On finding a cycle basis with a shortest maximal cycle

David M. Chickering, Dan Geiger *, David Heckerman

Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, USA

Communicated by D. Dolev; received 24 August 1993; revised 31 May 1994

Abstract

The Shortest Maximal Cycle Basis (SMCB) problem is that of finding a cycle basis B of a given graph G such that the length of the longest cycle included in B is the smallest among all bases of G. We show that any cycle basis B' of G such that the sum of the lengths of the cycles included in B' is the smallest among all cycle bases of G constitutes a solution to the SMCB problem. Finding a basis with the latter property requires at most $O(m^3 n)$ steps using Horton's algorithm where m is the number of edges and n is the number of vertices.

Keywords: Algorithms; Analysis of algorithms; Combinatorial problems; Coding theory

1. Introduction

The Shortest Cycle Basis (SCB) problem is that of finding a cycle basis B of a given graph G with the property that the sum of the lengths of the cycles included in B is the smallest among all bases of G. This problem and some variations of it were dealt in several articles (e.g., [2,3,6,7,9, 10,12]). The latest reference is Horton's work which establishes a polynomial algorithm having a time complexity of $O(m^3 n)$ where m is the number of edges and n is the number of vertices in G [5]. Horton states that his algorithm is the first polynomial algorithm that actually solves this problem and gives citations that contain counter-examples to the attempts made prior to his.

The Shortest Maximal Cycle Basis (SMCB) problem is that of finding a cycle basis B of a given graph G with the property that the length of the longest cycle included in B is the smallest among all bases of G. We show that any basis that constitutes a solution to the SCB problem is also a solution to the SMCB problem. This correspondence between the two problems is shown to hold in any vector space over GF_2, not necessarily those induced by cycles of a graph. Consequently, Horton's polynomial algorithm for the SCB problem (for graphs) solves the SMCB problem as well.

As far as we know, the SMCB problem has not been raised in the past. Our interest in this problem comes from a possible application as a preprocessing step in a Bayesian inference algorithm [8]. This inference algorithm heuristically generates a clique-tree of a directed acyclic graph (dag) before some probability computations are

---

* Corresponding author. Author's primary affiliation: Computer Science Department, Technion, Haifa 32000, Israel.

done. In this algorithm, if we limit the size of the largest cycle in a cycle basis of the underlying undirected graph of the dag, then the size of the largest clique in the clique-tree generated is often small which consequently reduces the probabilistic computations. However, the relationship found between the size of the largest clique and the size of the largest cycle in the cycle basis is heuristic with no theoretical support.

2. Definitions and basic properties

By a *graph* $G = (V(G), E(G))$ we mean an undirected graph with no self-loops or parallel edges where $V(G)$ is the set of vertices and $E(G)$ is the set of edges. A connected subgraph C is called a *simple cycle* if each vertex is incident to two edges in C. A subgraph C is called a *cycle* if each vertex has an even degree in C. Note that a cycle need not be a connected subgraph and that a simple cycle is a cycle. Each cycle C can be written as a vector of length $|E(G)|$ having 1 in each location that corresponds to an edge in $E(C)$ and having 0 otherwise. The *sum* $C_1 \oplus C_2$ of two cycles C_1 and C_2 is the subgraph induced by the edges

$$(E(C_1) \cup E(C_2)) \setminus (E(C_1) \cap E(C_2)).$$

Equivalently, the vector corresponding to $C_1 \oplus C_2$ is the sum mod 2 of the vectors corresponding to C_1 and C_2. The set of vectors corresponding to all cycles of a graph form a linear vector space over GF_2. (GF_2 is the field with constants $\{0, 1\}$ and addition taken mod 2.) We will call this linear vector space the *cycle space* and a basis for this vector space will be called a *cycle basis*. The dimension of the cycle space is $|E(G)| - |V(G)|$ plus the number of connected components of G [11].

We use the following basic properties of vector spaces.

