



Minimization of Graph Weighted Models over Circular Strings

Guillaume Rabusseau^(✉)

Reasoning and Learning Lab, School of Computer Science,
McGill University, Montreal, Canada
guillaume.rabusseau@mcgill.ca

Abstract. Graph weighted models (GWMs) have recently been proposed as a natural generalization of weighted automata over strings, trees and 2-dimensional words to arbitrary families of labeled graphs (and hypergraphs). In this paper, we propose polynomial time algorithms for minimizing and deciding the equivalence of GWMs defined over the family of circular strings on a finite alphabet (GWM^cs). The study of GWM^cs is particularly relevant since circular strings can be seen as the simplest family of graphs with cycles. Despite the simplicity of this family and of the corresponding computational model, the minimization problem is considerably more challenging than in the case of weighted automata over strings and trees: while linear algebra tools are overall sufficient to tackle the minimization problem for classical weighted automata (defined over a field), the minimization of GWM^cs involves fundamental notions from the theory of finite dimensional algebra. We posit that the properties of GWM^cs unraveled in this paper will prove useful for the study of GWMs defined over richer families of graphs.

1 Introduction

Functions defined over syntactical structures such as strings, trees and graphs are ubiquitous in computer science. Automata models allow one to succinctly represent such functions. In particular, *weighted automata* can efficiently model functions mapping structured objects to values in a semi-ring. Weighted automata have been defined to handle functions whose domain are e.g. strings [9, 26], trees [8, 16] and 2-dimensional words [11]. More recently, Bailly et al. [2] proposed a computational model for functions mapping labeled graphs (or hypergraphs) to values in a field (see also [22, Chap. 2]): Graph Weighted Models (GWMs). GWMs extend the notion of *linear representation* of a function defined over strings and trees to functions defined over graphs labeled by symbols in a ranked alphabet: loosely speaking, while string weighted automata can be defined by associating each symbol in a finite alphabet to a linear map and tree weighted automata by associating each symbol in a ranked alphabet to a multilinear map, GWMs are defined by associating each arity k symbol from a ranked alphabet to a k th order tensor. The computation of a GWM boils down to mapping each vertex in a graph to the tensor associated to its label and performing contractions directed by the edges of the input graph to obtain a value in the supporting

field. When restricted to the families of strings, trees or 2-dimensional words, GWMs are expressively equivalent to the classical notions of weighted automata over these structures.

Weighted automata have recently received interest from the machine learning community due to their ability to represent functions defined over structured objects. Efficient (and often consistent) learning algorithms have been developed for such computational models defined over sequences [3, 6, 10, 19] and trees [1, 4, 14]. Motivated by the relevance of learning functions defined over richer families of labeled graphs, our long term objective is to design efficient learning algorithms for GWMs. This is however a challenging task. Given the close relationship between minimization and learning for classical weighted automata (see e.g. [7, 21, 27]), we take a first step in this direction by tackling the problem of minimizing GWMs defined over the simple family of *circular strings*.

Circular strings are strings whose last symbol is connected to the first. A circular string can be seen as a directed graph where each vertex is labeled by a symbol from a finite alphabet and is connected to his unique successor (i.e. a labeled graph composed of a unique cycle). Circular strings are relevant in biology (see e.g. [20] and references therein) and have been studied from a formal language perspective in the non-quantitative setting in [24]. The study of GWMs defined over such graphs is particularly relevant since circular strings are in some sense the simplest family of graphs with cycles (and cycles can be seen as the key obstacle for going from strings and trees to general graphs). Moreover, GWMs defined over the family of circular strings—which we henceforth denote by GWM^c s to avoid confusions—take a simple form making them easily amenable to theoretical study: a GWM^c is given by a set of matrices \mathbf{A}^σ for each symbol σ in a finite alphabet, and maps any circular string $\sigma_1\sigma_2\cdots\sigma_k$ to the trace of the products of the matrices associated with the letters in the string¹. Despite the simplicity of this computational model and its strong connection with string weighted automata, the minimization problem is considerably more challenging than in the case of string or tree weighted automata. More precisely, while the minimization problem can easily be handled using notions from linear algebra for e.g. real-valued string weighted automata (see e.g. [7]), we show in this paper that the minimization of GWM^c s requires fundamental concepts from the theory of finite-dimensional algebras (such as the ones of radical and semi-simplicity).

Contributions. Throughout the paper, *we only consider automata defined over a field of characteristic 0*. After introducing notions on weighted automata, GWM^c s and finite-dimensional algebras in Sect. 2, we first tackle the problem of deciding the equivalence of GWM^c s in Sect. 3. The study of the equivalence problem is motivated by the simple observation that two minimal GWMs computing

¹ Note that this is a not a definition *per se* but rather a consequence of the definition of general GWMs (as introduced in [2, 22]): when restricted to the family of circular strings, a GWM is given by a set of matrices and its computation can be succinctly expressed using the trace operator (whereas a general GWM is given by a set of *tensors* and its computation relies on *partial traces*).

the same function are not necessarily related by a change of basis, which is in contrast with a classical result stating that two minimal string weighted automata are equivalent if and only if they are related by a change of basis. Building from this observation, we unravel the fundamental notion of semi-simple GWM^c and we show that *any function recognizable by a GWM^c can be computed by a semi-simple GWM^c* (Corollary 1) and that *two semi-simple GWM^cs of equal dimensions computing the same function are necessarily related by a change of basis* (Corollary 2). These two results naturally give rise to a *polynomial time algorithm to decide whether two GWM^cs are equivalent*. We then move on to the minimization problem in Sect. 4, where we give a *polynomial time minimization algorithm for GWM^cs* which fundamentally relies on the notion of semi-simple GWM^c (Corollary 3). While the problem of minimizing a GWM defined over the simple family of circular strings is central to this paper, we see it as a test bed for developing the theory of general GWMs: beyond the minimization and equivalence algorithms we propose, we believe that one of our main contributions is to illustrate how the theory of GWMs will rely on advanced concepts from algebra theory and to unravel fundamental properties that will surely be central to the study of GWMs defined over more general families of graphs (such as the one of semi-simple GWM^c).

1.1 Notations

For any integer n we let $[n] = \{1, 2, \dots, n\}$. We denote the set of integers by \mathbb{N} and the fields of real and rational numbers by \mathbb{R} and \mathbb{Q} respectively. Let \mathbb{F} be a field of characteristic 0, we denote by $\mathcal{M}_n(\mathbb{F}) = \mathbb{F}^{n \times n}$ the set of all $n \times n$ matrices over \mathbb{F} . We use lower case bold letters for vectors (e.g. $\mathbf{v} \in \mathbb{F}^{d_1}$) and upper case bold letters for matrices (e.g. $\mathbf{M} \in \mathbb{F}^{d_1 \times d_2}$). We denote by \mathbf{I}_n the $n \times n$ identity matrix (or simply \mathbf{I} if the dimension is clear from context). Given a matrix $\mathbf{M} \in \mathbb{F}^{d_1 \times d_2}$, we denote its entries by $\mathbf{M}_{i,j}$ and we use $\text{vec}(\mathbf{M}) \in \mathbb{F}^{d_1 d_2}$ to denote the column vector obtained by concatenating the columns of \mathbf{M} . We use $\ker(\mathbf{A})$ to denote the kernel (or null space) of a matrix \mathbf{A} . Given two matrices $\mathbf{A} \in \mathcal{M}_m(\mathbb{F})$ and $\mathbf{B} \in \mathcal{M}_n(\mathbb{F})$ we denote their Kronecker product by $\mathbf{A} \otimes \mathbf{B} \in \mathcal{M}_{mn}(\mathbb{F})$ and their direct sum by $\mathbf{A} \oplus \mathbf{B} \in \mathcal{M}_{m+n}(\mathbb{F})$: $\mathbf{A} \otimes \mathbf{B}$ is the block matrix with blocks $(\mathbf{A}_{i,j} \mathbf{B})_{i,j}$ and $\mathbf{A} \oplus \mathbf{B}$ is the block diagonal matrix with \mathbf{A} in the upper diagonal block and \mathbf{B} in the lower one. We denote by Σ^* the set of strings on a finite alphabet Σ and the empty string by λ . We denote by Σ^+ the set of non-empty strings and by Σ^k the set of all strings of length k .

