COMPUTER SCIENCE REVIEW 3 (2009) 199–243

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cosrev

Survey

Cycle bases in graphs characterization, algorithms, complexity, and applications

Telikepalli Kavitha[a], *Christian Liebchen*[b], *Kurt Mehlhorn*[c,*], *Dimitrios Michail*[d], *Romeo Rizzi*[e], *Torsten Ueckerdt*[f], *Katharina A. Zweig*[g]

[a] *Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India*
[b] *DB Schenker Rail Deutschland AG, Mainz, Germany*
[c] *Max-Planck-Institut für Informatik, Saarbrücken, Germany*
[d] *Department of Informatics and Telematics, Harokopio University of Athens 89, Harokopou Street 17671, Kallithea, Athens, Greece*
[e] *Dipartimento di Matematica ed Informatica (DIMI), Università degli Studi di Udine, Udine, Italy*
[f] *Technische Universität Berlin, Berlin, Germany*
[g] *Department of Biological Physics, Eötvös Lorand University, Budapest, Hungary*

ARTICLE INFO

Article history:
Received 3 April 2009
Received in revised form
25 August 2009
Accepted 25 August 2009

ABSTRACT

Cycles in graphs play an important role in many applications, e.g., analysis of electrical networks, analysis of chemical and biological pathways, periodic scheduling, and graph drawing. From a mathematical point of view, cycles in graphs have a rich structure. Cycle bases are a compact description of the set of all cycles of a graph. In this paper, we survey the state of knowledge on cycle bases and also derive some new results. We introduce different kinds of cycle bases, characterize them in terms of their cycle matrix, and prove structural results and a priori length bounds. We provide polynomial algorithms for the minimum cycle basis problem for some of the classes and prove $\mathcal{APX}$-hardness for others. We also discuss three applications and show that they require different kinds of cycle bases.

1. Introduction

Cycles in graphs play an important role in many applications, e.g., analysis of electrical networks, analysis of chemical and biological pathways, periodic scheduling, and graph drawing. From a mathematical point of view, cycles in graphs have a rich structure. Cycle bases are a compact description of the set of all cycles of a graph and cycle bases consisting of short cycles or, in weighted graphs, of small weight cycles are interesting both mathematically and from an application viewpoint. In the applications above, sparse descriptions are to be preferred.

The study of cycle bases dates back to the early days of graph theory; MacLane [1] gave a characterization of planar graphs in terms of cycle bases. Within the last ten years, many new results on cycle bases have been published, most notably a classification of different kinds of cycle bases, structural results, a priori bounds on the length and weight of minimum cycle bases, polynomial time algorithms for constructing exact or approximate minimum cycle bases for some kinds, and hardness results for other kinds of minimum cycle bases.

In this paper, we survey these results and also provide some new ones. Fig. 1 shows the landscape of cycle bases. We

* *Corresponding author.*
E-mail address: mehlhorn@mpi-inf.mpg.de (K. Mehlhorn).

1574-0137/$ - see front matter
doi:10.1016/j.cosrev.2009.08.001

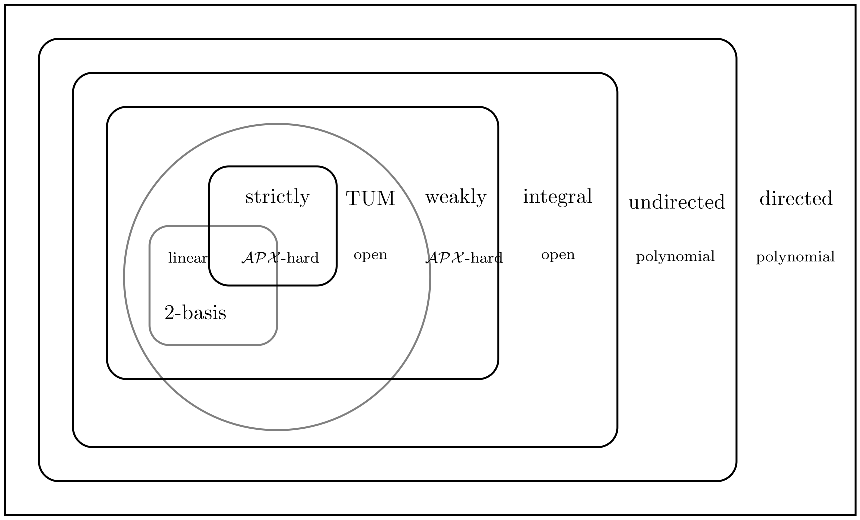

Fig. 1 – The inclusion diagram of cycle bases and the complexity status of their minimum weight cycle basis problems.

will review the different kinds of cycle bases in Sections 2 and 3: directed, undirected, integral, weakly fundamental, totally unimodular, and strictly fundamental bases, and 2-bases. In Section 3, we characterize the different kinds in terms of properties of their cycle matrices. For example, undirected cycle bases are characterized by the fact that the determinant of their cycle matrix is odd and integral cycle bases are characterized by the fact that their determinant is ± 1. We will establish the inclusion map and show that different classes lead to different minimum cycle basis problems. We will also establish many structural results.

Section 4 deals with a priori length and weight bounds on minimum cycle bases. We will prove results of the following kind: every graph of n nodes and m edges has a weakly fundamental cycle basis of length $O(m \log m / \log(m/n))$. We will also show that there are graphs for which every basis has length $\Omega(m \log m / \log(m/n))$.

In Section 5, we will give polynomial time algorithms for constructing minimum weight directed, undirected and planar cycle bases. We will also discuss approximation algorithms.

