Dartmouth Computer Science Technical Report TR2005-555 (August 11th)

The Theory of Trackability with Applications to Sensor Networks

Valentino Crespi
Department of Computer Science
California State University Los Angeles, ECST
5151 State University Drive, Los Angeles CA 90032
Email: vcrespi@calstatela.edu

George Cybenko
Thayer School of Engineering
Dartmouth College, Hanover, NH 03755
Email: gvc@dartmouth.edu

Guofei Jiang
NEC Laboratories America
4 Independence Way, Princeton, NJ 08540
Email: gfj@nec-labs.com

Abstract

In this paper, we formalize the concept of tracking in a sensor network and develop a rigorous theory of *trackability* that investigates the rate of growth of the number of consistent tracks given a sequence of observations made by the sensor network. The phenomenon being tracked is modelled by a nondeterministic finite automaton and the sensor network is modelled by an observer capable of detecting events related, typically ambiguously, to the states of the underlying automaton.

More formally, an input string, Z^t, of $t+1$ symbols (the sensor network observations) that is presented to a nondeterministic finite automaton, M, (the model) determines a set, $\mathcal{H}_M(Z^t)$, of state sequences (the tracks or hypotheses) that are capable of generating the input string Z^t. We study the growth of the size of this set, $|\mathcal{H}_M(Z^t)|$, as a function of the length of the input string, $t+1$. Our main result is that for a given automaton and sensor coverage, the worst-case rate of growth is either polynomial or exponential in t, indicating a kind of phase transition in tracking accuracy.

The techniques we use include the Joint Spectral Radius, $\rho(\Sigma)$, of a finite set, Σ, of $(0,1)$-matrices derived from M. Specifically, we construct a set of matrices, Σ, corresponding to M with the property that $\rho(\Sigma) \leq 1$ if and only if $|\mathcal{H}_M(Z^t)|$ grows polynomially in t. We also prove that for $(0,1)$-matrices, the decision problem $\rho(\Sigma) \leq 1$ is Turing decidable and, therefore, so is the problem of deciding whether worst case state sequence growth for a given automaton is polynomial or exponential. These results have applications in sensor networks, computer network security and autonomic computing as well as various tracking problems of recent interest involving detecting phenomena using noisy observations of hidden states.

1 Introduction

This paper studies the problem of tracking a nondeterministic finite automaton using indirect, possibly noisy, observations of the states that the automaton occupies over time. This kind of problem arises in many applications of current interest such as tracking using a sensor network, computer and networks security using various network- and host-based sensors, and autonomic computing systems in which the goal is to develop systems capable of estimating their own states from indirect observations. While many such applications use statistical models, our work addresses so-called weak models which are nondeterministic but not probabilistic. That is, we only consider what is *possible* not what is *probable*.

Research on probabilistic tracking in a sensor network, using tracking formalisms such as Bayesian maximum likelihood estimation and Multiple Hypothesis Tracking [14], is currently being investigated by other groups [11]. This paper does not require any probabilistic or statistical assumptions, thus allowing the results to be applicable to domains in which the estimation of probabilities is either inappropriate or difficult.

After developing a formal model for tracking a nondeterministic finite automaton, we introduce the notion of *trackability*. A problem is *trackable* if the rate of growth of the number of possible hypotheses, or state sequences, that are consistent with a sequence of observations of the state machine is subexponential. It is important to differentiate the state estimation problem (namely, what is the set of states that the model could currently be in) from the state sequence estimation problem (namely, what are the possible state trajectories over the whole observation time). In this paper, we are concerned with the latter and we demonstrate that the growth rate of the set of possible state sequences is either polynomial or exponential.

We now introduce some of the formal constructs which will allow us to develop these ideas rigorously. Let $G = (V, E)$ be a finite directed graph and A_G be the vertex adjacency matrix corresponding to the directed edges in E. We identify the graph G with an automaton in which the vertices represent states and directed edges are the allowable state transitions. Transitions from a state to itself are allowed so that A_G can have nonzero diagonal entries. Moreover, there is a finite set, Φ, of possible observable events related to the automaton and a mapping $L : V \to 2^{\Phi}$ of states to subsets of events. The automaton transitions from state to state according to the allowable transitions determined by E. When the automaton enters a state, say v, one of the events contained in $L(v) \subseteq \Phi$ is generated and observed.

We call such a construction a *weak model* by analogy with Hidden Markov Models (HMMs) [13]. Weak models have the structural ingredients of Hidden Markov Models, namely states, state transitions and observations related to states, but they do not have probabilities associated with those state transitions and observations. Weak models only describe the possible transitions and possible associations of transitions to states, without assigning any probabilities to them. Weak models have been used in a variety of process detection problems [8].

More formally, a weak model is an automaton defined by a quadruple, $W = (V, E, L, \Phi)$, with the following properties:

- corresponding to each state (vertex), $v \in V$, is a finite set of possible outputs, $L(v) \subseteq \Phi$, which are not necessarily unique to that state; that is, $L(v) \cap L(v') \neq \emptyset$ is possible;
- when occupying a state, the underlying automaton generates exactly one of the possible outputs associated with that state which an external observer can detect.

One weak model, $W' = (V, E', L', \Phi)$ is a *noisy* version of another weak model, $W = (V, E, L, \Phi)$, written as $W \leq W'$ if $L(v) \subseteq L'(v)$ for all $v \in V$ and $E \subseteq E'$. This notion of noise corresponds to the usual intuitive concept of noise in the sense that an observation, say ξ, is associated with a set of states of W' which are always a superset of the states of W that can be associated with ξ, since for every state $v \in V$, $\xi \in L(v)$ implies $\xi \in L'(v)$ due to the inclusions $L(v) \subseteq L'(v)$. Additionally, the condition $E \subseteq E'$ states that the legal state transitions of W are a subset of the state transitions allowed in W' so that W is a more specific model than W'.

A single observation, $\xi \in \Phi$, of W determines a set of possible states, $\mathcal{H}_W(\xi)$, trivially defined by $\mathcal{H}_W(\xi) = \{v|\xi \in L(v)\}$. The set $\mathcal{H}_W(\xi)$ is a collection of ***hypotheses*** about which state the automaton was in when ξ was observed. A pair of consecutive observations, $\xi_0\xi_1$, determines a set of pairs of states according to

$$\mathcal{H}_W(\xi_0\xi_1) = \{v_0v_1 \mid v_0 \in \mathcal{H}_W(\xi_0),\ v_1 \in \mathcal{H}_W(\xi_1) \text{ and } (v_0, v_1) \in E\}.$$

This allows a recursive construction of hypotheses for longer sequences of consecutive observations as follows. Suppose $Z^T = \xi_0\xi_1...\xi_{T-1}\xi_T$ are $T+1$ consecutive observations of W. Then

$$\mathcal{H}_W(Z^T) = \{\ v_0v_1...v_T \mid v_0v_1...v_{T-1} \in \mathcal{H}_W(Z^{T-1}),\ v_T \in \mathcal{H}_W(\xi_T) \text{ and } (v_{T-1}, v_T) \in E\}.$$

Evidently, if $W \leq W'$ then if follows from the definitions that $\mathcal{H}_W \subseteq \mathcal{H}_{W'}$ for all observation sequences because W has more restrictive state transitions and observation-to-state associations.

One of the main results of this paper is that $h_W(Z) = |\mathcal{H}_W(Z^T)|$, the cardinality of the set $\mathcal{H}_W(Z^T)$, grows either polynomially or exponentially in T. Note that for $W \leq W'$, we must have $h_W(Z^T) \leq h_{W'}(Z^T)$, so that the number of hypotheses relative to a sequence of models is monotonically increasing as the models get noisier in the sense of our definitions above. Accordingly, if we have a family of weak models, say $W \leq W^{(1)} \leq W^{(2)} \leq ... \leq W^{(k)}$, our results show that there is an abrupt change in the worst case growth rate of hypotheses, from polynomial to exponential, as the models get noisier. This abrupt change is a kind of phase transition in the modelling process.

Our main motivation for this work arises in applications in which the basic framework is similar to that of Hidden Markov Models but without the underlying probabilistic assumptions. That is, in weak models, state transitions and output-state associations are simply possible or not, as in state machines, but the outputs are associated with states as in an HMM, and not with state transitions (edges) as in the usual definition of a state machine. We will use "outputs" and "observations" synonymously, in the sense that the model produces an output while externally that output is an observation detectable by an observer.

A weak model as defined above is equivalent to a nondeterministic finite automaton (NDFA) with at most a polynomial growth in the number of states and state transitions. Another difference between weak models and the usual definition of finite state machines is that there are no initial or accepting states in a weak model. This minor difference allows observation of the automaton to start at any time, not just when the automaton is initialized and to continue indefinitely. In other words, all states of weak models can be considered initial states and none are accepting states. Specific relationships between weak models, nondeterministic finite automata, deterministic finite automata and other constructs are outlined in [18].

In this paper, we investigate the *trackability* of a weak model; that is, the extent to which a given sequence of observations or outputs can effectively determine the actual set of state sequences that could have produced the outputs. Given a sequence of observations, we consider the set of possible state sequences that could have produced the observed outputs. By analogy with classical computational complexity theory, we will say that a model is *trackable* if the size of that set grows polynomially as a function of the observation sequence length in the worst case for a fixed model. The model is *untrackable* if the set grows exponentially in the worst case.

One of the main results of this paper is a proof of the fact that these two cases are the only ones possible. Specifically, the number of possible state sequences that could have produced an observation sequence grows either polynomially or exponentially in the worst case output scenario. Furthermore, we demonstrate that the problem of determining whether a given weak model is trackable or untrackable can be effectively computed, although we are not aware of efficient (subexponential) algorithms for making that determination.

1.1 Paper Outline and Organization

We study the problem of state sequence growth in weak models by formulating the problem in terms of matrix norms of sequences of 0-1 matrices. The relevant mathematical machinery we invoke involves the notion of the *Joint Spectral Radius* of a set of matrices. The Joint Spectral Radius, $\rho(\Sigma)$, is a quantity introduced by Rota and Strang [16] as a generalization, to a set of matrices Σ, of the classical notion of the spectral radius of a matrix. The Joint Spectral Radius is a quantity that measures the worst-case rate of growth of the norm of products of matrices from a finite set of real matrices. The concept arises commonly in the study of the boundedness of time-varying dynamical systems of the form $x_{t+1} = A_t x_t$ [9, 4, 20].

