NP-complete even for sparse matrices. Moreover, we show that Physical Mapping with chimeric clones is NP-complete even for sparse matrices. Both problemes are modeled as variants of the Consecutives Ones problem which makes the results interesting for other application areas

Goldberg et al [10] considered the problem of Physical Mapping with chimeric clones, i.e. some of the clones may be the result of a concatenation from several clones from different parts of the DNA molecule. This problem can be modeled as the $k-$Consecutive Ones Problem with instance a Boolean matrix M. The question is: Does there exist an ordering of the rows of M such taht column contains at most k blocks of ones.

Throughout this paper, we will say that a matrix is in 2-C1P form or is a 2-C1P permutation if each column has at most two blocks of ones. We will say that a matrix has the 2-C1P property or is 2-C1P if its rows can be permuted to a 2-C1P permutation. The 2-Consecutive Ones Problem is to determine whether or not a 2-C1P permutation of a given matrix exists.

The two versions of physical Mapping Problem presented in [12] are NP-complete even for sparse matrices. Other important applications where found in several study like an application in cancer radiotherapy planning, computational problem, file organization, computer science and many optimization problems.

4. Previous Work

After booth and luecker invention of the PQ-trees, Meidanis and munuera created PQR-trees, a natural generalization of PQ-trees. The difference between them is that PQR exist for every set collection, even when there are no valid permutations. The R nodes encapsulate subsets where the C1P fails. Guilherme.Telles and Joan Meidanis [13] present an almost linear time algorithm to build a PQR-tree for an arbitrary set collection. This algorithm is not a simple extension of Boot and Lucker's algorithm but relies on deeper properties of the tree uncovred by the new theory

To reconstruct a tree reporting the evolutionary history of a group of texa (phylogeny problem) for which a binary matrix of characteristics is given. Gusfield [14] presented an $O(mn)$-time algorithm for this problem with n texa and m characteristics. In this case a linear time algorithm is possible provided the input is given as a list of the « 1 » positions in the matrix. PQ-trees are used here. More precisely, we show that a binary phylogeny exists if and only if the input admits a PQ-tree without Q-nodes. This immediately gives a linear time algorithm for the problem.

The work of Guilherme.Telles and Joan Meidanis [13], show that is for a long time it was unknown whether the extra nodes introduce enough complexity into the structure as to make a linear time algorithm impossible. Telles and Meidanis proposed how to implement the algorithm so as to never have to move uncolored nodes. Another important issue was to merge two Q or R nodes in time $O(1)$. Jonathan Atkins and Bruce Hendrickson [15], present a spectral algorithm for seriation problem and C1P, this problem is a generalization of the well-studied consecutive ones problem, the result of this algorithm is that correctly solves a nontrivial combinatoriel problem. In addition, spectral methods are being successfully applied as heuristics to a variety of sequencing problems. In applications ranging from *DNA* sequencing through archeological dating to sparse matrix reordering, a recurrent problem is the sequencing of elements in such a way that highly correlated pairs of elements are near each other. The result helps explain and justify these applications. In this paper authors present a spectral algorithm for this class of problems. Unlike traditional combinatorial methods, the approach uses an eigenvector of a matrix to order the elements. Although there may be an exponential number of such orderings, they can all be described in a compact data structure known as a $PQ-tree$ [7]. Not all correlation functions allow for a consistent sequencing. If a consistent ordering is possible we will say the problem is well posed. Determining an ordering from a correlation function is what we will call the seriation problem, reflecting its origins in archeology [3,16]. Spectral methods are closely related to the more general method of semidefinite programming, which has been applied successfully to many combinatorial problems and graph coloring for a survey of semidefinite programming with applications to combinatorial optimization. The result is important for several reasons. First, it provides new insight into the well-studied C1P. Second, some important practical problems like envelope reduction for matrices and genomic reconstruction can be thought of as variations on seriation. For example, if biological experiments were error-free, the genomic reconstruction problem would be precisely C1P. Unfortunately, real experimental data always contain errors, and attempts to generalize the consecutive ones concept to data with errors seems to invariably lead to NP-complete problems. According to the results of the study of Rui Wang, and Francis. C.M. Lau [17]. We show that the Hamiltonicity of a regular graph G can be fully characterized by the numbers of blocks of consecutive ones in the binary matrix $A+I$, where A is the adjacency matrix of G, I is the unit matrix, and the blocks can be either linear or circular. Concretely, a k-regular graph G with girth $g(G)\geq 5$ has a Hamiltonian circuit if and only if the matrix $A+I$ can be permuted on rows such that each column has at most (or exactly) $k-1$ circular blocks of consecutive ones; and if the graph G is k-regular except for two $(k-1)$-degree vertices a and b, then there is a Hamiltonian path from a to b if and only if the matrix $A+I$ can be permuted on rows to have at most (or exactly) $k-1$ linear blocks per column. Then we turn to the problem of determining whether a given matrix can have at most k blocks of consecutive ones per column by some row permutation.