Section 6 treats hardness results; in particular, $\mathcal{APX}$-hardness of the minimum cycle basis problem for weakly fundamental and strictly fundamental bases. Fig. 1 summarizes the complexity results. For two classes the complexity is open:

Open Problem 1. Resolve the complexity status of computing minimum weight integral and minimum weight totally unimodular bases.

Finally, Section 7 discusses three applications of cycle bases; we will see that they require different kinds of cycle bases. The analysis of electrical circuits does not require any particular kind of cycle basis, whereas periodic scheduling requires integral cycle bases, and graph drawing needs strictly fundamental bases.

The paper mostly surveys known results, but it also contains several new ones. In particular, we give additional structural and characterization results, we obtain tight length bounds for weakly fundamental cycle bases for the full spectrum of graph densities, we give a simplified algorithmic treatment of directed cycle bases, and we present the first algorithms for minimum cycle bases in the presence of negative edges. In each section, we also state open problems.

This survey is targeted at mathematicians and computer scientists. We give complete proofs for most results to make the survey self-contained. We wrote the survey because this area has developed quickly in the past decade and is still a rich source of open problems.

2. Definitions

An *(undirected) graph* is a pair $G = (V, E)$, where V is a finite set, and E is a family of unordered pairs of elements of V. The elements of V are called *vertices* or *nodes* and the elements of E are called *edges*. An edge $e = \{v, w\}$ is *incident to* the vertices v and w; v and w are the endpoints of e. The same pair $\{v, w\}$ may occur several times in E; we refer to a pair occurring more than once as a *multiple edge*. Graphs without multiple edges are called *simple*. An edge of the form $\{v, v\}$ is called a *loop*. The *degree* $deg(v)$ of a vertex v is the number of times v occurs as an endpoint of an edge. Observe that a loop $\{v, v\}$ contributes two to the degree of v. We use $\delta(v)$ to denote the set of edges incident to v; a loop $\{v, v\}$ appears twice in $\delta(v)$.

A *(directed) graph* is a pair $D = G = (V, A)$, where V is a finite set, and A is a family of ordered pairs of elements of V. The elements of V are called the vertices or nodes of G, and the elements of A are called the *(directed) edges* or *arcs* of G. We use $G = (V, E)$ to denote directed and undirected graphs and $D = (V, A)$ to denote directed graphs. The vertices v and w are called the *tail* and *head* of the arc $e = (v, w)$, respectively; e is said to leave v and to enter w; it is incident to v and w. The notions *multiple edge*, *simple graph*, and *loop* are defined analogously as for undirected graphs. The *outdegree* $outdeg(v)$ and *indegree* $indeg(v)$ of a vertex v are the number of times v occurs as the tail and head, respectively, of an edge. Observe that a loop (v, v) contributes one to both the indegree and the outdegree of v. We use $\delta^+(v)$ and $\delta^-(v)$ for the edges leaving and entering v, respectively.

We use n and m to denote the number of nodes and edges or arcs, respectively, i.e., $n = |V|$ and $m = |E|$ or $m = |A|$. We use the notation $e = vw$ to denote both directed and undirected edges, i.e., the notation stands for the directed edge (v, w) and the undirected edge $\{v, w\}$. Every directed graph D can be turned into an undirected graph $G(D)$ by ignoring the orientation of the edges and every undirected graph G can be turned into a directed graph by orienting the edges arbitrarily; we call D an *orientation* of G. In this way, we can view every graph as directed.

A *subgraph* $G' = (V', E')$ of G is a graph with $V' \subseteq V$ and $E' \subseteq E$. If V' is a subset of V, $G - V'$ denotes the graph obtained by removing all vertices in V' and their incident edges from G. A *path* P from v to w in G is a subgraph of G with $V' = \{v_0 = v, v_1, \ldots, v_k = w\}$ with $v_i \neq v_j$ and $E' = \{x_0x_1, x_1x_2, \ldots, x_{k-1}x_k\}$. We write $P(v, w)$ if we want to emphasize that P is a path from v to w. The *length of a path* is the number of its edges. An undirected graph is *connected* if there exists a path from any vertex to every other vertex. A vertex v in a connected graph G is called an *articulation point*, or *cut vertex*, if $G - v$ is disconnected. An undirected graph is *biconnected* if it has no articulation point. A directed graph is connected if the underlying undirected graph is connected. Any maximal connected subgraph of G is called a *connected component*. A graph T is a *tree* if it is connected and has $n - 1$ edges. A subgraph G' of a connected graph G is called a *spanning tree* if it constitutes a tree on all vertices in G. If G is not connected, any union of spanning trees for each connected component is called a *spanning forest*.

A *cycle in an undirected graph* is a subgraph in which every vertex has even degree. A cycle is a *circuit* if it is connected and every one of its vertices has degree two. If $C_1, \ldots, C_k$ are cycles, $C_1 + \cdots + C_k$ consists of all edges that are contained in an odd number of C_i's; the sum is again a cycle. An *undirected cycle basis* is a minimal set of circuits such that any cycle can be written as a sum of the circuits in the basis.

