



The Validity of Weighted Automata

Sylvain Lombardy¹ and Jacques Sakarovitch²(✉)

¹ LaBRI - UMR 5800 - Bordeaux INP - Bordeaux University - CNRS,
Bordeaux, France

² IRIF - UMR 8243 - CNRS/Paris Diderot University and Telecom ParisTech,
Paris, France
sakarovitch@enst.fr

Extended Abstract

(adapted from the introduction of [14])

This invited talk presents the work conducted on the problems that arise when dealing with weighted automata containing ε -transitions: how to define the behaviour of such automata in which the presence of ε -circuits results in infinite summations, and second how to eliminate the ε -transitions in an automaton whose behaviour has been recognised to be well-defined. The origin of this work is the implementation, in the AWALI platform [19], of an ε -transition removal algorithm for automata with weight in \mathbb{Q} or \mathbb{R} , a case that had never been treated before in the rich literature on the subject of ε -transition removal algorithms (*cf.* [16] for a survey). The results of this work have been published in [14].

The equivalence between non-deterministic finite automata (NFA) and non-deterministic finite automata with ε -transitions (ε -NFA) is a basic result in the theory of finite automata (*e.g.* in [10]). Besides the proof that ε -transitions do not increase the power of the computation model, this equivalence allows simple constructions; hence the interest in the *construction* of NFA from ε -NFA which amounts to the computation of the *transitive closure* of the graph of ε -transitions, more or less intertwined with the construction of the NFA itself.

Automata with weights taken in semirings have been considered since the beginning of the 60's and the recent investigations on quantitative evaluation of systems make them of current interest. Removal of ε -transitions is of equal importance for weighted automata as for Boolean ones, but even the definition of the behaviour of a weighted automaton with ε -transitions is not so obvious: if the graph of ε -transitions contains a circuit, then the sum of the weights along the paths following this circuit may well be not defined, as for instance in the automaton \mathcal{Q}_1 of Fig. 1(a). For that reason, some of the most mathematically oriented works on automata, such as [1] or [12], have ruled out the possibility of having circuits of ε -transitions in automata, either explicitly with the hypothesis of *cycle-free* automata, or implicitly by considering the *discrete topology* on the weight semiring. As a result, the automaton \mathcal{Q}_2 of Fig. 1(b) is also considered as an illegitimate object. Yet, the natural topology on \mathbb{Q} or \mathbb{R} should allow to consider that such weighted automata with circuits of ε -transitions may be *valid*.

J. Sakarovitch Joint work with S. Lombardy.

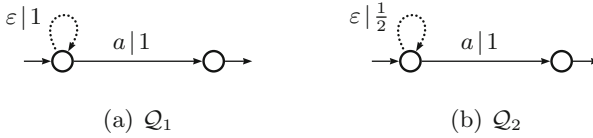


Fig. 1. A non-valid automaton \mathcal{Q}_1 , and \mathcal{Q}_2 that should be considered valid

The definition of the behaviour of weighted automata with ε -transitions requires a theoretical framework in which it is possible to define *infinite sums*, at least some particular ones. It appears indeed that these infinite sums are created by the paths along a circuit (as in the loops in Fig. 1) and may be expressed by the star operator.

A first approach to solving the issue of infinite sums, which we call *the axiomatic approach*, is the setting of a framework in which the star operator is always defined — making the previous evaluation always possible — and satisfies a number of axioms that guarantee the behaviour be well-defined and the evaluation correct. This yields on one hand the notions of *complete*, and *continuous*, semirings (cf. [2, 4, 7, 11]), and on the other the notion of *Conway* and *partial Conway semirings* (cf. [2, 3, 9]). The weighted automata we want to deal with, such as probabilistic or distance automata, are natural computational models, while their weight semirings (e.g. \mathbb{Q} , \mathbb{R} or \mathbb{Zmin}) are neither complete nor continuous nor partial Conway semirings. In order to address these cases, we have made the choice in our previous work ([17, 18]) of a *topological approach* to infinite sums and to the notion of *summability* of infinite families.

