Technische Universität Berlin

FAKULTÄT II
MATHEMATIK UND NATURWISSENSCHAFTEN
Institut für Mathematik

A Greedy Approach to Compute a Minimum Cycle Bases of a Directed Graph

by

Christian Liebchen
TU Berlin, Institut für Mathematik, Sekr. MA 6-1
Strasse des 17. Juni 136, D-10623 Berlin, Germany
liebchen@math.tu-berlin.de
Romeo Rizzi
Università degli Studi di Trento, Facoltà di Scienze
Dipartimento di Informatica e Telecomunicazioni
Via Sommarive, 14, I-38050 Povo - Trento, Italy

No. 2004/31

A Greedy Approach to Compute a Minimum Cycle Basis of a Directed Graph ⋆

Christian Liebchen [a,*] Romeo Rizzi [b]

[a] *Technische Universität Berlin, Institut für Mathematik, Combinatorial Optimization and Graph Algorithms, Straße des 17. Juni 136, D-10623 Berlin, Germany*

[b] *Università degli Studi di Trento, Facoltà di Scienze – Dipartimento di Informatica e Telecomunicazioni Via Sommarive, 14, I-38050 Povo - Trento, Italy*

Abstract

Given a directed graph $D = (V, A)$, we consider its cycle space $\mathcal{C}_D$, i.e. the vector subspace of $\mathbb{Q}^{|A|}$ spanned by the incidence vectors of the oriented cycles of D. An *oriented cycle* of D is just any cycle of the underlying undirected graph of D along with an orientation; its *incidence vector* is 0 on the arcs not included, while, for the included arcs, it is $+1$ on the arcs oriented according to the orientation and -1 on the arcs going backward. Assume a nonnegative weight $w_a \in \mathbb{R}_+$ is associated to each arc a of D. We can extend the weighting w to subsets F of A and to families $\mathcal{F}$ of such subsets by defining $w(F) := \sum_{f \in F} w(f)$ and $w(\mathcal{F}) := \sum_{F \in \mathcal{F}} w(F)$. Given the pair (D, w), we are interested in computing a minimum weight basis of $\mathcal{C}_D$.

This problem is strongly related to the classical problem of computing a minimum cycle basis of an undirected graph. In 1987, Horton developed the first polynomial time algorithm for computing a minimum cycle basis of an undirected graph. As for directed graphs, the first algorithm for computing a minimum directed cycle basis is due to Kavitha and Mehlhorn. Its asymptotic complexity is $\tilde{O}(m^4 n)$.

In this paper, we show how the original approach of Horton can be actually pursued also in the context of directed graphs, while retaining its simplicity. This both allows for a practical $\tilde{O}(m^4 n)$ adaptation of Horton's original algorithm requiring only minor modifications in the actual code and for a more involved $\tilde{O}(m^{\omega+1} n)$ solution. At the end, we discuss the applicability of this approach to more specialized classes of directed cycle bases, namely, integral cycle bases and generalized fundamental cycle bases.

Key words: Minimum Cycle Bases, Directed Graphs, Greedy Algorithm

Preprint submitted to Elsevier Science *20 January 2005*

1 Introduction

The task of computing a minimum cycle basis of a graph is well-studied. Besides its beauty, the problem is motivated by its practical relevance as a preprocessing step in various application fields, such as electric circuits [2] or chemical ring perception [5].

Undirected graphs. The *cycle space* of an undirected graph $G = (V, E)$ is the vector space $\mathcal{C}_G$ over $GF(2)$ generated by the incidence vectors of the cycles of G. Assume a nonnegative weight $w_e \in \mathbb{R}_+$ is associated to each edge e of G. We can extend the weighting w to subsets F of E and to families $\mathcal{F}$ of subsets of E by defining $w(F) := \sum_{f\in F} w(f)$ and $w(\mathcal{F}) := \sum_{F\in\mathcal{F}} w(F)$.

Horton [8] developed the first polynomial time algorithm for computing a minimum cycle basis of a graph. His approach was based on two main observations: First, the incidence vectors of the cycles of a graph form a matroid, when considering standard linear independence over $GF(2)$. Second, he identified a set of $O(mn)$ cycles [1], which includes the elements of *all* minimum cycle bases. These ideas delivered an $O(m^3n)$ greedy algorithm. Later, Golinsky and Horton [6,7] could blend them into a more sophisticated recursion scheme based on fast matrix multiplication. Hereby the running time is $O(m^\omega n)$, with ω being the constant of fast matrix multiplication, thus $\omega < 2.376$.

Recently, there have been published new algorithms to solve this problem. The approaches of Berger, Gritzmann, and de Vries [1] and Kavitha et al. [10] — which share in fact some ideas that can already be found in [3] — subsequently build up a minimum cycle basis by adding in each iteration a shortest cycle, which is in a sense orthogonal to the ones chosen in previous iterations. More technically spoken, these algorithms rely on certificates of independence to be updated meanwhile new cycles enter the basis. Along this line, the best running time culminated to only $O(m^2n + mn^2 \log n)$.