We next generalize the notion of an undirected cycle basis. Let κ be a field. A *κ-cycle* C in a directed graph D is a vector in κ^A such that for any vertex v we have

$$\sum_{e\in\delta^+(v)} C_e = \sum_{e\in\delta^-(v)} C_e;$$

here C_e denotes the component of C indexed by e. Instead of C_e, we will also sometimes write $C(e)$. We prefer the latter notation when $C = C_i$ belongs to an indexed family of cycles. In other contexts, cycles are sometimes referred to as *circulations* and the constraint $\sum_{e\in\delta^+(v)} C_e = \sum_{e\in\delta^-(v)} C_e$ is called *flow conservation*. The set

$$\mathcal{C}_\kappa(D) = \{C \mid C \text{ is a } \kappa\text{-cycle of } G\}$$

forms a vector space over κ, the *κ-cycle space* of G; if C_1 and C_2 are cycles and $\lambda \in \kappa$ is a constant, we have

$$(C_1 + C_2)(e) = C_1(e) + C_2(e) \quad \text{and} \quad (\lambda C)(e) = \lambda C(e)$$

for all edges e. The support of a cycle is the set of edges e with $C_e \neq 0$. A cycle is *simple* if $C_e \in \{-1, 0, +1\}$ for all e, and a simple cycle is a *circuit* if its support is connected and for any v there are at most two edges in the support incident to v. A *κ-cycle basis* is a set of circuits forming a basis of the cycle space. Any cycle basis consists of $\nu := m - n + 1$ circuits (see Theorem 2.3). If D and D' are orientations of the same undirected graph G, their cycle spaces $\mathcal{C}_\kappa(D)$ and $\mathcal{C}_\kappa(D')$ are isomorphic. Indeed, if $C \in \kappa^A$ is a cycle in D, the corresponding cycle in D' is obtained by reversing the sign of those components C_e, where e is oriented differently in D and D'. We conclude that the vector space $\mathcal{C}_\kappa(D)$ does not depend on the orientation D; it is uniquely defined by the underlying undirected graph G. Hence, we may also write $\mathcal{C}_\kappa(G)$.

Particularly interesting are the cases $\kappa = \mathbb{Z}_2 = GF(2)$, the field of two elements, and $\kappa = \mathbb{Q}$, the field of rationals. In these cases, the cycle space and cycle basis are referred to as *undirected* or *directed cycle space* and *basis*, respectively.

In $\mathbb{Z}_2$, $-1 = +1$ and $+1$ is the only non-zero element in the field. Thus a $\mathbb{Z}_2$-cycle or simply, a cycle, is a vector $C \in \mathbb{Z}_2^E$ such that $\sum_{e\in\delta(v)} C_e = 0$ for any vertex v. A cycle may alternatively be viewed as a set of edges; e belongs to C iff $C_e = 1$. We use C to denote the vector in $\mathbb{Z}_2^E$, the corresponding subset of E, and also the subgraph (V', C), where V' is the set of vertices having at least one edge in E incident to it. A cycle is an *even* or *Eulerian* subgraph, i.e., every vertex has even degree in C. Conversely, any even subgraph is a cycle.

A $\mathbb{Q}$-cycle C has components in $\mathbb{Q}$; we call it a *directed cycle* if all components of C are integral. Directed cycles may use arcs in *forward* ($C_e > 0$) or *backward* ($C_e < 0$) direction. If any arc is replaced by C_e copies of itself and, in addition, the direction of all arcs e with $C_e < 0$ is reversed, then we end up with a digraph in which the indegree of every vertex is equal to its outdegree.

Let D be a directed graph and let $G = G(D)$ be the underlying undirected graph. For any directed cycle C of D, let $\pi(C) := (C_e \bmod 2)_{e\in E}$. Then $\pi(C)$ is an undirected cycle in G. We call $\pi(C)$ the *projection* of C.

Fig. 2 illustrates these definitions. In addition, it provides a first example showing that directed cycle bases do not necessarily project into undirected cycle bases. However, a set of dependent cycles projects into a set of dependent cycles. Indeed, let C_i, $i \in I$, be a family of dependent directed cycles. Then $\sum_{i\in I} \lambda_i C_i = \mathbf{0}$, with $\lambda_i \in \mathbb{Q}$ not all zero. Here $\mathbf{0}$ denotes the zero-vector in $\mathbb{Q}^E$. We may assume $\lambda_i \in \mathbb{Z}$ not all even. Then $\sum_{i\in I}(\lambda_i \bmod 2)\ \pi(C_i) = \mathbf{0} \bmod 2$ and at least one coefficient $\lambda_i \bmod 2$ is non-zero. Thus the $\pi(C_i)$, $i \in I$, are dependent.

We use $+$ and Σ to denote addition in $\mathbb{Q}$ and in $GF(2)$ (and also in $GF(p)$ for prime p). The distinction will usually be clear from the context. If both fields occur in the same argument, as in the paragraph above, we will emphasize the difference by the additional operator " mod 2".

We may also lift undirected cycles from an undirected graph G to an orientation D of G. Let C' be any undirected cycle in G. We call $C \in \{-1, 0, +1\}^A$ a lifting of C' if C projects to C'. For a circuit C' the lifting is unique up to the sign. Clearly, if C' lifts to C then C projects to C'. Algorithmically, we may lift as follows: assume C' to be connected (components are lifted independently) and consisting of k edges. Since an undirected cycle is a Eulerian subgraph of G, there is a closed traversal $(e_0, \ldots, e_{k-1})$ of the edges of C', i.e., $e_i = \{v_i, v_{i+1}\}$ for $0 \le i < k$ and $v_0 = v_k$. This traversal defines a simple cycle C in D; we have $C_e = 0$ if C' does not contain e and $C_e = +1$ (-1) if the traversal uses e in forward (backward) direction.