For this problem, Booth and Lueker gave a linear algorithm for $k = 1$. Then based on this necessary and suffcient condition and the NP-completeness of hamiltonian circuit for cubic graphs, they prove that for every fixed $k \geq 2$ the $k-$Consecutive blocks problem reamins NP-complete even if restricted to symmetric matrices and matrices having at most 3 blocks of Consecutive ones per row. This result significantly generalizes the related results of [18], and gets its applications in [18], in proving the NP-completeness of all shortest paths interval Routing schemes with compacteness k for every $k \geq 3$.

Now, let us turn to blocks consecutive in *linear* sense, in this sense, the number of blocks of consecutive 1's in column j of a binary matrix $A(n,n)$ is defined to be the number of entries $A(i,j)$ such that $A(i,j)=1$ and $A(i+1,j)=0$ or $i=n$. i.e., we won't take the first row for a natural adjacent row of the last row. The theorem presented in the paper of Rui Wang, and Francis. C.M.Lau[19], is clear : Let $G(V,E)$ be a graph with $g(G) \geq 5$, and every vertex $v \in V$ has a degree of 3 except two special vertices $a,b \in V$ which are degree 2, then G has a Hamiltonian Path from a to b if and only if there is a row permutation for $B(G)=A(G)+I$ resulting in a matrix having at most 2 blocks of consecutive 1's per column.

In there study, they will only talk about blocks consecutive in linear sense. In next section we will point out that the NP-Completeness results here also hold in the sense of circular blocks. In 1976, Garey, Johnson and Tarjan [20] have proved that the well-known Hamiltonian Circuit problem remains NP-complete when restricted to graphs which are cubic, triple-connected, planar, and have no face with fewer than 5 edges.

In conclusion, Hamiltonian Path remains NP-Complete even if restricted to graphs which are each vertex is of degree 3 except the two specified vertices a and b which are of degree 2; 2-connected graph; and planar graph; and $g(G) \geq 5$ and therefore having no face with fewer than 5 edges.

Consecutive block submatrix is important because the matrices are highly restricted and in the question can ask *exactly?* as well as *at most?*. From it we can infer some other NP-Complete variants, such as, deciding whether a matrix can be row permuted such that the total number of blocks in columns is no more than (or exactly) k, or simultaneously permuted on rows and columns such that the total number of blocks in matrix is no more than (or exactly) k (21). It is conjecturable that deciding whether a given matrix can be row permuted such that each column has *exactly* k blocks is also NP-Complete for every $k \geq 2$.

Consecutive Blocks for binary matrix is a classical problem and has applications in Information Retrieve, DNA computing, and Interval Routing [18, 19, 21]. Many variants in asking about the total number of consecutive 1's blocks in all columns have been proved to be NP-Complete [22]. Concerning the maximum number k of blocks over all columns, Flammini et al. [18] showed its NP-Completeness for general k even if restricted to matrices having at most k blocks per row (they didn't state this restriction in the paper); Goldberg et al. [10] proved the problem is NP-Complete for every $k \geq 2$. The results in this paper significantly extend the related results in [18, 20, 21]. In [19]. by taking the advantage of the symmetry of matrices in k-CBS, we prove that for every $k \geq 3$ deciding if a given graph supports an all shortest paths Interval Routing Scheme with compactness k is NP-Complete, and so is for every $k \geq 4$ deciding if a given graph supports an all shortest paths Strict Interval Routing Scheme with compactness k.

Now we need to establish a connection between Hamiltonian circuit and circular blocks and a connection between Hamiltonian path and linear blocks.

In the paper of Rui Wang . Francis C.M. Lau,Yingchao Zhao [23], they prove that for any k-regular graph $G(V,E)$ with $g(G) \geq 5$ has a Hamiltonian circuit if and only if the matrix $A(G)+I$ can be row permuted to have at most $k-1$ circular blocks in each column.

