Approximability and Parameterized Complexity of Consecutive Ones Submatrix Problems

Michael Dom, Jiong Guo⋆, and Rolf Niedermeier

Institut für Informatik, Friedrich-Schiller-Universität Jena,
Ernst-Abbe-Platz 2, D-07743 Jena, Germany.
{dom,guo,niedermr}@minet.uni-jena.de

Abstract. We develop a refinement of a forbidden submatrix characterization of 0/1-matrices fulfilling the Consecutive Ones Property (C1P). This novel characterization finds applications in new polynomial-time approximation algorithms and fixed-parameter tractability results for the problem to find a maximum-size submatrix of a 0/1-matrix such that the submatrix has the C1P. Moreover, we achieve a problem kernelization based on simple data reduction rules and provide several search tree algorithms. Finally, we derive inapproximability results.

1 Introduction

A 0/1-matrix has the *Consecutive Ones Property (C1P)* if there is a *permutation* of its columns, that is, a finite series of column swappings, that places the 1's consecutive in every row[1]. The C1P of matrices has a long history and it plays an important role in applications from computational biology and combinatorial optimization. It is well-known that it can be decided in linear time whether a given 0/1-matrix has the C1P, and, if so, also a corresponding permutation can be found in linear time [1, 6, 10, 11, 14]. Moreover, McConnell [13] recently described a certifying algorithm for the C1P.

Often one will find that a given matrix M does not have the C1P. Hence, it is a natural and practically important problem to find a submatrix of M with maximum size that has the C1P [5, 8, 16]. Unfortunately, even for sparse matrices with few 1-entries this quickly turns into an NP-hard problem [8, 16]. In this paper, we further explore the algorithmic complexity of this problem, providing new polynomial-time approximability and inapproximability and parameterized complexity results. To this end, our main technical result is a structural theorem, dealing with the selection of particularly useful forbidden submatrices. Before we describe our results concerning algorithms and complexity in more detail, we introduce some notation.

We call a matrix that results from deleting some rows and columns from a given matrix M a *submatrix* of M. Then the decision version of the problem we study here is defined as follows:

---

[1] It can be defined symmetrically for columns; we focus on rows here.

Consecutive Ones Submatrix (COS)
Input: An $m \times n$ 0/1-matrix M and a nonnegative integer $n' \leq n$.
Question: Is there an $m \times n'$ submatrix of M that has the C1P?

We study two optimization versions of the decision problem: The minimization version of COS, denoted by Min-COS, asks for a minimum-size set of columns whose removal transforms M into a matrix having the C1P. The maximization version of COS, denoted by Max-COS, asks for the maximum number n' such that there is an $m \times n'$ submatrix of M having the C1P.

Whereas an $m \times n$ matrix is a matrix having m rows and n columns, the term (x, y)-*matrix* will be used to denote a matrix that has at most x 1's in a column and at most y 1's in a row. (This notation was used in [8, 16].) With $x = *$ or $y = *$, we indicate that there is no upper bound on the number of 1's in columns or in rows.

Hajiaghayi [7] observed that in Garey and Johnson's monograph [5] the reference for the NP-hardness proof of COS is not correct. Then, COS has been shown NP-hard for $(2, 4)$-matrices by Hajiaghayi and Ganjali [8]. Tan and Zhang showed that for $(2, 3)$- or $(3, 2)$-matrices this problem remains NP-hard [16]. COS is trivially solvable in $O(m \cdot n)$ time for $(2, 2)$-matrices.

Tan and Zhang [16] provided polynomial-time approximability results for the sparsest NP-hard cases of Max-COS, that is, for $(2, 3)$- and $(3, 2)$-matrices: Restricted to $(3, 2)$-matrices, Max-COS can be approximated within a factor of 0.5. For $(2, *)$-matrices, it is approximable within a factor of 0.5; for $(2, 3)$-matrices, the approximation factor is 0.8.

Let d denote the number of columns we delete from the matrix M to get a submatrix M' having the C1P. Besides briefly indicating the computational hardness (approximation and parameterized) of Max-COS and Min-COS on $(*, 2)$-matrices, we show the following main results.