Directed graphs. Given a directed graph $D = (V, A)$, and $F \subseteq A$, we denote by $F^\star$ the arc set obtained from F by reversing all arcs, that is, $F^\star := \{(u, v) \,|(v, u) \in F\}$. An *oriented cycle* C of $D = (V, A)$ is a pair (C^+, C^-) of disjoint subsets of A such that $C^+ \cup (C^-)^\star$ is a directed cycle, in which all arcs point into the same direction. The arcs in C^+ (resp., in C^-) are called the *forward* (resp., *backward*) arcs of C. The *incidence vector* χ_C *of an oriented cycle* C is a vector in $\{-1, 0, 1\}^m$, with entry 1 (-1) in component a, if and only if a is a forward (backward) arc of C. The *cycle space* of D is the vector

⋆ Supported by the DFG Research Center Matheon in Berlin

∗ Corresponding Author.
Email address: liebchen@math.tu-berlin.de (Christian Liebchen).

[1] As usual, we define $n := |V|$ and $m := |A|$.

space $\mathcal{C}_D$ over $\mathbb{Q}$ generated by the incidence vectors of oriented cycles of D. Assume a nonnegative weight $w_a \in \mathbb{R}_+$ is associated to each arc a of D. We can extend the weighting w to subsets F of A and to families $\mathcal{F}$ of such subsets by defining $w(F) := \sum_{f \in F} w(f)$ and $w(\mathcal{F}) := \sum_{F \in \mathcal{F}} w(F)$. Given a weighted digraph (D, w), our task is to compute a minimum weight basis of $\mathcal{C}_D$.

Notice that a set of oriented cycles, which projects onto a cycle basis for the underlying undirected graph, is already a cycle basis of the directed graph ([12]). Hence, by computing a minimum cycle basis for the underlying undirected graph, one can obtain a short cycle basis for the directed graph. But there exist directed graphs that have a minimum cycle basis which does not project onto a cycle basis for the underlying undirected graph. A node-minimal simple digraph for this phenomenon is any oriented version of K_6. This example has been introduced in [12], and cited recently in [9] to justify the proposal of algorithms specifically designed to compute directed cycle basis of minimum weight. Nevertheless, if one takes any minimum undirected cycle bases of K_6, then the corresponding directed cycles do still form a minimum directed cycle basis in every orientation of K_6. This is because in K_6 there exist undirected cycle bases whose weight is as small as the minimum weight of a directed one. Hence, the justification provided by this one example was only partial.

Kavitha and Mehlhorn [9] gave the first algorithm for computing a minimum cycle basis of a directed graph, by generalizing the ideas in [10]. But as the concept of orthogonality becomes much more complex when switching from $GF(2)$ to $\mathbb{Q}$, its running time can only be bounded by $\tilde{O}(m^4 n)$.

Contribution. We provide a directed graph no minimum cycle basis of which projects onto a cycle basis of the underlying undirected graph. Hence, for computing a minimum cycle basis of a directed graph, it is *not* an option simply to compute an MCB for the underlying undirected graph. Rather, new algorithmic approaches—as they can be found in [9] and in this paper—are necessary.

We adapt Horton's original ideas to the setting of directed graphs. Conceptually, this approach stays very simple. Concerning complexity, and by omitting polylogarithmic terms, we pay a multiplicative slow-down factor of m, as arithmetics are now performed on non-binary numbers with up to $O(m \log m)$ bits. In particular, if the test for linear independence is simply based on iterative Gaussian elimination as in [8], then a practical $\tilde{O}(m^4 n)$ algorithm is obtained. Moreover, if we build on the recursion scheme sketched in [7], then we achieve a complexity of $\tilde{O}(m^{\omega+1} n)$, thus in particular of $\tilde{O}(m^{3.376} n)$. Finally, we show that the greedy approach does not apply to more specialized classes of cycle bases, such as integral or generalized fundamental cycle bases.

2 A directed cycle basis smaller than any undirected one

Consider the generalized Petersen graph $P_{7,2}$, cf. Figure 1. We call an edge $e = \{u, v\}$ an *inner edge* if $\{u, v\} \subset \{0, \dots, 6\}$. Similarly, we call an edge $e = \{u, v\}$ an *outer edge* if $\{u, v\} \subset \{a, \dots, g\}$. The seven edges that remain are called *spokes*. We define the weight function w as follows. Assign weight two to the seven inner edges. The outer edges and the spokes get weight three.