We relate the State Sequence Growth Problem to the value of the Joint Spectral Radius $\rho(\Sigma)$ of a finite set Σ of $(0,1)$-matrices derived from G as above in the simple example, namely from the possible state transitions and the sensor model. We reduce the question of polynomial versus exponential rates of growth of the number of hypotheses to the Joint Spectral Radius of a derived set of matrices as above. In particular we establish that

- In a weak model, the worst-case number of hypotheses consistent with a given sequence of t observations, grows polynomially if and only if the Joint Spectral Radius of the induced finite set of $(0,1)$-matrices (see Section 4) is less than or equal to 1.
- Conversely, in a weak model the number of hypotheses consistent with a sequence of t observations grows exponentially if and only if the Joint Spectral Radius of the induced finite set of $(0,1)$-matrices is larger than 1. Namely there are no intermediate rates of growth.
- Finally, unlike the general case for rational matrices, the decision problem $\rho(\Sigma) \leq 1$ for $(0,1)$-matrices is (Turing) solvable. Accordingly, given a weak model, we can decide (in the Turing sense) whether there exists a family of sequences of observations such that the number of consistent hypotheses grows at least exponentially or for all families of sequences of observations the growth is at most polynomial.

In Section 2 we develop some simple examples to illustrate the main concepts and constructs as well as applications to noisy sensor networks. In Section 3 we present some definitions and results about the Joint Spectral Radius, $\rho(\Sigma)$, and the Generalized Spectral Radius. In Section 4, we study how the State Sequence Growth Problem relates to the Joint and Generalized Spectral Radius concepts and establish our main results. Section 5 contains a summary with several suggestions for future work. The Appendix contains four examples to illustrate how these ideas apply to different simple sensor networks and associated object kinematics.

2 Illustrative Examples

2.1 Tracking in a Simple Sensor Network

The framework described above applies to a simple version of tracking an object, say a vehicle or person, using a network of sensors for example. Figures 1, 2 and 3 are used to illustrate the example. The reader is encouraged to consider how these simple ideas apply to other domains as well. Our interest in this problem arose from the study of effectively using weak models for detecting and tracking processes in a variety of applications such as computer security [3], autonomic computing [15] and object tracking using sensor networks [6]. Examples showing the effects of process dynamics (the state machine model) and sensor coverage on hypothesis growth are developed in the Appendix.

In this simple tracking application, the states simply correspond to three rooms and the system is in one of the three states if an object (a person for example) is currently in the corresponding room. If a person is in Room 2, for example, the system is in state s_2 and so on.

Figure 1 also shows the sensor coverage which determines the possible observations that correspond to each state. In Sensor Coverage α, the observation will be 0 if the object is in states s_1 or s_2 and the observation will be 1 if the object is in state s_3. In Sensor Coverage β, the observation is a 1 if the state is s_2 and 0 otherwise.

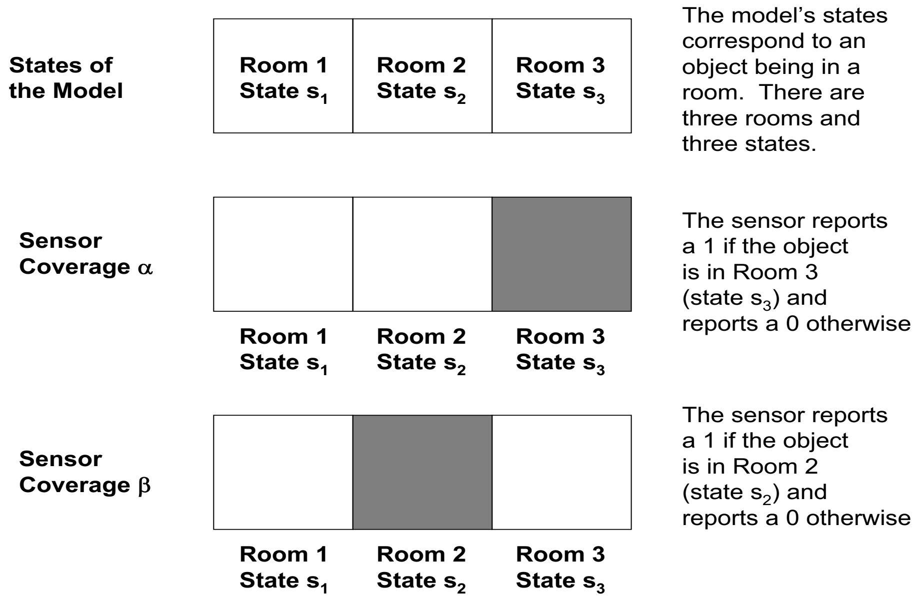

Figure 1: The underlying state space has three states (each corresponding to occupancy of a room by an object). In Sensor Coverage α, the sensor detects presence in state s_3 (room 3) and in Sensor Coverage β, the sensor detects presence in state s_2 (room 2). The sensors report a 0 if no object is present and a 1 if an object is present (in each of the corresponding states (rooms). The object moves according to the dynamical models shown in Figure 2.

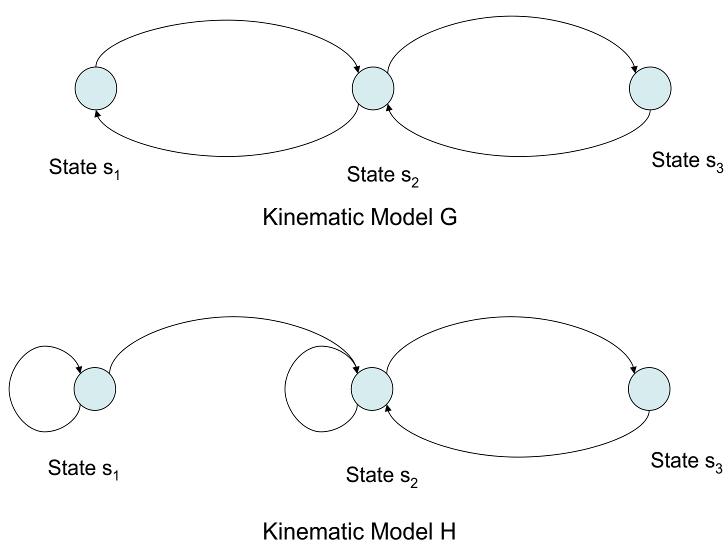

Figure 2: There are two dynamical models, G and H, that define the allowed movement between states (rooms) of the state space.

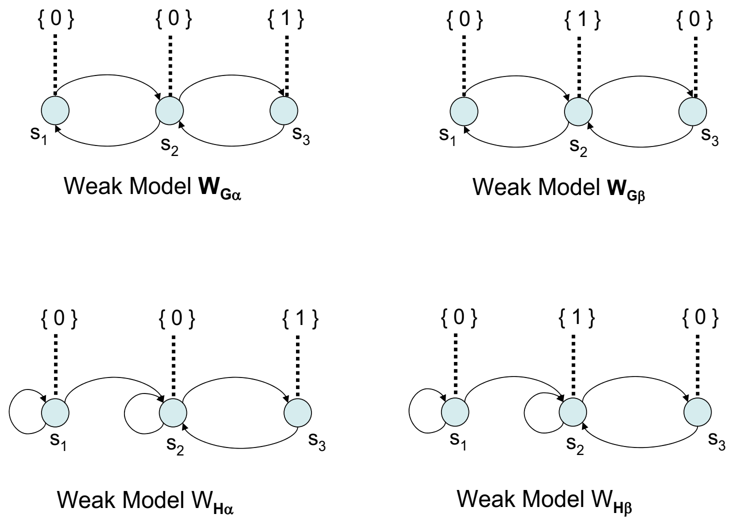

Figure 3: There are four resulting *weak models* corresponding to the possible combinations of two different sensor coverages and two different dynamical models. This figure depicts the four models. In these graphical depictions, the possible sensor reports for each state are shown in braces connected to the state with a dotted line.

Using the notation introduced above, we have $L_\alpha(s_1) = L_\alpha(s_2) = \{0\}$ and $L_\alpha(s_3) = \{1\}$ while $L_\beta(s_1) = L_\beta(s_3) = \{0\}$ and $L_\beta(s_2) = \{1\}$.

Figure 2 shows two possible kinematic models for the object moving between the three rooms. In model G, the object must move between adjacent rooms at each time step. In model H, the object can stay in rooms 1 and 2 indefinitely or can move from room 1 to room 2, room 2 to room 3 or room 3 to room 2. The object cannot stay in room 3 and cannot move from room 2 to room 1 in model H.

These two kinematic models combined with the two sensor models lead to four weak models as depicted in Figure 3, namely $W_{G\alpha}, W_{G\beta}, W_{H\alpha}$ and $W_{H\beta}$.

Now consider an observation, 0, in each of the four weak models. We have hypothesis sets: $\mathcal{H}_{W_{G\alpha}}(0) = \{s_1, s_2\}, \mathcal{H}_{W_{G\beta}}(0) = \{s_1, s_3\}, \mathcal{H}_{W_{H\alpha}}(0) = \{s_1, s_2\}$ and $\mathcal{H}_{W_{H\beta}}(0) = \{s_1, s_3\}$. Suppose that two consecutive 0's are observed. We then get hypothesis sets:

$$\begin{aligned}
\mathcal{H}_{W_{G\alpha}}(00) &= \{s_1s_2, s_2s_1\}, \\
\mathcal{H}_{W_{G\beta}}(00) &= \{\ \} = \emptyset, \\
\mathcal{H}_{W_{H\alpha}}(00) &= \{s_1s_1, s_1s_2, s_2s_2\}, \\
\mathcal{H}_{W_{H\beta}}(00) &= \{s_1s_1\}.
\end{aligned}$$

Similarly, $\mathcal{H}_{W_{G\alpha}}(1) = \{s_3\}, \mathcal{H}_{W_{G\beta}}(1) = \{s_2\}, \mathcal{H}_{W_{H\alpha}}(1) = \{s_3\}, \mathcal{H}_{W_{H\beta}}(1) = \{s_2\}$ and

$$\begin{aligned}
\mathcal{H}_{W_{G\alpha}}(10) &= \{s_3s_2\}, \\
\mathcal{H}_{W_{G\beta}}(10) &= \{s_2s_1, s_2s_3\}, \\
\mathcal{H}_{W_{H\alpha}}(10) &= \{s_3s_2\}, \\
\mathcal{H}_{W_{H\beta}}(10) &= \{s_2s_3\}.
\end{aligned}$$

In this article we are concerned about the worst-case growth of the hypothesis sets $\mathcal{H}_W$. By inspection, note that

$$\begin{aligned} |\mathcal{H}_{W_{G\alpha}}(0^t)| &= |\{s_1s_2..., s_2s_1...\}| = 2, & (1)\\ |\mathcal{H}_{W_{G\beta}}((01)^{2t})| &= |\{s_1s_2s_1..., s_1s_2s_3..., ..., s_3s_2s_3...\}| = 2^t, & (2)\\ |\mathcal{H}_{W_{H\alpha}}(0^t)| &= |\{s_1s_1s_1...s_1, s_1s_1....s_1s_2, s_1s_1...s_1s_2s_2\}| = t, & (3)\\ |\mathcal{H}_{W_{H\beta}}(01^t)| &= |\{s_1s_2....s_2, s_3s_2s_2...s_2\}| = 2 & (4)\end{aligned}$$

where a^t for a string a means repeating the string t times. The hypothesis growth in these cases is either constant (namely 2), polynomial (namely t), or exponential (namely $2^{t/2}$). The reader can verify that a model like $W_{H\alpha}$ but with k additional rooms like s_1 and s_2 for which the object can either stay in a room or move to rooms only on the right leads to a worst-case polynomial growth of order k. Such an example is worked out in more detail in the Appendix.