1. For $(*, \Delta)$-matrices with any constant $\Delta \geq 2$, Min-COS is approximable within a factor of $(\Delta + 2)$, and it is fixed-parameter tractable with respect to d. In particular, this implies a polynomial-time factor-4 approximation algorithm for Min-COS for $(*, 2)$-matrices. Factor-4 seems to be the best factor one can currently hope for, because a factor-δ approximation for Min-COS restricted to $(*, 2)$-matrices will imply a factor-$\delta/2$ approximation for Vertex Cover. It is commonly conjectured that Vertex Cover is not polynomial-time approximable within a factor of $2 - \epsilon$, for any constant $\epsilon > 0$, unless P=NP [12].

2. For $(*, 2)$-matrices, Min-COS admits a data reduction to a problem kernel consisting of $O(d^2)$ columns and rows.

3. For Min-COS on $(2, *)$-matrices, we give a factor-6 polynomial-time approximation algorithm and a fixed-parameter algorithm with a running time of $O(6^d \cdot \min\{m^4 n, mn^6\})$.

Due to the lack of space, several proofs are deferred to the full version of the paper.

1

2 Preliminaries

Parameterized complexity is a two-dimensional framework for studying the computational complexity of problems [3, 4, 15]. One dimension is the input size n (as in classical complexity theory) and the other one the *parameter* d (usually a positive integer). A problem is called *fixed-parameter tractable* (fpt) if it can be solved in $f(d) \cdot n^{O(1)}$ time, where f is a computable function only depending on d. A core tool in the development of fixed-parameter algorithms is polynomial-time preprocessing by *data reduction rules*, often yielding a *reduction to a problem kernel (kernelization)*. Here the goal is, given any problem instance x with parameter d, to transform it into a new instance x' with parameter d' such that the size of x' is bounded from above by some function only depending on d, the instance (x, d) has a solution iff (x', d') has a solution, and $d' \leq d$. A mathematical framework to show *fixed-parameter intractability* was developed by Downey and Fellows [3] who introduced the concept of *parameterized reductions*. A parameterized reduction from a parameterized language L to another parameterized language L' is a function that, given an instance (x, d), computes in time $f(d) \cdot n^{O(1)}$ an instance (x', d') (with d' only depending on d) such that $(x, d) \in L \Leftrightarrow (x', d') \in L'$. The basic complexity class for fixed-parameter intractability is W[1] as there is good reason to believe that W[1]-hard problems are not fixed-parameter tractable [3, 4, 15].

We only consider 0/1-matrices $M = (m_{i,j})$, that is, matrices containing only 0's and 1's. We use the term *line* of a matrix M to denote a row or column of M. A column of M that contains only 0-entries is called a 0-*column*. Two matrices M and M' are called *isomorphic* if M' is a permutation of the rows and columns of M. *Complementing* a line ℓ of a matrix means that all 1-entries of ℓ are replaced by 0's and all 0-entries are replaced by 1's.

Let $M = (m_{i,j})$ be a matrix. Let r_i denote the i-th row and let c_j the j-th column of M, and let M' be the submatrix of M that results from deleting all rows except for $r_{i_1}, \ldots, r_{i_p}$ and all columns except for $c_{j_1}, \ldots, c_{j_q}$ from M. Then M' *contains an entry* $m_{i,j}$ of M, denoted by $m_{i,j} \in M'$, if $i \in \{i_1, \ldots, i_p\}$ and $j \in \{j_1, \ldots, j_q\}$. A row r_i of M *belongs* to M', denoted by $r_i \in M'$, if $i \in \{i_1, \ldots, i_p\}$. Analogously, a column c_j of M belongs to M' if $j \in \{j_1, \ldots, j_q\}$. A matrix M is said to *contain a matrix* M' if M' is isomorphic to a submatrix of M.

3 Hardness Results

As observed by Tan and Zhang [16], MAX-COS with each row containing at most two 1's is equivalent to the MAXIMUM INDUCED DISJOINT PATHS SUBGRAPH problem (Max-IDPS), where, given an undirected graph $G = (V, E)$, we ask for a maximum-size set $W \subseteq V$ of vertices such that the subgraph of G induced by W, denoted by $G[W]$, is a set of vertex-disjoint paths. Since we may assume w.l.o.g. that the input matrix M has no two identical rows and no row of M contains only one 1, a 0/1-matrix M where each row contains exactly two 1's

can be interpreted as a graph $G_M = (V, E)$ with V corresponding to the set of columns and E corresponding to the set of rows. It is easy to verify that M has the C1P iff G_M is a union of vertex-disjoint paths.