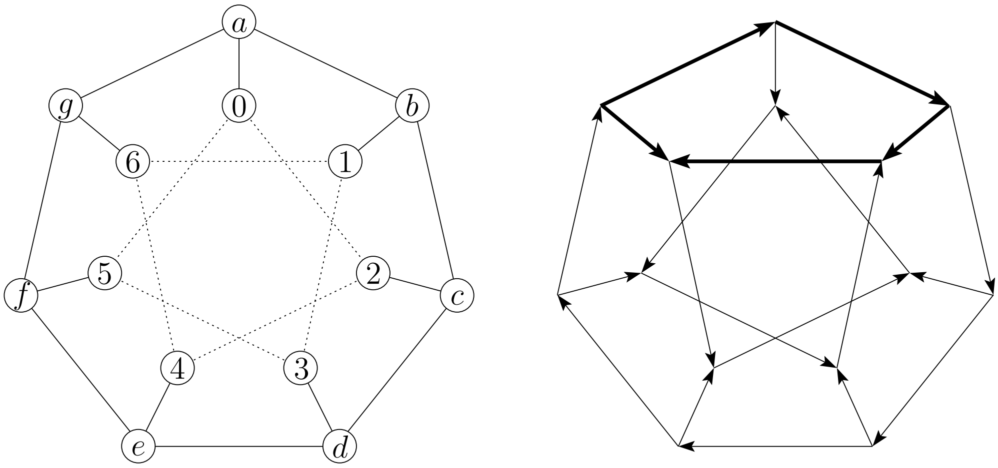

Fig. 1. The generalized Petersen graph $P_{7,2}$ (left) and a template cycle in an orientation of $P_{7,2}$ (right).

Claim 1 $(P_{7,2}, w)$ *has girth* 14*, and there are precisely eight cycles having weight* 14*.*

Proof. We will analyze the cycles of $P_{7,2}$ according to the number of spokes they involve.

There are precisely two cycles which do not contain any spoke. The outer cycle has weight 21, and the inner cycle is one of the eight cycles having weight 14.

Any other cycle uses an even number of spokes, and at least one of the outer edges and one of the inner edges. Hereby, any cycle using more than two spokes has weight at least $16 > 14$.

Consider therefore a cycle C taking precisely two spokes. If C contains precisely one outer edge, then C must traverse at least three of the inner edges, which yields $w(C) \geq 15$. If C contains at least three outer edges, already the weights of the outer edges plus the weights of the spokes sum up to 15. Among the cycles that use two outer edges and two spokes, the seven cycles that respect the template displayed in Figure 1 on the right have weight 14, and the remaining ones have weight 24. □

Observe that every edge of $P_{7,2}$ belongs to precisely two of the eight cycles with weight 14. Therefore, in the undirected case, these $8 = m - n + 1$ cycles are not independent over $GF(2)$. Thus, in the undirected case, every cycle basis of $(P_{7,2}, w)$ has weight at least 113.

Notice that in the directed case, linear independency over $\mathbb{Q}$ is invariant under re-orienting edges and cycles. This follows from the fact that the determinant of a directed cycle basis ([11]) is invariant under these operations, too. Hence, we may orient the edges of $P_{7,2}$ according to Figure 1 on the right. Further, we orient the template cycles as to follow the orientations of the outer arcs, and the inner cycle to traverse its arcs forwardly. We collect these eight oriented cycles in the set $\mathcal{B}$.

Claim 2 *$\mathcal{B}$ is a directed cycle basis.*

Proof. Assume to have a combination $\sum_{C \in \mathcal{B}} \lambda_C \chi_C$ that cancels out. Unless all multipliers λ_C are zero, there exists an oriented template cycle C' with multiplier $\lambda_{C'} = \alpha \neq 0$.

By considering the arcs that are spokes, we conclude that $\lambda_C = \alpha$ for *every* oriented cycle $C \in \mathcal{B}$ respecting the template. Then, the arcs in the outer cycle have charge amounting to 2α. But the multiplier of the remaining inner cycle cannot affect their charge. $\square$

Corollary 3 *For every orientation of $P_{7,2}$ there exists a directed cycle basis having weight* 112. *But every cycle basis of $(P_{7,2}, w)$ has weight at least* 113.

Alternatively, one may consider the determinant of $\mathcal{B}$. As the definition of the determinant of a directed cycle basis ([11]) generalizes to arbitrary sets of $m - n + 1$ oriented cycles, this would also be well-defined, if $\mathcal{B}$ was not independent over $\mathbb{Q}$. We compute $\det(\mathcal{B}) = 2$. Hence, we conclude that $\mathcal{B}$ is a directed cycle basis, which does not project onto an undirected one.

Notice that we may easily derive an unweighted graph to which our considerations apply as well. To that end, subdivide each edge e of $P_{7,2}$ precisely $w(e) - 1$ times. The result is an unweighted graph with 49 nodes and 56 edges.

3 Computing a Minimum Cycle Basis of a Directed Graph

Much like in the undirected case, the $\mu := m - n + 1$ dimensional vector subspace spanned by the incidence vectors of oriented cycles obviously leads to a linear matroid, applying standard linear independence over $\mathbb{Q}$. In order for the greedy algorithm to become a sufficiently fast procedure, we need to identify a small set of cycles including the union of all minimum cycle bases.