It is useful, for an understanding of the matrix formulation that follows below in this paper, to develop the matrix operations underlying the different hypothesis growths.

The kinematic models G and H can be represented by state transition matrices (or equivalently node-node adjacency matrices) A_G and A_H:

$$A_G = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad , \quad A_H = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The ijth entry of these matrices is 1 if a transition from state i to state j is possible in the model and 0 if the transition is not possible. Additionally, the sensor report to observation relationships in Figure 1 can be summarized similarly by

$$I_\alpha(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} , \; I_\alpha(1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \; I_\beta(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \; I_\beta(1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

These matrices have a 1 in the ith diagonal position if the observation could have been generated while the system is in state i and a 0 in the ith diagonal position otherwise.

The relationship between the number of hypotheses and these matrix representations goes as follows. Suppose that we are dealing with the kinematic model G and the sensor coverage α and that we observe the sequence of $t+1$ observations $\xi_0\xi_1\xi_2\ldots\xi_t$. Let $\underline{1} = [111]^T$ denote the column vector of all 1's. Having observed ξ_0, the possible states are $\underline{1}^T I_\alpha(\xi_0)$ where state s_i could have generated that observation if the ith coordinate of the product vector is 1 and state s_i could not have generated that output if the ith coordinate is 0. Now if we next observe ξ_1 then the possible states are determined by

$$\underline{1}^T I_\alpha(\xi_0) A_G I_\alpha(\xi_1).$$

In general, let $z_{G\alpha}(\xi_0\xi_1\xi_2\ldots\xi_k)$ be the row vector whose ith coordinate is the number of hypotheses (that is, possible state sequences) that end in state s_i and that are consistent with the observations $\xi_0\xi_1\ldots\xi_k$. Then

$$z_{G\alpha}(\xi_0\xi_1\xi_2\ldots\xi_k\xi_{k+1}) = z_{G\alpha}(\xi_0\xi_1\xi_2\ldots\xi_k)A_G I_\alpha(\xi_{k+1}) = \underline{1}^T I_\alpha(\xi_0)\left[\prod_{j=1}^{k+1}(A_G I_\alpha(\xi_j))\right].$$

A recursive argument easily establishes that if $z_{G\alpha}(\xi_0\xi_1\xi_2\ldots\xi_k)$ is as claimed, then

$$z_{G\alpha}(\xi_0\xi_1\xi_2\ldots\xi_k)A_G$$

is a row vector representing the number of state sequences consistent with $\xi_0\xi_1\ldots\xi_k$ as propagated to the next time step. Multiplying that vector by $I_\alpha(\xi_{k+1})$ on the right merely selects the state sequences that are further consistent with the observation ξ_{k+1}. These operations can be thought of as a prediction step followed by a filtering step, not unlike what is used in Kalman Filtering for continuous state space filtering.

As an example, consider again model G and sensor coverage α and suppose we observe the sequence 000. Then the number of hypotheses consistent with those observations is given by the expression

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}^T = z_{G\alpha}(000).$$

Similarly, if we use the kinematic model given by H, then the number of possible state sequences, or equivalently hypotheses, is given by

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}^T = z_{H\alpha}(000).$$

In each case, the ith coordinate of the resulting vector is a count of the number of hypotheses which are consistent with the given observations and end in the ith state. For example, given the observations 000, there is only one state sequence in model $W_{H\alpha}$ that ends in state s_1, namely $s_1 s_1 s_1$ while there are three consistent state sequences that end in state s_2, namely $s_1 s_1 s_2$, $s_1 s_2 s_2$ and $s_2 s_2 s_2$. There are no consistent state sequences that end in state s_3 because then the last observation would have had to have been a 1 not a 0.

The total number of hypotheses (possible state sequences) is clearly given by $z\underline{1}$. Moreover, letting $A_G(\xi_j) = A_G I_\alpha(\xi_j)$ we see that the number of hypotheses is given by

$$h_{W_{A_G\alpha}}(Z^t) = |\mathcal{H}_{W_{A_G\alpha}}(\xi_0\xi_1 \ldots \xi_t)| = \underline{1}^T I_\alpha(\xi_0) \prod_{j=1}^{t} A_G(\xi_j)\underline{1}.$$

Evidently, the number of hypotheses consistent with an observation sequence is related to the matrix norm of a product of matrices drawn from a finite set of 0-1 matrices, namely the possible $A_G(\xi_j)$ as determined by the kinematics of the underlying model and the associated sensor coverage. It should be noted that, in the general case, if the underlying state space has n states, then the total number of possible matrices of the form $A_G(\xi_j)$ is 2^n since there are 2^n possibilities for $I(\xi)$, namely all possible 0-1 possibilities for the diagonal entries.

2.2 An Application to Noisy Sensor Networks

For the purposes of this example, a ***sensor network*** is a collection of sensors each of which detects physical signals in the environment in which they are deployed. For simplicity, suppose the sensors are binary meaning that they only report either a 0 or a 1 at each sampling time instant, depending on whether they detect a signal in their neighborhood or not. For example, the sensor network could consist of acoustic microphones deployed in some geographical region. A single microphone sensor will report a 0 if it detects no acoustic signal (thresholded typically) and a 1 if it does. At each sampling instant, the sensor network consisting of this collection of m simple microphones reports a binary m-vector with as many coordinates as there are sensors.

The set of possible binary vectors is $\{0,1\}^m$ and the set of subsets of possible binary vectors is similarly denoted by $2^{\{0,1\}^m}$. Now suppose that a vehicle is moving through the region where the sensor network (microphones) is deployed. The vehicle's engine and tires make sounds that the microphones may or may not detect depending on their proximity to the vehicle and the local topography and ground cover, among other things.

Suppose that the region is discretized into n cells with each cell corresponding to a state of the vehicle in this model. The possible movements between these states then determines the dynamics of the model, namely the possible state transitions in a given time interval. The state space and associated dynamics are described by $G = (V, E)$.

The mapping, L, of the n states to the $2^{\{0,1\}^m}$ subsets of possible observations effectively associates a subset of possible sensor reports with each state of the model (in this example, the discretized location of the vehicle) . This abstraction captures the overall properties of the sensor network and vehicle model in this example .

Under ideal conditions, we could expect each state to be uniquely identified by an m-tuple bits, that is the mapping L associates a subset consisting of a single binary vector with each state. This would correspond to a *noiseless* sensor network in the sense that there is a one-to-one correspondence between the states of the system (vehicle location) and possible sensor observations. Noise can flip some bits in the observed binary m-vector. In such a case, the mapping may no longer be one-to-one.

Suppose we allow k bits to be flipped in this way, corresponding with noise in the measurements, where $0 \leq k \leq m$. The resulting models, M_k, form a hierarchy of weak models, each *noisier* than the other.

More formally, consider a weak model $M = (V, E, L, 2^{\{0,1\}^m})$ where $G = (V, E)$ is a directed graph capturing the kinematics of the underlying system and $L :\to \Phi$ is a mapping function that assigns a set of binary m-uples to each state:

$$L : V \to 2^{\{0,1\}^m} \ .$$

In ideal conditions (no noise), $|L(v)| = 1$, for all $v \in V$, and $L(u) \cap L(v) = \emptyset$, for $u \neq v$. We add *noise* by allowing k or fewer bits to flip in the m-tuple during the reporting of the sensors. Noise level k means that for each state $u \in V$, the set $L_k(u) = \{y \in \{0,1\}^m \mid d_H(L(u), y) \leq k\}$ is the set of m-tuples that can be reported by the sensor network when the system is in state u. Here $d_H(x, y)$ is the standard Hamming distance between binary m-vectors x and y.

For each noise level, k, we now have a mapping function, $L_k : V \to 2^{\{0,1\}^m}$, according to $L_k(u)$ defined above, for all $u \in V$. By varying the noise level k we obtain an ordered sequence of weak models

$$W_0 \leq W_1 \leq \ldots \leq W_m$$

where $W_k = (V, E, L_k, 2^{\{0,1\}^m})$ is a noisy version of $W_{k-1} = (V, E, L_{k-1}, 2^{\{0,1\}^m})$, in the sense made clear before, and W_0 is the model in ideal conditions. According to the various concepts we have introduced, we must have that

$$h_{W_0}(t) \leq h_{W_1}(t) \leq \ldots \leq h_{W_m}(t).$$

Moreover, one of the main results of this paper implies that there is some noise level, k^*, for which $h_{W_k}(t)$ grows polynomially in t if $k < k^*$ while $h_{W_{k^*}}(t)$ grows exponentially in t. It might be that $k^* = 0$ or $k^* = m + 1$, namely that all such models have hypothesis growth rates that are either all polynomial or all exponential.

We can also vary the sampling rate at which the sensor network reports its observations. This can change the possible state transitions of the underlying kinematic model. In particular if we reduce the sampling frequency, the system might be able to transition between states many times between consecutive sensor readings. This means adding new transitions to the kinematic model. Conversely, increasing the sampling frequency might remove some possible transitions and introduce the possibility of staying in the same state between sensor samples whereas that might not be possible at a lower sampling frequency.

