Georg-August-Universität
Göttingen

Set covering problems with consecutive ones property

Anita Schöbel

Preprint Nr. 2005-03

Preprint-Serie des
Instituts für Numerische und Angewandte Mathematik
Lotzestr. 16-18
D - 37083 Göttingen

Set covering problems with consecutive ones property

Anita Schöbel
Georg-August Universität Göttingen
e-mail: schoebel@math.uni-goettingen.de

July 8, 2005

Abstract

The paper deals with variants of set covering problems whose coefficient matrices have the consecutive ones property, (i.e. the ones in each row appear consecutively). We propose a new approach transforming such problems to shortest path problems in acyclic digraphs. The approach can also be used to solve the bicriterial variant of the problem in which we aim to minimize the costs and to maximize the weight of the cover simultaneously. Extensions to matrices with more blocks of consecutive ones per row are also given.

1 Introduction

Let A be a zero-one matrix with costs for each of its columns. A row m of A is *covered* by column j of A if the corresponding matrix entry a_{ij} is equal to one. The *set covering* problem then asks to choose a *cover*, i.e. a set of columns covering all rows of A, with minimal costs. In the *unweighted set covering* problem, all costs are one, i.e. the goal is to find a minimum cardinality set of columns covering all rows of A.

Set covering problems belong to the best studied combinatorial optimization problems; many exact and approximate solution algorithms have been published. We refer to the annotated bibliography [CNS97] or the survey [CFT00] on state-of-the art algorithms.

Among other reasons the interest in set covering problems is due to their large potential of modeling real-world problems such as scheduling, facility location, or production optimization problems. Unfortunately, the majority of set covering problems arising in practice are very large. For example, in crew scheduling one easily obtains set covering problems with thousands of variables and constraints as it is reported, e.g., in [CFT$^+$97] for railway and in [MS00] for airline crew scheduling problems. Since the set covering problem is NP-hard ([GJ79]) (even for no more than two non-zero entries in each row) and also difficult from the point of view of theoretical approximation ([LY94]), such large problem instances are hard to solve. This motivates the development of efficient heuristic procedures for solving large-scale problems, see e.g. the Lagrangian-based heuristic of [CFT99].

In this paper, we follow another line, namely we discuss new approaches for a special class of set covering problems. The type of problem we refer to are set covering problems with a coefficient matrix satisfying the *consecutive ones property*, i.e., the ones in each row appear consecutively. This property becomes, in fact, very important, since set covering problems appearing in real-world applications are often close to having the consecutive ones property in their coefficient matrices. An example is the problem of locating stations along an existing track system, such to cover a given set of demand points. For this problem it has been shown in [SHLW02, Sch03a] that the resulting coefficient matrix has the consecutive ones property, if only a linear part of the track system is considered, while for the complete data set of Germany, it *almost* has the consecutive ones property, see [RS04]. Looking closer at other types of set covering problems, e.g., in crew scheduling or line planning, it also turns out that the columns in real-world data sets often can be ordered to obtain a coefficient matrix that is close to a consecutive ones matrix. Note that matrices with consecutive ones property also play an important role in radiation theraphy planning, see [BEHW05] and references therein.
Given an arbitrary $m \times n$ matrix, a permutation of the columns making the ones appear consecutively can be found in $\mathrm{O}(mn)$ time using the approach of [BL76, MPT98].

Our approach for solving set covering problems with consecutive ones property can not only be used in the classical variant of the set covering problem, but it can also be applied to find efficient solutions in the *bicriteria set covering problem*. In this problem, we need not cover all rows, but our goal is to maximize the number of covered rows and to minimize the costs of the cover simultaneously. We also may allow a weight w_i for each row i and maximize the total weight covered instead of just counting the number of covered rows. A solution (i.e. a set of columns) is called *non-dominated*, if for all cheaper sets of columns the weight of the covered rows decreases.

The remainder of the paper is structured as follows. In Section 2 we introduce the notation needed and some basic properties. The new approach for solving set covering problems with consecutive ones property is developed in Section 3, including a few results for the unweighted set covering problem. The solution approach for the bicriteria variant of the set covering problem is presented in Section 4. Section 5 generalizes the results obtained to set covering problems with more than one block of ones per row. Conclusions and extensions follow in Section 6.

2 Basic properties

To formulate set covering problems as integer programs, let $\mathcal{M} = \{1, 2, \ldots, m\}$ be the set of rows and $\mathcal{N} = \{1, \ldots, n\}$ be the set of columns of the given matrix A. A *cover* of A is a set of columns $\bar{\mathcal{N}}$ such that for each row i, $i \in \mathcal{M}$ there exists a column $j \in \bar{\mathcal{N}}$ with $a_{ij} = 1$.
We use binary variables

$$x_j = \begin{cases} 1 & \text{if } j \in \bar{\mathcal{N}} \\ 0 & \text{otherwise} \end{cases},$$

to state the well known formulation

(SCP)

$$\begin{array}{rrcl} \min & cx & & \\ \text{s.t.} & Ax & \geq & \underline{1}_m \\ & x & \in & \{0,1\}^n, \end{array} \tag{1}$$

where $\underline{1}_m \in \mathbb{R}^m$ is a vector with a 1 in each component, and $c = (c_1, c_2, \ldots, c_n)$ contains the costs c_j of the columns j. We further need the notation

$$\begin{array}{rcl} \mathcal{M}_j & := & \{i \in \mathcal{M} : a_{ij} = 1\}, \text{ and} \\ \mathcal{N}_i & := & \{j \in \mathcal{N} : a_{ij} = 1\}. \end{array}$$

The following reduction rules proposed in [TR73] can be found in many textbooks (see, e.g. [NW88]). Nevertheless we collect them here since they will be important later.