We show the hardness of Max-IDPS by giving an approximation-preserving reduction from the NP-hard Independent Set problem to Max-IDPS[2].

Theorem 1. *There exists no polynomial-time factor-$O(|V|^{(1-\epsilon)})$ approximation algorithm, for any $\epsilon > 0$, for* Maximum Induced Disjoint Paths Subgraph *(Max-IDPS) unless NP-complete problems have randomized polynomial-time algorithms.*

With the equivalence between Max-COS, restricted to $(*, 2)$-matrices, and Max-IDPS and results from [2, 3, 9], we get the following corollary.

Corollary 1. *Even in case of $(*, 2)$-matrices,*

1. *there exists no polynomial-time factor-$O(|V|^{(1-\epsilon)})$ approximation algorithm, for any $\epsilon > 0$, for* Max-COS *unless NP-complete problems have randomized polynomial-time algorithms,*
2. Max-COS *is W[1]-hard with respect to the number of the columns in the resulting consecutive ones submatrix, and*
3. *assuming P$\neq$NP,* Min-COS *cannot be approximated within a factor of* 2.7212.

4 $(*, \Delta)$-Matrices

This section will present two positive results for the minimization version of Consecutive Ones Submatrix (Min-COS) restricted to $(*, \Delta)$-matrices, one concerning the approximability of the problem, the other one concerning its fixed-parameter tractability. To this end, we develop a refinement of the forbidden submatrix characterization of the C1P by Tucker [18], which may also be of independent interest.

Definitions and Observations. Every 0/1-matrix $M = (m_{i,j})$ can be interpreted as the adjacency matrix of a bipartite graph G_M: For every line of M there is a vertex in G_M, and for every 1-entry $m_{i,j}$ in M there is an edge in G_M connecting the vertices corresponding to the i-th row and the j-th column of M. In the following definitions, we call G_M the *representing graph* of M; all terms are defined in analogy to the corresponding terms in graph theory.

Let M be a matrix and G_M its representing graph. Two lines ℓ, ℓ' of M are *connected in* M if there is a path in G_M connecting the vertices corresponding to ℓ and ℓ'. A submatrix M' of M is called *connected* if each pair of lines belonging to M' is connected in M'. A maximal connected submatrix of M is called a *component* of M. A *shortest path* between two connected submatrices M_1, M_2 of M is the shortest sequence $\ell_1, \ldots, \ell_p$ of lines such that $\ell_1 \in M_1$ and $\ell_p \in M_2$

---

[2] A different reduction from Independent Set to Max-IDPS was independently achieved by Tan and Zhang [16].
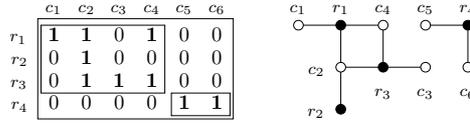
1

$$\begin{array}{c|cccccc}
 & c_1 & c_2 & c_3 & c_4 & c_5 & c_6 \\
\hline
r_1 & 1 & 1 & 0 & 1 & 0 & 0 \\
r_2 & 0 & 1 & 0 & 0 & 0 & 0 \\
r_3 & 0 & 1 & 1 & 1 & 0 & 0 \\
r_4 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}$$

Fig. 1. A matrix with two components and its representing bipartite graph.

$$M_{\mathrm{I}_k},\, k \ge 1 \qquad M_{\mathrm{II}_k},\, k \ge 1 \qquad M_{\mathrm{III}_k},\, k \ge 1$$

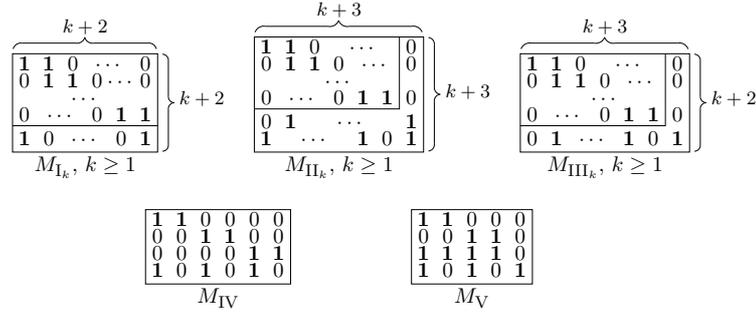$$M_{\mathrm{IV}} \qquad M_{\mathrm{V}}$$

Fig. 2. The forbidden submatrices due to Tucker mentioned in Theorem 2.

and the vertices corresponding to $\ell_1, \ldots, \ell_p$ form a path in G_M. If such a shortest path exists, the value $p-1$ is called the *distance* between M_1 and M_2.