For the sake of simplicity, let us assume that we can only decrease the sampling frequency in integer quantities: that is, decrease the sampling by $1/2, 1/3, \ldots$ from the original given timed kinematics with the understanding that decreasing the sampling rate by $1/s$ allows the system to make as many as s transitions allowed in the underlying kinematic model, $G = (V, E)$. We will

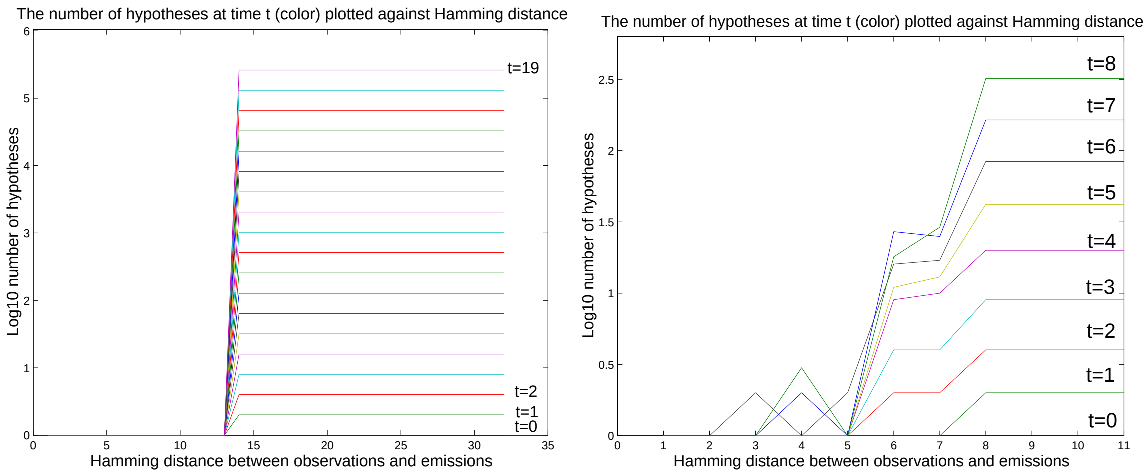

Figure 4: a) The transitions to exponential growth can be very abrupt (left). b) Transitions to exponential growth can be gradual, transitioning through a polynomial growth region (right).

simply say that the sampling rate is s if a sample is taken only every s time steps so that there are s possible state transitions in the underlying state machine.

Varying the sensor noise and changing the sampling rate as above creates a family of weak models doubly indexed by the noise level, k, and by the sampling rate, s. This collection of weak models is partially ordered according to the notion of noisy models we introduced above. Note,

$$W_{k,s} \leq W_{k',s'} \text{ if and only if } k \leq k' \text{ and } s \leq s' \ .$$

With the sampling frequencies discretized and interpreted this way, the partial order is a lattice where the minimum is determined by the model with noise level 0 and with the fewest possible number of transitions; whereas the maximum is represented by a fully connected (directed) graph whose states can emit all the possible m-tuples.

Figures 4a and 4b show how the number of hypotheses can grow both as a function of time, represented by the different lines in each graph with the number of hypotheses at different times with represented by a different line and increasing time generally corresponding to higher lines in the plot, and the Hamming Distance noise metric as defined in the text on the horizontal axis. The vertical axis is the number of hypotheses. Accordingly, each line, going from left to right depicts the growth in the number of hypotheses, for a fixed time, as the Hamming Distance noise measure increases. These examples were taken from randomly generated weak models. A random but legal observation sequence was generated and then the total number of consistent hypotheses was computed for that observation sequence.

For a given underlying kinematic model and sensor deployment, we can define ***strong Nyquist*** models to be those models, $W_{k,s}$ for which $h_{W_{k,s}}(t) = 1$ while $h_{W_{k',s'}}(t) > 1$ if $k' > k$ or $s' > s$. Similarly, we can define ***weak Nyquist*** models to be those models, $W_{k,s}$ for which $h_{W_{k,s}}(t)$ is bounded by a polynomial while $h_{W_{k',s'}}(t)$ is exponential if $k' > k$ or $s' > s$. The analogy with classical Nyquist sampling theory in signal processing should be apparent.

3 A Review of the Joint Spectral Radius

Let A be a matrix and $\sigma(A)$ be its spectrum. The spectral radius of A is defined by

$$\rho(A) = \max\{|\lambda| \mid \lambda \in \sigma(A)\} \ .$$

The Joint Spectral Radius [16] is a generalization of this concept to a set of matrices and is based on the following well known identity:

$$\rho(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$$

for any norm. Let Σ be a finite set of matrices in $R^{n\times n}$. Then the *Joint Spectral Radius* $\overline{\rho}(\Sigma)$ is defined by

$$\overline{\rho}(\Sigma) = \limsup_{k\to\infty} \overline{\rho}_k(\Sigma)$$

where, for $k \geq 1$

$$\overline{\rho}_k(\Sigma) = \sup\{\|A_1 A_2 \cdots A_k\|^{1/k} : A_i \in \Sigma\} \ .$$

Furthermore, if the norm satisfies $\|AB\| \leq \|A\| \cdot \|B\|$ (that is, an induced norm) then for all $k \geq 1$[1]

$$\overline{\rho}(\Sigma) \leq \overline{\rho}_k(\Sigma)$$

and so

$$\overline{\rho}(\Sigma) = \lim_{k\to\infty} \overline{\rho}_k(\Sigma) \ .$$

Another natural generalization of the notion of spectral radius for a set of matrices is the *Generalized Spectral Radius* which is defined as

$$\rho(\Sigma) = \limsup_{k\to\infty} \rho_k(\Sigma)$$

where

$$\rho_k(\Sigma) = \max\{\rho(A_k A_{k-1} \cdots A_1)^{1/k} : A_i \in \Sigma\} \ .$$

It can be shown that $\rho_k(\Sigma) \leq \rho(\Sigma)$ for all k, [10] and that for any finite set of matrices Σ, $\rho(\Sigma) = \overline{\rho}(\Sigma)$ [2].

The inequalities $\rho_k(\Sigma) \leq \rho(\Sigma) = \overline{\rho}(\Sigma) \leq \overline{\rho}_k(\Sigma)$ can be used to approximate the Joint Spectral Radius to arbitrary precision. However the crucial problem of determining whether $\rho(\Sigma) \leq 1$, for the case of matrices with real or rational entries was shown to be undecidable by Blondel and Tsitsiklis [4] who reduced to it the empty word problem in Probabilistic Automata theory, which was previously known to be undecidable [5, 12]. The critical case is $\rho(\Sigma) = 1$ because no matter how close the approximations are, it is not possible to definitively conclude that $\rho(\Sigma) = 1$. Additionally, the computation of approximations is a hard computational problem as was demonstrated by Tsitsiklis and Blondel also [19, 20]:

Theorem 1 (Tsitsiklis and Blondel) *Unless $P = NP$, the Joint Spectral Radius $\rho(\{A, B\})$ of two $(0,1)$-matrices cannot be approximated by any algorithm that takes as input A, B and a relative error ϵ and returns a result within relative error ϵ and running in time polynomial in the size of $\Sigma = \{A, B\}$ and in the size of ϵ: $\log(1/\epsilon)$.*

4 The Joint Spectral Radius and State Sequence Growth

Let $M = (V, E, L, \Phi)$ be a weak model and let Z^t be a sequence of $t+1$ observations. Let $A = A_G$ be the adjacency matrix of the kinematic specification, $G = (V, E)$, of the state machine M, and let $A(Z^t) = I(\xi_0)\prod_{j=1}^{t} A(\xi_j)$ as in the example given previously. Then, as defined in the introduction, the number of possible consistent hypotheses is given by

$$\begin{aligned} h_M(Z^t) &= \sum_{i,j} e_i^T I(\xi_0) \cdot A(\xi_1) \cdots A(\xi_t) e_j \\ &= \underline{1}^T A(Z^t) \underline{1} \\ &= \|A(Z^t) \cdot \underline{1}\|_1 \ . \end{aligned}$$

[1]Note that the values $\overline{\rho}_k(\Sigma)$ in general depend on the norm while the limiting value does not.

Noting that the 1-norm is an induced norm, we have

$$\begin{aligned} h_M(Z^t) &= \|A(Z^t)\cdot \underline{1}\|_1 \\ &\le \|A(Z^t)\|_1 \cdot \|\underline{1}\|_1 \\ &\le \|A(Z^t)\|_1 \cdot n \\ &\le n\left(\|A(Z^t)\|_1^{1/t}\right)^t \\ &\le n(\overline{\rho}_t(\Sigma(\Phi)))^t \end{aligned}$$

where $\Sigma(\Phi) = \{A(\xi) \mid \xi \in \Phi\} \cup \{I(\xi) \mid \xi \in \Phi\}$ which we abbreviate to Σ for simplicity. Recalling the definition of the ∞-norm of a matrix:

$$\|A\|_\infty = \max\left\{\sum_{j=1}^{n} |a_{i,j}| \ : \ 1 \le i \le n\right\} ,$$

and the fact that in our case all matrix entries are nonnegative, we have

$$\begin{aligned} h_M(Z^t) &= \|A(Z^t)\cdot \underline{1}\|_1 \\ &\ge \|A(Z^t)\cdot \underline{1}\|_\infty \\ &= \|A(Z^t)\|_\infty \\ &\ge \rho(A(Z^t)) . \end{aligned}$$

Thus

$$(\rho(A(Z^t))^{1/t})^t \le h_M(Z^t) \le n(\overline{\rho}_t(\Sigma))^t$$

and, taking the max over Z^t, we get

$$\rho_t(\Sigma)^t \le h_M(t) \le n\overline{\rho}_t(\Sigma)^t . \tag{5}$$

The reader can verify that

$$\max_{Z^t} \rho(A(Z^t)) = \max_{A_j \in \Sigma(\Phi)} \rho(\prod_{j=0}^{t} A_j)$$

which is used in the above. Furthermore,

$$(\rho(A(Z^t))^{1/t})^t \le \rho_t(\Sigma)^t \le \rho(\Sigma)^t \le \overline{\rho}_t(\Sigma)^t .$$

For t very large $\overline{\rho}_t(\Sigma)$ approaches $\rho(\Sigma)$ from above and $\rho_t(\Sigma)$ approaches $\rho(\Sigma)$ from below, independently of the norm.

It is evident that the growth of the number of hypotheses, $h_M(t)$, depends on whether the Joint Spectral Radius of Σ is less than or greater than 1. When it is strictly smaller than 1, the number of hypotheses is bounded from above by a quantity decreasing exponentially in t and must therefore be exactly 0 since $h_M(t)$ is always a non-negative integer. When the Joint Spectral Radius is strictly larger than 1, $h_M(t)$ has a lower bound that grows exponentially in t to infinity.

The difficult case therefore is when the Joint Spectral Radius is exactly 1. If $\rho(\Sigma) = 1$, the inequalities in (5) do not provide enough information, because of the indeterminacy in the upper bound. Specifically, the upper bound in (5) has a factor of the form $w(t)^t$ where $w(t)$ is decreasing to 1 and t is increasing to infinity.

For example, the number of hypotheses may remain bounded. This is the case when there is a one-to-one relationship between states and observations. Then each matrix $A(\xi)$ has exactly one nonzero column. In such cases, $\rho(A(\xi)) = 1$, for all ξ and so $\rho(\Sigma) \le 1$ since the product of two matrices whose only nonzero entries are in one single column is again a matrix whose only nonzero entries are in one single column. Also, it is clear that $\overline{\rho}_t(\Sigma) = 1$ for all t. Accordingly, the inequalities in (5) effectively bound $h_M(t)$ by a constant.