In the undirected case, where working over $GF(2)$, Horton [8] considers only the cycles of the form $C = P_{wu} + uv + P_{vw}$ for some edge uv and some node w in the input graph, where P_{ab} denotes a shortest path between a and b.

In principle, these could still be an exponential number of cycles. Horton observed however that if one perturbs the weights on the edges such that there exists a unique minimum path between any pair of nodes, then the Horton family contains at most mn different circuits.

An appropriate perturbation is easy to come along: simply assume the original weights were integers and add 2^{-i} to the original weight of the i-th edge of G for $i = 1, \ldots, m$. We assume such a perturbation has been performed and propose to consider the family of those oriented circuits which project onto cycles in the Horton family once the directions of the arcs are ignored.

Our first preparatory lemma concerns the underlying undirected graph G_D of the digraph D and the Horton family for G_D.

Lemma 4 *Let $\mathcal{H}$ be the Horton family of a weighted graph (G, w) in which there exists a unique minimum path between any pair of nodes. Let C be a cycle of G not in $\mathcal{H}$. Then there exists a minimum path $P_{u,v}$ between u and v, internally disjoint from C, and with u and v nodes of C.*

Proof. Let ab be an edge of maximum weight in C. Then there exists a node y of C such that the two unique paths $C_{y,a}$ and $C_{y,b}$ in C, which do not contain ab, are minimum paths in C between y and a, or b respectively.

Now, since $C \notin \mathcal{H}$, and w.l.o.g., we can assume that $C_{y,a}$ is not the minimum path between y and a in G. Let $P_{y,a}$ be the minimum path between y and a in G. Let P be a subpath of $P_{y,a}$ which is internally disjoint from C and with endpoints on C. Being a subpath of a minimum path, P is a minimum path between its endpoints. □

Lemma 5 *All oriented cycles which belong to some minimum cycle basis of the directed graph D are in the Horton family for G_D, once directions are disregarded.*

Proof. Let $\mathcal{B}^\circ = C_1^\circ, \ldots, C_t^\circ, \ldots, C_\mu^\circ$ be some minimum cycle basis of a given directed graph D. We may assume it to be obtained by applying the greedy algorithm to the set of all cycles of D. Assume that C_t° is the first cycle not contained in the Horton family, which the greedy algorithm selects.

By Lemma 4, there are two nodes u and v in C_t°, such that the shortest path $P_{u,v}$ between u and v is internally disjoint from C_t°. Hence, for the two cycles C_1 and C_2 in $C_t^\circ \cup P_{u,v}$ different from C_t°, we have $w(C_1) < w(C_t^\circ)$ and $w(C_2) < w(C_t^\circ)$. We choose their orientations as to disagree on $P_{u,v}$.

The greedy algorithm ensures that both C_1 and C_2 can be generated from $\{C_1^\circ, \ldots, C_{t-1}^\circ\}$ – otherwise they would have been added instead of C_t°. But then, since $C_t^\circ = C_1 + C_2$, C_t° can also be generated from $\{C_1^\circ, \ldots, C_{t-1}^\circ\}$, which contradicts the fact that C_t° was chosen. □

Hence, we propose two algorithms for computing a minimum cycle basis of a digraph. The first one is extremely simple and follows the ideas of Horton [8]:

1. Compute the Horton family $\mathcal{H}$;
2. Sort the elements of $\mathcal{H}$ according to their weights;
3. Perform the greedy algorithm to extract a minimum cycle basis out of $\mathcal{H}$.

Notice that Step 3 dominates the total runtime of this procedure. This remains true, even if we perform successive Gaussian elimination.

In more detail, assume the cycles $C_1, \ldots, C_t$ to be already selected by the greedy algorithm. Consider the matrix Γ_t, having their incidence vectors γ_i as columns. For testing linear independence, w.l.o.g. we may omit $n-1$ rows which correspond to the arcs of some spanning tree and hereby obtain Γ_t'. Let

$$R_t \Gamma_t' = \begin{bmatrix} U_t \\ 0 \end{bmatrix},$$

where U_t denotes a t dimensional regular upper triangular matrix.

The greedy algorithm has to decide for the next cycle C with incidence vector γ, whether $\{C_1, \ldots, C_t, C\}$ are independent. But this is equivalent to the property that $R_t \gamma'$ has a non-zero entry in one of the rows $t+1, \ldots, \mu$. The computation of $R_t \gamma'$ involves $O(t\mu)$ arithmetic operations and is performed once for every C in the Horton family $\mathcal{H}$. Since $|\mathcal{H}| \le mn$, testing independence requires $O(\mu^2 mn)$ arithmetic operations over the whole execution.