Lemma 1

1. *If $\mathcal{N}_{i_1} \subseteq \mathcal{N}_{i_2}$ then an optimal solution of problem (SCP) can be found by considering the reduced problem without row i_2.*

2. *If $\mathcal{M}_{j_1} \subseteq \mathcal{M}_{j_2}$ and $c_{j_1} \geq c_{j_2}$ then there exists an optimal solution of problem (SCP) with $x_{j_1} = 0$, i.e. it is sufficient to consider the reduced problem without column j_1.*

The set covering problem is NP-hard (even in the unweighted case and also if only two nonzero elements exist in each row), see [GJ79]. In this paper, however, we deal with a polynomially solvable variant of the set covering problem, namely if the matrix A has the *consecutive ones property*. An $m \times n$ matrix has this property, if in each row of A the ones appear consecutively, i.e.

$$a_{ik} = 1, a_{il} = 1 \text{ and } k \leq l \Longrightarrow a_{ij} = 1 \text{ for all } k \leq j \leq l.$$

It is well known that a matrix with consecutive ones property is totally unimodular (see, e.g., III.1 of [NW88]), and hence the linear programming relaxation of the set covering problem yields an integer solution, such that (SCP) is polynomially solvable by linear programming in this case.
In a more efficient approach we can use the fact that the transposed of a matrix with consecutive ones property is an interval matrix, and hence a network matrix. Since there exists an optimal solution satisfying $x_j \leq 1$ for all $j \in \{1, \ldots, N\}$, we omit these constraints and obtain the following packing program as the dual of the set covering problem.

(Dual-SCP)

$$\begin{array}{rrcl} \max & \underline{1}\eta & & \\ s.t. & A^T\eta & \leq & c \\ & \eta & \geq & 0. \end{array} \tag{2}$$

Note that, since A is totally unimodular, the optimal solution values of (SCP) and its dual formulation (Dual-SCP) are equal in this case.

Following the approach of Example 3.2. in Chapter III.1.3 of [NW88], this dual formulation can be reformulated as a network flow problem in an acyclic network. This network is constructed by interpreting the rows of A^T as arcs and the columns as paths. One starts by defining the set of nodes as

$$V_{\text{flow}} = \{0, 1, \ldots, n\},$$

and by constructing an arc $(j-1, j) \in E_{\text{flow}}$ for each row j of A^T. Furthermore, each column i of A^T can be interpreted as a path which is composed of edges $(j-1, j)$ with $a_{ij} = 1$. For such a path we add one additional arc to E_{flow}, namely the one replacing the respective path. These arcs correspond to the dual variables η_d. Since all entries in A are positive, the network is acyclic. Defining $d_0 = c_1$, $d_j = c_{j+1} - c_j$ for $s = 1, \ldots, n-1$, and $d_n = -c_n$ as the demand of the respective node in V_{flow}, and setting 0 as the cost of arc $(i-1, i)$, and 1 as the costs for all other arcs, one finally obtains an equivalent *min-cost flow problem* in an acyclic digraph with $n+1$ nodes, see [NW88] for more details.

In this paper, however, we develop an alternative approach for solving set covering problems with consecutive ones property. This approach transforms the set covering problem into a *shortest path problem* in a directed acyclic network with $n+2$ nodes.

3 A new approach for set covering problems with consecutive ones property

Let A be a matrix with consecutive ones property. Denoting

$$\begin{aligned} s_i &:= \min\{j \in \mathcal{N} : a_{ij} = 1\} \\ e_i &:= \max\{j \in \mathcal{N} : a_{ij} = 1\} \end{aligned}$$

we may rewrite $\mathcal{N}_i = \{j \in \mathcal{N} : s_i \leq j \leq e_i\}$ which is an interval for all $i \in \mathcal{M}$. We first show that a matrix A with consecutive ones property can be transformed into the following — more convenient — form.
To this end, let us call a matrix A with consecutive ones property *monotone* if $s_1 \leq s_2 \ldots \leq s_m$ and $e_1 \leq e_2 \ldots \leq e_m$ hold simultaneously. Furthermore, if $s_1 < s_2 \ldots < s_m$ and $e_1 < e_2 \ldots < e_m$, A we will be called *strictly monotone.*

Lemma 2 *Let A, c be the input data of a set covering problem (SCP) with consecutive ones property. Then there exists an equivalent set covering problem with input data $\bar{A}, c$ such that $\bar{A}$ is a strictly monotone matrix, possibly with less rows than A.*

Proof: The proof works by first sorting the rows of A according to s_m and then applying part 1 of Lemma 1 to eliminate rows until strictly monotonicity is obtained.

QED

This can be performed in $O(m \log m)$ time as follows:

Algorithm 1

Input: Matrix A with m rows and consecutive ones property.

Output: Strictly monotone matrix.

Step 1. Order the rows of A such that $s_1 \leq s_2 \leq \dots s_m$.

Step 2. If $s_1 < s_2 \dots < s_m$ set $i^0 = 1$ and goto Step 4. Otherwise choose i, i' such that $s_i = s_{i'}$.

Step 3. (Reduction 1) If $e_i \geq e_{i'}$: delete row i, otherwise delete row i'. Let $m := m - 1$, and rename to obtain $s_1 < s_2, \dots < s_m$. Goto 2.