2 Preliminaries

We first present notions on weighted automata, graph weighted models and finite dimensional algebras. The reader is referred to [9, 16, 25] for more details on weighted automata theory, to [2] and [22, Chap. 2] for an introduction to graph weighted models, and to [13, 17] for a thorough introduction to finite dimensional algebras.

2.1 Weighted Automata and GWMs over Circular Strings

Let Σ be a finite alphabet. A *weighted finite automaton* (WFA) over a field \mathbb{F} with n states is a tuple $M = (\alpha, \{\mathbf{M}^\sigma\}_{\sigma \in \Sigma}, \omega)$ where $\alpha, \omega \in \mathbb{F}^n$ are the initial and final weight vectors respectively, and $\mathbf{M}^\sigma \in \mathcal{M}_n(\mathbb{F})$ is the transition matrix for each symbol $\sigma \in \Sigma$. A WFA computes a function $f_M: \Sigma^* \rightarrow \mathbb{F}$ defined for each word $x = x_1x_2 \cdots x_k \in \Sigma^*$ by

$$f_M(x) = \alpha^\top \mathbf{M}^{x_1} \mathbf{M}^{x_2} \cdots \mathbf{M}^{x_k} \omega.$$

We will often use the shorthand notation $\mathbf{M}^x = \mathbf{M}^{x_1} \mathbf{M}^{x_2} \cdots \mathbf{M}^{x_k}$ for any word $x = x_1x_2 \cdots x_k \in \Sigma^*$. A WFA M with n states is *minimal* if its number of states is minimal, i.e. any WFA M' such that $f_M = f_{M'}$ has at least n states. We say that a function $f: \Sigma^* \rightarrow \mathbb{R}$ is *WFA-recognizable* if there exists a WFA computing it.

Graph weighted models (GWMs) have been introduced as a computational model over arbitrary labeled graphs and hypergraphs in [2]. In this paper, we focus on the simple model of GWMs defined over the family of circular strings. A *circular string* is a string without a beginning or an end, one can think of it as a string closed onto itself (see Fig. 1).

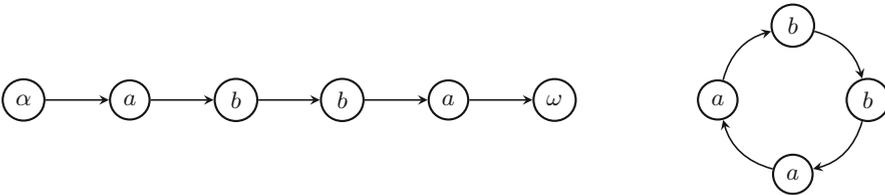


Fig. 1. (left) Graph representation of the string $abba$ where the special vertices labeled with α and ω denote the beginning and end of the string respectively. (right) In contrast, the circular string $abba$ has no beginning and no end, it is thus the same object as e.g. the circular string $baab$.

A d -dimensional GWM A over circular strings (GWM^c) on Σ is given by a set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$. It computes a function $f_A: \Sigma^+ \rightarrow \mathbb{F}$ defined² for each word $x = x_1x_2 \cdots x_k \in \Sigma^+$ by

$$f_A(x) = \text{Tr}(\mathbf{A}^{x_1} \mathbf{A}^{x_2} \cdots \mathbf{A}^{x_k}) = \text{Tr}(\mathbf{A}^x).$$

By invariance of the trace under cyclic permutation, we have $f_A(x_1x_2 \cdots x_k) = f_A(x_2x_3 \cdots x_kx_1) = f_A(x_3x_4 \cdots x_kx_1x_2) = \cdots$. This is in accordance with the

² Observe that we exclude the empty string from the domain of f_A . This is on purpose since $f_A(\lambda)$ would be the dimension of A (using the convention $\mathbf{A}^\lambda = \mathbf{I}$): given two GWM^cs of different dimensions computing the same function on Σ^+ , we want to consider them as equivalent even though they disagree on λ .

definition of a circular string: for any string x' obtained by cyclic permutation of the letters of a string x , both x and x' correspond to the same circular string. Similarly to WFAs, a GWM^c is *minimal* if its dimension is minimal and a function $f: \Sigma^+ \rightarrow \mathbb{F}$ is *GWM^c -recognizable* if it can be computed by a GWM^c .

It is immediate to see that there exist WFA-recognizable functions that are not GWM^c -recognizable, this is the case of any WFA-recognizable function that is not invariant under cyclic permutation of letters in a word³. In contrast, one can easily show that any GWM^c -recognizable function is WFA-recognizable. More precisely, we have the following result.

Proposition 1. *For any d -dimensional $\text{GWM}^c A = \{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ on Σ , the WFA M with d^2 states $(\boldsymbol{\alpha}, \{\mathbf{M}^\sigma\}_{\sigma \in \Sigma}, \boldsymbol{\omega})$ where $\boldsymbol{\alpha} = \boldsymbol{\omega} = \text{vec}(\mathbf{I}_d)$ and $\mathbf{M}^\sigma = \mathbf{I}_d \otimes \mathbf{A}^\sigma$ for each $\sigma \in \Sigma$, is such that $f_M(x) = f_A(x)$ for all $x \in \Sigma^*$.*

Proof. For any $w = w_1 \cdots w_n \in \Sigma^*$ we have $f_A(w) = \text{Tr}(\mathbf{A}^w) = \sum_{i \in [d]} \mathbf{A}_{i,i}^w = \sum_{i \in [d]} \mathbf{e}_i^\top \mathbf{A}^w \mathbf{e}_i$ where \mathbf{e}_i is the i -th vector of the canonical basis of \mathbb{F}^d . Since $\boldsymbol{\alpha} = \boldsymbol{\omega} = (\mathbf{e}_1^\top, \dots, \mathbf{e}_d^\top)^\top$ and $\mathbf{M}^\sigma = \mathbf{I} \otimes \mathbf{A}^\sigma$ is the block-diagonal matrix with \mathbf{A}^σ repeated d times on the diagonal, one can check that $f_M(w) = \boldsymbol{\alpha}^\top \mathbf{M}^w \boldsymbol{\omega} = \sum_{i \in [d]} \mathbf{e}_i^\top \mathbf{A}^w \mathbf{e}_i = f_A(w)$. \square

It follows from this proposition that the learning and equivalence problems for GWM^c s could be handled by using the corresponding algorithms for WFAs. We will nonetheless study the equivalence problem in the next section⁴ without falling back onto the theory of WFAs, which will allow us to unravel fundamental properties of GWM^c s that will be particularly relevant to further studies (moreover, the minimization problem obviously cannot be handled in such a way).

2.2 Finite-Dimensional Algebras

An *algebra* \mathcal{A} over a field \mathbb{F} (or \mathbb{F} -algebra) is a vector space over the field \mathbb{F} equipped with a bilinear operation (called multiplication or product). An algebra is *associative* if its product is associative and it is *finite-dimensional* if it is of finite dimension as a vector space over \mathbb{F} . *In this paper, we will only consider finite-dimensional associative algebras.* A sub-algebra \mathcal{B} of an algebra \mathcal{A} is a linear subspace of \mathcal{A} which is closed under product (i.e. \mathcal{B} equipped with the operations of \mathcal{A} is an algebra itself).

A classical example of finite-dimensional algebra is the set $\mathcal{L}(V)$ of linear operators on some finite-dimensional vector space V (where the product is composition). In this particular example, the algebra $\mathcal{L}(V)$ is isomorphic to the *full matrix algebra* $\mathcal{M}_d(\mathbb{F})$, where d is the dimension of V ; we will mainly focus on matrix algebras in this paper, i.e. sub-algebras of the full matrix algebra $\mathcal{M}_d(\mathbb{F})$ for some d (an example of such an algebra is the set of $d \times d$ upper triangular matrices). In particular, we will often consider the *algebra generated by a finite*

³ Note that this is not a necessary condition: the function f defined on $\{a, b\}^*$ by $f(x) = 1$ if $x = a$ and 0 otherwise is WFA-recognizable but not GWM^c -recognizable.