In the work of Hajianghayi and Ganjali [24] they showed the problem of C1P for a matrix like a special variant of consecutive ones submatrix problem (C1SP), in which a positive integer K is given and they want to know if there exists a submatrix B of A consisting of K columns of A with C1P property. In their work, they present an error in the proof of NP-Completeness for this problem in the reference cited in text by Garey and Johnson [20]. The generalized C1SP is NP-Complete. This NP-Completeness result, as well as the ones for many variants of this problem is used to show the NP-Completeness of some other problems. However, in the classic reference text of Garey and Johonson, the proof of NP-Completeness for this problem is referenced to [25], witch proves the NP-completeness of a slightly different problem rather than the original C1S as defined above and in [21]. Their paper presents an error, as well as another proof for this problem. They presents also some special cases of this problem like the COSP is NP-Complete and the COSP is NP-Complete even for matrices with at most two ones in each column and at most four ones in each row.

In the work of S. Istrail et Al [26]. Single, nucleotide polymorphisms (SNPs) are the most frequent form of human genetic variation. They are of fundamental importance for a variety of applications including medical diagnostic and drug design. They also provide the highest-resolution genomic fingerprint for tracking disease genes. This paper is devoted to algorithmic problems related to computational SNPs validation based on genome assembly of diploid organisms. In diploid genomes, there are two copies of each chromosome. A description of the SNPs sequence information from one of the two chromosomes is called SNPs haplotype. The basic problem addressed here is the Haplotyping, i.e., given a set of SNPs prospects inferred from the assembly alignment of a genomic region of a chromosome, find the maximally consistent pair of SNPs haplotypes by removing data \errors" related to DNA sequencing errors, repeats, and paralogous recruitment.

W.L.Hsu [27] developed an off-line linear time test for the consecutive ones property without using *PQ*-trees and the corresponding template matching, which is considerably simpler. A simplification of the consecutive ones test will immediately simplify algorithms (and computer codes) for interval graph and planar graph recognition. The approach is based on a decomposition technique that separates the rows into prime subsets, each of which admits essentially a unique column ordering that realizes the consecutive ones property. The success of this approach is based on finding a good "row ordering" to be tested iteratively.

A consecutive ones test (*COT*) can be used to recognize interval graphs and planar graphs. The algorithm takes $O(m + n\log n)$.

Meidanis, Porto and Telles [28], contributes to a better understanding of the consecutive ones property in several ways, following the steps of Meidanis and Munuera. First they develop a new theory that recasts the property in terms of collections of sets. Then, the class of PQ trees of Booth and Lueker is extended to include PQR trees. A new simpler algorithm for building a PQR tree for a given collection is given. The algorithm runs in polynomial time, but it depends on previous algorithms, and it would be interesting to find a linear time algorithm based on the ideas developed here. The linear time algorithms in the literature are hard to implement, and do not construct R nodes.

In their paper, D.Baatar, H.Hamacher and M.Ehrgott [29], consider the problem of decomposing an integer matrix into a weighted sum of binary matrices that have the strict consecutive ones property. This problem is motivated by an application in cancer radiotherapy planning, namely the sequencing of multileaf collimators to realize a given intensity matrix. In addition we also mention another application in the design of public transportation. We are interested in two versions of the problem, minimizing the sum of the coefficients in the decomposition (decomposition time) and minimizing the number of matrices used in the decomposition (decomposition cardinality). They present polynomial time algorithms for unconstrained and constrained versions of the decomposition time problem and prove that the (unconstrained) decomposition time problem is strongly NP-hard.

Meidanis and Munuera [30]. Present a linear time algorithm for binary phylogeny using PQ-trees. The binary phylogeny problem is to reconstruct a tree reporting the evolutionary history of a group of texa for which a binary of characteristics is given. In this paper we show that a linear algorith is possible provided the input is given as a list of the « 1 » positions in the matrix. The PQ -trees introduced by Booth and Lucker are used here. More precisely, we show that a binary phylogeny exists if and only if the input adlits a PQ -tree without Q nodes this immediately gives a linear time algorithm for the problem. An $O(nm^2)$ algorithm to solve this problem was proposed by Camin and Sokal [31]. Later, a simple property of the matrix allowed an $O(n^2m)$ algorithm. Gusfield [14] presented an $O(nm)$ algorithm. With this viewpoint, we are able to give a new upper bound for this problem: linear $O(m+n+r)$ time was obtained.