We now develop an example where $h_M(t)$ is not bounded although $\rho_t(\Sigma) \to \rho(\Sigma) = 1$ in the limit. Let $M = (V, E, L, \Phi)$ be the weak model defined by the adjacency matrix A_G given below, where $G = (V, E)$, and the mapping is defined by $L : \{1, 2\} \to \{0, 1\}$, $L(i) = \{0, 1\}$, $i = 1, 2$:

$$A_G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} .$$

Then

$$A_G^t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

and so we can see immediately that $h_M(t) = \|A^t \cdot \underline{1}\|_1 = t + 1$ which means that the number of hypotheses grows polynomially. On the other hand it is also true that

$$\|A_G^t\|_1 = t + 1$$

and

$$\overline{\rho}_t(A_G) = \|A_G^t\|_1^{1/t} = (t+1)^{1/t} \to 1 \text{ as } t \to \infty \quad .$$

That is, the Joint Spectral Radius in this case is exactly 1.

4.1 Rate of growth of the trajectories: polynomial versus exponential

We saw in the previous section that the case $\rho(\Sigma) = 1$ is critical and can correspond to models with unbounded hypothesis growth. We start by showing that the rate of growth when $\rho(\Sigma) = 1$ is bounded by a polynomial function whose order is related to the number of states in the system.

Consider a weak model $M = (V, E, L, \Phi)$, $V = \{1, 2, \ldots, n\}$, and a sequence of $T+1$ observations $Z^T = (\xi_0, \xi_1, \ldots, \xi_T)$. One way to describe all possible $h_M(Z^T)$ trajectories in M that are consistent with Z^T is to use an auxiliary *trellis* structure. This structure is a $(T+1)$-partite graph defined by $\mathcal{T}(M, Z^T) = (\{X_0, X_1, \ldots, X_T\}, Y)$ where $X_i = \{(v, i) \mid v \in V\}$ and $((u, i), (v, j)) \in Y$ if and only if both $j = i + 1$ and $(A(\xi_j))_{u,v} = 1$ (see Figure 5). Essentially the nodes in V are replicated T times and the connections between the nodes at time t and those at time $t+1$ are determined by the observation that was made at time $t+1$, ξ_{t+1}. We can graphically represent [2] $\mathcal{T} = (\{X_i\}_{i \in [T+1]}, Y)$ by drawing nodes in X_0 along a vertical direction (columns) and then replicate the array T times to obtain a lattice network, not unlike those commonly used for Hidden Markov Models for instance . So, for example, node (u, i) will occur in column i at level u and will be labelled simply by u.

A path, π, in $\mathcal{T}$ is a function $\pi : [t_1, t_2] \to V$ such that

$$\forall t \in [t_1, t_2 - 1] \;\; ((\pi(t), t), (\pi(t+1), t+1)) \in Y \ ,$$

where $[t_1, t_2] := \{t_1, t_1 + 1, \ldots, t_2\}$. It is clear from this construction that

$$h_M(Z^T) = |\{\pi : [0, T] \to V \mid \pi \text{ represents a path in } \mathcal{T}(M, Z^T)\}\ |.$$

The trellis structure, $\mathcal{T}$, defined above possesses an interesting and useful property described in the following theorem.

Theorem 2 *Let $M = (V, E, L, \Phi)$ and let $\mathcal{T}(M, Z^T) = (\{X_i\}, Y)$ be a trellis structure as defined above. Assume that for each node $u \in V$ and for all time intervals, $[t_1, t_2]$ where $0 \le t_1 < t_2 \le T$, there exists at most one path $\pi : [t_1, t_2] \to V$ such that $\pi(t_1) = \pi(t_2) = u$. Then*

$$h_M(Z^T) = O(T^{2n} = |\{\pi : [0, T] \to V \mid \pi \textit{ represents a path in } \mathcal{T}(M, Z^T)\}|) \ .$$

[2] Here we define $[k] := \{0, 1, \ldots, k-1\}$.

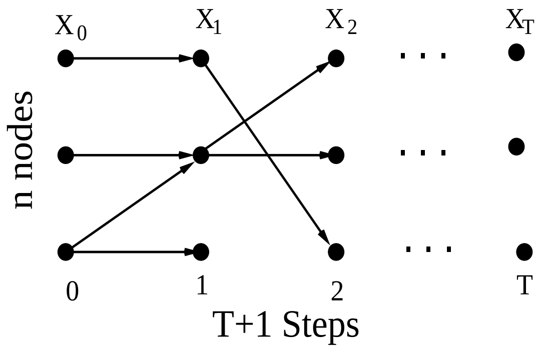

Figure 5: Trellis structure. The n nodes in X_i are drawn along a vertical direction.

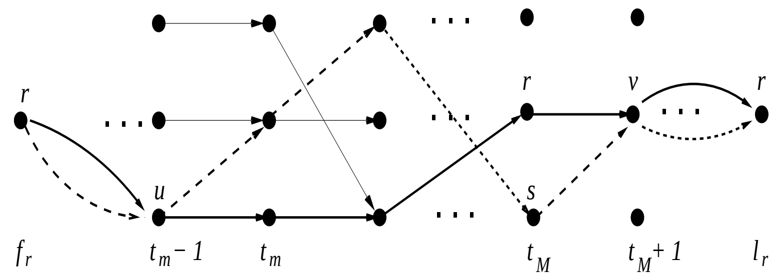

Figure 6: At time f_r (l_r) node r is visited by both π and π' for the first (last) time.

That is, $h_M(Z^T)$ is bounded by a polynomial in T of order $2n$ where n is the number of states (vertices) in the underlying kinematic model, $G = (V, E)$.

Proof. Under the assumptions of the theorem, we will show that for any path $\pi : [0, T] \to V$, $V = \{1, 2, \ldots, n\}$, there is a unique n-tuple of the form:

$$H(\pi) = ((f_1, l_1), (f_2, l_2), \ldots, (f_n, l_n)) \ ,$$

where f_i is the time at which node $i \in V$ is visited the first time in π and l_i is the time at which node i is visited the last time in π. We let $f_i = l_i = -1$ if node i is never visited and note that $f_i = l_i$ is entirely possible. By construction, $\pi(f_i) = \pi(l_i) = i$ and for all $t \in [0, T]$ if $t < f_i$ or $t > l_i$ then $\pi(t) \neq i$. We will also denote the u^{th} component of $H(\pi)$ with $(f_u(\pi), l_u(\pi))$ so that $f_u(\pi)$ is time when π first visits node u and $l_u(\pi)$ is the time when π last visits u with $f_u(\pi) = {}_u(\pi) = -1$ if π never visits u.

Now $H(\pi)$ is a hash function on paths through the trellis, $\mathcal{T}(M, Z^T) = (\{X_i\}, Y)$, with hash values in $([-1, T] \times [-1, T])^n = [-1, T]^{2n} = (T + 2)^{2n} = O(T^{2n})$. The proof consists of showing that the hash function H is actually a $1-1$ (injective) function: that is, there are no hash collisions.

To see this, suppose there are two different paths, $\pi, \pi' : [0, T] \to V$, that have the same hash values, $H(\pi) = H(\pi') = ((f_1, l_1), (f_2, l_2), \ldots, (f_n, l_n))$. Since π and π' are different, we can define t_m and t_M to be two times that delimit a maximal time interval on which they are different. Specifically, for all t such that $t_m \leq t \leq t_M$ we have $\pi(t) \neq \pi'(t)$ but $\pi(t_m - 1) = \pi'(t_m - 1)$ and $\pi(t_M + 1) = \pi'(t_M + 1)$.

Note that $0 < t_m \leq t_M < T$. In fact, at time $t = 0$ the two paths start from the same node since they both visit all nodes for the first time at the same times. Similarly, at time $t = T$ both paths visit the same node because both paths also visit all nodes for the last time at the same times.

By construction therefore, at time $t = t_m - 1$ and $t = t_M + 1$ the two paths must visit the same node, say u at time $t_m - 1$ and v at time $t = t_M + 1$.

Let r be the node visited by π at time t_M and s the node visited by π' at time t_M. Since $\pi \neq \pi'$ for $t \in [t_m, t_M]$ it must be $r \neq s$. This implies that neither r nor s are being visited for the first or last time by either of the two paths because $f_r(\pi) = f_r(\pi') \neq t_M$ and $l_r(\pi) = l_r(\pi') \neq t_M$.

Moreover it must be $f_r < t_m$ and $f_s < t_m$; that is, r and s were visited for the first time before time t_m since $f_r(\pi) = f_r(\pi')$ and so $\pi(f_r) = \pi'(f_r) = r$ while by assumption $\pi \neq \pi'$ on $[t_m, t_M]$. By the same argument it must be $l_r > t_M$ and $l_s > t_M$ where $\pi(l_r) = \pi'(l_r) = r$ and $\pi(l_s) = \pi'(l_s) = s$ (see Figure 6).

We now can conclude the proof by constructing two different paths between two equal nodes (that is, at the same level in the trellis graph) in different instants of time.

Let γ be the following path:

- $\forall t \in [0, f_r - 1]\ \gamma(t) := \pi(t)$
- $\forall t \in [f_r, t_m - 1]\ \gamma(t) := \pi(t)$
- $\forall t \in [t_m, t_M]\ \gamma(t) := \pi'(t)$
- $\forall t \in [t_M + 1, l_r]\ \gamma(t) := \pi(t)$
- $\forall t \in [t_r + 1, T]\ \gamma(t) := \pi(t)$

Clearly $\pi \neq \gamma$ on, at least, $[t_m, t_M]$ but both connect r at time f_r to itself at time l_r so there are two legal paths connecting a node to itself contradicting the assumptions of the theorem. Therefore, the hash function must be injective and so there must be no more than $(T+2)^{2n}$ total paths which are equivalent to consistent hypotheses.

□

An immediate consequence of the above result is

Corollary 3 *Let $M = (V, E, L, \Phi)$ and $\Sigma = \{A(\xi) \mid \xi \in \Phi\} \cup \{I(\xi) \mid \xi \in \Phi\}$ be the associated set of all the possible transition matrices derived from A_M and Φ. Suppose that any sequence of observations $Z^T \in \Phi^T$ induces a trellis structure $\mathcal{T}(M, Z^T) = (\{X_i\}_i, Y)$ that satisfies the conditions of theorem 2; that is, for each node $u \in V$ and for all instants of time $0 \leq t_1 < t_2 \leq T$ there exists at most one path $\pi : [t_1, t_2] \to V$ such that $\pi(t_1) = \pi(t_2)$. Then $\rho(\Sigma) \leq 1$.*

Proof. We saw in the previous section that if $\rho(\Sigma) > 1$ then the number of hypotheses must grow exponentially. Taking the contrapositive and remembering that in that case $h_M(T) = O(T^{2n})$ we can conclude that the Joint Spectral Radius cannot be greater than 1. □

It would be interesting to prove the converse also. In other words we would like to show that if $\rho(\Sigma) \leq 1$ then necessarily $h_M(T) = O(p(T))$ where $p(T)$ is a polynomial function in T.