⁴ The learning problem has been previously considered in [5, 22].

set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$ for some finite alphabet Σ , that is the set of all finite linear combinations of matrices of the form $\mathbf{A}^x = \mathbf{A}^{x_1} \mathbf{A}^{x_2} \dots \mathbf{A}^{x_k}$ for $x = x_1 x_2 \dots x_k \in \Sigma^*$. More formally, if we denote by \mathcal{A} this algebra, we have

$$\mathcal{A} = \left\{ \sum_{i=1}^n \alpha_i \mathbf{A}^{w_i} : n \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{F}, w_1, \dots, w_n \in \Sigma^* \right\}.$$

Let \mathcal{A} be a finite-dimensional algebra over \mathbb{F} . A sub-algebra \mathcal{X} of \mathcal{A} is called an *ideal* of \mathcal{A} if both $xa \in \mathcal{X}$ and $ax \in \mathcal{X}$ for any $x \in \mathcal{X}$, $a \in \mathcal{A}$ (i.e. \mathcal{X} is both left and right \mathcal{A} -invariant), which we will denote by $\mathcal{A}\mathcal{X} = \mathcal{X}\mathcal{A} = \mathcal{A}$. A sub-algebra \mathcal{X} of \mathcal{A} is *nilpotent* if there exists some integer k such that $\mathcal{X}^k = \{x_1 x_2 \dots x_k : x_i \in \mathcal{X}, i \in [k]\} = \{0\}$. The *factor algebra* \mathcal{A}/\mathcal{X} of an algebra \mathcal{A} by an ideal \mathcal{X} is the algebra consisting of all cosets $a + \mathcal{X}$ for $a \in \mathcal{A}$, in other words \mathcal{A}/\mathcal{X} is the quotient of \mathcal{A} by the equivalence relation ($a \sim b$ if and only if $a - b \in \mathcal{X}$). The *radical*⁵ of \mathcal{A} is the maximal nilpotent ideal of \mathcal{A} and will be denoted by $\text{Rad}(\mathcal{A})$ (the existence of $\text{Rad}(\mathcal{A})$ follows from the fact that \mathcal{A} is of finite dimension). An algebra \mathcal{A} is *semi-simple* if its radical is $\{0\}$.

Let us illustrate these definitions with a very simple example. Let $\mathcal{G} \subset \mathcal{M}_2(\mathbb{R})$ be the algebra generated by the matrix $\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. One can easily check that

$$\mathcal{G} = \left\{ \begin{bmatrix} \alpha & \beta \\ 0 & \alpha \end{bmatrix} : \alpha, \beta \in \mathbb{R} \right\} \text{ and is thus of dimension 2. Consequently, both}$$

$$\mathcal{G}_1 = \left\{ \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} : \alpha \in \mathbb{R} \right\} \text{ and } \mathcal{G}_2 = \left\{ \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} : \beta \in \mathbb{R} \right\} \tag{1}$$

are sub-algebras of \mathcal{G} . Moreover, \mathcal{G}_2 is a nilpotent ideal and one can check that it is maximal, i.e. $\text{Rad}(\mathcal{G}) = \mathcal{G}_2$ and hence \mathcal{G} is not semi-simple.

Intuitively, the radical of an algebra \mathcal{A} contains its *bad elements* (in the sense that these elements annihilate all simple \mathcal{A} -modules). In our previous example, this *bad* property translates into the fact that the non-zero elements of \mathcal{G}_2 cannot be diagonalized. We will use two fundamental results from the theory of finite dimensional algebra. The first one is the Wedderburn-Malcev theorem which states that (under some conditions on the ground field \mathbb{F}) the elements of the radical can be *filtered out* from the algebra, i.e. one can find a sub-algebra of \mathcal{A} that is isomorphic to $\mathcal{A}/\text{Rad}(\mathcal{A})$ (see e.g. [17, Theorem 6.2.3]).

Theorem 1 (Wedderburn-Malcev Theorem). *Let \mathcal{A} be a finite-dimensional algebra over a field of characteristic 0. There exists a semi-simple subalgebra $\tilde{\mathcal{A}}$ of \mathcal{A} which is isomorphic to $\mathcal{A}/\text{Rad}(\mathcal{A})$ and such that $\mathcal{A} = \tilde{\mathcal{A}} \oplus \text{Rad}(\mathcal{A})$ (direct sum of vector spaces).*

Going back to the example of the algebra \mathcal{G} described above, we showed that it is not semi-simple, however one can easily check that $\mathcal{G}/\text{Rad}(\mathcal{G})$ is isomorphic to the algebra \mathcal{G}_1 in Eq. (1) which is semi-simple, and furthermore that $\mathcal{G} = \mathcal{G}_1 \oplus \text{Rad}(\mathcal{G})$.

⁵ Note that this definition is specific to the finite-dimensional case; for general rings, there exist distinct non-equivalent definitions of radicals, which all agree with the one given here in the case of finite-dimensional algebras.

The second fundamental result we will need is related to the notion of representation of an algebra. A *representation* of an \mathbb{F} -algebra \mathcal{A} is a homomorphism of \mathcal{A} into the algebra $\mathcal{L}(V)$ of the linear operators on some vector space V (over \mathbb{F}). Two representations $\rho: \mathcal{A} \rightarrow \mathcal{L}(V)$ and $\tau: \mathcal{A} \rightarrow \mathcal{L}(W)$ are *similar* if there exists an isomorphism $\phi: V \rightarrow W$ such that $\rho(a) = \phi^{-1}\tau(a)\phi$ for all $a \in \mathcal{A}$. For semi-simple algebras, the notion of similar representations is fundamentally related to the trace operator, which will be particularly relevant to the present study. Formally, we have the following theorem (see e.g. [17, Corollary 2.6.3]).

Theorem 2. *Let ρ and τ be two representations of a semi-simple algebra \mathcal{A} over a field of characteristic 0. These representations are similar if and only if $\text{Tr}(\rho(a)) = \text{Tr}(\tau(a))$ for all $a \in \mathcal{A}$.*

3 Semi-Simple GWMs and the Equivalence Problem

In this section, we study the equivalence problem: given two GWMs over circular strings, how can we decide whether they compute the same function? In light of Proposition 1, one could solve this problem by simply *converting* the two GWM^cs into WFAs and checking whether these two WFAs compute the same function; indeed the equivalence problem for WFAs defined over a field is decidable in polynomial time [9]. Nonetheless, we will tackle this problem without relying on this proposition and, by doing so, we will unravel the notion of *semi-simple GWM^c* which will be relevant to the study of the minimization problem in the next section (and which should also be central to the study of GWMs defined over more general families of graphs).

3.1 Semi-Simplicity, Nilpotent Matrices and Traces

Let \mathcal{A} be a finite dimensional matrix algebra. Recall that the radical of \mathcal{A} is its maximal nilpotent ideal. A useful characterization of the elements of the radical relies on the notion of strongly nilpotent elements: $\mathbf{A} \in \mathcal{A}$ is *strongly nilpotent* if $\mathbf{A}\mathbf{X}$ is nilpotent for any $\mathbf{X} \in \mathcal{A}$. It turns out that the radical of \mathcal{A} is exactly the set of its strongly nilpotent elements [17, Corollary 3.1.10]. Since the computation of a GWM^c boils down to applying the trace operator, we will leverage this property to relate the notions of radical and semi-simplicity to simple properties of the elements of \mathcal{A} with respect to the trace operator. We start with a simple lemma relating nilpotency and trace.

Lemma 1. *Let \mathbb{F} be a field of characteristic 0 and let $\mathbf{A} \in \mathcal{M}_d(\mathbb{F})$. Then \mathbf{A} is nilpotent if and only if $\text{Tr}(\mathbf{A}^n) = 0$ for all $n \geq 1$.*

Proof. Let \mathbf{A} be a nilpotent matrix and let k be such that $\mathbf{A}^k = 0$. Suppose $\mathbf{A}\mathbf{v} = \gamma\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{0}$ (where γ could belong to an algebraically closed field extension of \mathbb{F}). Then $\mathbf{A}^k\mathbf{v} = \gamma^k\mathbf{v} = 0$ hence $\gamma = 0$ since \mathbb{F} is of characteristic 0, thus \mathbf{A} has only 0 eigenvalues and $\text{Tr}(\mathbf{A}^n) = 0$ for all $n \geq 1$.

Conversely, suppose that $\text{Tr}(\mathbf{A}^n) = 0$ for all $n \geq 1$. Then, we have $\text{Tr}(P(\mathbf{A})) = 0$ for any polynomial P with constant term 0. Suppose that \mathbf{A}

has a non-zero eigenvalue γ and let $m > 0$ be its multiplicity. Choose a polynomial P such that $P(\gamma) = 1$, $P(0) = 0$ and $P(\mu) = 0$ for any eigenvalue μ of \mathbf{A} distinct from γ . We then have $0 = \text{Tr}(P(\mathbf{A})) = m$, a contradiction. Hence \mathbf{A} has only zero eigenvalues and is nilpotent. \square

One can use the previous lemma to show that an element $\mathbf{A} \in \mathcal{A}$ is strongly nilpotent if and only if $\text{Tr}(\mathbf{A}\mathbf{X}) = 0$ for all $\mathbf{X} \in \mathcal{A}$, which leads to the following useful characterization of the semi-simplicity of an algebra.