Note that each submatrix M' of M corresponds to an induced subgraph of G_M and that each component of M corresponds to a connected component of G_M. An illustration of the components of a matrix is shown in Fig. 1. If the distance between two lines ℓ_1 and ℓ_p is a positive even number, then ℓ_1 and ℓ_p are either both rows or both columns; if the distance is odd, then exactly one of ℓ_1 and ℓ_p is a row and one is a column.

Observation 1 *Let M be a matrix and let ℓ be a line of M. Then ℓ belongs to exactly one component M' of M and M' contains all 1-entries of ℓ.*

The following corollary is a direct consequence of Observation 1.

Corollary 2. *Let M be a matrix and let $M_1, \ldots, M_i$ be the components of M. If the column sets $F_1, \ldots, F_i$ are optimal solutions for MIN-COS on $M_1, \ldots, M_i$, respectively, then $F_1 \cup \ldots \cup F_i$ is an optimal solution for MIN-COS on M.*

Matrices that have the C1P can be characterized by a set of *forbidden submatrices* as shown by Tucker [18].

Theorem 2 ([18, Theorem 9]). *A matrix M has the C1P iff it contains none of the matrices M_{I_k}, M_{II_k}, M_{III_k} (with $k \ge 1$), M_{IV}, and M_{V} (see Fig. 2).*

The matrix type M_{I} is closely related to the matrix types M_{II} and M_{III}; this fact is expressed, in a more precise form, by the following two lemmas. They are used in the proof of our main structural result presented in Theorem 4.

1

Fig. 3. An illustration of Lemma 1. Matrix M contains an M_{II_k} and a 0-column. Complementing rows r_2, r_{k+1}, and r_{k+2} of M leads to matrix M'. Complementing the rows of M' that have a 1 in column c_{k+3}, namely, r_2, r_{k+1}, and r_{k+3}, transforms M' to matrix M'' which contains an $M_{\mathrm{I}_{k+1}}$ and a 0-column, c_{k+3}.

Lemma 1. *For an integer $k \geq 1$, let M be a $(k+3) \times (k+4)$-matrix containing M_{II_k} and additionally a 0-column, and let M' be the matrix resulting from M by complementing a subset of its rows.*

Then complementing all rows of M' that have a 1 in column c_{k+3} results in a $(k+3) \times (k+4)$-matrix containing $M_{\mathrm{I}_{k+1}}$ and additionally a 0-column.

Proof. Let $R \subseteq \{1, 2, \ldots, k+3\}$ be the set of the indices of the complemented rows, that is, all rows r_i of M with $i \in R$ are complemented.

After complementing the rows $r_i, i \in R$, the column c_{k+3} of M' contains 1's in all rows r_i with $i \in (\{1, \ldots, k+1\} \cap R) \cup (\{k+2, k+3\} \setminus R)$. It is easy to see that complementing these rows of M' results in the described matrix. □

See Fig. 3 for an illustration of Lemma 1. The proof of the following lemma is analogous.

Lemma 2. *For an integer $k \geq 1$, let $M = M_{\mathrm{III}_k}$, and let M' be the matrix resulting from M by complementing a subset of its rows.*

Then complementing all rows of M' that have a 1 in column c_{k+3} results in a $(k+2) \times (k+3)$-matrix containing M_{I_k} and additionally a 0-column.

The Circular Ones Property, which is defined as follows, is closely related to the C1P, but is easier to achieve. It is used as an intermediate concept for dealing with the harder to achieve C1P.

Definition 1. *A matrix has the* Circular Ones Property (Circ1P) *if there exists a permutation of its columns such that in each row of the resulting matrix the 1's appear consecutively or the 0's appear consecutively (or both).*

Intuitively, if a matrix has the Circ1P then there is a column permutation such that the 1's in each row appear consecutively when the matrix is wrapped around a vertical cylinder. We have no theorem similar to Theorem 2 that characterizes matrices having the Circ1P; the following theorem of Tucker is helpful instead.