Proposition 1. *Let B be a basis of a vector space $\mathscr{V}$. If any vector v in B is replaced by the sum of v and a linear combination of the vectors in $B \setminus \{v\}$, then the resulting set of vectors is a basis of $\mathscr{V}$.*

Proposition 2. *Let $\{v_1, \ldots, v_k\}$ be a basis of a vector space $\mathscr{V}$ over GF_2 and let $\{u_1, \ldots, u_k\}$ be another basis of $\mathscr{V}$. Then, there exists a permutation θ of $\{1, \ldots, k\}$ such that for $i = 1, \ldots, k$ each $u_{\theta(i)}$ can be written as the sum of v_i and a linear combination of $\{v_1, \ldots, v_k\} \setminus \{v_i\}$.*

Proof. Let $M = (m_{i,j})$ be the non-singular matrix that maps the first basis to the second one and let ϕ denote a permutation of $\{1, \ldots, k\}$. Since M is non-singular its determinant $\Sigma_\phi m_{1,\phi(1)} m_{2,\phi(2)} , \ldots, m_{k,\phi(k)}$ (sum taken mod 2) must contain at least one addend in which all factors are non-zero. Let θ be a permutation that corresponds to any such addend. θ satisfies the condition of this proposition. $\square$

3. Main result

Define the *length* of a vector v over GF_2, denoted by $|v|$, to be the number of 1's that it contains. The *shortest basis* of a vector space is a basis B in which the sum of the lengths of all vectors in B is minimized. Consider the following exponential greedy algorithm for finding the shortest basis of a vector space over GF_2.

Algorithm FINDBASIS

Input: A basis $v_1, \ldots, v_k$ of a vector space $\mathscr{V}$;
Output: A shortest basis of $\mathscr{V}$;

Until no changes occur;
   For $i = 1$ to k **do**;
      $v_i \leftarrow v_i \oplus \alpha(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$
      where $\alpha(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$ is a linear combination of $\{v_1, \ldots, v_k\} \setminus \{v_i\}$ chosen such that the updated vector v_i will have the smallest possible length;
   end;
end;
Output$(v_1, \ldots, v_k)$;

The algorithm always terminates because in each step the sum of the lengths of the vectors $v_1, \ldots, v_k$ is reduced by at least one. The algorithm always outputs a basis because, due to

Proposition 1, after each step $\{v_1, \ldots, v_k\}$ remains a basis of $\mathscr{V}$. Furthermore, when the algorithm stops no v_i can be improved by adding to it a linear combination of the other $k-1$ vectors. We now argue that the algorithm always outputs a shortest basis. Let $u_1, \ldots, u_k$ be a shortest basis of $\mathscr{V}$ and $v_1, \ldots, v_k$ be the basis FINDBASIS generates. Then, by Proposition 2, there exists a permutation θ of $\{1, \ldots, k\}$ such that each $u_{\theta(i)}$ equals

$$v_i \oplus \alpha(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k).$$

Consequently, for each i, $|v_i| \leqslant |u_{\theta(i)}|$ because otherwise FINDBASIS would have performed an additional iteration. However, equality must hold for all i because $u_1, \ldots, u_k$ is a shortest basis. Thus, FINDBASIS always terminates and outputs a shortest basis. In fact, we can prove that FIND-BASIS modifies each vector only once at the first iteration of the external loop; however, the proof of this observation is not needed herein and is thus omitted.

This discussion implies the following property of every shortest basis.

Theorem 3. *Let $\{u_1, \ldots, u_k\}$ and $\{v_1, \ldots, v_k\}$ each be a shortest basis of a vector space $\mathscr{V}$ over GF_2 having lengths $|u_1| \leqslant |u_2| \leqslant \cdots \leqslant |u_k|$, and $|v_1| \leqslant |v_2| \leqslant \cdots \leqslant |v_k|$, respectively. Then, for $i = 1, \ldots, k$, $|u_i| = |v_i|$.*

Proof. Apply FINDBASIS to the basis $\{v_1, \ldots, v_k\}$. We first argue that FINDBASIS will not make any changes to the given basis. Assume it does. Consider the set $\{v_1, \ldots, v_k\}$ just after some v_i has been changed. The resulting new set of vectors is a basis whose length is shorter than the given basis which contradicts the minimality of the given basis. We have shown previously that the outcome of FINDBASIS satisfies $|v_i| \leqslant |u_{\theta(i)}|$ for $i = 1, \ldots, k$ where θ is a permutation on $\{1, \ldots, k\}$. Since $u_1, \ldots, u_k$ is a shortest basis, equality must hold for all i. □

It is worth noting that a shortest basis of a vector space of GF_2 contains as its smallest element a vector of $\mathscr{V}$ whose length is minimum

across all vectors of $\mathscr{V}$. The existence of a polynomial algorithm for finding a vector whose length is minimal in a given vector space over GF_2 is a major open question in coding theory [1].

