Minimizing Markov Chains Beyond Bisimilarity

Giovanni Bacci, Giorgio Bacci, Kim G. Larsen, and Radu Mardare

Department of Computer Science, Aalborg University, Denmark
{giovbacci,grbacci,kgl,mardare}@cs.aau.dk

Probabilistic bisimulation for Markov chains (MCs), introduced by Larsen and Skou [19], is widely considered a key concept for reasoning about the equivalence of probabilistic systems. This concept is intimately related to that of lumpability by Kemeny and Snell [15] and has been efficiently applied in state space reduction techniques for probabilistic systems [7]. Computing the bisimilarity quotient of a Markov chain (a.k.a. the optimal state-space lumping) yields *the smallest* Markov chain that behaves exactly as the original chain. The optimal lumping quotient has a (typically much) smaller state space, nevertheless, it may be still too large to yield significant computational improvements. This is a well known problem that state-of-the-art probabilistic model-checking tools [18, 14, 12] strive to overcome.

The main reason for this phenomenon arises from the fact that probabilistic bisimilarity is a notion of equivalence that is too "exact" for many purposes as it only relates processes with identical bahaviors. In a number of applications, such as systems biology [4], games [5] and machine learning [13], one is interested in knowing whether two probabilistic systems that may differ by a small amount in the real-valued parameters have "sufficiently" similar behaviours.

This motivated the study of behavioral metrics for Markov chains, initiated by Desharnais et al. [8] and greately developed in the last decade [20, 6, 1, 2]. The idea consists in using distances (more precisely, 1-bounded pseudometrics) to measure the behavioral dissimilarities of two MCs, that is, the farther two states are the more distinguishable their behaviors are. In this respect, bisimilar states are at distance zero with each other, while two states that have no significant similarity are at distance 1.

Analogously, also the concept of lumpability has been relaxed by considering partitioning of the state space whose elements are at most ε-apart from each other [9]. Thus, by suitably choosing the value of ε, one can significantly reduce the size of the resulting lumping at the cost of loosing in precision. However, the value of ε only bounds the relative distance between two given states in the original MC, whereas the distance between the original MC and its ε-lumping is (typically much) greater than ε. Generally, it is difficult to obtain a good tradeoff between the size and the precision of the resulting ε-lumping, and that often requires ad hoc adjustments based on a deep knowledge of the problem domain where the original MC originates.

These issues were also pointed out in [9], leaving open the following problem. Given a finite MC $\mathcal{M}$ and a positive integer k, what is its "best" k-state approximation? Here by "best" we mean a k-state MC of minimal distance to the original. In this paper we address the above question, that we call *the Markov chain Approximation problem* (MCA). The MCA problem is defined as the following optimization problem

$$\min \{\delta(\mathcal{M},\mathcal{N}) \mid \mathcal{N} \in \mathbb{F}(k)\} \ , \qquad \text{(MCA PROBLEM)}$$

where $\mathbb{F}(k)$ denotes the class of all the MCs with at most $k \in \mathbb{N}$ states. If an instance of the MCA problem admits an optimal solution, say $\mathcal{N}^*$, it represents a (suitably small) MC that best approximates the behavior of $\mathcal{M}$, and the *optimal value* $\delta(\mathcal{M},\mathcal{N}^*)$ certifies its quality.

Obviously, for values of k greater than or equal to the cardinality of the bisimilarity quotient, the MCA problem reduces to constructing the optimal lumping quotient of $\mathcal{M}$ which can be solved in $O(n \log m)$ where n is the number of states and m is the number of transitions [7].

As one may expect, the MCA problem is in general non-trivial to solve, moreover it seems unlikely that it can be solved efficiently. We formalize this conjecture by studying the computational complexity of its corresponding decision problem, namely, the *the Threshold MCA problem*. The Threshold MCA problem is defined as follows. Given $k \in \mathbb{N}$, a finite MC $\mathcal{M}$ and $\varepsilon \in \mathbb{Q} \cap [0,1]$, decide whether $\min\{\delta(\mathcal{M},\mathcal{N}) \mid \mathcal{N} \in \mathbb{F}(k)\} \leq \varepsilon$.

Theorem 1. *The Threshold MCA problem is NP-hard.*

The above complexity lower-bound follows from a stronger result on a specialized version of the Threshold MCA problem, called *the Qualitative MCA problem*.