In case $\{C_1, \ldots, C_t, C\}$ are linearly independent, we must also provide R_{t+1} in order to continue the above procedure. We have in fact $R_{t+1} = F_{t+1} R_t$, where the regular matrix F_{t+1} encodes the operations necessary to obtain zero entries in the last $(\mu - (t+1))$ columns of the vector $R_t \gamma'$. The computation of F_{t+1} involves $O(\mu - t)$ operations. Apart from an occasional row exchange, computing R_{t+1} touches at most two values per column of R_t, and hence costs $O(t\mu)$ arithmetic operations. Notice that these two types of operations appear no more than μ times.

Since every arithmetic operation costs $\tilde{O}(m)$, we already obtain an overall runtime of only $\tilde{O}(m^4 n)$, being asymptotically as fast as the more technical algorithm proposed in [9].

But we may even reduce this runtime down to $\tilde{O}(m^{\omega+1} n)$. For our second algorithm, we may follow the lines of the divide-and-conquer approach of Golynski and Horton [7]:

1. Compute the Horton family $\mathcal{H}$;
2. Encode the linear matroid with ground set $\mathcal{H} \subseteq \mathcal{C}_D$ by standard matrix representation;
3. Recursively decide for the two halves of the non-basic elements, which of its elements belong to a minimum basis.

Still, Step 3 dominates the total runtime. By analyzing the proposed recursion, which we shortly postpone until the next section, one can bound the number of arithmetic operations by $O(m^\omega n)$. Considering the coding length of numbers, we obtain the following corollary.

Corollary 6 *A minimum cycle basis of a digraph can be computed in* $\tilde{O}(m^{\omega+1}n)$.

Complexity of Golynski and Horton's Algorithm [7]

As a detailed analysis of the complexity of Golynski and Horton's algorithm for computing a minimum cycle basis of a linear matroid is not publicly available in the literature so far, we provide one in this report. To that end, we follow closely the notation of the original paper.

We start by recapitulating in detail the definition of the standard matrix representation of a linear matroid $M = (E', \mathcal{I})$ subject to a fixed basis B of M. Consider the matrix

$$C := [e]_{e \in E'}$$

in which the elements of the ground set E' are immediately given as vectors. We set $n' := \mathrm{rank}(M)$ and $m' := |E'| - n'$. Let $B = \{e_1, \ldots, e_{n'}\}$ be some basis of M. Finally, denote by C_B the submatrix of C corresponding to B. Then, the standard matrix representation A' of M with respect to B is defined by

$$C_B^{-1} C = [I_{n'} | A'].$$

Notice that the identity matrix $I_{n'}$ enables us to relate the basic elements of M to the rows of A', and the current cobase elements to its columns.

Before we are able to analyze the complexity of the algorithm of Golynski and Horton [7], we are going to present that recursive algorithm.

The input is a linear matroid M given in standard matrix representation A' with respect to a basis B of M. The output is a minimum basis B° of M. Start with an arbitrary basis B of M. Let X denote the set of the indices of the current basic elements, and let $Y = \{1, \ldots, m'\} \setminus X$ denote the indices of the current cobasic elements. Split Y into even parts Y_1 and Y_2. Use recursion on Y_1 to obtain a minimum basis B_1 of the matroid $M \setminus Y_2$. Compute the standard matrix representation $\overline{A'}$ of M with respect to B_1. Finally, use recursion to compute a minimum basis B° of M, i.e. to decide which elements of Y_2 must replace elements of B_1.

A key issue of this algorithm is to provide the standard matrix representation of M with respect to B_1 as input for the second recursion. Assume the cobasic elements $Y_{11} \subseteq Y_1$ to replace the basic elements $X_1 \subseteq X$ and consider four submatrices of A'

$$A' = \begin{array}{c||c|c||} & Y_{11} & Y_{12} \cup Y_2 \\ \hline X_1 & F & G \\ \hline X_2 & H & J \\ \hline \end{array}$$

Then, the standard matrix representation $\overline{A}$ of M with respect to $B_1 = B \cup Y_{11} \setminus X_1$ can be obtained by the following *group pivot*

$$\overline{A'} = \begin{array}{c||c|c||} & X_1 & Y_{12} \cup Y_2 \\ \hline Y_{11} & \overline{F} & \overline{G} \\ \hline X_2 & \overline{H} & \overline{J} \\ \hline \end{array} = \begin{array}{c||c|c||} & X_1 & Y_{12} \cup Y_2 \\ \hline Y_{11} & F^{-1} & F^{-1}G \\ \hline X_2 & -HF^{-1} & J - HF^{-1}G \\ \hline \end{array}.$$

This becomes most clear, when re-introducing columns for the current basic elements temporarily:

$$\begin{pmatrix} F^{-1} & 0 \\ -HF^{-1} & I_{|X_2|} \end{pmatrix} \cdot \begin{array}{||c|c|c|c||} X_1 & X_2 & Y_{11} & Y_{12} \cup Y_2 \\ \hline I_{|X_1|} & 0 & F & G \\ \hline 0 & I_{|X_2|} & H & J \\ \hline \end{array} =$$

$$\begin{array}{||c|c|c|c||} X_1 & X_2 & Y_{11} & Y_{12} \cup Y_2 \\ \hline F^{-1} & 0 & I_{|X_1|} & F^{-1}G \\ \hline -HF^{-1} & I_{|X_2|} & 0 & J - HF^{-1}G \\ \hline \end{array}.$$

Let $f(a, b, c)$ denote the number of arithmetic operations required to multiply an $a \times b$ matrix with a $b \times c$ matrix. If we assume b to be minimum among a, b, c, and if we assume both $\frac{a}{b}$ and $\frac{c}{b}$ to be integer, then we have

$$f(a, b, c) = \frac{ac}{b^2} O(b^\omega), \tag{1}$$

which can be seen easily by partitioning the two input matrices into $b \times b$ submatrices.