The problem of Circular-Ones Arrangements was established by Wen-Lian Hsu and Ross M. McConnell [32]. A 0-1 matrix has the circular-ones property if its columns can be ordered so that, in every row, either the ones or the zeros are consecutive. The classic PQ- trees are used for representing all consecutive-ones orderings of the columns of a matrix that has the consecutive-ones property. They give an analogous structure, called a PC tree, for representing all circular-ones orderings of the columns of a matrix that has the circular-ones property. No such representation has been given previously. In contrast to PQ- trees, PC trees are unrooted. We obtain a much simpler algorithm for computing PQ-trees that those that were previously available, by adding a zero column, x to a matrix, computing the PC tree, and then picking the PC tree up by x to root it.

Booth and Lueker showed that testing for the circular-ones property reduces in linear time to testing for the consecutive-ones property. If we add a zero column to the matrix, the PC tree is the PQ tree without the root, this implies that the update step for the pctree must do exactly what Booth and Lueker's algorithm does, except that it fails to keep track of where the root is.

Atkins Middendortf [12], observed that the two interesting versions of the physical Mapping problem are NP complete even for sparse matrices. In pratical terms, this means, that these problemes are still hard for fingerprints which contain only a constant number of probes and where each probe is contained in at most a constant number of fingerprints. Since our problems are modeled as Consecutive ones problem for sparse matrices, the results have implications also for other areas of applications.

In 1972, Ghosh [33] introduced the concept of the consecutive retrieval file organization. Easwaran [34] developed a graph theoretic approach for analyzing the consecutive retrieval property and established some necessary conditions for its existence. Obviously the problem is to find an organization with minimum redundancy. This problem is known as the Consecutive Retrieval with Minimum Redundancy problem. This problem is NP-complete. Therefore the focus is on polynomial time approximation algorithms. Heuristic algorithms for the same are presented. We suggest the development of efficient polynomial time approximation algorithm with a provable bound on redundancy as a direction for further research.

5.Approximation algorithms for the C1P and CBM-optimization

The structure of approximability classes by the introduction of approximation preserving reductions has been one of the main research programmes in theoretical computer science during the last thirty years. The issue of polynomial approximation theory is the construction of algorithms for **NP**-hard problems that compute .good.feasible solutions for them in polynomial time. The technique of transforming a problem into another in such a way that the solution of the latter entails, somehow, the solution of the former, is a classical mathematical technique that has found wide application in computer science since the seminal works of Cook and Karp who introduced particular kinds of transformations (called reductions) with the aim of studying the computational complexity of combinatorial decision problems. The interesting aspect of a reduction between two problems consists of its twofold application: on one side it allows to transfer positive results from one problem to the other one and, on the other side, it may also be used for deriving negative results.

5.1 Approximation for the C1P.

All of approximation algorithms presented here proves that the C1P problem remains NP-hard for (2, 3)-matrices and (3, 2)-matrices. It's also proved that the C1S problem is polynomial-time 0.8-approximable for (2, 3)-matrices in which no two columns are identical and 0.5-approximable for (2, ∞)-matrices in general. We also show that the C1S problem is polynomial time 0.5-approximable for (3, 2)-matrices. However there exists an ε > 0 such that approximating the C1S problem for (∞, 2)-matrices within a factor of n^{ε} (where n is the number of columns of the input matrix) is NP-hard. The general form is presented below:
Consecutive Ones Submatrix (C1S)
Instance A (0-1)-matrix A
Question : find the largest number of columns of A that form a submatrix with C1P property The decision version of this problem is one of the earliest NP-Complete problems appearing in the Garey and Johnson's book [22]. However, the NP-completeness proof was misreferred to in the book. Recently Hajiaghayi and Ganjali gave a proof [24]. Actually they proved that the C1S problem is NP-complete even for (2, 4)-matrices. On the other hand, the C1S problem can be solved in polynomial time for (2, 2)-matrices. Therefore their work raises the problem of weather the C1S problem remains NP-complete or not for (3, 2)-matrices and (2, 3)-matrices.
Studying the C1S problem for matrices with a small number of 1's in column and/or rows is not only theoretically inetresting, but also pratically important. Actually, this sparsity restriction was proposed by Lander and Istrail with hopes of making the mapping problem tractable.

However, as indicated in the following study, the C1S problem is still hard to solve for sparse matrices.