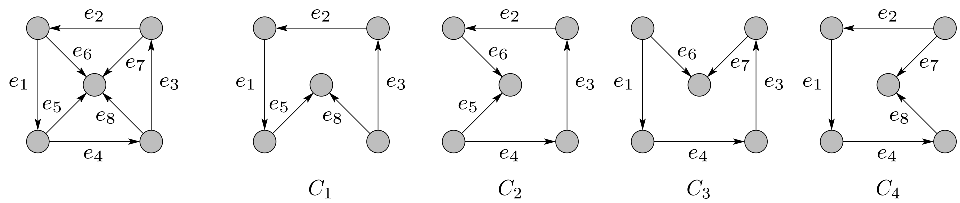

Fig. 2 – An orientation D of the undirected wheel graph W_5, and four circuits C_1 to C_4 in D. The edges of D are numbered from e_1 to e_8. The circuit C_1 uses the edges e_1, e_2, e_3, and e_5 in forward direction and the edge e_8 in backward direction. Thus $C_1 = (1, 1, 1, 0, 1, 0, 0, -1)$. The cycles C_1 to C_4 form a directed cycle basis of D. The cycle C consisting of edges 1 to 4 is represented as: $C = (1, 1, 1, 1, 0, 0, 0, 0) = (C_1 + C_2 + C_3 + C_4)/3$. Let G be the underlying undirected graph, let $\pi(C_i)$ be the undirected cycle corresponding to C_i, and let $\pi(C)$ be the undirected cycle corresponding to C. Then $\pi(C_1) = (1, 1, 1, 0, 1, 0, 0, 1)$ and $\pi(C) = \pi(C_1) \oplus \pi(C_2) \oplus \pi(C_3) \oplus \pi(C_4)$. The circuits $\pi(C_1)$ to $\pi(C_4)$ form an undirected cycle basis of G. The set $\{C_1, C_2, C_3, 2C_4\}$ is also a directed cycle basis of D. However, $\pi(2C_4) = 0$ and hence $\{\pi(C_1), \pi(C_2), \pi(C_3), \pi(2C_4)\}$ is *not* an undirected cycle basis of G. There are less trivial reasons for a directed cycle basis not projecting into an undirected cycle basis.

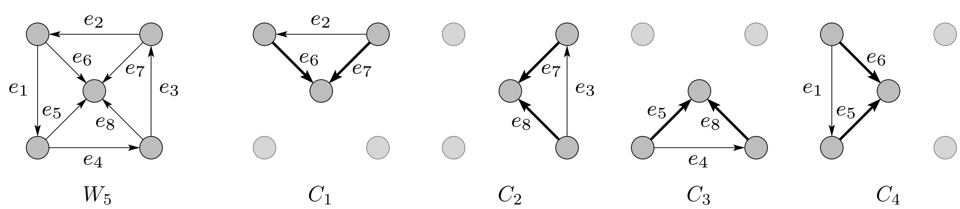

Fig. 3 – An orientation D of the undirected wheel graph W_5 and four circuits C_1 to C_4 in D. The edges are numbered from e_1 to e_8. The edges $\{e_5, e_6, e_7, e_8\}$ form a spanning tree T of D. Circuit C_1 is induced by non-tree edge e_2 and uses edges e_2 and e_6 in forward direction and edge e_7 in backward direction. Thus $C_1 = (0, 1, 0, 0, 0, 1, -1, 0)$. Cycles C_2, C_3, and C_4 are obtained in an analogous way. The set $\{C_1, C_2, C_3, C_4\}$ is a strictly fundamental cycle basis of D.

A *weighted graph* is a graph together with a weight function $w : E \to \mathbb{R}$. If the graph is unweighted, we set $w : E \to 1$ and call w the *uniform weight function*. The weight of a set of edges is the sum of the weights of its members. The *weight and length of a simple cycle* C are

$$w(C) := \sum_e |C_e| w(e) \quad \text{and} \quad |C| := \sum_e |C_e|, \quad \text{respectively.}$$

In an unweighted graph, weight and length are identical. The *weight of a cycle basis* B is the sum of the weights of its cycles, i.e.,

$$w(B) = \sum_{C \in B} w(C).$$

A *minimal κ-cycle basis*, or *MCB*, of G is a κ-cycle basis with minimal weight. We assume that there are no simple cycles of negative weight; such weight functions are called *conservative*. For most of our algorithms, we need to assume that weights are non-negative, i.e., $w : E \to \mathbb{R}^+$.

We close this section with a first theorem. Every graph has a κ-cycle basis and the dimension of the κ-cycle space is given by the graph's *cyclomatic number* $\nu := m - n + CC$, where CC denotes the number of connected components of G. On the way, we get to know a particularly simple set of cycles, the fundamental cycles with respect to a spanning forest. Let G be an (undirected or directed) graph and let T be a spanning forest of G. For any non-tree edge e, let C_T^e be the circuit consisting of e and the tree path connecting the endpoints of e. In the case of a directed graph, we use e in forward direction and traverse the tree path from the head of e to the tail of e; Fig. 3. We call C_T^e the *fundamental circuit* defined by T and e.

Lemma 2.1. *Let G be a graph and let T be any spanning forest of G. Let C be a cycle that uses only edges in T, i.e., $C_e = 0$ for $e \notin T$. Then $C = \mathbf{0}$.*

If C and C' are cycles with $C_e = C'_e$ for all $e \notin T$, then $C = C'$.

Proof. The support of C is contained in T. If the support were non-empty, there would have to be a vertex v having exactly one incident edge with $C_e \neq 0$. This is clearly impossible and hence the support must be empty.