Proposition 2. *Let $\mathcal{A} \subset \mathcal{M}_d(\mathbb{F})$ be a matrix algebra. We have*

$$\text{Rad}(\mathcal{A}) = \{\mathbf{A} \in \mathcal{A} : \text{Tr}(\mathbf{A}\mathbf{X}) = 0 \text{ for all } \mathbf{X} \in \mathcal{A}\}.$$

Consequently, \mathcal{A} is semi-simple if and only if for all $\mathbf{A} \in \mathcal{A}$ different from 0 there exists $\mathbf{X} \in \mathcal{A}$ such that $\text{Tr}(\mathbf{A}\mathbf{X}) \neq 0$.

Proof. We will show that $\mathbf{A} \in \mathcal{A}$ is strongly nilpotent if and only if $\text{Tr}(\mathbf{A}\mathbf{X}) = 0$ for all $\mathbf{X} \in \mathcal{A}$. The proposition will then directly follows from the fact that $\text{Rad}(\mathcal{A})$ is the set of strongly nilpotent elements of \mathcal{A} and from the fact that \mathcal{A} is semi-simple if and only if $\text{Rad}(\mathcal{A}) = \{0\}$.

Let $\mathbf{A} \in \mathcal{A}$ be such that $\text{Tr}(\mathbf{A}\mathbf{X}) = 0$ for all $\mathbf{X} \in \mathcal{A}$. Since $\mathbf{X}(\mathbf{A}\mathbf{X})^{n-1} \in \mathcal{A}$ for all $n \geq 1$ and all $\mathbf{X} \in \mathcal{A}$ we have $\text{Tr}((\mathbf{A}\mathbf{X})^n) = 0$ for all $n \geq 1$ and all $\mathbf{X} \in \mathcal{A}$, hence $\mathbf{A}\mathbf{X}$ is nilpotent for all $\mathbf{X} \in \mathcal{A}$ by Lemma 1, i.e. \mathbf{A} is strongly nilpotent. Conversely, let \mathbf{A} be a strongly nilpotent element of \mathcal{A} . By Lemma 1 we have $\text{Tr}((\mathbf{A}\mathbf{X})^n) = 0$ for all $\mathbf{X} \in \mathcal{A}$ and all $n \geq 1$, in particular $\text{Tr}(\mathbf{A}\mathbf{X}) = 0$. \square

3.2 Equivalence of GWMs

We now consider the problem of deciding whether two GWM^cs are equivalent. Let us first briefly show how one can decide whether two real-valued WFAs compute the same function. One way to address this problem relies on the following result: two minimal real-valued WFAs computing the same function are related by a change of basis. Note that it is easy to check that WFAs are invariant under a change of basis of their weight vectors and transition matrices. The following proposition show that such a change of basis is actually the only way for two minimal WFAs to compute the same function [26] (see also [6, Corollary 4.2]).

Proposition 3. *If two WFAs $A = (\alpha, \{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}, \omega)$ and $\tilde{A} = (\tilde{\alpha}, \{\tilde{\mathbf{A}}^\sigma\}_{\sigma \in \Sigma}, \tilde{\omega})$ with d states taking their values in \mathbb{R} are minimal and compute the same function, i.e. $f_A = f_{\tilde{A}}$, then there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_d(\mathbb{R})$ such that*

$$\alpha^\top = \tilde{\alpha}^\top \mathbf{P}, \quad \omega = \mathbf{P}^{-1} \tilde{\omega} \quad \text{and} \quad \mathbf{A}^\sigma = \mathbf{P}^{-1} \tilde{\mathbf{A}}^\sigma \mathbf{P} \text{ for each } \sigma \in \Sigma.$$

Hence, to decide whether two WFAs compute the same function one can simply minimize them and check whether the weight vectors and transition matrices obtained after minimization are related by a change of basis (which can both be done in polynomial time). In contrast, one can easily find an example of two minimal GWM^cs whose matrices are not related by a change of basis. Consider

the constant function $f(x) = 2$ for all $x \in \Sigma^+$. One can check that the two GWM^cs G and \tilde{G} with 2 states defined by the matrices

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

respectively are minimal and compute f , however \mathbf{G} and $\tilde{\mathbf{G}}$ are not similar.

Let us now introduce the notion of *semi-simple GWM^c*. We say that a GWM^c A defined by a set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$ is *semi-simple* if the algebra \mathcal{A} generated by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ is semi-simple. It follows from the example presented in Sect. 2.2 that G is not semi-simple while \tilde{G} is a semi-simple GWM^c computing the *GWM^c-recognizable function* f . We will now show that this simple example can be generalized: *any GWM^c-recognizable function can be computed by a semi-simple GWM^c*. This non-trivial result relies on the following theorem which is a direct consequence of the Wedderburn-Malcev theorem.

Theorem 3. *Let $\mathcal{A} \subset \mathcal{M}_d(\mathbb{F})$ be a matrix algebra over a field of characteristic 0. Then there exist a semi-simple sub-algebra $\tilde{\mathcal{A}}$ of \mathcal{A} and a surjective homomorphism $\pi: \mathcal{A} \rightarrow \tilde{\mathcal{A}}$ such that $\text{Tr}(\mathbf{A}) = \text{Tr}(\pi(\mathbf{A}))$ for all $\mathbf{A} \in \mathcal{A}$.*

Proof. By Theorem 1 there exists a semi-simple sub-algebra $\tilde{\mathcal{A}}$ of \mathcal{A} which is isomorphic to $\mathcal{A}/\text{Rad}(\mathcal{A})$ and such that $\mathcal{A} = \tilde{\mathcal{A}} \oplus \text{Rad}(\mathcal{A})$ (direct sum of vector spaces). Let $\pi: \mathcal{A} \rightarrow \tilde{\mathcal{A}}$ be the projection associated with this direct sum. Then for any $\mathbf{A} \in \mathcal{A}$ we have

$$\text{Tr}(\mathbf{A}) = \text{Tr}(\pi(\mathbf{A}) + (1 - \pi)(\mathbf{A})) = \text{Tr}(\pi(\mathbf{A})) + \text{Tr}((1 - \pi)(\mathbf{A})) = \text{Tr}(\pi(\mathbf{A})).$$

Indeed, since $(1 - \pi)(\mathbf{A}) \in \text{Rad}(\mathcal{A})$, it is nilpotent, hence its trace is zero. □

Using the notations from Theorem 3, it follows that for any d -dimensional GWM^c A given by a set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$ generating the algebra \mathcal{A} , the d -dimensional \tilde{A} given by the matrices $\{\tilde{\mathbf{A}}^\sigma = \pi(\mathbf{A}^\sigma)\}_{\sigma \in \Sigma}$ is a semi-simple GWM^c computing the function f_A , hence the following corollary.

Corollary 1. *Any function that can be computed by a GWM^c can be computed by a semi-simple GWM^c of the same dimension.*

Given a finite dimensional algebra \mathcal{A} , one can compute the surjective homomorphism π from Theorem 3 in polynomial time when \mathbb{F} allows efficient arithmetic computations (e.g. $\mathbb{F} = \mathbb{Q}$) [12, 15]. The algorithm takes as input a basis a_1, \dots, a_n of \mathcal{A} (as a vector space) and the structure coefficients of the algebra (which are the scalars $c_{i,j}^k \in \mathbb{F}$ satisfying $a_i a_j = \sum_k c_{i,j}^k a_k$). Since one can easily compute a basis and the structure coefficients of a matrix algebra \mathcal{A} given a set of generators $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ in polynomial time, it follows that any GWM^c can be transformed in polynomial time into a semi-simple GWM^c (of the same dimension) computing the same function.

We now show that a result similar to Proposition 3 holds for semi-simple GWM^cs: *two semi-simple d -dimensional GWM^cs are equivalent if and only if they are related by a change of basis*. This result relies on the following theorem.