We start by proving the converse of theorem 2.

Theorem 4 *Let $M = (V, E, L, \Phi)$ and Z^{T_0} a sequence of $T_0 + 1$ observations. Let $\mathcal{T}(M, Z^{T_0}) = (\{X_i\}_i, Y)$ be the induced trellis structure. Assume that there exists a node $u \in V$ and two instants of time $0 \leq t_1 < t_2 \leq T_0$ such that there exist at least two different paths $\pi_1, \pi_2 : [t_1, t_2] \to V$ satisfying $\pi_1(t_1) = \pi_1(t_2) = \pi_2(t_1) = \pi_2(t_2)$. Then $h_M(T) = \Omega(2^{cT})$ for some constant $c > 0$. Moreover it must be that $\rho(\Sigma) > 1$.*

Proof. The idea is to build a sequence of observations Z' by repeatedly concatenating the subsequence of observations $Z^{t_1:t_2} = (\xi_{t_1}, \xi_{t_1+1}, \ldots, \xi_{t_2})$. It should be clear that over every $t_2 - t_1 + 1$ instants of time the number of paths on the trellis structure at least doubles. So, asymptotically $h_M(T)$ is bounded from below by 2^{cT} for some $c > 0$ and infinitely many values of T.

Let $Z^{T_0} = (\xi_0, \xi_1, \ldots, \xi_{T_0})$ and $B = [b_{i,j}] = A(\xi_{t_1}) \cdot A(\xi_{t_1+1}) \cdot \ldots \cdot A(\xi_{t_2})$. Since there are at least two paths from u to u of length $t_2 - t_1 + 1$ in the segment of the trellis graph delimited by

t_1 and t_2 it must be that $b_{u,u} \geq 2$. Consider the following matrix $B' = [b'_{i,j}]$ defined by $b'_{u,u} = 2$ and $b'_{i,j} = 0$ if $i \neq u$ or $j \neq u$. Obviously $\rho(B') = 2$. Moreover B' is essentially a matrix that has nonnegative entries and is component-wise smaller than B and so $\rho(B') \leq \rho(B)$ since any increase in one of the entries does not decrease the spectral radius of a positive matrix [7]. Recalling the definitions and properties of the Generalized Spectral Radius we have

$$\rho(\Sigma) \geq \rho_{t_2-t_1+1}(\Sigma) \geq \rho(B)^{1/(t_2-t_1+1)} \geq \rho(B')^{1/(t_2-t_1+1)} \geq 2^{1/(t_2-t_1+1)} > 1 \ .$$

Letting $c = 1/(t_2 - t_1 + 1) > 0$ so that $2^c > 1$, note that the number of hypotheses can be made to double every $t_2 - t_1 + 1$ time steps so that over a period of T time steps, we can get at least $2^{T/(t_2-t_1+1)}$ hypotheses over time T as claimed. □

We can now state a complete characterization of the hypothesis growth rates in terms of the Joint Spectral Radius of Σ.

Corollary 5 *Let $M = (V, E, L, \Phi)$ and Σ be the set of matrices derived from A_M and Φ. Then*

$$\begin{array}{rcl} h_M(T) = p(T) & \text{if and only if} & \rho(\Sigma) \leq 1 \\ h_M(T) = \Omega(2^{cT}) & \text{if and only if} & \rho(\Sigma) > 1 \ , \end{array}$$

where $c > 0$ and p a polynomial function.

Proof. If $\rho(\Sigma) \leq 1$ then by Theorem 4 (specifically, its contrapositive) and Theorem 2 it must be $h_M(T) = p(T)$ for p a polynomial (of possibly degree 0). If, on the other hand, $\rho(\Sigma) > 1$ then by inequality 5 together with the properties of the limit (see also Corollary 3) it must be $h_M(T) = \Omega(2^T)$. □

Corollary 5 does not tell us what happens exactly when $\rho(\Sigma) = 1$ as we have already seen examples in which $h_M(T) = 1$ and $h_M(T)$ grows as a polynomial.

Lemma 6 *Let $Z^t = (\xi_0, \xi_1, \ldots, \xi_t)$ be a sequence of $t + 1$ observations,*

$$A(Z^{u:v}) \ = \ A(\xi_u) \cdot A(\xi_{u+1}) \cdots A(\xi_v)$$

the transition matrix associated with the subsequence $(\xi_u, \xi_{u+1}, \ldots, \xi_v)$ and $A(Z^t) = I(\xi_0) \cdot A(Z^{1:t})$. Assume that there exist i and j with $i \neq j$, such that

1. $e_i^T \cdot A(Z^t) \cdot e_j = c \geq 1$,
2. $e_i^T \cdot A(Z^t) \cdot e_i \geq 1$.
3. $e_j^T \cdot A(Z^t) \cdot e_j \geq 1$.

Then $f(t) := \max\{e_i^T A(Z^t) e_j \mid Z^t \in \Phi^t\}$ is unbounded and so is $h_M(t) \geq f(t)$.

Proof. Recall that $A(Z^t)$ is a matrix with nonnegative entries and so

- $h_M(Z^t) = \|A(Z^t) \cdot \underline{1}\|_1 = \underline{1}^T \cdot A(Z^t) \cdot \underline{1} = \sum_{i,j} e_i^T A(Z^t) e_j = \sum_{i,j} (A(Z^t))_{i,j}$.
- $e_i^T \cdot A(Z^t) \cdot e_j = \#\{\pi : [0, t] \to V \mid \pi \text{ path in } \mathcal{T}(M, Z^t),\ \pi(0) = i \text{ and } \pi(t) = j\}$.
- $e_i^T \cdot A(Z^t) \cdot \underline{1} = \#\{\pi : [0, t] \to V \mid \pi \text{ path in } \mathcal{T}(M, Z^t),\ \pi(0) = i \text{ and } \pi(t) \in V\}$.
- $\underline{1}^T \cdot A(Z^t) \cdot e_j = \#\{\pi : [0, t] \to V \mid \pi \text{ path in } \mathcal{T}(M, Z^t),\ \pi(0) \in V \text{ and } \pi(t) = j\}$.

Now, let $A_t = A(Z^{1:t})$. We will show by induction on k that $e_i^T I(\xi_0) A_t^k e_j \geq ck$, $e_i^T I(\xi_0) A_t^k e_i \geq 1$ and $e_j^T I(\xi_0) A_t^k e_j \geq 1$. And that will prove that $f(t)$ is not bounded.

- *Base:* For $k = 1$, the statement is true by the assumptions of the lemma.
- *Induction:* Assume that the statement is true for some $k \geq 1$. Recall that $I = \sum_{u=1}^{n} e_i e_i^T$ and using inequality 1 in the lemma's statement we can write:

$$
\begin{aligned}
e_i^T I(\xi_0) A_t^{k+1} e_j &= e_i^T I(\xi_0) A_t^k \cdot A_t e_j && (6)\\
&\geq e_i^T I(\xi_0) A_t^k \cdot I \cdot I(\xi_0) A_t e_j && (7)\\
&\geq e_i^T I(\xi_0) A_t^k \cdot \left(\sum_{u=1}^{n} e_u e_u^T \right) \cdot I(\xi_0) A_t e_j && (8)\\
&\geq e_i^T I(\xi_0) A_t^k \cdot (e_i e_i^T + e_j e_j^T) \cdot I(\xi_0) A_t e_j && (9)\\
&\geq e_i^T I(\xi_0) A_t^k e_i \cdot e_i^T I(\xi_0) A_t e_j + e_i^T I(\xi_0) A_t^k e_j \cdot e_j^T I(\xi_0) A_t e_j && (10)\\
&\geq e_i^T I(\xi_0) A_t e_j + e_i^T I(\xi_0) A_t^k e_j && (11)\\
&\geq c + ck && (12)\\
&\geq c(k+1) && (13)
\end{aligned}
$$

The induction hypothesis has been applied in going from (10) to (11) and from (11) to (12). With similar reasoning we can verify the other two inequalities for $k + 1$.

□

A model that satisfies the conditions of the above lemma is specified by $\Sigma = \{A\}$, $i = 1, j = 2$ and

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} .$$

Notice that $\rho(\Sigma) = 1$.

On the other hand, if either condition 2 or 3 of Lemma 6 is false for all t, then the number of trajectories will remain bounded as shown by the model Σ represented by the following matrix:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} .$$

In this case $\rho(\Sigma) = \rho(A) = 1$, $h_M(t) = 4$ and condition 3 is not satisfied.

We have seen so far that the growth of $h_M(t)$ depends on the Joint Spectral Radius of a finite set of $(0,1)$ matrices Σ. In particular, $h_M(t) = O(p(t))$ if and only if $\rho(\Sigma) \leq 1$. In the following we show that, unlike the general case for real or rational matrices, the problem of deciding whether the Joint Spectral Radius of a finite set of $(0,1)$ matrices is less than or equal to 1 is decidable, meaning there is an effective (although perhaps inefficient) algorithm for answering the question. Consequently we can decide, given a weak model, whether the number of hypotheses grows polynomially or exponentially in the worst-case.

Theorem 7 *Let Σ be a finite set of $(0,1)$ matrices of order n. Then $\rho(\Sigma) > 1$ if and only if $\exists w \exists k, 1 \leq k \leq 2^n + 1$ such that*

$$e_w^T A_{j_1} A_{j_2} \cdots A_{j_k} e_w \geq 2 ,$$

for $A_{j_i} \in \Sigma$ for all i.

Proof. The if-part is immediate since, as in Theorem 4, we can say that

$$\rho(\Sigma) \geq \rho(A_{j_1} A_{j_2} \cdots A_{j_k})^{1/k} \geq 2^{1/k} > 1 .$$

We need to prove the only-if part. Assume that $\rho(\Sigma) > 1$. Then there must exist an integer v and indices $j_1, j_2, \ldots, j_T$ such that $e_v^T A_{j_1} A_{j_2} \cdots A_{j_T} e_v \geq 2$. If $T \leq 2^n + 1$ we are done. So, assume that $T > 2^n + 1$.

Let $\mathcal{T}[j_1, j_2, \ldots, j_T]$ be the trellis graph determined by the transition matrices $A_{j_1}, A_{j_2}, \ldots, A_{j_T}$ in the obvious way: $\mathcal{T}[j_1, j_2, \ldots, j_T] := (\{X_i\}_{i=0,1,\ldots,T}, Y)$, where for all i, $X_i = \{1, 2, \ldots, n\}$ and for all $i = 1, \ldots, T$, $((u, i-1), (v, i)) \in Y$ if and only if $(A_{j_i})_{u,v} = 1$. Then the condition $e_v^T A_{j_1} A_{j_2} \cdots A_{j_T} e_v \geq 2$ implies that there are at least two different paths in the trellis that join node v at time 0 to itself at time T.