Assume next that C and C' are cycles with $C_e = C'_e$ for all $e \notin T$. Then $C - C'$ is a cycle with $(C - C')_e = 0$ for all $e \notin T$. Thus $C - C' = \mathbf{0}$. □

Lemma 2.2. *Let B be a set of cycles in G and let T be any spanning forest of G. For any cycle $C \in B$, let C' be its restriction to $N := E \setminus T$. The cycles are linearly independent if and only if their restrictions to N are linearly independent.*

Proof. Clearly, linear dependence of the cycles implies linear dependence of their restrictions. Conversely, assume that there is a non-trivial linear combination of the restrictions that yields the zero vector, i.e., $\sum_{C \in B} \lambda_C C' = \mathbf{0}_N$. Here $\mathbf{0}_N$ denotes the zero vector over index set N. Then $\sum_{C \in B} \lambda_C C$ is a cycle that uses only tree edges and hence is equal to $\mathbf{0}$. □

Thus, we may restrict attention to the restricted incidence vectors when discussing questions of linear independence.

Theorem 2.3 (*Dimension of the Cycle Space of a Graph*). *The dimension of the κ-cycle space of a graph G is given by its* cyclomatic number

$$\nu = m - n + CC,$$

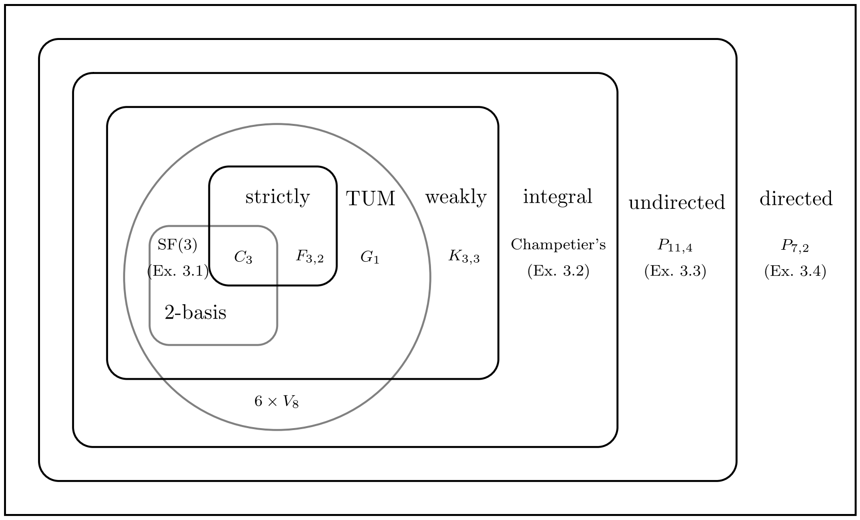

Fig. 4 – The Venn diagram of directed cycle bases: Ex. 3. X refers to examples that are discussed in detail later in this section, $K_{3,3}$ refers to a weighted version of the complete bipartite graph on 3 × 3 vertices, $P_{7,2}$ is a weighted version of a generalized Petersen graph, V_8 is Wagner's graph (see Section 6), $F_{3,2}$ is a fan graph on five vertices, and G_1 is a simple graph on eight vertices; see [2].

where CC denotes the number of connected components of G. Moreover, if T is any spanning forest of G, the set of fundamental circuits with respect to T forms a basis.

Proof. The number of fundamental circuits is equal to ν, because a connected component with m' edges and n' vertices contributes $m' - (n' - 1)$ fundamental cycles. Let N be the set of non-tree edges. The fundamental cycles are clearly independent since any edge $e \in N$ is contained in C_T^e and in no other circuit. It remains to prove that the set of fundamental circuits spans all cycles. Let C be an arbitrary cycle. Consider the cycle

$$\tilde{C} := \sum_{e \in N} C_e C_T^e.$$

We claim that $C = \tilde{C}$. Indeed, for any $e \in N$, we have $\tilde{C}_e = C_e$ and hence $C - \tilde{C}$ is a cycle using only edges of T. Thus $C - \tilde{C} = \mathbf{0}$. □

The following lemma is a first step towards clarifying the relation between directed and undirected cycle bases.

Lemma 2.4. *Let D be a directed graph, let $B = \{C_1, \ldots, C_\nu\}$ be a set of circuits in D, let G be the underlying undirected graph, and let $\pi(B) = \{\pi(C_1), \ldots, \pi(C_\nu)\}$. If $\pi(B)$ is an undirected cycle basis of G then B is a directed cycle basis of D.*

Proof. We have already shown that a set of dependent cycles projects into a set of dependent cycles. Hence $\pi(B)$, being an undirected cycle basis, implies that the cycles in B are independent. Also, ν must be equal to the cyclomatic number of D since $\pi(B)$ is a basis. □

3. Classification of cycle bases

We present seven classes of cycle bases and provide characterizations for them. We will show that each class gives rise to its own minimum cycle basis problem. The complexity of the minimum cycle basis problem differs widely. For three classes the problem is polynomial time, for two classes the problem is NP-complete, and for two classes the status is unknown. This section is mainly based on [2]; the missing proofs can be found there.