A natural definition for a *valid* automaton is then that the family of all computations be summable, that is, that for every word w the family of weights of paths with label w be summable. One may be more cautious, and require that for every pair of states the family of weights of paths between these states and with label w be summable. This definition, taken in [17, 18], yields a consistent theory; it however conceals the weakness of *not being effective* as it says nothing on how one can compute or decide that an infinite family is summable. As noted above, the evaluation of infinite sums is made through the star operator, which implies *block summations*. And then, computations are likely to meet two pitfalls, different, and almost opposite:

1. A subfamily of a summable family is not necessarily summable, and an evaluation may fail on what is considered as a ‘valid’ automaton. Such a situation occurs with the automaton \mathcal{T}_2 of Fig. 2(a) whose weights are in \mathbb{N}_∞ , the semiring $\mathbb{N} \cup +\infty$ equipped with the discrete topology, where an infinite sum is defined if, and only if, almost all terms are 0, or at least one term is $+\infty$. The sum of the weights of paths in this automaton is therefore defined, but the evaluation of this sum through the star operator may require to compute the star of 1 which is not defined.
2. On the other side, it is well-known that block summations may give values to non summable families. Automaton \mathcal{Q}_3 of Fig. 2(b) yields an example of such

a phenomenon. Weights of \mathcal{Q}_3 are in \mathbb{Q} where, for every x in $] - 1; 1[$, the star of x is defined and is equal to $(1 - x)^{-1}$. The usual formula for the evaluation of the computations labelled by ε gives $\left(\frac{-1}{2} + \frac{1}{2} \left(\frac{-1}{2}\right)^* \frac{1}{2}\right)^* = \frac{3}{5}$ but a direct computation on the *transition matrix* M yields $M^2 = -M$ which shows that computations grouped by length are not summable.

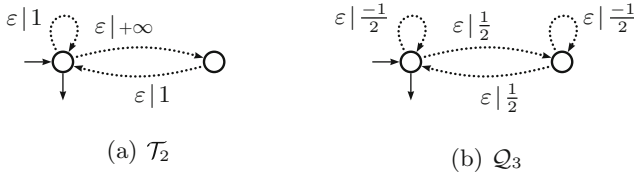


Fig. 2. The two pitfalls of the ε -removal computations

We therefore take a *stronger definition of validity*: we will say that a weighted automaton is *valid* if the sum of weighted labels of *every rational set of paths* is well-defined. This definition has several outcomes. It insures first that in any kind of semirings, any (reasonable) ε -removal algorithm will succeed on every valid weighted automaton and turn it into an equivalent proper automaton. Second, this definition is not as restrictive as it may appear. It rules out tricky examples in awkward semirings (such as \mathbb{N}_∞) but coincides with the former definition in all cases of usual semirings. It encompasses the cases treated with the axiomatic approach where indeed the decision of validity is trivial. Finally, this definition provides a framework in which the closure algorithm yields a decision procedure for the validity of automata in the cases we wanted to treat. All in all, this definition of validity is more complex than the other ones but fulfills our goals without narrowing its domain of application.

A byproduct of this definition of validity is the stability and consistency with natural transformations of automata: a subautomaton of a valid automaton is valid and so is a *covering* (kind of unfolding) of a valid automaton. The validity of automata is also consistent with that of rational expressions, provided the transformation of the latter into the former does not artificially introduce ε -transitions. This last condition brings to light a weakness of Thompson’s construction compared to the other classical ones.

After the definition of validity, the next step is the computation of the behaviour of an automaton, which amounts to describing algorithms for ε -transition removal. It begins with the study of the relationships between the validity of an automaton and the starability of its transition matrix. In the axiomatic approach, the behaviour of an automaton is defined by the star of the transition matrix and the problem is to set up conditions that guarantee its starability. The topological approach goes in the opposite direction and an automaton which is not valid may well have a starable transition matrix. It is proved that under the

condition that an automaton is valid, its transition matrix is starable and the classical formulas can be applied to compute the star. This leads to a first family of algorithms that have been abundantly described in the literature ([13, 15, 16]).

Another family of algorithms, hardly less classical, is based on the iteration of elementary operations of suppression of ε -transitions. The *correctness* comes from the correctness of each elementary operation, but the *termination* of the algorithm has to be carefully checked.

The final task is to address the problem of deciding the validity of an automaton. The success of an algorithm on a given automaton is not a sufficient condition for the automaton be valid. There is no general and uniform answer to the problem and that the existence, and the details, of a solution will depend on the properties of the weight semiring. It is for instance classical that the only valid automata with weights in \mathbb{N} or \mathbb{Z} are those with no circuits of ε -transitions, a decidable property. The opposite case, that is, the one in which every automaton is valid, and which meets the class of *rationaly additive semirings* of [8], is also characterised.

The cases in-between, that is, semirings such that some, but not all, automata with circuits of ε -transitions are valid, is a new area of research. In this work, it is established that the validity of \mathbb{K} -automata is decidable — decided by the success of the algorithm quoted above — when \mathbb{K} is a *topological (partially) ordered positive semiring* (TOPS) with the property that the domain of the star operator is downward closed. This class of semirings contains for instance \mathbb{Q}_+ , \mathbb{R}_+ , \mathbb{Z}_{\min} (and all the tropical semirings). Moreover, in starable TOPS — such as for instance the semiring $\text{Rat } A^*$ of rational languages over A^* — every weighted automaton is valid and this class contains the *continuous* semirings (*cf.* [5]).