Theorem 4. *Let Σ be a finite alphabet and let $\mathcal{A}, \mathcal{B} \subset \mathcal{M}_d(\mathbb{F})$ be the algebras generated by the sets of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ and $\{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}$ respectively.*

If \mathcal{A} and \mathcal{B} are semi-simple and $\text{Tr}(\mathbf{A}^w) = \text{Tr}(\mathbf{B}^w)$ for all $w \in \Sigma^$ then \mathcal{A} is isomorphic to \mathcal{B} . Moreover, the mapping $\tilde{\phi}: \mathcal{A} \rightarrow \mathcal{B}$ defined by extending the mapping*

$$\phi: \mathbf{A}^x \mapsto \mathbf{B}^x \quad \text{for all } x \in \Sigma^*$$

by linearity is well-defined and is an isomorphism.

Proof. The mapping ϕ is by construction a trace-preserving surjective semi-group homomorphism. We first show⁶ that ϕ can be extended to a homomorphism $\tilde{\phi}: \mathcal{A} \rightarrow \mathcal{B}$. By definition, any $\mathbf{A} \in \mathcal{A}$ can be written as $\mathbf{A} = \sum_{i=1}^n \alpha_i \mathbf{A}^{x_i}$ for some $n \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{F}, x_1, \dots, x_n \in \Sigma^*$. We will show that the mapping

$$\tilde{\phi}: \sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} \mapsto \sum_{i=1}^n \alpha_i \phi(\mathbf{A}^{x_i})$$

is well-defined. By construction of $\tilde{\phi}$, it suffices to show that if $\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} = 0$ for some $\alpha_i \in \mathbb{F}, x_i \in \Sigma^*$, then $\tilde{\phi}(\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i}) = 0$. Suppose $\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} = 0$, then $\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} \mathbf{A}^x = 0$ for any $x \in \Sigma^*$. By linearity of the trace and since ϕ is a trace-preserving morphism, it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^n \alpha_i \text{Tr}[\mathbf{A}^{x_i} \mathbf{A}^x] = \sum_{i=1}^n \alpha_i \text{Tr}[\phi(\mathbf{A}^{x_i} \mathbf{A}^x)] = \sum_{i=1}^n \alpha_i \text{Tr}[\phi(\mathbf{A}^{x_i}) \phi(\mathbf{A}^x)] \\ &= \text{Tr} \left[\left(\sum_{i=1}^n \alpha_i \phi(\mathbf{A}^{x_i}) \right) \phi(\mathbf{A}^x) \right] = \text{Tr} \left[\tilde{\phi} \left(\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} \right) \phi(\mathbf{A}^x) \right] \end{aligned}$$

for all $x \in \Sigma^*$. By linearity of the trace and since ϕ is surjective, we thus have $\text{Tr} \left[\tilde{\phi} \left(\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} \right) \mathbf{B} \right] = 0$ for any $\mathbf{B} \in \mathcal{B}$, hence $\tilde{\phi} \left(\sum_{i=1}^n \alpha_i \mathbf{A}^{x_i} \right)$ belongs to $\text{Rad}(\mathcal{B})$ by Proposition 2 and must be 0 since \mathcal{B} is semi-simple.

One can easily check that $\tilde{\phi}$ is trace-preserving, is surjective and is a homomorphism. It remains to show that $\tilde{\phi}$ is injective. Let $\mathbf{A} \in \mathcal{A}$ be such that $\tilde{\phi}(\mathbf{A}) = 0$. Since $\tilde{\phi}$ is a homomorphism we have $\tilde{\phi}(\mathbf{A}\mathbf{X}) = 0$ for any $\mathbf{X} \in \mathcal{A}$, and thus $0 = \text{Tr}(\tilde{\phi}(\mathbf{A}\mathbf{X})) = \text{Tr}(\mathbf{A}\mathbf{X})$ for all $\mathbf{X} \in \mathcal{A}$. Hence $\mathbf{A} \in \text{Rad}(\mathcal{A})$ by Proposition 2 and must be 0 since \mathcal{A} is semi-simple. □

The previous theorem can be leveraged to show that if two semi-simple GWM^cs of the same dimension compute the same function, then they are related by a change of basis (note that the converse of this statement is immediate since the trace is a basis independent operator). Let A and B be two d -dimensional semi-simple GWM^cs computing the same function and let $\mathcal{A}, \mathcal{B} \subset \mathcal{M}_d$ be the algebras generated by their respective sets of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ and $\{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}$. First observe that the identity mapping $\rho: \mathcal{A} \rightarrow \mathcal{L}(\mathbb{F}^d)$ defined by $\rho(\mathbf{A}) = \mathbf{A}$ for all $\mathbf{A} \in \mathcal{A}$ is (trivially) a representation of the algebra \mathcal{A} . Now, since A and B

⁶ This part of the proof is adapted from the proof of Proposition 3.1 in [18].

compute the same function and are semi-simple, we have $\text{Tr}(\mathbf{A}^w) = \text{Tr}(\mathbf{B}^w)$ for all $w \in \Sigma^*$ and it follows from Theorem 4 that \mathcal{A} is isomorphic to \mathcal{B} ; let $\tilde{\phi}: \mathcal{A} \rightarrow \mathcal{B}$ be the isomorphism defined in this theorem. Then, the mapping $\tau: \mathcal{A} \rightarrow \mathcal{L}(\mathbb{F}^d)$ defined by $\tau(\mathbf{A}) = \tilde{\phi}(\mathbf{A})$ for all $\mathbf{A} \in \mathcal{A}$ is also a representation of \mathcal{A} , and since \mathcal{A} is semi-simple it follows from Theorem 2 that ρ and τ are similar. That is, there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_d(\mathbb{F})$ such that $\rho(\mathbf{A}) = \mathbf{P}^{-1}\tau(\mathbf{A})\mathbf{P}$ for all $\mathbf{A} \in \mathcal{A}$. In particular we have

$$\mathbf{A}^\sigma = \rho(\mathbf{A}^\sigma) = \mathbf{P}^{-1}\tau(\mathbf{A}^\sigma)\mathbf{P} = \mathbf{P}^{-1}\tilde{\phi}(\mathbf{A}^\sigma)\mathbf{P} = \mathbf{P}^{-1}\mathbf{B}^\sigma\mathbf{P}$$

for all $\sigma \in \Sigma$, hence the following corollary.

Corollary 2. *Two d -dimensional semi-simple GWM^cs A and B compute the same function if and only if they are related by a change of basis, i.e. there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_d(\mathbb{F})$ such that $\mathbf{A}^\sigma = \mathbf{P}^{-1}\mathbf{B}^\sigma\mathbf{P}$ for all $\sigma \in \Sigma$.*

In the case where \mathbb{F} allows for efficient arithmetic computations (e.g. $\mathbb{F} = \mathbb{Q}$), it follows that the equivalence of GWM^cs can be decided in polynomial time. Indeed, given two GWM^cs A and B of the same dimension defined by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ and $\{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}$ respectively, one can first transform them into semi-simple GWM^cs using Theorem 3 and the algorithm in [12, 15], and then check whether the resulting matrices are related by a change of basis. The case where the two GWM^cs are not of the same dimension can be easily handled. Without loss of generality, suppose that A and B are semi-simple GWM^cs of dimension d and d' respectively with $d' < d$. One can construct a d -dimensional GWM^c \tilde{B} computing the same function as B by considering the block-diagonal matrices $\tilde{\mathbf{B}}^\sigma = \mathbf{B}^\sigma \oplus \mathbf{0}$ for each $\sigma \in \Sigma$ (where $\mathbf{0}$ is the $(d - d') \times (d - d')$ matrix with all entries equal to 0). It is easy to check that \tilde{B} is semi-simple if B is semi-simple, hence one can decide if A is equivalent to B by checking whether the matrices \mathbf{A}^σ and $\tilde{\mathbf{B}}^\sigma$ are related by a change of basis.

4 Minimization of GWMs over Circular Strings

We now consider the minimization problem: given a GWM^c A , can we find a minimal GWM^c computing f_A ? We will show that the answer is in the positive and that such a minimal GWM^c can be computed in polynomial time. We start with a technical lemma that generalizes the classical result stating that for any $d \times d$ matrix \mathbf{A} , the kernel of \mathbf{A}^d is equal to the kernel of \mathbf{A}^{d+k} for any $k \geq 0$.