Definition 3.1 (*Classes of Cycle Bases*). A directed cycle basis (D-*basis*) $B = \{C_1, \ldots, C_\nu\}$ of a graph D is called $a(n)$:

1. *undirected* or U-*basis*, if the projections $\pi(C_i)$ of the basic circuits C_i onto the underlying undirected graph $G(D)$ constitute a cycle basis of $G(D)$;
2. *integral* or I-*basis*, if each cycle C of D can be written as an *integer* linear combination of circuits in B, i.e.

   $$\exists \lambda_i \in \mathbb{Z} : C = \lambda_1 C_1 + \cdots + \lambda_\nu C_\nu;$$
3. *zero-one* or[1] TUM-*basis*, if each cycle C' of $G(D)$ has an orientation C that can be written as a linear combination with coefficients in $\{-1, 0, +1\}$ of circuits in B, i.e.

   $$\exists \lambda_i \in \{-1, 0, +1\} : \gamma_C = \lambda_1 C_1 + \cdots + \lambda_\nu C_\nu;$$
4. *weakly fundamental* or W-*basis*, if there exists some permutation σ such that

   $$C_{\sigma(i)} \setminus (C_{\sigma(1)} \cup \cdots \cup C_{\sigma(i-1)}) \neq \emptyset, \quad \forall i = 2, \ldots, \nu;$$
5. *strictly fundamental* or F-*basis*, if there exists some spanning forest $T \subseteq E$ such that $B = \{C_T(e) \mid e \in E \setminus T\}$, where $C_T(e)$ denotes the unique circuit in $T \cup \{e\}$; and
6. *planar*, or 2-*basis*, if each arc is contained in at most two basic circuits and the basis is undirected.

Fig. 4 depicts the relationship between these classes: The inclusions are established in Theorem 3.4, and examples for the non-emptiness of the regions will be provided in Section 3.4.

[1] It will become clear in Theorem 3.4 why zero-one bases are called totally unimodular (TUM).

3.1. Existence

Except for 2-bases, every graph has a basis of each type. This follows from the fact that every graph has a strictly fundamental cycle basis and that all other classes generalize fundamental cycle bases. In contrast, MacLane [1] established that a graph has a 2-basis if and only if it is planar.

3.2. Characterizations

We define the cycle matrix corresponding to a basis and show that the different classes of cycle bases can be characterized in terms of simple properties of this matrix. An important property is the determinant of the cycle basis. The *cycle matrix* corresponding to a D-basis B of D is an $m \times \nu$ matrix whose columns are the incidence vectors of the basic circuits. The cycle matrix is determined up to a permutation of the rows and columns.

The cycle matrix Φ of a fundamental basis has a particularly simple form. Let T be a spanning forest and let N be the set of non-tree arcs. Then, for a suitable permutation of the columns, the $\nu \times \nu$ submatrix Φ' selected by the rows corresponding to non-tree arcs is the identity matrix.

Lemma 3.1 ([3]). *Let B be a directed cycle basis of a directed graph G and let Γ be the corresponding cycle matrix. A $\nu \times \nu$ submatrix Γ' of Γ is nonsingular if and only if the rows of Γ' correspond to the non-tree arcs of some spanning forest of G.*

Proof. To prove sufficiency, consider a spanning forest T of D, and let Φ be the cycle matrix of the fundamental basis with respect to T. Because$B$ is a directed cycle basis, any fundamental cycle is a linear combination of cycles in B. Thus there is a matrix $R \in \mathbb{Q}^{\nu\times\nu}$ with $\Phi = \Gamma R$. The restriction of Φ to the non-tree arcs of T is the identity matrix. Hence, R is the inverse of Γ'.

Conversely, assume that the rows which are *not* in Γ' do not form a spanning forest. Then there is a circuit C consisting only of such arcs. As B is a D-basis, we have $C = \Gamma x_C$ for some x_C and clearly $x_C \neq \mathbf{0}$. Restriction to the rows indexing Γ' yields $\mathbf{0} = \Gamma' x_C$, and hence Γ' is singular. □

Lemma 3.2 ([3]). *Let B be a D-basis, let Γ be its cycle matrix, and let A_1 and A_2 be two nonsingular $\nu \times \nu$ submatrices of Γ. Then* $\det A_1 = \pm \det A_2$.

Proof. By Lemma 3.1, the rows of A_i correspond to the non-tree arcs of some spanning forest T_i. It suffices to prove the claim for the case where $T_2 = T_1 + e - f$, for some edges e and f. Let Φ be the cycle matrix of the fundamental basis with respect to T_1. Then $\Gamma = \Phi N$ for some matrix N. Let Φ_i be the submatrix of Φ selected by the non-tree arcs of T_i. Then $\det A_i = \det \Phi_i \cdot \det N$ and therefore it suffices to prove $\det \Phi_2 = \pm \det \Phi_1$. We have $\Phi_1 = I$ and hence $\det \Phi_1 = 1$. Also, since e must lie on the path in T_1 connecting the endpoints of f (otherwise, T_2 would not be a spanning tree), the entry of Φ in row e and column C_f is either $+1$ or -1. Developing $\det \Phi_2$ according to column C_f shows $\det \Phi_2 = \pm 1$. □

The above lemma allows us to define the determinant of a directed cycle basis.

Definition 3.2 (*Determinant of a Set of ν Oriented Circuits*). Let B denote a set of ν circuits in a directed graph D. Consider the matrix Γ with the incidence vectors of B as columns. Let Γ' be the $\nu \times \nu$ submatrix of Γ that arises when deleting the arcs of some spanning forest of D. We define

$$\det B := |\det \Gamma'|.$$

Determinants of directed cycle bases are positive integers. The value of the determinant is invariant under reorienting arcs of D or reorienting circuits of B, because this simply translates to multiplying a row or column by minus one. Thus, starting with a cycle basis of an undirected graph G, orienting the edges of G arbitrarily, and choosing one of the two orientations for each circuit, always results in the same determinant.

How large can the determinant of a cycle basis be? Hadamard's bound gives an upper bound of $\sqrt{n}^{\nu}$, since we are dealing with the determinant of a $\nu \times \nu$ matrix with entries in $\{-1, 0, +1\}$ in which every column has at most n non-zero entries.

