On the Metric-based Approximate Minimization of Markov Chains

Giovanni Bacci
Aalborg University, DK
giovbacci@cs.aau.dk

Giorgio Bacci
Aalborg University, DK
grbacci@cs.aau.dk

Kim G. Larsen
Aalborg University, DK
kgl@cs.aau.dk

Radu Mardare
Aalborg University, DK
mardare@cs.aau.dk

Minimization of finite automata, i.e., the process of transforming a given finite automaton into an equivalent one with minimum number of states, has been a major subject since the 1950s due to its fundamental importance for any implementation of finite automata tools.

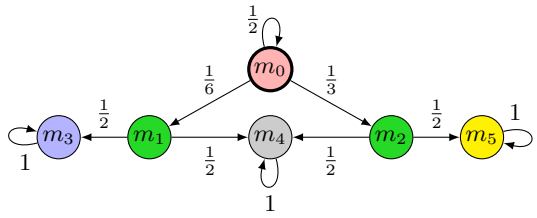
The first algorithm for the minimization of deterministic finite automata (DFAs) is due to Moore [1], with time complexity $O(n^2 s)$, later improved by the now classical Hopcroft's algorithm [2] to $O(ns \log n)$, where n is the number of states and s the size of the alphabet. Their algorithms are based on a partition refinement of the states into equivalence classes of the *Myhill-Nerode equivalence relation*. Partition refinement has been employed in the definition of efficient minimization procedures for a wide variety of automata: by Kanellakis and Smolka [3], [4] for the minimization of labelled transition systems (LTSs) w.r.t. Milner's strong bisimulation [5]; by Baier [6] for the reduction of Markov Chains (MCs) w.r.t. Larsen and Skou's probabilistic bisimulation [7]; by Alur et al. [8] and by Yannakakis and Lee [9], respectively, for the minimization of timed transition systems and timed-automata.

In [10], Jou and Smolka observed that for reasoning about the behavior of probabilistic systems (and more in general, all type of quantitative systems), rather than equivalences, a notion of distance is more reasonable in practice, since it permits "*a shift in attention from equivalent processes to probabilistically similar processes*". This observation motivated the development of metric-based semantics for quantitative systems, that consists in proposing 1-bounded pseudometrics capturing the similarities of the behaviors in the presence of small variations of the quantitative data. These pseudometrics generalize behavioral equivalences in the sense that, two processes are at distance 0 iff they are equivalent, and at distance 1 if no significant similarities can be observed between them.

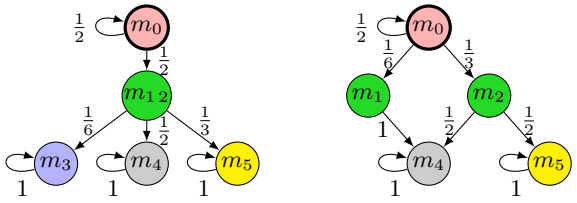
The first proposal of a behavioral pseudometric is due to Desharnais et al. [11] on labelled MCs, a.k.a. *probabilistic bisimilarity distance*, with the property that two MCs are at distance 0 iff they are probabilistic bisimilar. Its definition is parametric on a discount factor $\lambda \in (0, 1]$ that controls the significance of the future steps in the measurement. This pseudometric has been greatly studied by van Breugel and Worrell [12]–[14] who noticed, among other notable results, its relation with the Kantorovich distance on probability distributions and provided a polynomial-time algorithm for its computation.

The introduction of metric-based semantics motivated the interest in the approximate minimization of quantitative systems. The goal of approximate minimization is to start from a minimal automaton and produce a smaller automaton that is close to the given one in a certain sense. The desired size of the approximating automaton is given as input. Inspired by the aggregation of equivalent states typical of partition refinement techniques, Ferns et al. [15] undertake the approximate minimization problem by aggregating states having relative smaller distance. An example of this approach on MCs using the λ-bisimilarity distance of Desharnais et al. is shown below.

$\mathcal{M} =$



Let $\mathcal{M}$ be the MC above and assume we want to approximate it by an MC with at most 5 states. Since m_1, m_2 are the only two states at distance less than 1, the most natural choice for an aggregation shall collapse (via convex combination) m_1 and m_2, obtaining the MC on the left, which has distance $\frac{4}{9}(\frac{\lambda^2}{2-\lambda})$ from $\mathcal{M}$. Approximate aggregation of states does not necessarily yield the closest optimal solution. Indeed, the MC on the right is a closer approximant of $\mathcal{M}$, at distance $\frac{1}{6}(\frac{\lambda^2}{2-\lambda})$.