Lemma 2. *Let $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$ be a finite set of matrices. Then for all $k \geq 0$ we have*

$$\bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x) = \bigcap_{y \in \Sigma^{d+k}} \ker(\mathbf{A}^y).$$

Proof. For any integer i , let $E_i = \bigcap_{x \in \Sigma^i} \ker(\mathbf{A}^x)$. We start by showing that if $E_i = E_{i+1}$ for some i then $E_{i+1} = E_{i+2}$. The inclusion $E_{i+1} \subseteq E_{i+2}$ is immediate. Suppose $E_i = E_{i+1}$ for some integer i . If $\mathbf{v} \in E_{i+2}$ then $\mathbf{A}^\sigma \mathbf{v} \in$

$\ker(\mathbf{A}^x)$ for all $x \in \Sigma^{i+1}$ and all $\sigma \in \Sigma$, i.e. $\mathbf{A}^\sigma \mathbf{v} \in E_{i+1} = E_i$ for all $\sigma \in \Sigma$, which implies $\mathbf{A}^\sigma \mathbf{v} \in \ker(\mathbf{A}^y)$ for all $y \in \Sigma^i$ and all $\sigma \in \Sigma$ from which $\mathbf{v} \in E_{i+1}$ follows directly. To conclude, since each E_i is a linear subspace of \mathbb{F}^d , $E_i \subsetneq E_{i+1}$ implies $\dim E_i < \dim E_{i+1}$, hence there must exist an i for which $E_i = E_{i+1}$ and this i cannot be greater than d . \square

We show in the following theorem that the linear space $E = \bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x)$ is not relevant to the computation of a GWM^c A with matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$, i.e. one can project each matrix \mathbf{A}^x onto the orthogonal complement of E without changing the function computed by A .

Theorem 5. *Let A be a GWM^c given by the set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$. Consider the linear space*

$$E = \bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x) = \{\mathbf{v} \in \mathbb{F}^d : \mathbf{A}^x \mathbf{v} = \mathbf{0} \text{ for all } x \in \Sigma^d\}$$

and let $\mathbf{\Pi} \in \mathbb{F}^{d \times d}$ be the matrix of the orthogonal projection onto E .

Then, the GWM^c \hat{A} given by the matrices $\hat{\mathbf{A}}^\sigma = \mathbf{A}^\sigma(\mathbf{I} - \mathbf{\Pi})$ for each $\sigma \in \Sigma$ is such that $f_A = f_{\hat{A}}$.

Proof. Let \mathcal{A} be the algebra generated by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$. Let us first observe that E is \mathcal{A} -invariant, which follows from Lemma 2. Indeed, if $\mathbf{v} \in E$ and $y \in \Sigma^*$ we have $\mathbf{A}^x \mathbf{A}^y \mathbf{v} = \mathbf{0}$ for any $x \in \Sigma^d$ (since $|xy| \geq d$), hence $\mathbf{A}^y \mathbf{v} \in E$; the extension to an arbitrary element of \mathbf{A} is immediate by linearity. This implies that for any $\mathbf{A} \in \mathcal{A}$, we have

$$\mathbf{\Pi} \mathbf{A} \mathbf{\Pi} = \mathbf{A} \mathbf{\Pi} \quad \text{and} \quad (\mathbf{I} - \mathbf{\Pi}) \mathbf{A} \mathbf{\Pi} = \mathbf{0}. \tag{2}$$

Now, let $k \geq 1$, let $x = x_1 x_2 \cdots x_k \in \Sigma^k$ and let $\mathbf{P}_1 = \mathbf{\Pi}$ and $\mathbf{P}_2 = \mathbf{I} - \mathbf{\Pi}$. We can decompose \mathbf{A}^x into

$$\begin{aligned} \mathbf{A}^x &= \prod_{i=1}^k \mathbf{A}^{x_i} = \prod_{i=1}^k \mathbf{A}^{x_i} (\mathbf{P}_1 + \mathbf{P}_2) = \sum_{j_1, \dots, j_k \in \{1,2\}} \mathbf{A}^{x_1} \mathbf{P}_{j_1} \mathbf{A}^{x_2} \mathbf{P}_{j_2} \cdots \mathbf{A}^{x_k} \mathbf{P}_{j_k} \\ &= \hat{\mathbf{A}}^x + \mathbf{A}^{x_1} \mathbf{\Pi} \mathbf{A}^{x_2} \mathbf{\Pi} \cdots \mathbf{A}^{x_k} \mathbf{\Pi} + \sum_{\substack{j_1, \dots, j_k \in \{1,2\} \text{ s.t.} \\ \exists r, r' : j_r \neq j_{r'}}} \mathbf{A}^{x_1} \mathbf{P}_{j_1} \mathbf{A}^{x_2} \mathbf{P}_{j_2} \cdots \mathbf{A}^{x_k} \mathbf{P}_{j_k}. \end{aligned}$$

We will show that the traces of all the summands in this last expression, except for the first one, are equal to 0. First, using Eq. (2) we have $\mathbf{A}^{x_1} \mathbf{\Pi} \mathbf{A}^{x_2} \mathbf{\Pi} \cdots \mathbf{A}^{x_k} \mathbf{\Pi} = \mathbf{A}^x \mathbf{\Pi}$. Moreover, for any integer s such that $sk \geq d$ we have $(\mathbf{A}^x \mathbf{\Pi})^s = \mathbf{A}^{x^s} \mathbf{\Pi} = \mathbf{0}$ by definition of E and by Lemma 2, thus $\mathbf{A}^x \mathbf{\Pi}$ is nilpotent and its trace is 0 by Lemma 1. For the remaining terms, let $j_1, \dots, j_k \in \{1,2\}$ not all equal. Let $l \in [k]$ be an index such that $j_l = 2$ and $j_{l+1} = 1$ where $l+1 = l+1$ if $l < k$ and 1 otherwise. Using the invariance of the trace under cyclic permutations of a matrix product, we obtain

$$\begin{aligned} \text{Tr}(\mathbf{A}^{x_1} \mathbf{P}_{j_1} \mathbf{A}^{x_2} \mathbf{P}_{j_2} \cdots \mathbf{A}^{x_k} \mathbf{P}_{j_k}) &= \text{Tr}(\mathbf{A}^{x_l} \mathbf{P}_{j_l} \mathbf{A}^{x_{l+1}} \mathbf{P}_{j_{l+1}} \cdots) \\ &= \text{Tr}(\mathbf{A}^{x_l} (\mathbf{I} - \mathbf{\Pi}) \mathbf{A}^{x_{l+1}} \mathbf{\Pi} \cdots) = 0 \end{aligned}$$

where we used Eq. 2 again for the last equality. To conclude, we have shown that $\text{Tr}(\mathbf{A}^x) = \text{Tr}(\hat{\mathbf{A}}^x)$ for all $x \in \Sigma^*$, hence A and \hat{A} compute the same function on circular strings. \square

Moreover, we now show that the subspace E from the previous theorem can be used to obtain a characterization of the minimality of a GWM^c .

Theorem 6. *Let A be a GWM^c given by the set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$. Then, A is minimal if and only if the linear space*

$$E = \bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x) = \{\mathbf{v} \in \mathbb{F}^d : \mathbf{A}^x \mathbf{v} = \mathbf{0} \text{ for all } x \in \Sigma^d\}$$

is trivial, i.e. $E = \{\mathbf{0}\}$.

Proof. Suppose that E is not trivial and let $\mathbf{\Pi}$ be the matrix of the orthogonal projection onto E . Then, the rank R of $\mathbf{I} - \mathbf{\Pi}$ is strictly less than d and there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{d \times R}$ such that $\mathbf{I} - \mathbf{\Pi} = \mathbf{U}\mathbf{U}^\top$. It follows from the previous proposition that, for any non-empty word $x = x_1 \cdots x_k$, we have

$$\begin{aligned} \text{Tr}(\mathbf{A}^x) &= \text{Tr}(\mathbf{A}^{x_1}(\mathbf{I} - \mathbf{\Pi})\mathbf{A}^{x_2}(\mathbf{I} - \mathbf{\Pi}) \cdots \mathbf{A}^{x_k}(\mathbf{I} - \mathbf{\Pi})) \\ &= \text{Tr}(\mathbf{A}^{x_1}\mathbf{U}\mathbf{U}^\top \mathbf{A}^{x_2}\mathbf{U}\mathbf{U}^\top \cdots \mathbf{A}^{x_k}\mathbf{U}\mathbf{U}^\top) = \text{Tr}((\mathbf{U}^\top \mathbf{A}^{x_1}\mathbf{U})(\mathbf{U}^\top \mathbf{A}^{x_2}\mathbf{U}) \cdots (\mathbf{U}^\top \mathbf{A}^{x_k}\mathbf{U})). \end{aligned}$$

Hence, the R -dimensional GWM^c given by the matrices $\hat{\mathbf{A}}^\sigma = \mathbf{U}^\top \mathbf{A}^\sigma \mathbf{U}$ computes the same function as A , showing that A is not minimal.