Lemma 3.3 ([4]). *Consider the generalized Petersen graph*[2] *$P_{n,2}$ with $n \geq 5$ and n odd. Let $\mathcal{C}$ denote the set of circuits, each of which contains exactly one inner edge, $n - 2$ outer edges and two spokes. $\mathcal{C}$, together with the inner circuit C_I, forms a cycle basis of $P_{n,2}$ and its determinant equals $n - 2$.*

Proof. $P_{n,2}$ consists of $2n$ vertices and $3n$ edges. Therefore every cycle basis has to consist of $n+1$ cycles, which is indeed the number of considered circuits. Additionally, it should be mentioned that the inner circuit C_I is indeed a simple cycle since n is odd.

Now let T be a spanning tree of $P_{n,2}$ made up of all but one inner edge and all spokes. Consider the square submatrix Γ' of the cycle matrix Γ obtained by deleting the rows corresponding to T. The non-tree edges and the circuits in $\mathcal{C} \cup \{C_I\}$ can be oriented and permuted such that

$$\Gamma' = \left(\begin{array}{cccccc|c}
1 & \cdots & \cdots & 1 & 0 & 0 & 0 \\
0 & 1 & \ddots & \ddots & 1 & 0 & \\
0 & 0 & 1 & \ddots & \ddots & 1 & \vdots \\
1 & 0 & 0 & 1 & \ddots & \vdots & \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \\
1 & \cdots & 1 & 0 & 0 & 1 & 0 \\
\hline
* & & \cdots & & & * & 1
\end{array}\right)$$

where the last column and the last row correspond to the inner circuit and the inner edge, respectively. The determinant of Γ' equals the determinant of its $n \times n$ submatrix obtained by deleting the last row and column. The resulting matrix is a *circulant matrix* whose first row has $n - 2$ consecutive ones

[2] The generalized Petersen graph $P_{k,\ell}$ consists of $2k$ vertices $\{I_i, O_i \mid 0 \leq i \leq k-1\}$ and edges $\{O_iO_{i+1}, O_iI_i, I_iI_{i+\ell} \mid 0 \leq i \leq k-1\}$. All indices are modulo k. The edges O_iO_{i+1} are called *outer edges*, the edges $I_iI_{i+\ell}$ are called *inner edges*, and the edges O_iI_i are called *spokes*.

followed by two zeros. The entries of every other row result from the row above by a circular shift to the right. We have

$\det \Gamma' = n - 2;$

see [4] for the calculation of the determinant. □

Open Problem 2. Provide better upper and/or lower bounds on the maximal determinant of cycle bases.

Theorem 3.4 ([2]). *Let B be a directed cycle basis with cycle matrix Γ. Then:*

1. *B is undirected, if and only if* $\det B$ *is odd.*
2. *B is integral, if and only if* $\det B$ *is one.*
3. *B is zero-one if and only if Γ is totally unimodular.*[3]
4. *B is weakly fundamental, if and only if Γ can be permuted so as to have a regular upper triangular* $\nu \times \nu$ *matrix in its last ν rows.*
5. *B is strictly fundamental, if and only if Γ can be permuted so as to have the* $\nu \times \nu$ *unit matrix in its last ν rows.*
6. *B is a 2-basis, if and only if B is an undirected cycle basis and Γ has at most two non-zero entries per row.*

Proof. *Case* 1. The projections $\pi(C_i)$ of the basic circuits are linearly independent if $\pi(\Gamma)$ has full rank, i.e., if there is a square submatrix $\pi(\Gamma')$ with non-zero determinant over $GF(2)$. The value of the determinant is $(\det \Gamma') \bmod 2$. We conclude that B is undirected if and only if $\det B$ is odd.

Case 2. Let T be some spanning forest, and let Γ' be the square submatrix of Γ indexed by the non-tree arcs of T.

Let Φ be the cycle matrix of the fundamental basis with respect to T. Since B is integral, there is an integral $\nu \times \nu$ matrix R such that $\Phi = \Gamma R$. Restriction to the non-tree arcs of T yields $I = \Gamma' R$. We have $\det \Gamma' \in \mathbb{Z}$ and $\det R \in \mathbb{Z}$, because both matrices are integral. Thus $(\det \Gamma') \cdot (\det R) = 1$ implies $\det \Gamma' = \pm 1$.

Let C be an arbitrary circuit. The representation x_C of C in terms of B satisfies $C = \Gamma x_C$. Restriction to the non-tree arcs of T yields $C' = \Gamma' x_C$ or $x_C = (\Gamma')^{-1} C$. The inverse of Γ' is integral, by Cramer's rule, and since $\det \Gamma' = \pm \det B = \pm 1$. Thus $x_C \in \mathbb{Z}^{\nu}$.

Case 3. A matrix is totally unimodular if and only if for any subset I of its columns there are coefficients $\lambda_i \in \{-1, +1\}$ such that $\sum_{i \in I} \lambda_i C_i$ is a vector with entries in $\{-1, 0, +1\}$, see [5, Theorem 19.3].

Assume first that B is a zero-one basis. Since zero-one bases are integral, B is an integral cycle basis and hence $\pi(B) = \{\pi(C_i) \mid C_i \in B\}$ is an undirected basis of $G(D)$. Let I be an arbitrary subset of the columns of Γ and consider the $\mathbb{Z}_2$-sum of the projections of the circuits in I, and call the resulting cycle C',

$$\sum_{i \in I} \pi(C_i) = C'.$$

Since B is a zero-one basis, C' has an orientation C that can be written as a linear combination with coefficients $\lambda_i \in \{-1, 0, +1\}$ of the circuits in B, i.e.,

$$\sum_{i=1}^{\nu} \lambda_i C_i = C.$$

[3] This item is a new result.