Step 4. $\bar{i} := \text{argmin}\{e_{i'} : i' \geq i^0\}$. If the minimum is not unique, choose the one with the larger row index i.

Step 5. (Reduction 2) Delete all rows i' with $i^0 \leq i' < \bar{i}$.

Step 6. If $\bar{i} \geq m - 1$ STOP, otherwise set $i^0 := \bar{i} + 1$ and return to 4.

Due to Lemma 2 we may assume that the covering matrix A already is in strictly monotone form. To find an optimal solution we transform the set covering problem to a shortest path problem in a cycle-free digraph.

First, we need some more notation. Analogously to s_i, e_i for rows we define for columns

$$\begin{aligned} \bar{s}_j &:= \min\{i \in \mathcal{M} : a_{ij} = 1\} \\ \bar{e}_j &:= \max\{i \in \mathcal{M} : a_{ij} = 1\}. \end{aligned} \tag{3}$$

The following observations are obvious.

Lemma 3

1. *Let A be a strictly monotone matrix. Then A^T is monotone.*
2. *If $\bar{\mathcal{N}} = \{j_1, j_2, \dots, j_p\} \subseteq \mathcal{N}$ is a cover of A with $j_1 < j_2 \dots < j_p$ then $\bar{s}_{j_1} = 1$ and $\bar{e}_{j_p} = m$.*

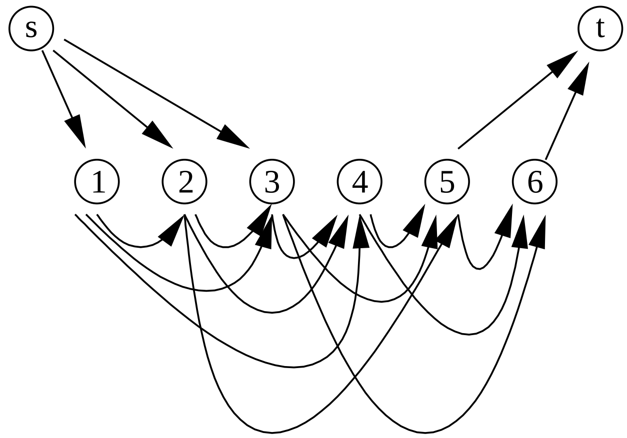

Figure 1: The digraph G_1 for the example.

Given the matrix A we are now in the position to define a digraph $G_1 = (V, E_1)$ by

$$V := \mathcal{N} \cup \{s, t\}$$

and

$$E_1 = \{(j,k) : j < k \text{ and } \bar{s}_k \leq \bar{e}_j + 1\} \cup \{(s,j) : \bar{s}_j = 1\} \cup \{(j,t) : \bar{e}_j = m\}.$$

For each edge (j, k) we associate a cost

$$c_{jk} := \begin{cases} c_j & \text{if } j \neq t \\ 0 & \text{if } j = t \end{cases}$$

Obviously, G_1 is a directed cycle-free graph. As an example, consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

The digraph G_1 corresponding to A is shown in Figure 1.

Now consider any s-t-path in G_1 with its set of nodes P. Since G_1 contains no cycles, the path belonging to a node set P is uniquely defined. This justifies the notation of the next theorem.

Theorem 1 *Let $P \subseteq V$. Then P is a cover of $\mathcal{M}$ if and only if $P \cup \{s, t\}$ is an s-t-path in G_1.*

Proof: Let $P = \{j_1, j_2, \dots j_p\}$ with $j_1 < j_2 < \dots j_p$.

1. Let $P \cup \{s, t\}$ be a path in G_1, and assume that P is not a cover. Choose an uncovered row $\bar{i}$ with minimal index. Consequently, row $\bar{i} - 1$ is covered, say by $j_k \in \mathcal{N}$, and choose j_k with maximal index. Then
$$\bar{e}_{j_k} = \bar{i} - 1. \tag{4}$$
We distinguish two cases, namely if j_k is the last node on the path before t or if it is followed by another node $j_{k+1} \neq t$.

   $k = p$: Then $(j_k, t) \in E$, such that $\bar{e}_{j_k} = m$ yielding $\bar{i} - 1 = m$, a contradiction.

   $k < p$: Then $j_{k+1} \neq t$ is the next node behind j_k within the path. Hence, $(j_k, j_{k+1}) \in E$. According to the definition of E, this means that $\bar{s}_{j_{k+1}} \leq \bar{e}_{j_k} + 1$. From (4) we consequently conclude
$$\bar{s}_{j_{k+1}} \leq \bar{i}. \tag{5}$$
Furthermore, note that we have chosen j_k as the column with maximal index covering row $\bar{i} - 1$. Hence j_{k+1} does not cover row $\bar{i} - 1$. Due to the monotonicity of A^T we hence know
$$\bar{s}_{j_{k+1}} \geq \bar{i}. \tag{6}$$
(5) and (6) together show that j_{k+1} covers row $\bar{i}$, a contradiction.