5.1.1 C1S for (2, 3)-matrices is NP-Complete

To prove the C1S problem NP-complete for (2, 3)-matrices, we first define a special spanning tree problem and prove it NP-complete. Formally it is defined as follows: A tree is caterpillar if each node of degree $c \geq 3$ is adjacent to at least $c-2$ leaf nodes.
We know that the Spanning caterpillar tree in degree-3 graph problem is NP-complet. The proof is given by the reduction from the Hamiltonian Path problem for cubic graphs in which every node has degree 3

5.1.2 C1S for (3, 2)-matrices is NP-Complete.

Now we prove the NP-completenesse of the C1S problem for (3, 2)-matrices by reduction from the following NP-complete problem for cubic graphs. The result is given by a simple reduction from the Induced Disjoint-Path-Union subgraph problem for cubic graphs.

5.1.3 Approximation Algorithms for $(2,\infty)-$ matrices

We shall present a 0.8-approximation algorithm for C1S problem for (2, 3)-matrices: given a (2, 3)-matrix A find a largest submatrix B of A consisting of a subset of A's columns with the C1P property. We also show that a direct generalization of the algorithm turns out to have an approximation ratio 0.5 for $(2,\infty)$-matrices.

5.1.4 A 0.8-approximation for (2, 3)-matrices.

We know in [35] that there is a polynomial time 0.8-approximation algorithm for the C1P problem and (2, 3)-matrix A. To prove that, let the (2, 3)-matrix A has m rows and n columns. Then $G(A)=(V,E)$ has m nodes and n edges. Since each row contains at most three 1's in A , each node has degree at most 3.

To find a 0.8-approximating solution to the C1S problem for A, we first find a spanning tree T in $G(A)$. Then we apply algorithm presented in [35], to find a union of caterpillar subtrees that at least $0.8(m-1)$ edges. The set of edges in the union gives a desired solution.

5.1.5 A 0.5-Approximation for $(2,\infty)$-matrices.

Now we show that a direct generalized of algorithm, has approximation ratio for $(2,\infty)$-matrices and there is a polynomial time 0.5-approximation algorithm for the C1S problem when the input matrix has at most two 1's in each column

5.1.6 A 0.5-Approximation for (3, 2)-matrices.

In this section, we use a partition theorem of Lovâsz in graphe theory to obtain a 0.5-approximation algorithm for the C1S problem for (3, 2)-matrices. The idea can also be generalized to $(\infty,2)$-matrices

Such a problem has application in various domains, it's proved that the C1S problem remain NP-hard for:

(2, 3) matrices with at most two 1's in each column and at most three 1's in each row.

(3, 2) matrices with at most three 1's in each column and at most two 1's in each row.

Therefore, their work raises the problem of weather the C1S problem remains NP-complete or not for :

(3,2)-matrices which have at most two 1's in each row and at most three 1's in each column.

(2,3)-matrices which have at most three 1's in each row and at most two 1's in each column.

The work of Jisong Tan and Louxin Zhang [36], answer the open problem by proving it's NP-completeness for the two special cases just mentioned. Furthermore, they prove that the C1S problem is 0,8-approximatable for (2,3)-matrices, 0,5-approximatable for the matrices in which each column contains at most two 1's and for (3,2)-matrices.

Next step is the generalization for (Δ,2)-Matrices, there is a polynomial time algorithm that output a subset C of at least $n/\lceil(\Delta+1)/2\rceil$ columns such that $A(C)$ has the C1P property given an input $(\Delta,2)$-matix A of n columns. But it is open whether the C1S problem can be approximatable with constant factor for matrices in which there are at most two 1 in each row.

5.2 Approximation for CBM:

The working paper S.Haddadi and Z. Layouni [36], present a polynomial-time heuristic for the Consecutive Block Minimization(CBM) problem such that the solution generated do not differ from optimal by more than 50%.

This problem is a well known in algorithmic research community. It is NP-complete. Morever it was shown that it is NP-complete even when restricted to binary matrices having two 1's per row.

5.2.1 Definition of CBM:

Instance : positive integers m,n,k and binary $m\times n$-matrix A.

Question: Is there a permutation π of $1,\ldots\ldots\ldots,n$ such that the number of the entries A_j^i of A_π verifying $A_j^i=1$ and either $j=1$ or $A_{j+1}^i=0$ is k?

CBM-optimization seeks for a permutation of the columns of A that give the minimum numbers of the entries referred to in the above definition of CBM.