Let $\Pi = \{\pi : [0, T] \to \{1, 2, \ldots, n\} \mid \pi(0) = v, \pi(T) = v\}$ be the set of all such paths and let $V_t \subseteq V$ be set of nodes that are touched by those paths at time t: $V_t = \{\pi(t) \mid \pi \in \Pi\}$.

Since $V_t \in 2^V - \{\emptyset\}$, $|2^V - \{\emptyset\}| = 2^n - 1$ and $T > 2^n + 1$ then by the Pigeonhole principle there must exist times t, s, $0 < t < s < T$, such that $V_t = V_s$.

Consider the two new trellises $\mathcal{T}_1 = \mathcal{T}[j_1, j_2, \ldots, j_t, j_{s+1}, \ldots, T]$ and $\mathcal{T}_2 = \mathcal{T}[j_{t+1}, j_{t+2}, \ldots, j_s]$, obtained by *cutting off* the portion delimited by the instants of time t and s ($\mathcal{T}_2$) and then *pasting* the external pieces together ($\mathcal{T}_1$).

Then only the following two cases may arise:

1. $|V_t| = 1$ so that $V_t = V_s = \{u\}$ for some u. In this case, paths in Π restricted to $\mathcal{T}_1$ may be identified. However if this happens to all the paths in Π then it must be that there exist at least two paths in Π that are different on $[t, s]$ and so are their restrictions to $\mathcal{T}_2$.

2. $|V_t| > 1$ so that there are at least two paths in Π that are different at time t and so are their restrictions to $\mathcal{T}_1$.

The idea is now to iterate this cutting process until we obtain a sufficiently small trellis with, at least, two different paths connecting the same node at the same instants of time. If the first case occurs then we check whether $e_v^T A_{j_1} A_{j_2} \cdots A_{j_t} e_u \geq 2$ or $e_u^T A_{j_{s+1}} A_{j_{s+2}} \cdots A_{j_T} e_v \geq 2$. If so it means that there are at least two paths in Π that are different either on $[0, t]$ or on $[s, T]$ and so we take $\mathcal{T}_1$ otherwise we take $\mathcal{T}_2$ with the assurance that there are at least two different paths joining node u to itself. In the second case we always take $\mathcal{T}_1$. At each cutting operation we obtain a trellis that is defined by a strictly shorter product of matrices and on which we can identify at least two different paths joining the same node at extreme instants of time (beginning and end of the resulting trellis). So after a finite number of cuts we must obtain a product of at most $2^n + 1$ matrices that verify the assertion.

To decide the problem it will be enough to consider $f(n) = \sum_{k=1}^{2^n+1} |\Sigma|^k$ alignments of matrices from the set Σ, compute their products and check whether the elements in the main diagonal of the results are all less than or equal to 1. □

5 Summary and Future Work

This paper has introduced a formal model, called a weak model, for tracking dynamic phenomena by observing a sequence of noisy observations. Weak models use nondeterministic automata as the modelling formalism. The worst-case growth of the number of the automaton's state sequences that are consistent with a given observation sequence has been shown to be equivalent to the problem of determining the Joint Spectral Radius of a set of 0-1 matrices. We have shown that this worst-case growth rate is either polynomial or exponential and that transition can happen abruptly, illustrating a kind of *phase transition* in the complexity of the modelling and tracking problem. We have shown that the decision problem $\rho(\Sigma) \leq 1$ for Σ a finite set of $(0, 1)$-matrices is in fact decidable in a Turing equivalent model of computation, unlike the general problem for sets of rational matrices which is known to be undecidable.

The equivalence between deciding whether worst-case hypothesis growth in a given *weak* model is polynomial or exponential and the Joint Spectral Radius decision problem suggests some interesting related problems for future work.

For example, *generalized* Nyquist sampling problems can be formulated in terms of finding a least-cost sensor coverage which guarantees hypothesis growth that is: a. bounded by 1; b. bounded by a constant; or c. bounded by a polynomial. The cost function can be articulated in terms of the combined sensor coverage in various ways. Similarly, we can seek to find a least-cost weak model variant, given an initial weak model, that achieves a desired hypothesis growth rate where cost is measured by the modelling effort required to remove possible transitions in the underlying weak model dynamics.

Another interesting direction for future research is the question of state estimation as it relates to state sequence estimation, which has been the focus of this paper. The various examples presented in this paper indicate that the number of state sequences can grow quickly while the number of actual possible states at any given time can be small or even unique. One possible formulation of the state estimation problem can be expressed in terms of efficiently computable and interesting *properties* of the hypothesis set. Such a property could be, for example, whether any of the hypothesized state sequences include a transition through specific states of the underlying automaton. Preliminary work in this direction has been done in [17].

Acknowledgements: This work was supported in part by ARDA under Grant F30602-03-C-0248, DARPA Projects F30602-00-2-0585 and F30602-98-2-0107, and the National Institute of Justice, Department of Justice Award number 2000-DT-CX-K001.

References

[1] V. K. Balakrishnan. *Combinatorics, including Concepts of Graph Theory.* McGraw-Hill, New York, 1995.

[2] M. Berger and Y. Wang. Bounded semigroups of matrices. *Linear Algebra and its Applications*, 166:21–27, 1992.

[3] V. H. Berk and N. Fox. Process query systems for network security monitoring. In *Proceedings of the SPIE Vol. 5403, Defense and Security Symposium, Orlando, FL*, March/April 2005.

[4] V. Blondel and J. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems & Control Letters*, 41(2), 2000.

[5] A. Condon and R.J. Lipton. On the complexity of space bounded interactive proofs. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 462–467, 1989.

[6] V. Crespi, W. Chung, and A. B. Jordan. Decentralized sensing and tracking for UAV scheduling. In *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III Orlando, FL*, April 2004.

[7] Dragos Cvetkovic. *Spectra of Graphs: Theory and Application.* Academic Press, 1980.

[8] G. Cybenko, R. S. Gray V. H. Berk, and G. Jiang. An overview of process query systems. In *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III, Orlando FL*, April 2004.

[9] I. Daubechies and J.C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra and its Applications*, 162:227–263, 1992.

[10] J.C. Lagarias and Y. Wang. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra and its Applications*, 214:17–42, 1995.

[11] S. Oh, L. Schenato, P. Chen, and S. Sastry. A scalable real-time multiple-target tracking algorithm for sensor networks. University of California, Berkeley, Technical Report UCB//ERL M05/9, Feb. 2005.

[12] A. Paz. *Introduction to Probabilistic Automata.* Academic New York, 1971.

[13] L.R. Rabiner. A tutorial on hidden Markov models. *Proceedings of the IEEE*, 77:257–286, 1989.

[14] D.B. Reid. An algorithm for tracking multiple targets. *IEEE Trans. on Automatic Control*, 24(6):843–854, Dec. 1979.

[15] C. Roblee, V.H. Berk, and G. Cybenko. Large-scale autonomic server monitoring using process query systems. In *IEEE International Conference on Autonomic Computing, Seattle, WA*, June 2005.

[16] G.-C. Rota and G. Strang. A note on the joint spectral radius. *Nederl. Akad. Wetensch. Indagationes Math.*, 22:379–381, 1960.

[17] Y. Sheng and G. Cybenko. Computation of regular property probabilities in HMM state sequences (extended abstract). Submitted to the Symposium on Discrete Algorithms 2006, July 2005.

[18] Y. Sheng and G. Cybenko. Distance measures for nonparameteric weak process models. To appear in Proceedings of IEEE Systems, Man and Cybernetics Conference, Hawaii, October 2005.

[19] J.N. Tsitsiklis and V.D. Blondel. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate. *Mathematics of Control, Signals, and Systems*, 10:31–40, 1997. Correction in 10, 381.

[20] J.N. Tsitsiklis and V.D. Blondel. A survey of computational complexity results in systems and control. *Automatica*, 36:1249–1274, 2000.

Appendix

The results of this paper demonstrate that the number of hypotheses for an observation sequence is a delicate function of the model dynamics and state-to-observation mapping. In this appendix, we develop a number of simple tracking problems that illustrate the effect of dynamics and sensor coverage on hypothesis growth.

The different examples are all instances of a general scenario. A ground vehicle moves on a two-dimensional grid in both vertical and horizontal directions. The grid is abstracted as a matrix of cells (x, y) that are traversed by the moving vehicle.

The position of the vehicle at any instant of time is specified by the coordinates of the cell that it is traversing. A number of binary sensors are deployed in the grid region. Each sensor detects the presence of the vehicle in a cell or in some aggregation of cells. When the vehicle traverses a monitored cell any sensor that has coverage in that cell emits the coordinates of all the cells that it covers and those are collected as a single observation. If sensors overlap then traversal of a cell will trigger all the sensors that cover that cell. The effective observation therefore consists of the coordinates of the cells that belong to the intersection of all sets of cells corresponding to the coverage of the triggered sensors.

We consider a number of specific instances of this general scenario. The instances differ according to the vehicles' kinematics and the coverage of the sensors: where the sensors are placed and the sensors' ranges (the number and geometry of the cells covered by a sensor). For each of these instances, we explain the trackability properties, namely the state sequence growth, of the problem of determining the vehicle's position given a temporal sequence of sensor observations. A *binary* sensor means that the sensor either detects the presence of a vehicle or not. If a sensor does detect a vehicle, it reports the coordinates of the cells that it covers.

Example 1

- **State:** Occupied cell (x, y).
- **Kinematics:** At each time step the vehicle is allowed to move in any of the four possible vertical or horizontal directions (up, down, left and right). Formally: $(x, y) \to (x \pm 1, y)$ or $(x, y) \to (x, y \pm 1)$. At the border cells of the region, the vehicle can only move to cells within the region. The vehicle cannot stand still and stay in the same cell for more than one time step.
- **Sensor placement and coverage:** Sensors are placed in cells (x, y) with x even and y odd (see Figure 7, where coordinates $(1, 1)$ denote the upper left cell, x is the column index and y the row index). Sensors cover only the cell in which they are placed. Each sensor observation is simply a pair, (x, y), of cell coordinates. When the vehicle is in a cell not covered by a sensor, the observation is called a *null* observation below to distinguish it from *sensor* observations which come from sensors in covered cells.