In this paper we address the issue of finding *optimal* solutions to the approximate minimization problem. Specifically we aim to answer to the following problem, left open in [15]: "*given a finite MC and a positive integer k, what is its 'best' k-state approximant? Here by 'best' we mean a k-state MC at minimal distance to the original*". We refer to this problem as *Closest Bounded Approximant* (CBA) and we present the following results related to it.

1) We characterize CBA as a bilinear optimization problem, proving the existence of *optimal* solutions.

2) We show that closest bounded approximants for Markov chains with rational transition probabilities may not have rational probabilities.

3) We study the complexity of the threshold problem of CBA, called *Bounded Approximant* problem (BA), that asks whether there exists a k-state approximant with distance from the original MC bounded by a given rational threshold. Specifically, we show that BA is in PSPACE and NP-hard.

4) We introduce the *Significant Bounded Approximant problem* (SBA) problem, that asks whether there exists an approximant of size k having some significant similarity to the original MC (i.e., at distance strictly less than 1). We show that this problem is NP-complete when one considers the undiscounted bisimilarity distance.

5) Finally, we propose an algorithm for finding suboptimal solutions of CBA that is inspired by Expectation Maximization (EM) techniques [16], [17].

Related Work: In [18], the approximate minimization of MCs is addressed via the notion of *quasi-lumpability*. An MC is quasi-lumpable if the given aggregations of the states can be turned into actual bisimulation-classes by a small perturbation of the transition probabilities. This approach differs from ours since there is no relation to a proper notion of behavioral distance (the approximation is w.r.t. the supremum norm of the difference of the stochastic matrices) and we do not consider any approximate aggregation of states. In [19], Balle et al. consider the approximate minimization of weighted finite automata (WFAs). Their method is via a truncation of a canonical normal form for WFAs that they introduced for the SVD decomposition of infinite Hankel matrices. Both [18] and [19] do not consider the issue of finding the *closest* approximant, which is the main focus of this paper, instead they give upper bounds on the distance from the given model.

I. Markov Chains and Bisimilarity Pseudometric

For $R \subseteq X \times X$ an equivalence relation, $X/_R$ denotes its quotient set. $\mathcal{D}(X)$ denotes the set of discrete probability distributions on X, i.e., functions $\mu\colon X \to [0,1]$, s.t. $\mu(X) = 1$, where $\mu(E) = \sum_{x \in E} \mu(x)$ for $E \subseteq X$. We fix a countable set L of labels representing properties that hold in certain states.

Definition 1: A *Markov chain* is a tuple $\mathcal{M} = (M, \tau, \ell)$ consisting of a finite *set of states* M, a *transition distribution function* $\tau\colon M \to \mathcal{D}(M)$, and a *labelling function* $\ell\colon M \to L$.

Intuitively, if $\mathcal{M}$ is in state m it moves to state m' with probability $\tau(m)(m')$. The set of labels of $\mathcal{M}$ is denoted by $L(\mathcal{M}) = \{\ell(m) \mid m \in M\}$.

Hereafter, we use $\mathcal{M} = (M, \tau, \ell)$ and $\mathcal{N} = (N, \theta, \alpha)$ to range over MCs and we refer to their constituents implicitly.

Probabilistic bisimulation of Larsen and Skou [7] is a key concept for reasoning about the behavioral equivalence of probabilistic systems. Two states are bisimilar if they have the same label and equal probability of moving to any bisimilarity class. Any probabilistic bisimulation $R \subseteq M \times M$ induces another MC, called *R-quotient* of $\mathcal{M}$ (a.k.a. R-lumping [20]), having states in $M/_R$. An MC is said *minimal* if it is isomorphic to its quotient w.r.t. probabilistic bisimilarity.

A function $d\colon X \times X \to [0,1]$ is a (1-bounded) *pseudometric* on X if for any $x, y, z \in X$, $d(x,x) = 0$, $d(x,y) = d(y,x)$, and $d(x,y) + d(y,z) \geq d(x,z)$. 1-bounded pseudometrics on X form a complete lattice under the point-wise partial order $d \sqsubseteq d'$ iff $d(x,y) \leq d'(x,y)$ for all $x, y \in X$.