Theorem 3 ([17, Theorem 1]). *Let M be a matrix. Form the matrix M' from M by complementing all rows with a 1 in the first column of M. Then M has the Circ1P iff M' has the C1P.*

Corollary 3. *Let M be an $m \times n$-matrix and let j be an arbitrary integer with $1 \le j \le n$. Form the matrix M' from M by complementing all rows with a 1 in the j-th column of M. Then M has the Circ1P iff M' has the C1P.*

Algorithms. In order to derive a constant-factor polynomial-time approximation algorithm or a fixed-parameter algorithm for Min-COS on $(*, \Delta)$-matrices, we exploit Theorem 2 by iteratively searching and destroying in the given input matrix every submatrix that is isomorphic to one of the forbidden submatrices given in Theorem 2: In the approximation scenario all columns belonging to a forbidden submatrix are deleted, whereas in the fixed-parameter setting a search tree algorithm branches recursively into several subcases—deleting in each case one of the columns of the forbidden submatrix.

Observe that a $(*, \Delta)$-matrix cannot contain submatrices of types M_{II_k} and M_{III_k} with arbitrarily large sizes. Therefore, for both algorithms, the main difficulty is that every problem instance can contain submatrices of type M_{I_k} of unbounded size—the approximation factor or the number of cases to branch into would therefore not be bounded from above by Δ.

To overcome this difficulty, we use the following approach:

1. We first destroy only those forbidden submatrices that belong to a certain finite subset X of the forbidden submatrices given by Theorem 2 (and whose sizes are upper-bounded, therefore).
2. Then, we solve Min-COS for each component of the resulting matrix. As we will show in Lemma 3, this can be done in polynomial time. According to Corollary 2 these solutions can be combined into a solution for the whole input matrix.

The finite set X of forbidden submatrices is specified as follows.

Theorem 4. *Let $X := \{M_{\mathrm{I}_k} \mid 1 \le k \le \Delta-1\} \cup \{M_{\mathrm{II}_k} \mid 1 \le k \le \Delta-2\} \cup \{M_{\mathrm{III}_k} \mid 1 \le k \le \Delta-1\} \cup \{M_{\mathrm{IV}}, M_{\mathrm{V}}\}$. If a $(*, \Delta)$-matrix M contains none of the matrices in X as a submatrix, then each component of M has the Circ1P.*

Proof. Let M be a $(*, \Delta)$-matrix containing at most Δ ones per row, and let X be the set of submatrices mentioned in Theorem 4. We use Corollary 3 to show that if a component of a matrix M does not have the Circ1P, then this component contains a submatrix in X.

Let A be a component of M that does not have the Circ1P. Then, by Corollary 3, there must be a column c of A such that the matrix A' resulting from A by complementing those rows that have a 1 in column c does not have the C1P and, therefore, contains one of the submatrices given in Theorem 2.

In the following, we will make a case distinction based on which of the forbidden submatrices is contained in A' and which rows of A have been complemented, and show that in each case the matrix A contains a forbidden submatrix from X.

We denote the forbidden submatrix contained in A' with B' and the submatrix of A that corresponds to B' with B. Note that the matrix A' must contain a 0-column due to the fact that all 1's in column c have been complemented.

Fig. 4. Illustration for Case 1 in the proof of Theorem 4. Complementing the second row of an M_{IV} generates an M_V. (The rows and columns of the M_V are labeled with numbers according to the ordering of the rows and columns of the M_V in Fig. 2.)

Fig. 5. Illustration for Case 2 in the proof of Theorem 4. Suppose that only the third row of B is complemented. Then B together with the complementing column forms an M_{IV}.

Because no forbidden submatrix given by Theorem 2 contains a 0-column, column c cannot belong to B' and, hence, not to B. We will call this column the *complementing column* of A.

When referencing to row or column indices of B', we will always assume that the rows and columns of B' are ordered as shown in Fig. 2.

Case 1: The submatrix B' is isomorphic to M_{IV}.

If no row of B has been complemented, then $B = B'$, and A also contains a submatrix M_{IV}, a contradiction to the fact that M contains no submatrices from X.