Suppose now that A is not minimal. Let B be a GWM^c of dimension $d' < d$, given by the matrices $\{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}$, such that $f_B = f_A$. Let \mathcal{A} (resp. \mathcal{B}) be the algebra generated by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$ (resp. $\{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}$). By Corollary 1, we can assume that both A and B are semi-simple GWM^c s, i.e. that the algebras \mathcal{A} and \mathcal{B} are semi-simple. For each $\sigma \in \Sigma$, let $\hat{\mathbf{B}}^\sigma = \mathbf{B}^\sigma \oplus \mathbf{0} \in \mathbb{R}^{d \times d}$ be the block diagonal matrix having \mathbf{B}^σ in the upper diagonal block and 0's elsewhere. Let $\hat{\mathcal{B}}$ be the algebras generated by the matrices $\{\hat{\mathbf{B}}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$. It is easy to check that the GWM^c \hat{B} computes the same function as A and B and that the algebra $\hat{\mathcal{B}}$ is semi-simple (it is indeed isomorphic to the semi-simple algebra \mathcal{B}). It then follows from Corollary 2 that there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_d(\mathbb{F})$ such that $\mathbf{A}^\sigma = \mathbf{P}\hat{\mathbf{B}}^\sigma\mathbf{P}^{-1}$ for all $\sigma \in \Sigma$. Let \mathbf{e}_d be the d th vector of the canonical basis of \mathbb{F}^d , by definition of $\hat{\mathbf{B}}^\sigma$ we have $\hat{\mathbf{B}}^\sigma \mathbf{e}_d = \mathbf{0}$ for any $\sigma \in \Sigma$, and consequently $\mathbf{A}^\sigma \mathbf{P}\mathbf{e}_d = \mathbf{0}$ for any symbol σ , showing that $\mathbf{P}\mathbf{e}_d \in E$ and $E \neq \{\mathbf{0}\}$. \square

It follows from the two previous theorems that by restricting the linear operators \mathbf{A}^σ of a GWM^c A to the subspace E^\perp , one can obtain a minimal GWM^c computing f_A . We formally state this result in the following corollary.

Corollary 3. *Let A be a GWM^c given by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$ and let $\mathbf{\Pi}$ be the matrix of the orthogonal projection onto the space $E = \bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x)$. For any orthogonal matrix $\mathbf{U} \in \mathbb{F}^{d \times R}$ such that $\mathbf{I} - \mathbf{\Pi} = \mathbf{U}\mathbf{U}^\top$ (where R is the dimension of E^\perp), the R -dimensional GWM^c \hat{A} given by the matrices $\hat{\mathbf{A}}^\sigma = \mathbf{U}^\top \mathbf{A}^\sigma \mathbf{U}$ is a minimal GWM^c computing f_A .*

Proof. Using the invariance of the trace under cyclic permutations of a matrix product, it directly follows from Theorem 5 that $f_{\hat{A}} = f_A$. Moreover, one can check that $\hat{E} = \bigcap_{x \in \Sigma^d} \ker(\hat{\mathbf{A}}^x) = \{\mathbf{0}\}$ by construction of the matrices $\hat{\mathbf{A}}^\sigma$, hence \hat{A} is minimal by Theorem 6. \square

We showed that a GWM^c can be minimized by restricting its matrices to the subspace E^\perp . In order to do so, one needs to compute a basis of $E = \bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x)$. This can naively be done by first computing $\ker(\mathbf{A}^x)$ for each $x \in \Sigma^d$ and then computing a basis for the intersection of these linear subspaces, however the complexity of this approach is exponential in the dimension d . We show in the following proposition that for semi-simple GWM^c s, one simply needs to compute a basis of the space $\bigcap_{\sigma \in \Sigma} \ker(\mathbf{A}^\sigma)$, which can be done in polynomial time (provided that the field \mathbb{F} admits efficient symbolic arithmetic, e.g. $\mathbb{F} = \mathbb{Q}$).

Proposition 4. *Let $\mathcal{A} \subset \mathcal{M}_d(\mathbb{F})$ be the finite dimensional algebra generated by the set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$. Then if \mathcal{A} is semi-simple we have*

$$\bigcap_{x \in \Sigma^d} \ker(\mathbf{A}^x) = \bigcap_{\sigma \in \Sigma} \ker(\mathbf{A}^\sigma).$$

Proof. For any integer $i \geq 1$, let $E_i = \bigcap_{x \in \Sigma^i} \ker(\mathbf{A}^x)$. Recall from the proof of Lemma 2 that $E_i \subset E_{i+1}$ for all i and that $E_i = E_{i+1}$ implies $E_i = E_{i+k}$ for any integer $k \geq 0$, hence it will be sufficient to show that $E_1 = E_2$. One can check that each E_i is \mathcal{A} -invariant, i.e. each E_i is an \mathcal{A} -module. Since \mathcal{A} is semi-simple, any \mathcal{A} -module is semi-simple [17, Theorem 2.6.2], which implies that if M is an \mathcal{A} -module, every submodule U of M has a complement [17, Proposition 2.2.1], i.e. there exists an \mathcal{A} -module V such that $M = U \oplus V$. Now since E_1 is a submodule of the \mathcal{A} -module E_2 , E_1 has a complement U in E_2 , i.e. U is \mathcal{A} -invariant and $E_2 = E_1 \oplus U$. Let $\mathbf{v} \in U$. We show $\mathbf{v} = \mathbf{0}$. Since $\mathbf{v} \in E_2$, we have $\mathbf{A}^{\sigma_1} \mathbf{A}^{\sigma_2} \mathbf{v} = \mathbf{0}$ for all $\sigma_1, \sigma_2 \in \Sigma$, hence $\mathbf{A}^\sigma \mathbf{v} \in E_1$ for all $\sigma \in \Sigma$. Moreover, we have $\mathbf{A}^\sigma \mathbf{v} \in U$ for all $\sigma \in \Sigma$ since U is \mathcal{A} -invariant. It follows that $\mathbf{A}^\sigma \mathbf{v} \in E_1 \cap U = \{\mathbf{0}\}$ and $\mathbf{A}^\sigma \mathbf{v} = \mathbf{0}$ for all $\sigma \in \Sigma$, hence $\mathbf{v} \in E_1$ and since $\mathbf{v} \in U$ we have $\mathbf{v} = \mathbf{0}$. To conclude, we have $U = \{\mathbf{0}\}$, hence $E_1 = E_2$. \square

Since a GWM^c can be transformed into an equivalent semi-simple GWM^c in polynomial time (see Corollary 1 and the following discussion), the minimization of a GWM^c defined over circular strings can be achieved in polynomial time by first converting it to a semi-simple GWM^c and then applying Corollary 3 with Proposition 4. The overall minimization algorithm is summarized in Algorithm 1.

Algorithm 1. Minimization of a GWM defined over circular strings

Input: A d -dimensional GWM^c A given by a set of matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma} \subset \mathcal{M}_d(\mathbb{F})$.

Output: A minimal GWM^c \hat{A} computing f_A .