The *bisimilarity distance* of Desharnais et al. [21] extends probabilistic bisimilarity in the sense that two states are at distance zero iff they are bisimilar. Its definition is based on the *Kantorovich (pseudo)metric* on $\mathcal{D}(X)$ for a finite set X, defined as $\mathcal{K}(d)(\mu,\nu) = \min\left\{\int d \,\mathrm{d}\omega \mid \omega \in \Omega(\mu,\nu)\right\}$, where d is a (pseudo)metric on X and $\Omega(\mu,\nu)$ denotes the set of *couplings* for (μ,ν), i.e., distributions $\omega \in \mathcal{D}(X \times X)$ such that, for all $E \subseteq X$, $\omega(E \times X) = \mu(E)$ and $\omega(X \times E) = \nu(E)$.

Definition 2 (Bisimilarity Distance): Let $\lambda \in (0,1]$. The λ-discounted *bisimilarity pseudometric* on $\mathcal{M}$, denoted by δ_λ, is the least fixed-point of the following functional operator on 1-bounded pseudometrics over M (ordered point-wise)

$$\Psi_\lambda(d)(m,n) = \begin{cases} 1 & \text{if } \ell(m) \neq \ell(n) \\ \lambda \cdot \mathcal{K}(d)(\tau(m),\tau(n)) & \text{otherwise}\,. \end{cases}$$

The operator Ψ_λ is monotonic, hence, by Tarski fixed-point theorem, δ_λ is well defined.

Intuitively, if two states have different labels δ_λ considers them as "incomparable" (i.e., at distance 1), otherwise their distance is given by the Kantorovich distance w.r.t. δ_λ between their transition distributions. The *discount factor* $\lambda \in (0,1]$ controls the significance of the future steps in the measurement of the distance; if $\lambda = 1$, the distance is said *undiscounted*.

Usually, MCs are associated with an initial state to be thought of as their initial configurations. In the rest of the paper when we talk about the distance between two MCs, written $\delta_\lambda(\mathcal{M},\mathcal{N})$, we implicitly refer to the distance between their initial states computed over the disjoint union of their MCs.

II. The Closest Bounded Approximant Problem

We introduce the *Closest Bounded Approximant* problem w.r.t. δ_λ (CBA-λ), and study it as an optimization problem.

Definition 3 (Closest Bounded Approximant): Let $k \in \mathbb{N}$ and $\lambda \in (0,1]$. The *closest bounded approximant problem* w.r.t. δ_λ for an MC $\mathcal{M}$ is the problem of finding an MC $\mathcal{N}$ with at most k states minimizing $\delta_\lambda(\mathcal{M},\mathcal{N})$.

When k is greater than or equal to the number of bisimilarity classes of $\mathcal{M}$, CBA-λ is efficiently solved by computing the bisimilarity quotient of $\mathcal{M}$ (*cf.* [6], [22]). Therefore, w.l.o.g., we will assume $1 \leq k < |M|$ and $\mathcal{M}$ to be minimal.

Let $\mathrm{MC}(k)$ denote the set of MCs with at most k states and $\mathrm{MC}_A(k)$ the subset of those using only labels in $A \subseteq L$.

In the following, we fix $\langle \mathcal{M}, k \rangle$ as the instance of CBA-λ. CBA-λ can be reformulated as finding $\mathcal{N}^* \in \mathrm{MC}(k)$ such that

$$\delta_\lambda(\mathcal{M},\mathcal{N}^*) = \inf\{\delta_\lambda(\mathcal{M},\mathcal{N}) \mid \mathcal{N} \in \mathrm{MC}(k)\}\,. \qquad (1)$$

A useful property of CBA-λ is that an optimal solution can be found among MCs which use labels from the given MC.

Lemma 1: Let $\mathcal{M}$ be an MC. Then, for any $\mathcal{N}' \in \mathrm{MC}(k)$ there exists $\mathcal{N} \in \mathrm{MC}_{L(\mathcal{M})}(k)$ s.t. $\delta_\lambda(\mathcal{M},\mathcal{N}) \leq \delta_\lambda(\mathcal{M},\mathcal{N}')$.