2. Now let P be a cover.
   - Then $\bar{s}_{j_1} = 1$ (Part 2 of Lemma 3) and hence $(s, j_1) \in E$.
   - Analogously, $\bar{e}_{j_p} = m$ yielding $(j_p, t) \in E$.
   - Assume $(j_k, j_{k+1}) \notin E$. Then $\bar{s}_{j_{k+1}} > \bar{e}_{j_k} + 1$. This means neither column j_k nor column j_{k+1} cover row $\bar{e}_{j_k} + 1$. Due to the monotonicity of A^T we further get (see Figure 2):
     - $\bar{s}_{j_{k'}} > \bar{e}_{j_k} + 1$ for all $k' \geq k + 1$, and
     - $\bar{e}_{j_{k'}} \leq \bar{e}_{j_k}$ for all $k' \leq k$.

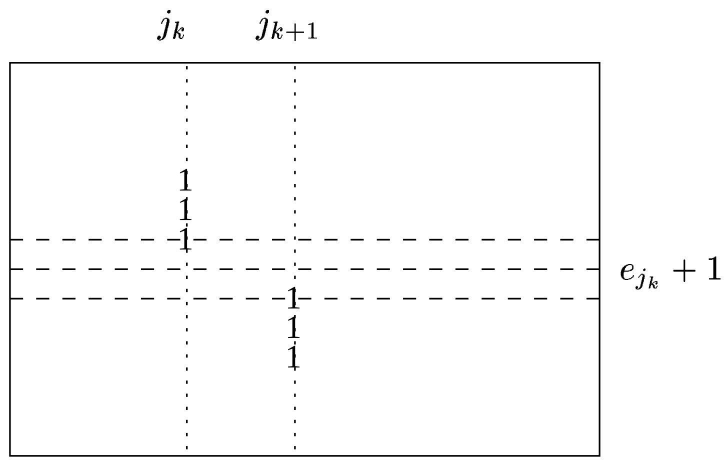

Figure 2: Illustration of second part of proof for Theorem 1

Together, row $\bar{e}_{j_k}+1$ is not covered by any column, a contradiction.

QED

Since the cost of a cover equals the cost of the corresponding path and vice versa, we finally get the following result.

Corollary 1 *A shortest s-t-path in G_1 represents a minimal cover and vice versa.*

This justifies the correctness of the following shortest path algorithm for solving set covering problems with consecutive ones property.

Algorithm 2

Input: Set covering problem where A has the consecutive ones property.

Output: An optimal solution $\bar{\mathcal{N}}$.

Step 1. Use Algorithm 1 to transform A into a strictly monotone matrix.

Step 2. Derive the graph $G_1 = (V, E_1)$.

Step 3. Find a shortest s-t-path P' in G_1 by a shortest path algorithm.
Output: $\bar{\mathcal{N}} := P' \setminus \{s, t\}$.

Note that Algorithm 2 is still correct if we replace E_1 by

$$E_1' = \{(j,k) : \bar{e}_j + 1 \in \mathcal{M}_k\} \cup \{(s,j) : 1 \in \mathcal{M}_j\} \cup \{(j,t) : m \in \mathcal{M}_j\},$$

since E_1' still contains all minimal covers, i.e. all covers $\bar{\mathcal{N}}$ which satisfy that no $\tilde{N} \subset \bar{\mathcal{N}}$ also is a cover. Minimal covers will also be needed in Section 5.

The unweighted case

Although set covering problems are still NP-hard in the unweighted case, this case is significantly easier to solve if the matrix has the consecutive ones property. The reason is due to the next lemma which is only true in the unweighted case.

Lemma 4 *A matrix A with consecutive ones property can be reduced to a (smaller) unit matrix by applying a finite sequence of the two reduction rules of Lemma 1.*

Proof: Let A' be the reduced matrix. Suppose that no further application of the rules of Lemma 1 is possible, and A' is not a unit matrix. From Lemma 2 we know that A' can be assumed to be strictly monotone. Then take i minimal such that

Case 1: either $a_{ii} = 1$ and a_{ii} is not the only non-zero entry of column i

Case 2: or $a_{ii} = 1$ and a_{ii} is not the only non-zero entry of row i

Case 3: or $a_{ii} = 0$.

Since A' is strictly monotone, Case 1 cannot occur. Case 2 means $a_{ii} = 1$ and $a_{ik} = 1$ for $k > i$, yielding $a_{ii+1} = 1$ due to the consecutive ones property. Below a_{ii} only zero entries occur (since A' is strictly monotone). This means, $\mathcal{M}_i \subseteq \mathcal{M}_{i+1}$ and column i could have been deleted. In the last case, if $a_{ii} = 0$ we get $s_i > i$, leading to $s_l > i$ for all $l > i$ and column i again could have been deleted.

QED

Note that the remaining columns of the unit matrix obtained by Lemma 4 are the optimal solution. For the sake of completeness, we also repeat the following well-known *"folklore"* algorithm which can also be used for solving an unweighted set covering problem with consecutive ones property. It can be performed in linear time $\mathrm{O}(m \log m + n)$.

Algorithm 3

Input: Unweighted set covering problem where A has the consecutive ones property.

Output: An optimal solution $\bar{\mathcal{N}}$.