With a last step, the decidability problem for \mathbb{Q} - or \mathbb{R} -automata is reduced to the one for automata with weights in TOPS by considering the *absolute value* of such automata. The initial goal of solving the problem of ε -transition removal algorithm for \mathbb{Q} - or \mathbb{R} -automata is then reached: we have both the theoretical justification and the algorithm that allows to transform the automaton \mathcal{Q}_4 of Fig. 3(a) into the automaton \mathcal{Q}_5 of Fig. 3(b).

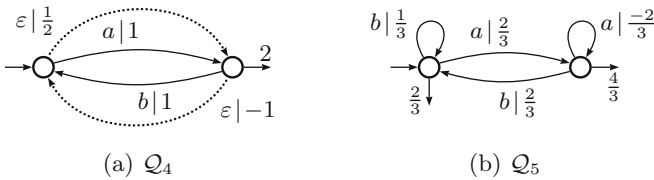


Fig. 3. A transformation which needs theoretical foundation

As a last remark, it can be mentioned that, for sake of simplicity, the only automata considered in this work are the weighted automata over free monoids, but the same definitions, constructions and results hold for automata over *graded monoids* (that is, monoids with a *length*), and thus in particular for *transducers*.

This work fills the *effectivity* gap that was left open by our definition of valid automata in [17, 18]. In general, the success of any ε -removal algorithm may never insure the validity of an automaton. In order to reach a decidability procedure, we have been led to a strong definition of validity, which implies stability under automata transformations that are paths preserving. The algorithms described in this work are *implemented* in AWALI [19].

References

1. Berstel, J., Reutenauer, C.: Noncommutative Rational Series with Applications. Cambridge University Press, Cambridge (2011)
2. Bloom, S.L., Ésik, Z.: Iteration Theories. Springer, Heidelberg (1993). <https://doi.org/10.1007/978-3-642-78034-9>
3. Bloom, S.L., Ésik, Z., Kuich, W.: Partial Conway and iteration semirings. *Fundam. Inform.* **86**, 19–40 (2008)
4. Conway, J.H.: Regular Algebra and Finite Machines. Chapman and Hall, London (1971)
5. Droste, M., Kuich, W.: Semirings and formal power series. In: Droste et al. [6], pp. 3–28
6. Droste, M., Kuich, W., Vogler, H. (eds.): Handbook of Weighted Automata. Springer, Heidelberg (2009). <https://doi.org/10.1007/978-3-642-01492-5>
7. Eilenberg, S.: Automata, Languages and Machines, vol. A. Academic Press, New York (1974)
8. Ésik, Z., Kuich, W.: Locally closed semirings. *Monatshefte für Mathematik* **137**, 21–29 (2002)
9. Ésik, Z., Kuich, W.: Finite automata. In: Droste et al. [6], pp. 69–104
10. Hopcroft, J.E., Motwani, R., Ullman, J.D.: Introduction to Automata Theory, Languages and Computation, 3rd edn. Addison-Wesley, Boston (2006)
11. Kuich, W.: Automata and languages generalized to ω -continuous semirings. *Theoret. Comput. Sci.* **79**, 137–150 (1991)
12. Kuich, W., Salomaa, A.: Semirings, Automata, Languages. Springer, Heidelberg (1986). <https://doi.org/10.1007/978-3-642-69959-7>
13. Lehmann, D.J.: Algebraic structure for transitive closure. *Theoret. Comput. Sci.* **4**, 59–76 (1977)
14. Lombardy, S., Sakarovitch, J.: The validity of weighted automata. *Int. J. Algebra Comput.* **23**(4), 863–914 (2013)
15. Mohri, M.: Generic ε -removal and input ε -normalization algorithms for weighted transducers. *Int. J. Found. Comput. Sci.* **13**, 129–143 (2002)
16. Mohri, M.: Weighted automata algorithms. In: Droste et al. [6], pp. 213–254
17. Sakarovitch, J.: Elements of Automata Theory. Cambridge University Press, Cambridge (2009). corrected English translation of *Éléments de théorie des automates*, Vuibert, Paris (2003)
18. Sakarovitch, J.: Rational and recognisable power series. In: Droste et al. [6], pp. 105–174
19. Awali: Another Weighted Automata Library. vaucanson-project.org/AWALI