- 1: Let \mathcal{A} be the algebra generated by the matrices $\{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}$.
 - 2: Compute a basis $(\mathbf{A}_1, \dots, \mathbf{A}_n)$ of \mathcal{A} (as an \mathbb{F} -vector space) and the structure coefficients $c_{i,j}^k \in \mathbb{F}$ for $i, j, k \in [n]$ satisfying $\mathbf{A}_i \mathbf{A}_j = \sum_{k=1}^n c_{i,j}^k \mathbf{A}_k$.
 - 3: Compute the sub-algebra $\tilde{\mathcal{A}}$ and the corresponding surjective homomorphism $\pi: \mathcal{A} \rightarrow \tilde{\mathcal{A}}$ satisfying $\mathcal{A} = \text{Rad}(\mathcal{A}) \oplus \tilde{\mathcal{A}}$ and $\tilde{\mathcal{A}} \cong \mathcal{A}/\text{Rad}(\mathcal{A})$ (using the algorithm from [15], see Theorem 3).
 - 4: Let \tilde{A} be the semi-simple GWM given by the set of matrices $\{\tilde{\mathbf{A}}^\sigma = \pi(\mathbf{A}^\sigma)\}_{\sigma \in \Sigma}$.
 - 5: Compute a basis of $E_1 = \{\mathbf{v} \in \mathbb{F}^d : \tilde{\mathbf{A}}^\sigma \mathbf{v} = \mathbf{0} \text{ for all } \sigma \in \Sigma\} = \bigcap_{\sigma \in \Sigma} \ker(\tilde{\mathbf{A}}^\sigma)$.
 - 6: Let $\mathbf{\Pi} \in \mathbb{F}^{d \times d}$ be the matrix of the orthogonal projection onto E_1 .
 - 7: Let R be the rank of $\mathbf{I} - \mathbf{\Pi}$ and let $\mathbf{U} \in \mathbb{F}^{d \times R}$ be an orthogonal matrix such that $\mathbf{I} - \mathbf{\Pi} = \mathbf{U}\mathbf{U}^\top$.
 - 8: **return** The R -dimensional GWM^c given by the matrices $\{\hat{\mathbf{A}}^\sigma = \mathbf{U}^\top \tilde{\mathbf{A}}^\sigma \mathbf{U}\}_{\sigma \in \Sigma}$.
-

5 Conclusion

We proposed polynomial time algorithms to handle both the minimization and the equivalence problems for GWMs defined over circular strings. By doing so, we unraveled fundamental notions from algebra theory that will be central to the study of GWMs. In particular, the notion of *semi-simple* GWM^c was paramount to our analysis. Intuitively, semi-simplicity can be thought of as a weak form of minimality: components from the radical do not contribute to the final computation of a GWM^c (semi-simplification thus corresponds to annihilating these irrelevant components from the algebra, i.e. from the GWM^c 's dynamics).

The next step is of course to try to extend the results obtained in this paper to GWMs defined over more general families of graphs. One promising direction we are currently investigating relies on extending the central notion of semi-simple GWM^c to GWMs defined over arbitrary families of labeled graphs: by opening any edge e in a graph G one obtains a graph G_e with two *free ports* (i.e. edges having one end that is not connected to any vertex) which would be mapped by a d -dimensional GWM A to a matrix $\mathbf{A}^{G_e} \in \mathcal{M}_d(\mathbb{F})$ (indeed, a GWM naturally maps any graph with k free ports to a k th order tensor; see [22, Sect. 2.2.3] for more details). For circular strings, opening an edge corresponds to choosing a particular position in the circular string leading to an actual string $x \in \Sigma^*$ which is mapped to \mathbf{A}^x by the GWM. For arbitrary labeled graphs, we have $f_A(G) = \text{Tr}(\mathbf{A}^{G_e})$ similarly to the case of circular strings. One can then consider the algebra \mathcal{A} generated by the matrices \mathbf{A}^{G_e} for any graph G in some family of graphs and any edge e in G , and define a semi-simple GWM as a GWM for which this algebra \mathcal{A} is semi-simple (note that one exactly recovers the notion of semi-simple GWM introduced here in the special case of circular strings). Hence, the fundamental results from algebra theory we leveraged in this paper should be directly relevant to the study of general GWMs. Beyond minimization, we intend to study the problem of approximate minimization (such as the ones considered

in [7, 23] for string and tree weighted automata) along with the closely related problem of learning GWMs defined over richer families of graphs than the one of circular strings.

Acknowledgements. The author acknowledges support of an IVADO postdoctoral fellowship and would like to thank the reviewers for their helpful comments as well as Philip Amortila, François Denis, Clara Lacroce, Prakash Panangaden and Joelle Pineau for fruitful discussions.

References

1. Bailly, R., Carreras Pérez, X., Luque, F.M., Quattoni, A.J.: Unsupervised spectral learning of WCFG as low-rank matrix completion. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 624–635. ACL (2013)
2. Bailly, R., Denis, F., Rabusseau, G.: Recognizable series on hypergraphs. In: Proceedings of the 9th International Conference on Language and Automata Theory and Applications, pp. 639–651 (2015)
3. Bailly, R., Denis, F., Ralaivola, L.: Grammatical inference as a principal component analysis problem. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 33–40. ACM (2009)
4. Bailly, R., Habrard, A., Denis, F.: A spectral approach for probabilistic grammatical inference on trees. In: Hutter, M., Stephan, F., Vovk, V., Zeugmann, T. (eds.) ALT 2010. LNCS (LNAI), vol. 6331, pp. 74–88. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16108-7_10
5. Bailly, R., Rabusseau, G., Denis, F.: Recognizable series on graphs and hypergraphs. *J. Comput. Syst. Sci.* (2017, in press)
6. Balle, B., Carreras, X., Luque, F.M., Quattoni, A.: Spectral learning of weighted automata. *Mach. Learn.* **96**(1–2), 33–63 (2014)
7. Balle, B., Panangaden, P., Precup, D.: A canonical form for weighted automata and applications to approximate minimization. In: 30th Annual Symposium on Logic in Computer Science, pp. 701–712. IEEE (2015)
8. Berstel, J., Reutenauer, C.: Recognizable formal power series on trees. *Theor. Comput. Sci.* **18**(2), 115–148 (1982)
9. Berstel, J., Reutenauer, C.: *Rational Series and Their Languages*. Springer, Heidelberg (1988)
10. Boots, B., Siddiqi, S.M., Gordon, G.J.: Closing the learning-planning loop with predictive state representations. *Int. J. Robot. Res.* **30**(7), 954–966 (2011)
11. Bozagalidis, S., Grammatikopoulou, A.: Recognizable picture series. *J. Automata Lang. Comb.* **10**(2/3), 159–183 (2005)
12. Bremner, M.R.: How to compute the Wedderburn decomposition of a finite-dimensional associative algebra. *Groups Complex. Cryptol.* **3**(1), 47–66 (2011)
13. Brešar, M.: *Introduction to Noncommutative Algebra*. Springer, Cham (2014). <https://doi.org/10.1007/978-3-319-08693-4>
14. Cohen, S.B., Stratos, K., Collins, M., Foster, D.P., Ungar, L.H.: Spectral learning of latent-variable PCFGs: algorithms and sample complexity. *J. Mach. Learn. Res.* **15**(1), 2399–2449 (2014)
15. de Graaf, W.A., Ivanyos, G., Küronya, K., Rónyai, L.: Computing Levi decompositions in lie algebras. *Appl. Algebra Eng. Commun. Comput.* **8**(4), 291–303 (1997)

16. Droste, M., Kuich, W., Vogler, H. (eds.): Handbook of Weighted Automata. Springer Science & Business Media, Heidelberg (2009). <https://doi.org/10.1007/978-3-642-01492-5>
17. Drozd, Y.A., Kirichenko, V.V.: Finite Dimensional Algebras. Springer Science & Business Media, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-76244-4>
18. Hladnik, M., Omladic, M., Radjavi, H.: Trace-preserving homomorphisms of semi-groups. *J. Funct. Anal.* **204**(2), 269–292 (2003)
19. Hsu, D.J., Kakade, S.M., Zhang, T.: A spectral algorithm for learning Hidden Markov Models. In: Proceedings of the Conference on Learning Theory (2009)
20. Lee, T., Na, J.C., Park, H., Park, K., Sim, J.S.: Finding consensus and optimal alignment of circular strings. *Theoret. Comput. Sci.* **468**, 92–101 (2013)
21. Marusic, I., Worrell, J.: Complexity of equivalence and learning for multiplicity tree automata. *J. Mach. Learn. Res.* **16**, 2465–2500 (2015)
22. Rabusseau, G.: A Tensor Perspective on Weighted Automata, Low-Rank Regression and Algebraic Mixtures. PhD thesis, Aix-Marseille Université (2016)
23. Rabusseau, G., Balle, B., Cohen, S.: Low-rank approximation of weighted tree automata. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, pp. 839–847 (2016)
24. Rittaud, B., Vivier, L.: Circular words and applications. In: WORDS, pp. 31–36 (2011)
25. Sakarovitch, J.: Elements of Automata Theory. Cambridge University Press, Cambridge (2009)
26. Tzeng, W.-G.: On the definition of a family of automata. *Inf. Control* **4**(2–3), 245–270 (1961)
27. Tzeng, W.-G.: A polynomial-time algorithm for the equivalence of probabilistic automata. *SIAM J. Comput.* **21**(2), 216–227 (1992)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