Step 1. Use Algorithm 1 to transform A into a strictly monotone matrix, set $\bar{i} := 1$, $\bar{\mathcal{N}} := \emptyset$

Step 2. $\bar{\mathcal{N}} := \bar{\mathcal{N}} \cup \{e_{\bar{i}}\}$

Step 3. If $\{i' : s_{i'} > e_{\bar{i}}\} \neq \emptyset$ choose $\bar{i} := \min\{i' : s_{i'} > e_{\bar{i}}\}$ and goto 2, otherwise STOP. Output: $\bar{\mathcal{N}}$

4 Bicriteria set covering problems

In practical optimization problems, one may not be interested in covering *all* rows with minimal cost, but may wish to cover as many rows as possible spending as few costs as possible. This leads to a bicriteria optimization problem. To formulate this problem, we allow positive weights w_m for each row $m \in \mathcal{M}$. For a given set $\bar{\mathcal{N}} \subseteq \mathcal{N}$ we define

$$\text{cover}\bar{\mathcal{N}} = \bigcup_{j \in \bar{\mathcal{N}}} \mathcal{M}_j$$

as the set of rows which are covered by $\bar{\mathcal{N}}$.
We consider the following two objective functions:

$$\begin{aligned} \min f_1(\bar{\mathcal{N}}) &= |\bar{\mathcal{N}}|, \quad \text{and} \\ \max f_2(\bar{\mathcal{N}}) &= \sum_{m \in \text{cover}(\bar{\mathcal{N}})} w_m. \end{aligned}$$

To state the bicriteria problem accurately, we need the following basic definition from multicriteria optimization. For a recent introduction into multicriteria optimization, see e.g. [Ehr05].

Let $\mathcal{F} = \{(f_1, f_2) : \text{ there exists some } \bar{\mathcal{N}} \subseteq \mathcal{N} : f_1(\bar{\mathcal{N}}) = f_1 \text{ and } f_2(\bar{\mathcal{N}}) = f_2\}$ be the set of feasible points in the objective space. Then the set of *efficient points* is given by

$$\begin{aligned} \text{Eff} \quad = \quad & \{(f_1, f_2) \in \mathcal{F} : \text{ There does not exist some } (g_1, g_2) \in \mathcal{F} \\ & \text{with } g_1 \leq f_1, g_2 \geq f_2, \text{ and } (f_1, f_2) \neq (g_1, g_2)\}. \end{aligned}$$

Furthermore, each $\bar{\mathcal{N}}$ with $(f_1(\bar{\mathcal{N}}), f_2(\bar{\mathcal{N}})) \in \text{Eff}$ is called a *Pareto solution.* The goal of the *bicriteria set covering problem* is to determine a set of Pareto solutions, from which the decision-maker can choose the most appropriate. In this section we show how the complete set Eff can be found in the bicriteria set covering problem. The idea of the *ϵ-constraint method* (see [HC83]) is to bound one of the objectives and to solve the restricted problem optimizing the remaining objective. Since in the unweighted case f_1 only can take the values $1, 2, \ldots, N$ we consider

$$\begin{aligned} \max \quad & f_2(\bar{\mathcal{N}}) \\ \text{s.t.} \quad & |\bar{\mathcal{N}}| \leq k \end{aligned} \tag{7}$$

for some fixed natural number $k \leq k^*$, where k^* denotes the cardinality of an optimal cover of the unweighted set covering problem (which can be found by applying Algorithm 3 of Section 3). Due to Haimes and Chankong [HC83] we have the following result:

Theorem 2 *If $\bar{\mathcal{N}}$ is a unique solution of (7) for some $k \leq k^*$ then $\bar{\mathcal{N}}$ is a Pareto solution belonging to the efficient point $(f_1(\bar{\mathcal{N}}), f_2(\bar{\mathcal{N}}))$. If more than one optimal solution to problem (7) exists, the solutions with the smallest f_1-values are Pareto solutions.*

In our case, if all $w_i > 0$ all solutions of (7) with $k \leq k^*$ are Pareto solutions.

Consequently, we want to tackle problem (7). To this end, we define a cycle-free digraph $G_2 = (V, E_2)$ by

$$V := \mathcal{N} \cup \{s, t\}$$

and

$$E_2 = \{(j, k) : j, k \in \mathcal{N} \text{ and } j < k\} \cup \{(s, j) : j \in \mathcal{N}\} \cup \{(j, t) : j \in \mathcal{N}\}.$$

For each edge (j,k) we define weights

$$w_{jk} := \begin{cases} \sum_{i\in\mathcal{M}_k\setminus\mathcal{M}_j} w_i & \text{if } j\neq s, k\neq t \\ \sum_{i\in\mathcal{M}_k} w_i & \text{if } j=s, k\neq t \\ 0 & \text{if } j\neq s, k=t \end{cases}$$

Furthermore, for an s-t-path P in G_2 let $W(P)$ denote its length according to the edge weights w_{jk}. We obtain the following result.

Theorem 3 *Let $P \subseteq \mathcal{N}$. Then $W(P\cup\{s,t\}) = f_2(P)$*

Proof: We use induction according to $p := |P|$.
For $p=1$ the claim is true. Now assume that

$$W(P'\cup\{s,t\}) = f_2(P')$$

for all P' with $|P'|\leq p$. Take some $P=\{j_1,j_2,\ldots,j_p,j_{p+1}\}$ and assume that $j_1<j_2<\ldots<j_{p+1}$. Define $P'=\{j_1,j_2,\ldots j_p\}$. Then we get

$$\begin{aligned} f_2(P) &= \sum_{i\in\text{cover}(P)} w_i \\ &= \sum_{i\in\text{cover}(P')} w_i + \sum_{i\in\mathcal{M}_{j_{p+1}}\setminus\text{cover}(P')} w_i \\ &= W(P'\cup\{s,t\}) + \sum_{i\in\mathcal{M}_{j_{p+1}}\setminus\mathcal{M}_{j_p}} w_i \qquad (8) \\ &= W(P\cup\{s,t\}), \end{aligned}$$

where it remains to prove

$$\mathcal{M}_{j_{p+1}}\setminus\text{cover}\{j_1,\ldots,j_p\} = \mathcal{M}_{j_{p+1}}\setminus\mathcal{M}_{j_p}$$

to show that (8) holds.