The following assertions about this tracking instance are true:

1. A sensor observation determines the state exactly. After a sensor observation, there are four possible states.
2. Let (x, y) be a monitored cell (that is, containing a sensor). There are two ways to move from (x, y) to $(x + 1, y + 1)$ (see Fig. 7):

$$(x, y) \to (x + 1, y) \to (x + 1, y + 1)$$
$$(x, y) \to (x, y + 1) \to (x + 1, y + 1)$$

and so there are four 4-step trajectories from (x, y) to itself consistent with the following sensor observation sequence: $(x, y), (x + 1, y + 1), (x, y)$. (Note that because there is a null

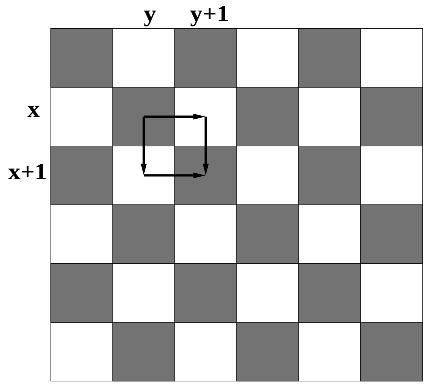

Figure 7: The black cells are monitored by binary sensors.

observation between any pair of sensor observations, this sequence of 3 sensor observations is technically a sequence of 5 observations.) According to Theorem 4 the worst-case hypothesis growth is exponential in the number of observations. Even for certain very short sequences of such observations, the growth is exponential.

Notice that, in this case, the initial state (cell) of the vehicle does not in general affect the complexity of the state estimation problem. However there are certain sequences of observations that are uniquely associated with a single state sequence. This is the case for example, when the vehicle starts in a monitored cell and moves along a straight line in one of the allowed directions. Although the state of the vehicle is ambiguous after a sensor observation, the ambiguity disappears if the next sensor observation is either two cells away vertically or horizontally from the last observation (because the vehicle can move only one cell at a time).

One conclusion of this analysis is that the kinematics in this instance are under-constrained for a vehicle to be trackable (according to our definition of trackable) given the assumed sensors and their coverage. While the number of state sequences can grow exponentially in time, the error in the current state estimate is always bounded.

Example 2

- **State:** Occupied cell (x, y) plus "momentum".
- **Kinematics:** The vehicle is allowed to change direction only after taking two steps in the same direction (up, down, left, right). In other words, at any instant of time the state of the vehicle is described by the occupied cell together with its *momentum*, namely whether the vehicle needs to move up (u_1), down (d_1), left (l_1), right (r_1) on this step or is free to change direction on this step (one of u_2, d_2, l_2, r_2). Formally, the state of the finite state machine that models these kinematics is a triple (x, y, p) where $p \in \{u_1, d_1, r_1, l_1, u_2, d_2, r_2, l_2\}$ and the transition function is defined by:

$$
\begin{array}{rcl}
(x, y, l_1) & \rightarrow & (x, y-1, l_2) \\
(x, y, r_1) & \rightarrow & (x, y+1, r_2) \\
(x, y, d_1) & \rightarrow & (x+1, y, d_2) \\
(x, y, u_1) & \rightarrow & (x-1, y, u_2) \\
(x, y, *_2) & \rightarrow & \{(x, y-1, l_1), (x, y+1, r_1), (x-1, y, u_1), (x+1, y, d_1)\} \\
\vdots & &
\end{array}
$$

Here $*$ is a wild-card which can be one of l, r, u, d. The meaning of state $(2, 2, d_1)$, for example, is that the vehicle is in cell $(2, 2)$ and is required to move down one cell on the next step. Accordingly, the next cell is uniquely determined, namely $(3, 2)$. It is in state $(3, 2, d_2)$ and can next move in any direction on the next step. Several successor states are possible because after two consecutive moves in the same direction the vehicle is allowed to choose one of the possible four directions (apart from the boundary conditions). The above description of the transition function must be completed in the obvious way to account for states along the boundaries (that is, the vehicle cannot move up or left from cell $(1, 1)$).

- **Sensor placement and coverage:** The sensors are placed as before (see Fig. 7) and the observations are again simple pairs of cell coordinates. There are *null* observations, as before, when the vehicle is in an unmonitored cell.

The following assertions about this tracking instance can be made. (Their verification is left to the reader.)

1. Every other observation does not necessarily determine the state uniquely any more because the momentum is not directly observable or uniquely inferable in many cases. However, if the initial state of the vehicle is (x, y, p) where (x, y) are the (interior) cell coordinates of a monitored cell then (x, y) will be the sensor observation and there will be no more than eight states (boundary cells will have fewer than eight) that are consistent with that initial observation.

2. If the next sensor observation is either vertically or horizontally related to (x, y), namely $(x \pm 2, y)$ or $(x, y \pm 2)$, there are still 8 possible current states since we do not know a priori whether the current state has four of the possible $*_2$ or four of the possible $*_1$ momentum coordinates. Of the 8 initial possible states and 8 possible current states there are only 5 hypotheses or state sequences that include those 8 possible initial states and 8 possible current states consistently.

3. On the other hand, if the next sensor observation is diagonally related to (x, y), namely one of $(x-1, y-1)$, $(x-1, y+1)$, $(x+1, y-1)$ or $(x+1, y+1)$, then we know that the direction must have changed in one of two possible adjacent, but unmonitored, cells. See Figure 8. This leads to two possible state sequences for any diagonal movement. Further movement along the same diagonal must be consistent with those two possibilities and so there are no more hypotheses generated through continued diagonal movement of the vehicle.

4. The reader can verify that by combining any nontrivial horizontal or vertical movements with diagonal observations such as outlined above there will never be more than four possible state sequences for any observation sequence that begins in a monitored cell.

5. If the initial state of the vehicle is (x, y, p) where (x, y) is *not* a monitored cell, then there are about $4N^2$ states consistent with this initial non-observation, where the grid has $N \times N$ cells. (We know it must be in one of the $N^2/2$ cells not being monitored and the momentum can be any of 8 possible values for interior cells and fewer than 8 for boundary cells.)

6. There can never be two different state sequences with the same observation sequence that start and end in the same state of the underlying state machine. Therefore, this example cannot have exponential hypothesis growth.

7. The worst-case hypothesis growth in this example is bounded by $4N^2$ on the first time step and subsequently by 20 on later time steps. Moreover there are possible observation sequences where the number of hypotheses can oscillate over time (when the first cell is not monitored and the vehicle moves in a straight line) and other possible observation sequences where the number of hypotheses is identically 1 (the movement (x, y) to $(x+2, y)$ to $(x+2, y+2)$ over 4 time steps is accountable by a unique state sequence in this model).

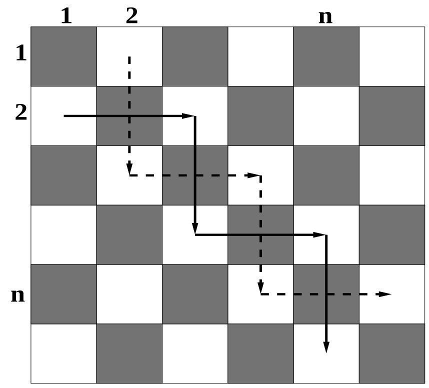

Figure 8: The only two paths consistent with observations $X, (2, 2), X, (3, 3), X, \ldots, (n, n), X$. Character X stands for "no-observation". When the vehicle deviates from this diagonal pattern, the ambiguity disappears completely.

The conclusion of this analysis is that by trivially constraining the kinematics of the vehicle relative to Example 1, the problem becomes trackable according to our definition of trackability. In particular the number of paths consistent with any given sequence of observations greater than 1 is less than or equal to a constant, namely 20, at all times and can in fact oscillate. The number of possible states if the initial observation is null is about $4N^2$.

Example 3

- **State:** As in example 1.
- **Kinematics:** At each step the vehicle is allowed to make only one of the following moves: up, right or stay in its current cell. The northern boundary prevents further upward moves and the right boundary prevents rightward moves. When the upper right corner is reached the vehicle stays in that cell for all subsequent times.
- **Sensors:** As in example 1.

We can state the following facts:

1. There is only one trajectory from any state to itself at different instants of time on the trellis graph. So, asymptotically, the number of hypotheses grows at most polynomially.
2. Curiously if we study the problem over the finite time horizon, $t \in [0, T]$, for T of the same order as the number of cells in the grid, then we can find sequences of observations that lead to an exponential growth in the number of hypotheses as long as $t \leq T$. That happens for example when the vehicle moves up along a diagonal so we cannot distinguish between two possible unobserved states between two sensor observations along the diagonal. This means there will be 4 possible state sequences for 3 consecutive sensor observations (that have two interleaved null observations), 8 state sequences for 4 consecutive sensor observations and so on. When the vehicle reaches the boundary of the region, this growth ends and the number of possible hypotheses then remains constant.

This example shows how the asymptotic behavior (in this case bounded) may be very different from transitional behavior (exponential). Unlike Example 1, the worst-case number of hypotheses grows exponentially initially but is asymptotically bounded by a constant when the length of the sequence of observations is sufficiently large.

Example 4

- **State:** As in the previous example, Example 3.
- **Kinematics:** As in the previous example, Example 3.
- **Sensors:** There is a sensor in each row of the grid and it can only report its row number which is equivalent to the distance of the vehicle from the northern boundary. Accordingly, a sensor observation consists of the first of the two coordinates of the cell occupied by the vehicle. This is clearly modelled in our framework by the mapping, $L((x, y)) = x$ where x is the sensor report of the sensor in row x.

We can make the following assertions about this example:

1. Because the coordinates of the vehicle cannot decrease, it is impossible to have more than one trajectory on the trellis graph that joins the same state at different instants of time. Accordingly, the asymptotic growth of the hypotheses cannot be exponential.
2. We can say more about the finite time behavior of the number of hypotheses. Assume that the coordinates of the initial cell position of the vehicle are (a, b) and that there are $n - 1$ cells separating this position from the right border, i.e., $(a, b + n - 1)$ is a cell on the right border. Consider the following sequence of $T + 1$ observations: $\zeta_t = a$, for $0 \leq t \leq T$. Thus, according to the sensor observations the vehicle is moving horizontally but we do not know how many times it stays in the same cell along this route. We can count the number of possible paths in the following way. Let m_i be a variable that denotes the number of steps the vehicle spends in cell $(a, b + i - 1)$. There is a one-to-one correspondence between the number of possible paths consistent with the observations, $h(T)$, and integer solutions to

$$\begin{cases} m_1 + m_2 + \ldots + m_n & = & T + 1 & \\ m_i & \geq & 1 & \text{for all } i \ . \end{cases}$$

   The number of solutions, $h(T)$, can be determined using standard methods from combinatorics [1]:

$$h(T) = \binom{T}{n-1} = \Theta(T^{n-1}) \ .$$

   We can also see that any other possible sequence of $T + 1$ constant sensor observations will give rise exactly to the same number of hypotheses.

This is an example of polynomial growth but this growth is over all time since the vehicle can stay in the same horizontal line indefinitely.