If exactly one of the first three rows of B has been complemented, then B contains one 0-column, and B with the 0-column deleted forms an M_V, independent of whether the fourth row of B also has been complemented (see Fig. 4). Again, we have a contradiction.

If two or three of the first three rows of B have been complemented, then A contains an M_{I_1} as a submatrix: Assume w.l.o.g. that the first two rows have been complemented. If the fourth row has also been complemented, there is an M_{I_1} consisting of the rows r_1, r_2, r_4 and the columns c_2, c_4, c_5 of B. Otherwise, there is an M_{I_1} consisting of the rows r_1, r_2, r_4 and the columns c_1, c_3, c_6 of B. This is again a contradiction.

Case 2: The submatrix B' is isomorphic to M_V.

Analogously to Case 1 we can make a case distinction on which rows of A have been complemented, and in every subcase we can find a forbidden submatrix from X in A. In some of the subcases the forbidden submatrix can only be found in A if in addition to B also the complementing column of A is considered. We will present only one example representing all subcases of Case 2. If e.g. only the third row of B has been complemented, then the complementing column of A

1

contains a 0 in all rows that belong to B except for the third. Then B forms an M_{IV} together with the complementing column of A (see Fig. 5).

We omit the details for the other cases. Herein, the case that B' is isomorphic to M_{I_k}, $k \geq \Delta$, is the most complicated one; one has to distinguish two subcases corresponding to the parity of the distance between the complementing column and B. Lemma 1 and Lemma 2 are decisive for the cases that B' is isomorphic to M_{II_k}, $k \geq 1$, and that B' is isomorphic to M_{III_k}, $k \geq 1$, respectively. □

If all submatrices mentioned in Theorem 4 are destroyed, then every component of the resulting matrix has the Circ1P by Theorem 4. The next lemma shows that Min-COS on these components then is polynomial-time solvable.

Lemma 3. Min-COS *can be solved in* $O(n^{\Delta} \cdot m)$ *time, when restricted to* $(*, \Delta)$-*matrices that have the Circ1P. Herein, n denotes the number of columns and m the number of rows.*

We can now state the algorithmic results of this section.

Theorem 5. Min-COS *on* $(*, \Delta)$-*matrices for constant* Δ *can be approximated in polynomial time within a factor of* $\Delta + 2$ *if* $\Delta = 2$ *or* $\Delta \geq 4$, *and it can be approximated in polynomial time within a factor of* 6 *if* $\Delta = 3$.

Proof. Our approximation algorithm searches in every step for a forbidden submatrix of X given by Theorem 4 and then deletes all columns belonging to this submatrix. An optimal solution for the resulting matrix can be found in polynomial time: Due to Corollary 2 an optimal solution for a matrix can be composed of the optimal solutions for its components, and due to Theorem 4 after the deletion of all submatrices from X all components have the Circ1P, which means that an optimal solution for every component can be found in polynomial time as shown in Lemma 3.

Because an optimal solution has to delete at least one column of every forbidden submatrix of X, the approximation factor is the maximum number of columns of a forbidden submatrix from X.

The running time of the algorithm is dominated by searching for submatrix from X with the largest number of rows and columns. Such a submatrix M' with R rows and C columns can be found in $O(\min\{mn^C RC, m^R nRC\})$ time.

If $\Delta \neq 3$, every submatrix of X has at most $\Delta + 1$ rows and $\Delta + 2$ columns. If $\Delta = 3$, the matrix M_{IV} with four rows and six columns is the biggest matrix of X. This yields the claimed approximation factors and running times. □

Theorem 6. *Restricted to* $(*, \Delta)$-*matrices,* Min-COS *with parameter* $d = $ *"number of deleted columns" can be solved in* $O((\Delta + 2)^d \cdot (\min\{m^{\Delta+1}n, mn^{\Delta+2}\} + mn^{\Delta}))$ *time if* $\Delta = 2$ *or* $\Delta \geq 4$ *and in* $O(6^d \cdot (\min\{m^4n, mn^6\} + mn^3))$ *time if* $\Delta = 3$.