Let $L(B)$ denote the length of the longest vector in a basis B. A basis of $\mathscr{V}$ with minimum longest vector is a basis B such that $L(B)$ is minimized over all bases of $\mathscr{V}$.

Theorem 4. *Let B be a shortest basis of a vector space $\mathscr{V}$ over GF_2 and let B' be a basis of $\mathscr{V}$ with minimum longest vector. Then, $L(B) = L(B')$.*

Proof. Let $B = \{u_1, \ldots, u_k\}$ and $B' = \{w_1, \ldots, w_k\}$. Apply FINDBASIS to B' and suppose the algorithm outputs $B'' = \{v_1, \ldots, v_k\}$. The algorithm never increases the length of an updated vector. Thus $L(B'') \leqslant L(B')$. Since FINDBASIS generates a shortest basis, by Theorem 3, $L(B'') = L(B)$. Hence, $L(B) \leqslant L(B')$. Since B' is a basis with minimum longest vector, equality is implied. □

Consequently, every algorithm that finds a shortest basis also finds a basis with minimum longest vector. Furthermore, in the case of cycle spaces, finding the shortest (cycle) basis requires only polynomial number of steps using Horton's algorithm. The algorithm can be roughly described as follows. First, for each pair of a vertex x and an edge (a, b) find a shortest path $p(a, x)$ between a and x and a shortest path $p(b, x)$ between b and x. Then form a cycle using $p(a, x)$, $p(b, x)$ and (a, b) unless $p(a, x)$ and $p(b, x)$ share any vertex other than x. In the latter case the degenerated cycle is ignored. Consequently, the number of cycles generated, r, is bounded by mn where m is the number of edges and n is the number of nodes. Finally a greedy algorithm is used which selects in each step the shortest cycle among the r cycles generated in the previous phase such that the newly selected cycle is independent of the previously selected ones. The last step is implemented by applying Gaussian elimination to a 0–1 matrix whose rows are the vectors corresponding to the r cycles generated in the first phase of the algorithm.

Theorem 4 shows that this algorithm also solves the SMCB problem.

Acknowledgment

References

[1] E.R. Berlekamp, R.J. McEliece and H.C.A. van Tilborg, On the inherent intractability of certain coding problems, *IEEE Trans. Inform. Theory* **24** (1978) 384–386.

[2] D.W. Cribb, R.D. Ringeisen and D.R. Shier, On cycle bases of a graph, *Congr. Numer.* **32** (1981) 221–229.

[3] N. Deo, Minimum length fundamental cycle set, *IEEE Trans. Circuits and Systems* **26** (1979) 894–895.

[4] N. Deo, G.M. Prabhu and M.S. Krishnamoorthy, Algorithms for generating fundamental cycles in a graph, *ACM Trans. Math. Software* **8** (1982) 26–42.

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

[6] E. Hubicka and M.M. Syslo, Minimal bases of cycles of a graph, in: M. Fiedler, ed., *Proc. Symp. on Recent Advances in Graph Theory*, Prague, 1974 (Academia Praha, Prague, 1975) 283–293.

[7] E. Kolasinska, On a minimum cycle basis of a graph, *Zastos. Mat.* **16** (1980) 631–639.

[8] S.L. Lauritzen and D.J. Spiegelhalter, Local comomputations with probabilities on graphical structures and their application to expert systems, J. Roy. Statist. Soc. Ser. B **50** (1988) 157–224.

[9] G.F. Stepanec, Basis systems of vector cycles with extremal properties in graphs, *Uspekhi Mat. Nauk* **19** (1964) 171–175 (in Russian).

[10] M.M. Syslo, On cycle bases of a graph, *Networks* **9** (1979) 123–132.

[11] K. Thulasiraman and M.N.S. Swamy, Graphs: Theory and Algorithms (Wiley, New York, 1992).

[12] A.A. Zykov, Theory of Finite Graphs (Nauka, Novosibirsk, 1969) (in Russian).