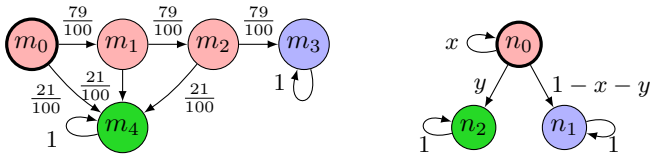
In general, it is not obvious that for arbitrary instances $\langle \mathcal{M}, k \rangle$ a minimum in (1) exists. The following result states that any instance of CBA-λ admits a minimal solution $\mathcal{N}^*$ in $\mathrm{MC}(k)$.

Theorem 1: CBA-λ always admits an optimal solution.

proof sketch: By Lemma 1 the search space of Equation (1) can be restricted to chains in $\mathrm{MC}_{L(\mathcal{M})}(k)$ with exactly k states. Since $L(\mathcal{M})$ is finite, one can model the optimization problem in Equation (1) as a bilinear program with a *linear cost function* representing the value $d(\mathcal{M}, \mathcal{N})$ for a generic prefixpoint d of Ψ_λ, and a *compact* feasible set of solutions modeling a generic approximant $\mathcal{N} \in \mathrm{MC}_{L(\mathcal{M})}(k)$ and d. ■

The following example shows that even when $\mathcal{M}$ has rational transition probabilities, optimal solutions for CBA-λ may have irrational transition probabilities.

Example 1: Consider the MC $\mathcal{M}$ on the left, with initial state m_0 and labels represented by colors. An optimal solution of CBA-1 on $\langle \mathcal{M}, 3 \rangle$ is the MC $\mathcal{N}_{xy}$ on the right with initial state n_0 and $x = \frac{1}{30}\left(10 + \sqrt{163}\right)$, $y = \frac{21}{100}$.



Since the distance $\delta_1(\mathcal{M}, \mathcal{N}_{xy}) = \frac{436}{675} - \frac{163\sqrt{163}}{13500} \approx 0.49$ is irrational, by [14, Proposition 13], any optimal solution must have some irrational transition probability. ■

III. Complexity Results

The *Bounded Approximant problem* w.r.t. δ_λ (BA-λ) is the threshold decision problem of CBA-λ, that, the given MC $\mathcal{M}$, integer $k \geq 1$, and rational $\epsilon \geq 0$, asks whether there exists $\mathcal{N} \in \mathrm{MC}(k)$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}) \leq \epsilon$.

The first result on BA-λ establishes the following complexity upper-bound.

Theorem 2: For any $\lambda \in (0, 1]$, BA-λ is in PSPACE.

proof sketch: An instance $\langle \mathcal{M}, k, \epsilon \rangle$ of BA-λ can be encoded as a decision problem for the existential theory of the reals, precisely, a bilinear matrix inequality (*cf.* Theorem 1). The encoding is polynomial in the size of $\langle \mathcal{M}, k, \epsilon \rangle$, thus it can be solved in PSPACE (*cf.* Canny [23]). ■

We will also show that BA-λ is NP-hard via a polynomial-time many-one reduction to Vertex Cover. Recall that, a vertex cover of an undirected graph G is a subset C of vertices such that every edge in G has at least one endpoint in C. Given a graph G and a positive integer h, the Vertex Cover problem asks if G has a cover of size at most h.

Theorem 3: For any $\lambda \in (0, 1]$, BA-λ is NP-hard.

proof sketch: Let $\langle G = (V, E), h \rangle$ be an instance of Vertex Cover and let $e = |E|$. W.l.o.g. we assume $e \geq 2$ and $k < n$. From G we construct the MC $\mathcal{M}_G = (M, \tau, \ell)$ as follows. The set of states M is given as the union of V and E to which we add two extra states: a root r (thought of as the initial state) and a sink s. Each node of $\mathcal{M}_G$ is associated with a unique label (i.e., ℓ is injective). The sink state s and all $v \in V$ loop to themselves with probability 1.

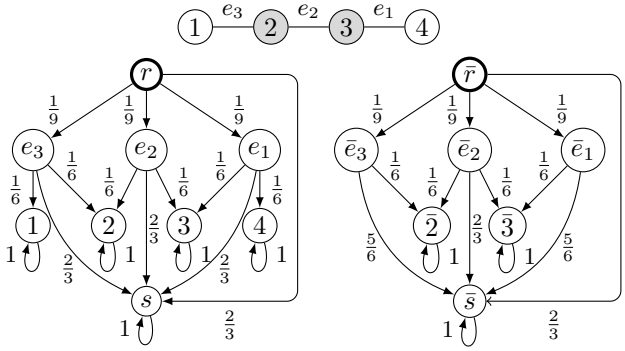


Fig. 1. (Top) An undirected graph G; (Left) The MC $\mathcal{M}_G$ associated with G; (Right) The MC $\mathcal{M}_C$ associated with the vertex cover $C = \{1, 2\}$ of G.