Since "$\subseteq$" is trivial, we only need to verify "$\supseteq$".
To this end, let $i\in\mathcal{M}_{j_{p+1}}\setminus\mathcal{M}_{j_p}$. We show that $i\notin\mathcal{M}_{j_k}$ for all $k\leq p$. Assume the contrary, i.e. $i\in\mathcal{M}_{j_k}$ and $i\in\mathcal{M}_{j_{p+1}}$. This means that $a_{ij_k}=a_{ij_{p+1}}=1$, and, since A has the consecutive ones property also $a_{ij_p}=1$, a contradiction to $i\notin\mathcal{M}_{j_p}$.

QED

Corollary 2 *A longest s-t-path in G_2 (with respect to the length W) with no more than k edges is a maximal cover with cardinality less than k, i.e. a solution to problem (7).*

Note that, since the digraph G_2 contains no cycles, the longest path problem is equivalent to a shortest path problem. To find a longest path with no more than k edges, we can hence use the shortest path algorithm of Bellmann-Ford (see, e.g. [NW88]). This algorithm needs $O(kN^2)$ time to find shortest paths with cardinality less than k from one specified starting node to all other nodes in the graph. Since in our case, the graph contains no cycles, this complexity reduces to $O(q^3)$ where $q = \max\{k, n-k\}$, and that is also the overall complexity of the next algorithm.

Algorithm 4

Input: Unweighted set covering problem where A has the consecutive ones property.

Output: All efficient points, and a Pareto solution for each of them.

Step 1. Use Algorithm 1 to transform A into a strictly monotone matrix.

Step 2. Solve the unweighted set covering problem by Algorithm 3, let k^* be the cardinality of the optimal solution.

Step 3. Derive G_2.

Step 4. Use the algorithm of Bellmann-Ford to find all longest paths from s to any other nodes with $k = 1, 2, \ldots k^*$ edges. Let h^k denote the length of a longest s-t-path P^k with at most k edges.

Step 5. Let Eff $= \{(h^1, 1)\}$ For $k = 2, \ldots, k^*$: If $h^k > h^{k-1}$ set Eff $=$ Eff $\cup \{h^k, k)\}$. Output: Eff .

In the context of stop location, a special case of this approach has been formulated in [Sch03b].

5 More than one block of consecutive ones

In this section we investigate if the proposed approach can be extended to set covering problems with coefficient matrices having more than one block of consecutive ones per row. Unfortunately, this is not the case: Even for two blocks per row the set covering problem is NP hard. This can be shown by reduction from *vertex cover*: Formulating the respective problem as integer program yields a set covering problem with exactly two non-zeros per row.

The next theorem shows a special case with at most two non-zeros per row, in which the set covering problem is already NP-hard.

Lemma 5 ([SM05]) *The set covering problem is NP-hard even for the case that the covering matrix can be written as $A = (A^1|A^2)$, where A^1 and A^2 both have the consecutive ones property and $(A^1|A^2)$ has exactly two ones per row, A^1 has no more than one one per row and A^2 has no more than two ones per row.*

However, we will show that the problem is polynomially solvable if A_1 and A_2 both have the *strong* consecutive ones property, (independent from the number of ones per row) and that it can be solved by extending our approach of Section 3.

More specific, we now consider *decomposable* set covering problems of the following form:

(SCP-dec)

$$\begin{array}{lrcl} \min & cx & & \\ \text{s.t.} & (A^1|A^2|\dots|A^G)x & \geq & \underline{1}_m \\ & x & \in & \{0,1\}^n, \end{array} \tag{9}$$

with $m \times n_g$ matrices A^g, $g = 1, \dots, G$.

Theorem 4 *(SCP-dec) with matrices A^g, $g = 1, \dots, G$ all satisfying the strong consecutive ones property is polynomially solvable.*

<u>Proof:</u> From part 1 of Lemma 3 we know that

$$A = (A^1|A^2|\dots|A^G)$$

is an interval matrix, i.e. A^T has the consecutive ones property. Hence, A is totally unimodular (see, e.g., III.1 of [NW88]), and hence the linear programming relaxation of (SCP-dec) yields an integer solution, such that (SCP-dec) is polynomially solvable.

QED

Let us for the following assume that all matrices A^g have the strong consecutive ones property. We show that instead of just using linear programming we can transform (SCP-dec) to a shortest path problem in an acyclic digraph with $\sum_{g=1}^{G} n_g + m$ nodes.

To this end, we need the following notations.
Let $\mathcal{N}_g = \{1, \ldots, n_g\}$ and $\mathcal{N} = \cup_{g=1,\ldots,G}\{(g,j) : j \in \mathcal{N}_g\}$ be the set of indices for all columns in the set covering problem. We denote the elements of matrix A^g by a_{ij}^g. Let A_j^g be the jth column of matrix A^g. Its index hence is given by (g,j).