Proof. We use a search tree approach, which searches in every step for a forbidden submatrix of X given by Theorem 4 and then branches on which column belonging to this submatrix has to be deleted. The solutions for the resulting

matrices without submatrices from X can be found, without branching, in polynomial time due to Lemma 3. The number of branches depends on the maximum number of columns of a forbidden submatrix from X; the running time of the algorithm can therefore be determined analogously to Theorem 5. □

5 $(*, 2)$- and $(2, *)$-Matrices

$(*, 2)$-*Matrices.* By Theorem 6 we know that MIN-COS, restricted to $(*, 2)$-matrices, can be solved in $O(4^d \cdot m^3 n)$ time with d denoting the number of deleted columns. Here, we show that MIN-COS, restricted to $(*, 2)$-matrices, admits a quadratic-size problem kernel. Using the equivalence between MIN-COS restricted to $(*, 2)$-matrices and the minimization dual of MAXIMUM INDUCED DISJOINT PATHS SUBGRAPH (Max-IDPS) (see Sect. 3), we achieve this by showing a problem kernel for the dual problem of Max-IDPS with the parameter d denoting the number of allowed vertex deletions.

We use Min-IDPS to denote the dual problem of Max-IDPS and formulate Min-IDPS as a decision problem: The input consists of an undirected graph $G = (V, E)$ and an integer $d \geq 0$, and the problem is to decide whether there is a vertex subset $V' \subseteq V$ with $|V'| \leq d$ whose removal transforms G into a union of vertex-disjoint paths. W.l.o.g., we assume that G is a connected graph.

Given an instance $(G = (V, E), d)$ of Min-IDPS, we perform the following polynomial-time data reduction:

Rule 1: If a degree-two vertex v has two degree-at-most-two neighbors u, w with $\{u, w\} \notin E$, then remove v from G and connect u, w by an edge.

Rule 2: If a vertex v has more than $d + 2$ neighbors, then remove v from G, add v to V', and decrease d by one.

A graph to which none of the two rules applies is called *reduced*.

Lemma 4. *The data reduction rules are correct and a graph $G = (V, E)$ can be reduced in $O(|V| + |E|)$ time.*

Theorem 7. *Min-IDPS with parameter d denoting the allowed vertex deletions admits a problem kernel with $O(d^2)$ vertices and $O(d^2)$ edges.*

Proof. Suppose that a given Min-IDPS instance (G, d) is reduced w.r.t. Rules 1 and 2 and has a solution, i.e., by deleting a vertex subset V' with $|V'| \leq d$ the resulting graph $H = (V_H, E_H)$ is a union of vertex-disjoint paths. Then H has only degree-one and degree-two vertices, denoted by V_H^1 and V_H^2, respectively. Note that $V_H = V_H^1 \cup V_H^2 = V \setminus V'$.

On the one hand, since Rule 2 has removed all vertices of degree greater than $d + 2$ from G, the vertices in V' are adjacent to at most $d^2 + 2d$ V_H-vertices in G. On the other hand, consider a V_H^1-vertex v. If v is not a degree-one vertex in G, then v is adjacent to at least one vertex from V'; otherwise, due to Rule 1, v's neighbor or the neighbor of v's neighbor is adjacent to at least one vertex from V'. Moreover, due to Rule 1 and the fact that H is a union of vertex-disjoint

paths, at least one of three consecutive degree-two vertices on a path from H is adjacent in G to at least one vertex from V'. Hence, at least $|V_H|/5$ vertices in V_H are adjacent in G to V'-vertices. Thus, the number of V_H-vertices can be upper-bounded by $5 \cdot (d^2 + 2d)$.

Since H is a union of vertex-disjoint paths, there can be at most $|V_H| - 1 = 5d^2 + 10d - 1$ edges in H. As shown above, each vertex from V' can have at most $d + 2$ incident edges. $\qquad\square$

$(2, *)$-*Matrices.* Here, we consider matrices M that have at most two 1's in each column but an unbounded number of 1's in each row. From Theorem 2, if M does not have the C1P, then M can only contain an M_{IV} or an M_{I_k} in Fig. 2.

The following lemma can be shown in a similar way as Theorem 4.