All the other states go with probability $1 - \frac{1}{e}$ to the sink state s. The rest of their transition probability mass is assigned as follows. The root r goes with probability $\frac{1}{e^2}$ to each $a \in E$, and all $(u, v) \in E$ go with probability $\frac{1}{2e}$ to their endpoints u, v. An example of construction for $\mathcal{M}_G$ is given in Figure 1. One can show that $\langle G, h \rangle \in$ Vertex Cover if and only if $\langle \mathcal{M}_G, e + h + 2, \frac{\lambda^2}{2e^2} \rangle \in$ BA-λ. ■

Recall that, two MCs are at distance 1 from each other when there is no significant similarity between their behaviors. Thus an MC $\mathcal{N}$ is said to be a *significant approximant* for the MC $\mathcal{M}$ w.r.t. δ_λ if $\delta_\lambda(\mathcal{M}, \mathcal{N}) < 1$. Given an MC $\mathcal{M}$ and a positive integer k, we call *Significant Bounded Approximant problem* w.r.t. δ_λ (SBA-λ), the decision problem that asks whether there exists $\mathcal{N} \in \mathrm{MC}(k)$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}) < 1$.

SBA-λ is trivial for $\lambda < 1$: an MC $\mathcal{N}$ with a single state labelled as the initial state of $\mathcal{M}$ ensures that $\delta_\lambda(\mathcal{M}, \mathcal{N}) \leq \lambda$.

Interestingly, the significant bounded approximant problem w.r.t. the undiscounted distance (i.e., SBA-1) is NP-complete.

This result is based on the following technical lemma.

Lemma 2: Let $\mathcal{M}$ be a MC (assumed to be minimal) with initial state m_0 and $G(\mathcal{M})$ its underlying directed graph. Then, $\langle \mathcal{M}, k \rangle \in$ SBA-1 iff there exists a bottom strongly connected component $G' = (V, E)$ in $G(\mathcal{M})$ and a path $m_0 \dots m_h$ in $G(\mathcal{M})$ such that $m_h \in V$ and $|\{\ell(m_i) \mid i \leq p\}| + |V| \leq k$ where $p < h$ is the smallest index such that there exists a path $v_i \dots v_{h-1} m_h$ in G' with $\ell(m_j) = \ell(v_j)$ for all $p \leq j < h$.

Theorem 4: SBA-1 is NP-complete.

proof sketch: Membership in NP can be proven by using Lemma 2 and exploiting Tarjan's algorithm for generating bottom SCCs. As for the NP-hardness, we provide a polynomial-time many-one reduction from Vertex Cover. Let $G = (V, E)$ be a graph with $E = \{e_1, \dots, e_n\}$. We construct the MC $\mathcal{M}_G$ as follows. The set of states is given by the set of edges E along with two states e_i^1 and e_i^2, for each edge $e_i \in E$, representing the two endpoints of e_i and one extra sink state e_0. The initial state is e_n. The transition probabilities are given as follows. The sink state e_0 loops with probability 1 to itself. Each edge $e_i \in E$ goes with probability $\frac{1}{2}$ to e_i^1 and e_i^2, respectively. For $1 \leq i \leq n$, the states e_i^1 and e_i^2 go with probability 1 to the state e_{i-1}.

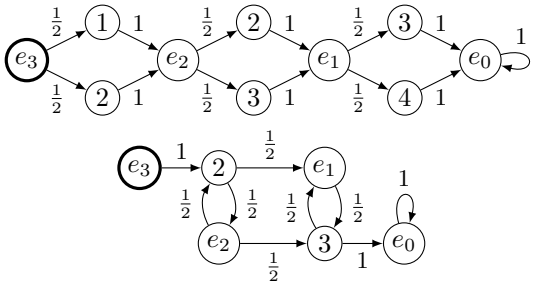


Fig. 2. (Top) The MC $\mathcal{M}_G$ associated to the graph G in Figure 1 and (Bottom) an MC $\mathcal{N}$ associated to the vertex cover $C = \{1, 2\}$ of G such that $\delta_1(\mathcal{M}_G, \mathcal{N}) < 1$ (cf. Theorem 4).