Formally, the Qualitative MCA problem is defined as follows. Given $k \in \mathbb{N}$ and a finite MC $\mathcal{M}$, decide whether $\min\{\delta(\mathcal{M},\mathcal{N}) \mid \mathcal{N} \in \mathbb{F}(k)\} < 1$. Recall that, two MCs are at distance 1 from each other when no significant similarity can be observed between their behaviors. In this respect, the Qualitative MCA problem corresponds to the question "Is it worth searching for a k-states approximation of $\mathcal{M}$?".

Theorem 2. *The Qualitative MCA problem is NP-complete.*

The NP-hardness of the Qualitative MCA problem is obtained by means of a polynomial-time many-one reduction from *vertex cover*. Interestingly, the same reduction can be easily adapted to prove the NP-hardness of the Thresold MCA problem w.r.t. *the total variation distance* (a.k.a. trace distance) or w.r.t. any multi-step bisimilarity distance as defined in [2].

Theorem 2 provides tight complexity bounds for the Qualitative MCA problem, however, for generic error bound ε, the Threshold MCA problem appears to be much harder to solve.

Theorem 3. *The Threshold MCA problem is in PSPACE.*

The above complexity upper-bound has been obtained by reducing the Threshold MCA problem to checking the validity of a formula in the existential theory of reals [3].

Notably, our reduction involves only conjunctions of bilinear constraints, therefore their feasibility may be more efficiently checked viewing them as bilinear matrix inequalities (BMIs) for which general algorithms [11, 10] and tools [17, 16] have been developed.

The fact that the MCA problem can be reduced to an optimization problem with bilinear matrix inequalities makes us hope for the possibility to adapt well studied non linear optimization methods to efficiently construct good locally optimal solutions for the MCA problem.

References

[1] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In *TACAS*, volume 7795 of *LNCS*, pages 1–15, 2013.

[2] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Converging from Branching to Linear Metrics on Markov Chains. In *ICTAC*, volume 9399 of *LNCS*, pages 349–367. Springer, 2015.

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

[4] Luca Cardelli and Attila Csikász-Nagy. The Cell Cycle Switch Computes Approximate Majority. *Scientific Reports*, 2, 2012.

[5] Krishnendu Chatterjee, Luca de Alfaro, Rupak Majumdar, and Vishwanath Raman. Algorithms for Game Metrics. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2008*, volume 2 of *LIPIcs*, pages 107–118. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2008.

[6] Di Chen, Franck van Breugel, and James Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In *FoSSaCS*, volume 7213 of *LNCS*, pages 437–451. Springer, 2012.

[7] Salem Derisavi, Holger Hermanns, and William H. Sanders. Optimal state-space lumping in Markov chains. *Inf. Process. Lett.*, 87(6):309–315, 2003.

[8] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Compututer Science*, 318(3):323–354, 2004.

[9] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov Decision Processes. In *UAI*, pages 162–169. AUAI Press, 2004.

[10] Mituhiro Fukuda and Masakazu Kojima. Branch-and-cut algorithms for the bilinear matrix inequality eigenvalue problem. *Comp. Opt. and Appl.*, 19(1):79–105, 2001.

[11] Keat-Choon Goh, Michael G. Safonov, and George P. Papavassilopoulos. Global optimization for the Biaffine Matrix Inequality problem. *J. Global Optimization*, 7(4):365–380, 1995.

[12] Ernst Moritz Hahn, Yi Li, Sven Schewe, Andrea Turrini, and Lijun Zhang. IscasMC: A web-based probabilistic model checker. In *Nineteenth international symposium of the Formal Methods Europe association (FM)*, volume 8442 of *LNCS*, pages 312–317. Springer, 2014.

[13] Manfred Jaeger, Hua Mao, Kim Guldstrand Larsen, and Radu Mardare. Continuity properties of distances for markov processes. In *Quantitative Evaluation of Systems - 11th International Conference, QEST 2014*, volume 8657 of *Lecture Notes in Computer Science*, pages 297–312. Springer, 2014.

[14] Joost-Pieter Katoen, Maneesh Khattri, and Ivan S. Zapreev. A Markov Reward Model Checker. In *Second International Conference on the Quantitative Evaluation of Systems (QEST 2005)*, pages 243–244. IEEE Computer Society, 2005.

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

[16] Michal Kočvara and Michael Stingl. PENBMI 2.0. http://www.penopt.com/penbmi.html. Accessed: 2016-08-28.

[17] Michal Kočvara and Michael Stingl. PENNON: A code for convex nonlinear and semidefinite programming. *Optimization Methods and Software*, 18(3):317–333, 2003.

[18] M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.

[19] Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991.

[20] Franck van Breugel, Babita Sharma, and James Worrell. Approximating a Behavioural Pseudometric without Discount for Probabilistic Systems. *Logical Methods in Computer Science*, 4(2), 2008.