Depending on $k := |X_1|$, computing the four submatrices of $\overline{A'}$ involves

$$\begin{array}{c||c|c||}
 & X_1 & Y_{12} \cup Y_2 \\
\hline\hline
Y_{11} & O(k^\omega) & f(k,k,m'-k) \\
\hline
X_2 & f(n',k,k) & f(n',k,m'-k) \\
\hline\hline
\end{array} \tag{2}$$

arithmetic operations, where we compute $\overline{J}$ as $J - H\overline{G}$. We know that $k \le n'$ and $k \le (m'+1) - k$, thus in particular $k \le m'$. Notice that if one is not interested in the standard matrix representation of M with respect to B°, then neither $\overline{H}$ nor the entries of the columns of Y_{12} have to be computed at any node of the recursion.

Under asymptotic notation, plugging (1) into (2) yields a total complexity $T(k, m', n')$ for the group pivot of

$$\begin{aligned}
T(k,m',n') &= O(k^\omega) + O((m'-k)k^{\omega-1}) + O(n'k^{\omega-1}) + O((m'-k)n'k^{\omega-2}) \\
&= O(m'(k+n')k^{\omega-2}) \\
&= O(m'n' \min\{m',n'\}^{\omega-2}).
\end{aligned}$$

In total, the number of arithmetic operations required for computing a minimum cycle basis of a linear matroid with m' elements and rank n' is bounded by the following recursive function

$$\begin{aligned}
T(m') &= 2T(\lceil \tfrac{m'}{2} \rceil) + O(m'n' \min\{m',n'\}^{\omega-2}), \\
T(1) &= O(n').
\end{aligned}$$

Hence, in total we obtain

$$T(m') = O(m'n' \min\{m',n'\}^{\omega-2}). \tag{3}$$

In order to apply (3) to the MCB problem of a directed graph, observe that we omit the $n-1$ rows of C, which correspond to some spanning tree. Further notice that we may restrict the ground set of M to the Horton family $\mathcal{H}$ in order to compute an MCB of a given directed graph. With these settings, we have the following correspondences

$$\mu \;\rightarrow\; n' \qquad \text{and} \qquad mn \;\rightarrow\; m'.$$

Thus, the number of arithmetic operations required to compute an MCB of a directed graph is $O(m^\omega n)$.

4 Is the Greedy Algorithm Suited for Other Classes of Cycle Bases?

There are four important subclasses of cycle bases of directed graphs, where each one is a subset of its predecessors ([11]):

(1) Cycle bases projecting onto bases of the underlying undirected graph;
(2) Integral cycle bases; [12]
(3) Generalized fundamental cycle bases; [14]
(4) (Strictly) Fundamental cycle bases. [14]

Since cycle bases of both, directed and undirected graphs form a matroid, the greedy algorithm provides a simple polynomial time algorithm for finding such minimum cycle bases. We may ask, whether this approach can also be used for more specialized classes of cycle bases. But "unfortunately, integral cycle bases do not form a matroid" [13].

A cycle basis of a directed graph D is called *integral*, if every cycle in D can be expressed as an integer linear combination of the basic cycles. Equivalently, the regular $\mu \times \mu$ submatrices of its cycle matrix, i.e. its arc-cycle incidence matrix, have absolute value one ([11]). Integral cycle bases play an important role in cyclic railway timetabling ([12]). A cycle basis $\{C_1, \ldots, C_\mu\}$ is called *generalized fundamental*, if there is a permutation σ, such that

$$C_{\sigma(i)} \setminus \{C_{\sigma(1)}, \ldots, C_{\sigma(i-1)}\} \neq \emptyset, \ \forall i = 2, \ldots, \mu.$$

A direct way to define an independence system related to integral cycle bases is to consider the set of oriented cycles of a directed graph as the ground set E, and the subsets of integral cycle bases as the set of independent sets $\mathcal{I}$.

Proposition 7 *The independence system* $(E, \mathcal{I})$ *is not a matroid.*

Proof. We provide two integral cycle bases with cycle matrices Γ' and Γ, such that we can select one cycle to leave Γ', but none of the cycles of Γ can complete this $m \times (\mu - 1)$ matrix to another integral cycle basis.