We furthermore extend (3) to our new case.

$$\begin{aligned} \bar{s}_j^g &:= \min\{i \in \mathcal{M} : a_{ij}^g = 1\} \\ \bar{e}_j^g &:= \max\{i \in \mathcal{M} : a_{ij}^g = 1\}. \end{aligned}$$

We are finally in the position to define a digraph $G_3 = (V_3, E_3)$.

$$V_3 := \mathcal{N} \cup \mathcal{M} \cup \{0\}$$

and

$$E_3 = \{((g,j),i) : \bar{e}_j^g = i\} \cup \{(i,(g,j)) : \bar{s}_j^g \leq i+1 \leq \bar{e}_j^g\}$$

The first class of edges connects each column A_j^g to the last row it covers, i.e., to $i = \bar{e}_j^g$. The second class of edges connects a row i to all columns (g,j) covering row $i+1$. Since all rows with larger indices than i are not covered by columns (g,j) belonging to incoming edges at i, the resulting digraph G_3 does not have any directed cycles.

For edges $((g,j),i)$ from $\mathcal{N}$ to $\mathcal{M}$ we associate costs $c_{(g,j),i} = 0$ while we define $c_{i,(g,j)} = c_j^g$ for edges from $\mathcal{M}$ to $\mathcal{N}$.
Our goal is to show that a minimal cover can be found by calculating a shortest path P from 0 to m and taking the nodes of P corresponding to columns of A, i.e., $\bar{\mathcal{N}} = P \cap \mathcal{N}$.
While we can show that each $0-m$ path corresponds to a feasible cover, there exist covers $\bar{\mathcal{N}}$ which can not be represented by a suitable path in G_3. The correspondence between paths and covers hence is not a one-to-one relation as in Theorem 1. But fortunately we are able to shown that covers which cannot be represented as $0-m$ paths will never be optimal. Let us first consider the transformation from paths to covers.

Theorem 5 *Let P be a path from 0 to m in G_3. Then $P \cap \mathcal{N}$ is a cover.*

Proof: Let P be a $0-m$-path in G_3 and $\bar{\mathcal{N}} = P \cap \mathcal{N}$. We have to show that all rows $i = 1, \ldots, m$ are covered by a column $(g, j) \in \bar{\mathcal{N}}$.
If $i \in P \cap \mathcal{M}$ this is clear since the predecessor (g, j) of i in P corresponds to a column A_j^g covering i (since $a_{ij}^g = 1$ according to the definition of E_3).

Now consider a row $i \in \mathcal{M} \setminus P$. Since each path ends at m we know that $i < m$. Consequently, there exist two nodes $\bar{i}$, $\underline{i} \in P \cap (\mathcal{M} \cup \{0\})$ such that

$$\underline{i} < i < \bar{i} \tag{10}$$

and the sequence $(\underline{i}, (g, j), \bar{i})$ is a subpath of P for some $(g, j) \in \mathcal{N}$. We want to show that (g, j) also covers i.

- From $(\underline{i}, (g, j)) \in E$ we know that $\bar{s}_j^g \leq \underline{i} + 1 \leq \bar{e}_j^g$.
- From $((g, j), \bar{i}) \in E$ we know that $\bar{e}_j^g = \bar{i}$.

Combining these results with (10) we obtain

$$\bar{s}_j^g \leq \underline{i} + 1 < i < \bar{i} = \bar{e}_j^g,$$

i.e., i is covered by (g, j).

QED

We now have to deal with the set of covers which can be represented as paths. To this end, we deal with *(inclusion) minimal* covers, i.e., covers $\bar{\mathcal{N}}$ which satisfy that $\bar{\mathcal{N}} \setminus \{j\}$ is not a cover for all $j \in \bar{\mathcal{N}}$.

In case of the strong consecutive ones property we obtain the following result for (SCP-dec).

Lemma 6 *Let $\bar{\mathcal{N}}$ be a minimal cover of (SCP-dec). Then $\bar{e}_j^g \neq \bar{e}_{j'}^{g'}$ for all $(g, j) \neq (g', j')$.*

Proof: If $\bar{e}_j^g = \bar{e}_{j'}^{g'}$ for two columns $(g, j) \neq (g', j')$ both in $\bar{\mathcal{N}}$, one of them can be deleted from $\bar{\mathcal{N}}$ resulting in a smaller cover.

QED

Theorem 6 *Let $\bar{\mathcal{N}}$ be a minimal cover. Then there exists a path P from 0 to m in G_3 with $P \cap \mathcal{N} = \bar{\mathcal{N}}$.*

Proof: Let $\bar{\mathcal{N}}$ a minimal cover. From Lemma 6 we know that no two columns of $\bar{\mathcal{N}}$ end at the same row. We consequently sort the elements of $\bar{\mathcal{N}}$ according to $\bar{e}_g^j$ resulting in a sequence

$$(g_1, j_1) < (g_2, j_2) < \ldots < (g_p, j_p).$$

Then the following properties hold:

(*) If row i satisfies $\bar{e}_{j_k}^{g_k} < i < \bar{e}_{j_{k+1}}^{g_{k+1}}$ then i is covered by (g_{k+1}, j_{k+1}).

(**) All rows i with $i < \bar{e}_{j_1}^{g_1}$ are covered by (g_1, j_1).

We now construct P as

$$P = (0, (g_1, j_1), \bar{e}_{j_1}^{g_1}, (g_2, j_2), \bar{e}_{j_2}^{g_2}, \ldots, (g_p, j_p), \bar{e}_{j_p}^{g_p}).$$

Clearly, $((g_k, j_k), \bar{e}_{j_k}^{g_k}) \in E_3$ since this is the first type of edges in E_3.
It remains to show that also the other edges exist:

- $(\bar{e}_{j_k}^{g_k}, (g_{k+1}, j_{k+1})) \in E_3$ for all $k = 1, \ldots, p-1$: From (*) we know that $i := \bar{e}_{j_k}^{g_k} + 1$ is covered by (j_{k+1}, g_{k+1}) Consequently,

$$\bar{s}_{j_{k+1}} \leq \bar{e}_{j_k}^{g_k} + 1 \leq \bar{e}_{j_{k+1}}^{g_{k+1}},$$

  hence $(\bar{e}_{j_k}^{g_k}, (g_{k+1}, j_{k+1})) \in E_3$.