Algorithm 1 Approximate Minimization – EM heuristic

Input: $\mathcal{M} = (M, \tau, \ell)$, $\mathcal{N}_0 = (N, \theta_0, \alpha)$, and $h \in \mathbb{N}$.

1. $\mathcal{C}_0 \in \Omega(\mathcal{M}, \mathcal{N}_0)$ s.t. $\delta_\lambda(\mathcal{M}, \mathcal{N}_0) = 1 - \beta(\mathcal{C}_\lambda)(\mathcal{M}, \mathcal{N}_{i-1})$
2. **for** $i = 0$ **to** $h - 1$ **do**
3. compute $\mathbf{E}[Z^{m,n}_{u,v} | \mathcal{C}_i]$ for all $m, u \in M$ and $n, v \in N$
4. get θ_{i+1} and $\mathcal{C}_{i+1}$ from an optimal solution of $\mathbf{M}\langle\mathcal{C}_i\rangle$
5. $\mathcal{N}_{i+1} \leftarrow (N, \theta_{i+1}, \alpha)$ ▷ *Update*
6. **end for**
7. **return** $\mathcal{N}_h$

The edge states and the sink state are labelled by pairwise distinct labels, while the endpoints states e^1_i and e^2_i are labelled by the node in V they represent. An example of construction for $\mathcal{M}_G$ is shown in Figure 2. By Lemma 2 one can easily show that $\langle G, h\rangle \in$ VERTEX COVER if and only if $\langle \mathcal{M}_G, h + n + 1\rangle \in$ SBA-1. ■

IV. An Expectation Maximization Heuristic

We present an algorithm for computing suboptimal solutions of CBA-λ. Given an initial approximant $\mathcal{N}_0 \in \mathrm{MC}(k)$, the algorithm produces a sequence of MCs $\mathcal{N}_0, \ldots, \mathcal{N}_h$ in $\mathrm{MC}(k)$ having successively decreased distance from $\mathcal{M}$. The procedure is described in Algorithm 1. Intuitively, the algorithm updates the current MC $\mathcal{N}_i$ by assigning relatively greater probability to transitions that are most representative of the behavior of the MC $\mathcal{M}$ w.r.t. δ_λ.

The update procedure leverages on a characterization of δ_λ based on the notion of coupling structure (*cf.* Lemma 3).

Definition 4: A function $\mathcal{C} \colon M \times N \to \mathcal{D}(M \times N)$ is a *coupling structure* for $(\mathcal{M}, \mathcal{N})$ if $\mathcal{C}(m, n) \in \Omega(\tau(m), \theta(n))$. $\Omega(\mathcal{M}, \mathcal{N})$ denotes the set of coupling structures of $(\mathcal{M}, \mathcal{N})$. Intuitively, a coupling structure can be thought of as an MC on the cartesian product $M \times M$, obtained as the probabilistic combination of two copies of $\mathcal{M}$. Consider the MC $\mathcal{C}_\lambda$ obtained by extending $\mathcal{C}$ with a "sink" state $\perp$ to which any other state moves with probability $(1 - \lambda)$. Let $\beta(\mathcal{C}_\lambda)$ be the probability that $\mathcal{C}_\lambda$ never reaches a pair of states with different labels. The following result gives an alternative characterization of the bisimilarity distance in terms of $\beta(\mathcal{C}_\lambda)$.

Lemma 3: $\delta_\lambda(\mathcal{M}, \mathcal{N}) = 1 - \max_{\mathcal{C} \in \Omega(\mathcal{M}, \mathcal{N})} \beta(\mathcal{C}_\lambda)$.

From equation (1) and Lemma 3, we can reduce the problem CBA-λ to that of taking the right marginal of a coupling structure $\mathcal{C}$ that maximizes $\beta(\mathcal{C}_\lambda)$, namely

$$\operatorname{argmax} \{\beta(\mathcal{C}_\lambda) \mid \mathcal{N} \in \mathrm{MC}(k),\, \mathcal{C} \in \Omega(\mathcal{M}, \mathcal{N})\} \,. \qquad (2)$$

This change in perspective allows us to apply Expectation Maximization (EM) techniques in a way similar to [17].