The question is, can we approximate CBM within a *provable* performance guarantee?

We provide a one negative result and we say that the CBM is NP-complete even for binary matrix with at most two ones per row. On the other hand we provide an other positive result with $\alpha=1.5$. the approximation precess begins with a polynomially transformation of the CBM to CIR(circular block minimization), then we polynomially transform CIR to TS (traveling salesman) obeying the triangle inequality. As long as these two transformations are approximation preserving, and the heuristic of Christofides is 1.5-

approximation for TS, so it is the heuristic for CBM-optimization.

The approximation of the CBM is shown by the correspondence between the entries of CBM and the consecutive blocks, so that the question of problem CBM could be recast like that: is there a permutation π of $1,......n$ such that the number of consecutive blocks in A_π is k? In the same manner, a circular block in row i is either a consecutive block or a concatenation of two distinct consecutive blocks, one beginning in $A_0^i = 1$, the other ending in $A_n^i = 1$.there are also a one-to-one correspondence between the entries evoked in the definition of Circular Block Minimization and the circular blocks, so that the question of problem CIR becomes : is there a permutation π of $0,....n$ such that the number of circular blocks in A_π is K?

This algorithm is based on these transformations is given in (37) and clearly this heuristic for CBM-optimization is polynomial-time.

The CBM-optimization is proven to be 1.5-approximable. However, it is possible to find heuristics with a better relative performance. Recall that the symmetric matrix, instance of Traveling salesman, is not ordinary. Of course, it obeys the triangle inequality, but it also has two nice properties: every entry of matrix does not exceed the number m of rows of A, and every cycle has an even length. The question became: can these two properties permit us to do better.

6. Conclusion

We present results obtained with the majority of algorithms mentioned in this paper for consecutive ones problems. We consider, that all results presented in this survey concern approximation algorithms, where approximation ratio is considered. However, one can consider completeness notions for other approximation classes, and we obtain completeness results for a class of problems admitting a better performance ratio. Note that, there has been several research papers attempting to generate solutions to the C1P and C1S. A proposed polynomial time algorithms for CBM have been stated and presented a linear time algorithm that solves several kind of problems.

References

[1] D. G. Kandall., "Incidence matrices, interval graphs and seriation in archaeology". Pacific J. Math., vol 28 , pp. 565-570, 1969

[2] D. R. Fulkerson and O. A. Gross., "Incidence matrices and interval graphs". Pacific J. Mathematics, vol 15, pp. 835-855, 1965

[3] W.S.Robinson., "A method for chronologically ordering archaeological deposits". American Antiquity vol 16, pp. 293-301, 1951

[4] A.C.Tucker., "A structure theorem for the consecutive ones property". Journal combinatorial theory, vol 12(B), pp. 153-162, 1972

[5] S. Benzer., "On the topology of the genetic fine structure", *Proc. Nat. Acad. Sci.* U.S.A. vol 45,pp. 1607-1620., 1959

[6] K. S. Booth and G. S. Lueker., "Test for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms". J. Comput. Systems Sci. Vol 13, pp. 335-379, 1976

[7] W.F.Lu and W.L.Hsu., " A test for consecutive ones on noise data-application to physical mapping and sequence assembly. J.Comput.Bilo, Vol 10,pp. 219-273,1995.

[8] J.Meidanis and E.G Munuera., "A theory for the consecutive ones property". In N.Ziviani, Ricardo baeza-yates and katia crumaraes, editors. proc. Of the 3 south America workshop on string processing Brazil pp. 194-202, 1996

[9] Cedric Chauve, Jan Manuch and Murray Patterson., "On the gapped consecutives ones property". Electronic notes in discret Mathematics. vol 34, pp. 121-125, 2009

[10] P.Goldberg et Al., " Four strikes against physical Mapping of DNA, J.comput.Biol, vol 2,pp. 139-152,
1995

[11] S.Haddadi., "A note on the NP-Hardness of consecutive Block minimization problem", int Trans.Oper.Res, vol 9, pp. 775-777, 2002

[12] J. Atkins and M. Middendorf., "On physical mapping and the consecutive ones property for sparse matrices". Discrete Applied Mathematics, vol 71, pp. 23-40, 1996

[13] G.P. Telles and J. Meidanis., "Building PQR trees in almost-linear time". Technical Report TR-03-26, Instituto de Computacao, UNICAMP, 2003.