- Analogously, $(0, (g_1, j_1)) \in E_3$ follows from (**).

- Finally, $(g_p, j_p, m) \in E_3$, otherwise row m would not have been a covered.

QED

We hence have shown the following result.

Corollary 3 *A shortest path from 0 to m in G_3 represents a minimal cover and vice versa.*

The resulting algorithm works analogously to Algorithm 2.

We remark that all nodes in $\mathcal{M} \setminus \{m\}$ can be deleted by the standard *node reduction* of project planning (see, e.g., [Elm77]) such that a digraph $G_3' = (V_3', E_3')$ with only $|\mathcal{N}| + 2$ nodes (but the same number of edges) remains. Note that for the case of only one block per row, the resulting digraph G_3' is exactly the same as the digraph (V, E_1') defined in (3) on page 10 (identifying s with 0 and t with m).

6 Conclusion

In this paper we discussed a new approach for solving set covering problems with consecutive ones property, or decomposable set covering problems with strong consecutive ones property. Some of the ideas might be transferred to dynamic set covering problems and multi-covering problems.
From a practical point of view it is a challenging task to be able to solve large instances of set covering problems using the block structure of the covering matrix. An approach for problems obtaining *almost* the consecutive ones property is presented in [RS04] and an approximation algorithm whose ratio depends on the number of blocks of consecutive ones in a row is suggested in [SM05].

References

[BEHW05] D. Baatar, M. Ehrgott, H.W. Hamacher, and G. Woeginger. Decomposition of integer matrices and multileaf collimator sequencing. *Discrete Applied Mathematics*, 2005. to appear.

[BL76] K.S. Booth and G.S. Lueker. Testing for the consecutive ones property, interval graphs and graph planarity using pq-tree algorithms. *Journal of Computing Systems Science*, 13(3):335–379, 1976.

[CFT+97] A. Caprara, M. Fischetti, P. Toth, D. Vigo, and P.L. Guida. Algorithms for railway crew management. *Mathematical Programming*, 79, 1997.

[CFT99] A. Caprara, M. Fischetti, and P. Toth. A heuristic method for the set covering problem. *Operations Research*, 47(5):730–743, 1999.

[CFT00] A. Caprara, M. Fischetti, and P. Toth. Algorithms for the set covering problem. *Annals of Operations Research*, 98:353–371, 2000.

[CNS97] S. Ceria, P. Nobili, and A. Sassano. Set covering problem. In M. Dell'Amico, F. Maffioli, and S. Martello, editors, *Annotated* *bibliographies in combinatorial optimization*, pages 415–428. J. Wiley and Sons, 1997.

[Ehr05] M. Ehrgott. *Multiple Criteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, 2005.

[Elm77] S.E. Elmaghraby. *Activity Networks*. Wiley Interscience Publication, 1977.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability — A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.

[HC83] Y.Y. Haimes and V. Chankong. *Multiobjective Decision Making — Theory and Methodology*. North Holland, New York, 1983.

[LY94] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41:960–981, 1994.

[MPT98] J. Meidanis, O. Porto, and G.P. Telles. On the consecutive ones property. *Discrete Applied Mathematics*, 88:325–354, 1998.

[MS00] E. Marchiori and A. Steenbeek. An evolutionary algorithm for large set covering problems with application to airline crew scheduling. In *EvoWorkshops*, pages 367–381. 2000.

[NW88] G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. Wiley, 1988.

[RS04] N. Ruf and A. Schöbel. Set covering problems with almost consecutive ones property. *Discrete Optimization*, 1(2):215–228, 2004.

[Sch03a] A. Schöbel. Customer-oriented optimization in public transportation, 2003. Habilitationsschrift, Universität Kaiserslautern.

[Sch03b] A. Schöbel. Locating stops along bus or railway lines — a bicriterial problem. *Annals of Operations Research*, 2003. to appear.

[SHLW02] A. Schöbel, H.W. Hamacher, A. Liebers, and D. Wagner. The continuous stop location problem in public transportation. Technical report, Universität Kaiserslautern, 2002. Report in Wirtschaftsmathematik Nr. 81/2001.

[SM05] D. Wagner S. Mecke, S. Schöbel. Stop location - complexity and approximation issues. Technical report, University of Karlsruhe, 2005.

[TR73] C. Toregas and C. ReVelle. Binary logic solutions to a class of location problems. *Geographical Analysis*, 5:145–155, 1973.

Institut für Numerische und Angewandte Mathematik
Universität Göttingen
Lotzestr. 16-18
D - 37083 Göttingen

Telefon: 0551/394512
Telefax: 0551/393944

Email: trapp@math.uni-goettingen.de URL: http://www.num.math.uni-goettingen.de

Verzeichnis der erschienenen Preprints:

2005-01	A. Schöbel, S. Scholl:	Line Planning with Minimal Traveling Time
2005-02	A. Schöbel	Integer programming approaches for solving the delay management problem
2005-03	A. Schöbel	Set covering problems with consecutive ones property