Let $R_\mathcal{C} \subseteq M \times N$ be the set of nodes in the underlying graph of $\mathcal{C}$ from which there is no path passing from nodes (m, n) such that $\ell(m) \neq \alpha(n)$. $\beta(\mathcal{C}_\lambda)$ is the probability that $\mathcal{C}_\lambda$ emits a path with prefix in $\cong^*(R_\mathcal{C} \cup \perp)$ starting from the pair (m_0, n_0) of initial states, where $\cong \,= \{(m, n) \notin R_\mathcal{C} \mid \ell(m) = \alpha(n)\}$. The idea is to count the expected number of occurrences of transitions in $\mathcal{C}_i$ in $\cong^*(R_\mathcal{C} \cup \perp)$ and update $\mathcal{C}_i$ by increasing the weight of the transitions that contribute the most.

Let $Z^{m,n}_{u,v} \colon (M \times N)^\omega \to \mathbb{N}$ be the random variable that counts the number of occurrences of the edge $((m, n)(u, v))$ in a prefix in $\cong^*(R_\mathcal{C} \cup \perp)$ of the given path. We denote by $\mathbf{E}[Z^{m,n}_{u,v} | \mathcal{C}]$ the expected value of $Z^{m,n}_{u,v}$ w.r.t. the probability distribution induced by $\mathcal{C}_\lambda$. We define $\mathbf{M}\langle\mathcal{C}\rangle$ as

$$\begin{array}{ll} \text{maximize} & \sum_{m,u \in M} \sum_{n,v \in N} \mathbf{E}[Z^{m,n}_{u,v} | \mathcal{C}] \cdot \ln(c^{m,n}_{u,v}) \\ \text{such that} & \sum_{v \in N} c^{m,n}_{u,v} = \tau(m)(u) \quad (m, u \in M,\, n \in N) \\ & \sum_{u \in M} c^{m,n}_{u,v} = \theta_{n,v} \quad (m \in M,\, n, v \in N) \\ & c^{m,n}_{u,v} \geq 0 \quad (m, u \in M,\, n, v \in N) \end{array}$$

An optimal solution of $\mathbf{M}\langle\mathcal{C}\rangle$ induces a coupling structure (with transitions $c^{m,n}_{u,v}$) and a Markov chain (with transitions $\theta_{n,v}$) that improves the coupling structure $\mathcal{C}$ in the sense of (2).

Theorem 5: Let $\beta(\mathcal{C}_\lambda) > 0$. Then, an optimal solution for $\mathbf{M}\langle\mathcal{C}\rangle$ describes a Markov chain $\mathcal{N}'$ and a coupling structure $\mathcal{C}' \in \Omega(\mathcal{M}, \mathcal{N}')$ such that $\beta(\mathcal{C}_\lambda) \leq \beta(\mathcal{C}'_\lambda)$.

The choice of the initial approximant $\mathcal{N}_0$ may have a significant effect on the quality of the solution. For the labeling of the states, one should follow Lemma 1. As for the choice of the underlying structure one shall be guided by Lemma 2. However, Theorem 3, suggests that its hard to select a good starting approximant candidate. Nevertheless, good selections for $\mathcal{N}_0$ may be suggested by looking at the problem instance.

V. Conclusion and Future Work

To the best of our knowledge, this is the first work addressing the complexity of the *optimal* approximate minimization of MCs w.r.t. a behavioral metric semantics.

Preliminary experiments on Algorithm 1 show that the maximization step is inefficient to perform. To tame this problem, we explored other heuristics based on relaxed versions of $\mathbf{M}\langle\mathcal{C}\rangle$ which admit analytic solutions. Prototype implementations[1] of two different "relaxed" variants of Algorithm 1 have given promising results.

In the light of [14], [24], relating the probabilistic bisimilarity distance to the LTL-model checking problem as $\delta_1(\mathcal{M}, \mathcal{N}) \geq |\mathbb{P}_\mathcal{M}(\varphi) - \mathbb{P}_\mathcal{N}(\varphi)|$, for all $\varphi \in$ LTL, our results might be used to lead saving in the overall model checking time. Membership of BA-λ in NP is left open as future work.

[1]The prototype is available at http://people.cs.aau.dk/giovbacci/tools.html.

REFERENCES

[1] E. F. Moore, "Gedanken Experiments on Sequential Machines," in *Automata Studies*. Princeton University, 1956, pp. 129–153.