Consider the directed "envelope graph" shown in Figure 2. The four oriented

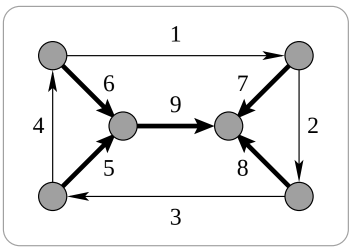

Fig. 2. The envelope graph.

cycles, whose incidence vectors form the rows of Γ'^T, cf. Equation (4), form an integral cycle basis, because the regular 4×4 submatrices of Γ' have determinants of absolute value one

$$\Gamma' := \left(\begin{array}{cccc|ccccc} 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & -1 \\ 1 & 0 & 1 & 0 & 1 & -1 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 & 1 & 0 & -1 \end{array}\right)^T . \tag{4}$$

Another integral cycle basis is obtained by the cycles whose incidence vectors compose the cycle matrix

$$\Gamma := \left(\begin{array}{cccc|ccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & -1 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 & 1 & 0 & -1 \end{array}\right)^T .$$

Since Γ' is an integral cycle basis, the integer matrix

$$U = \begin{pmatrix} -2 & -3 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix},$$

is the unique solution to the system $\Gamma' U = \Gamma$. Since the det function is distributive, U must be an unimodular matrix, in fact $\det U = 1$.

Now, choose the first column of Γ' to exit that basis. Of course, neither the third nor the fourth column of Γ can become its substitute, because they already appear in Γ'. But putting the first or the second column of Γ to Γ' results in a matrix having a determinant of absolute value different from one, which is induced by the entry in the first row of column one or two, resp., of matrix U. Thus, the first column of Γ' cannot be replaced by any of the columns of Γ, providing that integral cycle bases do not form a matroid. □

Corollary 8 *For $\mathcal{I}$ being family set of subsets of generalized fundamental cycle bases, the independence system $(E, \mathcal{I})$ is not a matroid, either.*

Proof. The two cycle bases, which we consider in the proof of Proposition 7 are in fact generalized fundamental cycle bases. Replacing the first column of Γ' with the first or the second column of Γ results in a directed cycle basis, which hits every arc at least twice. □

Remark 9 *Finding a minimum (strictly) fundamental cycle basis is MAX-SNP hard ([4]).*

5 Conclusions

We investigated the problem of computing a minimum cycle basis of a directed graph. We provided a digraph no minimum cycle basis of which projects onto a cycle basis of the underlying undirected graph. Hence Horton's original algorithm can not be employed as a black-box to solve the above problem. We also showed however that Horton's general approach can be adapted as to work with the directed case as well. This leads to a very simple algorithm, which is asymptotically as fast as the one previously known, as well as to an even faster one. Due to this relationship, the directed case may profit from further improvements obtained for the undirected case. Finally, we showed that this approach cannot be applied to more specialized classes of cycle bases of directed graphs.

References

[1] Berger, F., Gritzmann, P, and de Vries, S. (2004) Minimum Cycle Bases for Network Graphs. Algorithmica **40** (1), 51–62

[2] Bollobás, B. (1998) Modern Graph Theory. Springer, GTM 184

[3] de Pina, J.C. (1995) Applications of Shortest Path Methods. Ph.D. Thesis, University of Amsterdam, The Netherlands

[4] Galbiati, G. and Amaldi, E. (2003) On the Approximability of the Minimum Fundamental Cycle Basis Problem. In: WAOA 2003, Springer LNCS 2909, edited by Klaus Jansen and Roberto Solis-Oba

[5] Gleiss, P.M. (2001) Short cycles: minimum cycle bases of graphs from chemistry and biochemistry. Ph.D. Thesis, Universität Wien

[6] Golynski, A. (2002) A polynomial time algorithm to find the minimum cycle basis of a regular matroid. Master's thesis, University of New Brunswick

[7] Golynski, A., Horton, J.D. (2002) A Polynomial Time Algorithm to Find the Minimum Cycle Basis of a Regular Matroid. In: SWAT 2002, Springer LNCS 2368, edited by M. Penttonen and E. Meineche Schmidt

[8] Horton, J.D. (1987) A polynomial-time algorithm to find the shortest cycle basis of a graph. SIAM Journal on Computing **16**, 358–366

[9] Kavitha, T. and Mehlhorn, K. (2004) A polynomial time algorithm for minimum cycle basis in directed graphs. Kurt Mehlhorn's List of Publications, 185, MPI Saarbrücken,
`http://www.mpi-sb.mpg.de/~mehlhorn/ftp/DirCycleBasis.ps`,
to appear in Proceedings of STACS 2005

[10] Kavitha, T., Mehlhorn, K., Michail, D., and Paluch, K. (2004) A faster algorithm for Minimum Cycle Basis of graphs. In: ICALP 2004, Springer LNCS 3142, edited by J. Diaz et al., 846–857