Lemma 5. *Let M be a $(2, *)$-matrix without identical columns. If M does not contain M_{IV} and M_{I_1} and does not have the C1P, then the matrices of type M_{I_k} that are contained in M are pairwise disjoint, that is, they have no common row or column.*

Based on Lemma 5, we can easily derive a search tree algorithm for MIN-COS restricted to $(2, *)$-matrices:
1. Merge identical columns of the given matrix M into one column and assign to this column a weight equal to the number of columns identical to it;
2. Branch into at most six subcases, each corresponding to deleting a column from an M_{IV}- or M_{I_1}-submatrix found in M. By deleting a column, the parameter d is decreased by the weight of the column.
3. Finally, if there is no M_{IV} and M_{I_1} contained in M, then, by Lemma 5, the remaining M_{I}-submatrices contained in M are pairwise disjoint. Then, MIN-COS is solvable in polynomial time on such a matrix: Delete a column with minimum weight from each remaining matrix of type M_{I_k}.

Clearly, the search tree size is $O(6^d)$, and a matrix of type M_{IV} can be found in $O(\min\{m^4 n, mn^6\})$ steps, which gives the following theorem.

Theorem 8. MIN-COS, *restricted to $(2, *)$-matrices with m rows and n columns, can be solved in $O(6^d \cdot \min\{m^4 n, mn^6\})$ time where d denotes the number of allowed column deletions.*

By removing all columns of every M_{IV}- and M_{I_1}-submatrix found in M, one can show the following.

Corollary 4. MIN-COS, *restricted to $(2, *)$-matrices, can be approximated in polynomial time within a factor of 6.*

6   Future Work

Our results mainly focus on MIN-COS with no restriction on the number of 1's in the columns; similar studies should be undertaken for the case that we have no restriction for the rows. Moreover, it shall be investigated whether Δ

can be eliminated from the exponents in the running times of the algorithms for Min-COS on $(*, \Delta)$-matrices—in the parameterized case the question is if Min-COS on $(*, \Delta)$-matrices is fixed-parameter tractable with d and Δ as a combined parameter, or if Min-COS on matrices without restrictions is fixed-parameter tractable with d as parameter. Finally, we focused only on deleting columns to achieve a maximum-size submatrix with C1P. It remains to consider the NP-hard symmetrical case [5] with respect to deleting rows; our structural characterization theorem should be helpful here as well.

References

1. K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer and System Sciences*, 13:335–379, 1976.
2. I. Dinur and S. Safra. On the hardness of approximating Minimum Vertex Cover. *Annals of Mathematics*, 162(1):439–485, 2005.
3. R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Springer, 1999.
4. J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer, 2006.
5. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.
6. M. Habib, R. McConnell, C. Paul, and L. Viennot. Lex-BFS and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing. *Theoretical Computer Science*, 234(1–2):59–84, 2000.
7. M. T. Hajiaghayi. Consecutive ones property, 2000. Manuscript, University Waterloo, Canada.
8. M. T. Hajiaghayi and Y. Ganjali. A note on the consecutive ones submatrix problem. *Information Processing Letters*, 83(3):163–166, 2002.
9. J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.
10. W.-L. Hsu. A simple test for the consecutive ones property. *Journal of Algorithms*, 43:1–16, 2002.
11. W.-L. Hsu and R. M. McConnell. PC trees and circular-ones arrangements. *Theoretical Computer Science*, 296(1):99–116, 2003.
12. S. Khot and O. Regev. Vertex Cover might be hard to approximate to within $2 - \epsilon$. In *Proc. 18th IEEE Annual Conference on Computational Complexity*, pages 379–386. IEEE, 2003.
13. R. M. McConnell. A certifying algorithm for the consecutive-ones property. In *Proc. 15th ACM-SIAM SODA*, pages 768–777. SIAM, 2004.
14. J. Meidanis, O. Porto, and G. P. Telles. On the consecutive ones property. *Discrete Applied Mathmatics*, 88:325–354, 1998.
15. R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.
16. J. Tan and L. Zhang. The consecutive ones submatrix problem for sparse matrices. To appear in *Algorithmica*. Preliminary version titled "Approximation algorithms for the consecutive ones submatrix problem on sparse matrices" appeared in *Proc. 15th ISAAC*, Springer, volume 3341 of *LNCS*, pages 836–846, 2004.
17. A. C. Tucker. Matrix characterizations of circular-arc graphs. *Pacific Journal of Mathematics*, 2(39):535–545, 1971.
18. A. C. Tucker. A structure theorem for the consecutive 1's property. *Journal of Combinatorial Theory (B)*, 12:153–162, 1972.