[2] J. Hopcroft, "An $n \log n$ algorithm for minimizing states in a finite automaton," in *Theory of Machines and Computations*, Z. Kohavi and A. Paz, Eds. Academic Press, 1971, pp. 189–196.

[3] P. C. Kanellakis and S. A. Smolka, "CCS expressions, finite state processes, and three problems of equivalence," in *Proceedings of the 2nd Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing*. ACM, 1983, pp. 228–240.

[4] ——, "CCS expressions, finite state processes, and three problems of equivalence," *Information and Computation*, vol. 86, no. 1, pp. 43–68, 1990.

[5] R. Milner, *A Calculus of Communicating Systems*, ser. Lecture Notes in Computer Science. Springer, 1980, vol. 92.

[6] C. Baier, "Polynomial time algorithms for testing probabilistic bisimulation and simulation," in *CAV*, ser. Lecture Notes in Computer Science, vol. 1102. Springer, 1996, pp. 50–61.

[7] K. G. Larsen and A. Skou, "Bisimulation through probabilistic testing," *Information and Computation*, vol. 94, no. 1, pp. 1–28, 1991.

[8] R. Alur, C. Courcoubetis, N. Halbwachs, D. L. Dill, and H. Wong-Toi, "Minimization of timed transition systems," in *CONCUR*, ser. Lecture Notes in Computer Science, vol. 630. Springer, 1992, pp. 340–354.

[9] M. Yannakakis and D. Lee, "An efficient algorithm for minimizing real-time transition systems," *Formal Methods in System Design*, vol. 11, no. 2, pp. 113–136, 1997.

[10] C.-C. Jou and S. A.Smolka, "Equivalences, congruences, and complete axiomatizations for probabilistic processes," in *CONCUR'90 Theories of Concurrency: Unification and Extension*, ser. LNCS, vol. 458, 1990, pp. 367–383.

[11] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden, "Metrics for Labeled Markov Systems," in *CONCUR*, ser. LNCS, vol. 1664. Springer, 1999, pp. 258–273.

[12] F. van Breugel and J. Worrell, "Towards Quantitative Verification of Probabilistic Transition Systems," in *ICALP*, ser. LNCS, vol. 2076, 2001, pp. 421–432.

[13] ——, "Approximating and computing behavioural distances in probabilistic transition systems," *Theoretical Computer Science*, vol. 360, no. 3, pp. 373–385, 2006.

[14] D. Chen, F. van Breugel, and J. Worrell, "On the Complexity of Computing Probabilistic Bisimilarity," in *FoSSaCS*, ser. LNCS, vol. 7213. Springer, 2012, pp. 437–451.

[15] N. Ferns, P. Panangaden, and D. Precup, "Metrics for finite Markov Decision Processes," in *UAI*. AUAI Press, 2004, pp. 162–169.

[16] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. Wiley-Interscience, 2008.

[17] M. Benedikt, R. Lenhardt, and J. Worrell, "LTL Model Checking of Interval Markov Chains," in *TACAS*, ser. Lecture Notes in Computer Science, vol. 7795. Springer, 2013, pp. 32–46.

[18] G. Franceschinis and R. R. Muntz, "Bounds for quasi-lumpable markov chains," *Perform. Eval.*, vol. 20, no. 1-3, pp. 223–243, 1994.

[19] B. Balle, P. Panangaden, and D. Precup, "A canonical form for weighted automata and applications to approximate minimization," in *LICS*. IEEE Computer Society, 2015, pp. 701–712.

[20] J. G. Kemeny and J. L. Snell, *Finite Markov chains*, ser. Undergraduate texts in mathematics. New York: Springer, 1976, reprint of the 1960 ed. published by Van Nostrand, Princeton, N.J., in the University series in undergraduate mathematics.

[21] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden, "Metrics for labelled Markov processes," *Theoretical Compututer Science*, vol. 318, no. 3, pp. 323–354, 2004.

[22] S. Derisavi, H. Hermanns, and W. H. Sanders, "Optimal state-space lumping in Markov chains," *Inf. Process. Lett.*, vol. 87, no. 6, pp. 309–315, 2003.

[23] J. F. Canny, "Some Algebraic and Geometric Computations in PSPACE," in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC'88)*. ACM, 1988, pp. 460–467.

[24] G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare, "Converging from Branching to Linear Metrics on Markov Chains," in *ICTAC*, ser. LNCS, vol. 9399. Springer, 2015, pp. 349–367.