[14] D.Gusfield., "Efficient algorithms for inferring evolutionary trees networks", vol 21, pp. 19-28, 1991.

[15] E.Jonathan, E.G.Boman, and B.hendrickson.," A spectral algorithm for seriation and the consecutive ones problem" SIAM J.Comput Vol 28. N°1, pp. 297-310, 1998.

[16] E.M Wilkinson.,"Techniques of data analysis and seriation theory", Cologne, germany, pp. 1-142, 1974

[17] Rui Wang, Francis G.M.Lau. Y.Zhao., "Hamiltonicity of regular graphs and blocks of Consecutives ones in symmetric Matrices" Discrete App. Math., Vol 155, pp. 2312-2320, 2007.

[18] M. Flammini, G. Gambosi and S. Salomone., "Interval routing schems",12th Ann. Symp. On theoretical Aspects of Computer Science. Lecture Notes in Comput. Sci, vol 900.Springer.berlin, pp. 279-290, 1995.

[19] R.Wang.F.C.M.Lau, and Y.Liu., " NP-Complete results for all shortest paths internal Routing". In:Proceeding of the 11th Internat Colloq on Struct Info and Com complexity, june 21-23, 2004.

[20] M. R. Garey and D. S. Johnson, and R.E.tarjan., "The planar Hamiltonian problem" SIAM J Comput. Vol 5, pp. 704-714, 1976.

[21] L.T.Kou., "Polynomial complete consecutive information retrieval problem" SIAM J.Comput, vol 6, pp. 67-75, 1977.

[22] M. R. Garey and D. S. Johnson., "Computers and intractability: A guide to the theory of NP-completeness". San Francisco: W. H. Freeman, 1979.

[23] Rui. Wang, C.M. Francis.Lau,Ying chao., "Hamiltonicity of regular graphs and blocks of consecutive ones in symmetric matrices". ZhaoDiscrete Applied Mathematics, vol 155, pp. 2312 – 2320, 2007.

[24] M. T. Hajiaghayi and Y. Ganjali., "A note on the consecutive ones submatrix problem". Information Processing Letters, vol 83, pp. 163-166, 2002.

[25] K.S.Booth., "PQ tree Algorithm". Doctoral thesis. Dept of electrical engeneering and comuter Science.university of California Berkeley, CA. 1975.

[26] G.Lancia. S.Istrail, and R.schwarta., "SNP's problems, Complexity and algorithms". Algorithms, lecture note in Comp.Sci, vol 2161, pp .182-193,2001.

[27] W.Hsu., "A simple test for consecutive ones property". In T.Ibaraki et Al., editors ISAAC92, vol 650 of LNCS, Nagoya, Japan., pp. 459-468, 1992.

[28] J. Meidanis, O. Porto and G. P. Telles., "On the consecutive ones property". Discrete Appl. Math, vol 88, pp. 325-3 54, 1998.

[29] D.Baatar et Al.," Decompsition of integer Matrices and Multileaf Collimator sequencing". Discrete Applied Mathematics, vol 152, pp. 6-34, 2005.

[30] J. Meidanis and E.G. Munuera., "A simple linear time algorithm for binary phylogeny". *In*

Proceedings of the XV Int. Conf. of the Chilean Computer Science Society, Arica, Chile, pp. 275-283, 1995.

[31] J.Camin and R.Sokal.,"A method for deducing branching sequences in phylogeny evolution". Evolution vol 19, pp. 311-326, 1965.

[32] W.L.Hsu,and R.MMcconnell.," PC trees and Circular ones arrangements". Theoritical Computer science, vol 296, pp. 59-74. 2003.

[33] S. P. Ghosh., "File organization: the consecutive retrieval property". Commun. ACM, vol 15, pp. 802-808, 1972.

[34] K. P. Easwaran., " Placement of records in a file and file allocation in a computer network". Inform. Processing, pp. 304–307, 1974.

[35] Jinsong Tan and Louxin Zhang.," Approximation Algorithms for the Consecutive Ones Submatrix Problem on Sparse Matrices" Algorithmica, vol 48, pp. 287-299, 2007.

[36] Salim Haddadi, Zoubir Layouni.," Consecutive block minimization is 1.5-approximable. Information Processing Letters, vol 108, pp. 132–135, 2008.

[37] Zoubir.Layouni., "Consecutive Block minimization is 1.5-Approximable". Doctoral thesis. Dept. Computer Siences. Badji Mokhtar Annaba.Algeria. 2010.