

# Identification of partially observed deterministic and stochastic cellular automata

Witold Bołt

Warsaw, 2023

Cellular Automata (CAs) are dynamical systems in which space, time and the state set are discrete (very often finite). Despite this simplistic characterization, CAs exhibit many complex properties. Due to these two factors – simple characterization and interesting properties, CAs are an important tool for physical systems modelling as well as an important object of the dynamical systems theoretical studies.

This dissertation is devoted to the CA identification problem – an automated method of finding CAs matching a given set of observations. In other, less formal words, the identification algorithm takes a set of, potentially incomplete, “images” of an unknown CA and tries to find a CA that is able to generate the same image while filling all the missing information. These observations are to be understood as measurements or recordings of the evolution of some abstract, unknown system which is to be modelled with CAs.

Two classes of binary CAs are considered in this context: deterministic and non-deterministic. For both of these classes an identification algorithm is proposed, which allows to successfully find CAs based on a set of incomplete observations. In this work incompleteness (or partiality) of observations relates to missing state information of some cells at specific time steps and also to the absence of entire time steps. In the latter case, the exact number of missing time steps is also unknown. Such a problem setting is motivated by practical aspects of modelling of real-world phenomena.

Although the goal and problem setting are very similar for deterministic and non-deterministic CAs, the solution strategies discussed here differ greatly. For identifying deterministic CAs an evolutionary algorithm is used, which starts from a random set of potential solutions that are further evolved with the use of genetic operators. This algorithm can be time consuming and is non-deterministic. Yet, the results of large-scale computational experiments

confirm that it solves the problem even in the challenging setting of partial observations with many missing cells and time steps. Moreover, the presented analysis of these results shows a strong relationship between dynamical properties of CAs and the effort of identification.

In contrast, for the non-deterministic case, a direct solution construction algorithm is given, which is based on statistical inference and parameter estimation. It yields high-quality results and is computationally efficient, but lacks the ability to deal with missing time steps – only missing cells’ states are considered as a model of incompleteness. Similarly to the deterministic CA case, here also a rich set of results from computational experiments is given confirming the efficiency of the algorithm and analyzing the impact of some of the dynamical properties of CAs on the identification quality.

The discussed problem setting and corresponding solution algorithms are studied in a theoretical setting and relate only to specific classes of 1D, binary CAs. Yet, this theoretical setting allows for a very detailed analysis of results. Moreover, the way in which the identification algorithms are constructed allows for straightforward extensions to more general cases, which may be closer to real-world modelling tasks.

This dissertation is concluded with a discussion of future research directions including an outline of the plans for certain generalizations of the presented algorithms as well as new solution approaches. These methods are based on modern machine learning techniques which can be used both as an extension of the presented algorithms (to reduce the computational cost) or as alternative solution strategies.