[11] Liebchen, C. (2003) Finding Short Integral Cycle Bases for Cyclic Timetabling. In: Algorithms 2003, Springer LNCS 2832, edited by G. di Battista and U. Zwick, 715–726

[12] Liebchen, C., Peeters, L. (2002) On Cyclic Timetabling and Cycles in Graphs. Technical Report 761/2002, TU Berlin

[13] Schrijver, A. (2002) Personal Communication

[14] Whitney, H. (1935) On the Abstract Properties of Linear Dependence. American Journal of Mathematics **57**, 509–533

Reports from the group

"Combinatorial Optimization and Graph Algorithms"

of the Department of Mathematics, TU Berlin

2004/31 *Christian Liebchen and Romeo Rizzi:* A Greedy Approach to Compute a Minimum Cycle Bases of a Directed Graph

2004/235 *Laura Heinrich-Litan and Marco E. Lübbecke:* Rectangle Covers Revisited Computationally

2004/27 *Ekkehard Köhler and Rolf H. Möhring and Gregor Wünsch:* Minimizing Total Delay in Fixed-Time Controlled Traffic Networks

2004/26 *Rolf H. Möhring and Ekkehard Köhler and Ewgenij Gawrilow and Björn Stenzel:* Conflict-free Real-time AGV Routing

2004/21 *Christian Liebchen and Mark Proksch and Frank H. Wagner:* Performance of Algorithms for Periodic Timetable Optimization

2004/20 *Christian Liebchen and Rolf H. Möhring:* The Modeling Power of the Periodic Event Scheduling Problem: Railway Timetables — and Beyond

2004/19 *Ronald Koch and Ines Spenke:* Complexity and Approximability of k-splittable flow problems

2004/18 *Nicole Megow, Marc Uetz, and Tjark Vredeveld:* Stochastic Online Scheduling on Parallel Machines

2004/09 *Marco E. Lübbecke and Uwe T. Zimmermann:* Shunting Minimal Rail Car Allocation

2004/08 *Marco E. Lübbecke and Jacques Desrosiers:* Selected Topics in Column Generation

2003/050 *Berit Johannes:* On the Complexity of Scheduling Unit-Time Jobs with OR-Precedence Constraints

2003/49 *Christian Liebchen and Rolf H. Möhring:* Information on MIPLIB's timetab-instances

2003/48 *Jacques Desrosiers and Marco E. Lübbecke:* A Primer in Column Generation

2003/47 *Thomas Erlebach, Vanessa Kääb, and Rolf H. Möhring:* Scheduling AND/OR-Networks on Identical Parallel Machines

2003/43 *Michael R. Bussieck, Thomas Lindner, and Marco E. Lübbecke:* A Fast Algorithm for Near Cost Optimal Line Plans

2003/42 *Marco E. Lübbecke:* Dual Variable Based Fathoming in Dynamic Programs for Column Generation

2003/37 *Sándor P. Fekete, Marco E. Lübbecke, and Henk Meijer:* Minimizing the Stabbing Number of Matchings, Trees, and Triangulations

2003/25 *Daniel Villeneuve, Jacques Desrosiers, Marco E. Lübbecke, and François Soumis:* On Compact Formulations for Integer Programs Solved by Column Generation

2003/24 *Alex Hall, Katharina Langkau, and Martin Skutella:* An FPTAS for Quickest Multicommodity Flows with Inflow-Dependent Transit Times

2003/23 *Sven O. Krumke, Nicole Megow, and Tjark Vredeveld:* How to Whack Moles

2003/22 *Nicole Megow and Andreas S. Schulz:* Scheduling to Minimize Average Completion Time Revisited: Deterministic On-Line Algorithms

2003/16 *Christian Liebchen:* Symmetry for Periodic Railway Timetables

2003/12 *Christian Liebchen:* Finding Short Integral Cycle Bases for Cyclic Timetabling

762/2002 *Ekkehard Köhler and Katharina Langkau and Martin Skutella:* Time-Expanded Graphs for Flow-Dependent Transit Times

761/2002 *Christian Liebchen and Leon Peeters:* On Cyclic Timetabling and Cycles in Graphs

752/2002 *Ekkehard Köhler and Rolf H. Möhring and Martin Skutella:* Traffic Networks and Flows Over Time

739/2002 *Georg Baier and Ekkehard Köhler and Martin Skutella:* On the k-splittable Flow Problem

736/2002 *Christian Liebchen and Rolf H. Möhring:* A Case Study in Periodic Timetabling

Reports may be requested from:

Sekretariat MA 6–1
Fakultt II – Institut fr Mathematik
TU Berlin
Straße des 17. Juni 136
D-10623 Berlin – Germany

e-mail: klink@math.TU-Berlin.DE

Reports are also available in various formats from

`http://www.math.tu-berlin.de/coga/publications/techreports/`

and via anonymous ftp as

`ftp://ftp.math.tu-berlin.de/pub/Preprints/combi/Report-`*number*`-`*year*`.ps`