Projecting this equation into $\mathbb{Z}_2$, we obtain

$$\sum_{i=1}^{\nu} |\lambda_i| \pi(C_i) = C'.$$

Since the representation of C' with respect to $\pi(B)$ is unique, λ_i is non-zero if and only if $i \in I$. Thus, in the TUM characterization, C is the desired linear combination of the columns selected by I.

Assume conversely that Γ is totally unimodular. Then $\det B = 1$ and hence $\{\pi(C_i) \mid C_i \in B\}$ is a basis of $G(D)$. Let C' be any cycle in $G(D)$. Then $C' = \sum_{i \in I} \pi(C_i) \bmod 2$ for some index set $I \subseteq \{1, \ldots, \nu\}$. Since Γ is totally unimodular, there are coefficients $\lambda_i \in \{-1, +1\}$ such that $\sum_{i \in I} \lambda_i C_i$ is a vector C with components in $\{-1, 0, +1\}$. Clearly, $\pi(C) = C'$ and hence C is the desired orientation of C'.

Case 4. Order the columns of Γ such that $C_{\sigma(i)}$ is in the i-th column for $1 \le i \le \nu$. Order the rows of Γ such that an arc a with $a \in C_{\sigma(i)} \setminus (C_{\sigma(1)} \cup \cdots \cup C_{\sigma(i-1)})$ corresponds to row $\nu - 1 + i$.

Case 5. This is nothing but a reformulation of Sysło's characterization [6] of a strictly fundamental cycle basis B, namely that every circuit in the basis contains an arc that is contained in no other circuit of the basis.

Case 6. This is but a reformulation of the definition of 2-bases. □

The determinant of a set of ν circuits can be computed over any field κ. For directed bases the determinant is non-zero in $\mathbb{Q}$, for undirected bases the determinant is non-zero in $GF(2)$. We therefore also call directed bases $\mathbb{Q}$-bases and undirected bases $GF(2)$-bases. We call a directed basis a $GF(p)$-basis, where p is a prime, if its determinant is non-zero modulo p.

Theorem 3.4 establishes most of the inclusions shown in Fig. 4: Every fundamental basis is both weakly fundamental and totally unimodular, every weakly fundamental or totally unimodular basis is integral, every integral basis is undirected, and every undirected basis is directed. We shall next relate 2-bases to the other classes.

Lemma 3.5. *Every 2-basis is totally unimodular and weakly fundamental.*

Proof. Let $B = \{C_1, \ldots, C_\nu\}$ be a 2-basis of G. MacLane [1] showed that a graph having a 2-basis is planar and that, moreover, the basic circuits correspond to the bounded face cycles of some planar embedding of G. Orient the edges of G arbitrarily and let the C_i's correspond to counterclockwise traversals of the face cycles. Then every row of Γ has at most two non-zero entries; if there are two non-zero entries, one is $+1$ and one is -1. Thus Γ is totally unimodular [5, page 274].

We next show that B is weakly fundamental. Let $C = \{e_1, \ldots, e_k\}$ be the boundary of the infinite face of G. For $i = 1, 2, \ldots, k$, denote by C_{e_i} the unique circuit in B that contains $e_i \in C$. In the first iteration, we define

$$C_{\sigma(\nu)} = C_{e_1}, \qquad C_{\sigma(\nu-1)} = C_{e_2}, \ldots, C_{\sigma(\nu-k+1)} = C_{e_k}.$$

Then, we remove the edges of C from G and proceed in the same way for the 2-connected components of the remaining graph. □

Fig. 5 – A graph and a directed cycle basis. For each of the four circuits, the arcs belonging to the circuit are shown in bold. Arcs used in reversed direction are shown dotted. Every arc is used in exactly two circuits. The determinant of this basis is two. Thus the basis is not totally unimodular. Also, since each arc is used in exactly two circuits, the basis is not weakly fundamental.

Fig. 6 – Examples of a strictly fundamental cycle basis that is also a 2-basis, a weakly fundamental cycle basis, and a non-integral cycle basis in the wheel graph W_5. The last of these originates from [7].

We required a 2-basis to use every arc at most twice and to be undirected. Fig. 5 shows a graph and a directed basis that uses every arc exactly twice and is neither totally unimodular nor weakly fundamental (Tomasz Jurkiewicz, personal communication).

Open Problem 3. The definition of zero-one bases may seem strange. It would be equally natural to require that every circuit (every simple cycle) is a linear combination of the basic circuits with coefficients in $\{-1, 0, +1\}$. How do these definitions relate?

3.3. *Simple examples*

Fig. 6 presents three cycle bases for the wheel graph W_5: the strictly fundamental cycle basis $B_1 = \{C_{11}, C_{12}, C_{13}, C_{14}\}$, which is also a 2-basis, the weakly fundamental cycle basis $B_2 = \{C_{21}, C_{22}, C_{23}, C_{24}\}$, and the undirected basis $B_3 = \{C_{31}, C_{32}, C_{33}, C_{34}\}$; the lattermost is *not* integral. The strictly fundamental cycle basis B_1 corresponds to the spanning tree $T = \{e_1, e_2, e_3, e_4\}$. The corresponding cycle matrices are as follows:

$$\Gamma_1 = \left(\begin{array}{cccc} -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right), \quad \Gamma_2 = \left(\begin{array}{cccc} -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ \hline 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right),$$

$$\Gamma_3 = \left(\begin{array}{cccc} 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right).$$