The Theory of Trackability with Applications to Sensor Networks

VALENTINO CRESPI
California State University at Los Angeles
GEORGE CYBENKO
Dartmouth College
and
GUOFEI JIANG
NEC Laboratories America

In this article, we formalize the concept of tracking in a sensor network and develop a quantitative theory of *trackability* of *weak models* that investigates the rate of growth of the number of consistent tracks given a temporal sequence of observations made by the sensor network. The phenomenon being tracked is modelled by a nondeterministic finite automaton (a weak model) and the sensor network is modelled by an observer capable of detecting events related, typically ambiguously, to the states of the underlying automaton. Formally, an input string of symbols (the sensor network observations) that is presented to a nondeterministic finite automaton, M, (the weak model) determines a set of state sequences (the tracks or hypotheses) that are capable of generating the input string. We study the growth of the size of this candidate set of tracks as a function of the length of the input string. One key result is that for a given automaton and sensor coverage, the worst-case rate of growth is either polynomial or exponential in the number of observations, indicating a kind of phase transition in tracking accuracy. These results have applications to various tracking problems of recent interest involving tracking phenomena using noisy observations of hidden states

16

This work results from research programs at the Institute of Security Technology Studies at Dartmouth College, the Department of Computer Science at the California State University Los Angeles and the NEC Laboratories America, supported by the U.S. Department of Homeland Security Grant Award Number 2006-CS-001-000001, ARDA Grant Award Number F30602-03-C-0248, AFOSR Grant FA9550-07-1-0421, DARPA Project HR001-06-1-0033, and the National Institute of Justice, Department of Justice Award 2000-DT-CX-KOO1. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or other funding agencies.
Authors' addresses: V. Crespi, Department of Computer Science, California State University Los Angeles, ECST 5151, State University Drive, Los Angeles, CA 90032; email: vcrespi@calstatela.edu; G. Cybenko, Thayer School of Engineering, Dartmouth College Hanover, NH 03755; email: gvc@dartmouth.edu. G. Jiang, NEC Laboratories America, 4 Independence Way, Princeton, NJ 08540; email: gfj@nec-labs.com.

such as: sensor networks, computer network security, autonomic computing and dynamic social network analysis.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

General Terms: Performance

Additional Key Words and Phrases: Sensor networks, tracking, multiple hypotheses

ACM Reference Format:

Crespi, V., Cybenko, G., and Jiang, G. 2008. The theory of trackability with applications to sensor networks. ACM Trans. Sens. Netw. 4, 3, Article 16 (May 2008), 42 pages. DOI = 10.1145/1362542.1362547 http://doi.acm.org/10.1145/1362542.1362547

1. INTRODUCTION

This article 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 object tracking using a sensor network, computer and networks security using various network- and host-based sensors, autonomic computing systems and dynamic social network analysis. While it is possible to develop statistical models for such applications, the present work addresses so-called weak models that are nondeterministic but not probabilistic. That is, we 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 [Reid 1979a], is currently being investigated by other groups [Oh et al. 2005]. This article 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.

In this section, we develop a formal model for tracking a nondeterministic finite automaton and 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 article, 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.

In Section 2 we develop several simple examples to illustrate the main concepts and constructs as well as the 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 the main results there. Section 5 discussed specific applications and issues related to sensor networks. Section 6 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. A Glossary of Symbols is provided at the end of this article to help readers navigate notation.

1.1 Problem Formulation

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 $L : V \rightarrow 2^{\Phi}$ 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) [Rabiner 1989]. 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 [Cybenko et al. 2004].

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 both: $L(v) \subseteq L'(v)$ 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 restrictive 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

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

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

$$\mathcal{H}_W(Z^T) = \{v_0 v_1 \ldots v_T \mid v_0 v_1 \ldots v_{T-1} \in \mathcal{H}_W(Z^{T-1}),\ v_T \in \mathcal{H}_W(\xi_T) \text{ and } (v_{T-1}, v_T) \in E\}.$$

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 article is that $h_W(T) = \max_{Z^T} h_W(Z^T) = \max_{Z^T} |\mathcal{H}_W(Z^T)|$, the maximum cardinality of the set $\mathcal{H}_W(Z^T)$, grows either polynomially or exponentially in T and this property can be efficiently decided. 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 $W \leq W^{(1)} \leq W^{(2)} \leq \cdots \leq W^{(k)}$, 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 we have defined it is equivalent to a nondeterministic finite automaton (NDFA) although the number of states and state transitions in an NDFA that is equivalent to a weak model may be larger than the number of states and state transition in the original weak model [Sheng 2006]. 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 [Sheng 2006; Sheng and Cybenko 2005b].

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 article is a proof 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 in polynomial time.

1.2 Technical Approach and Related Work

We study the problem of state sequence growth in weak models by formulating it in terms of matrix norms of sequences of 0-1 matrices. The relevant mathematical machinery we invoke involves the *Joint Spectral Radius* of a set of matrices. The Joint Spectral Radius, $\rho(\Sigma)$, is a quantity introduced by Rota and Strang [1960] 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 $x_{t+1} = A_t x_t$ [Daubechies and Lagarias 1992; Blondel and Tsitsiklis 2000; Tsitsiklis and Blondel 2000].

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. (By $(0, 1)$-matrices, we mean matrices whose entries are either 0 or 1.) 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 the following results:

—For 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, the worst-case number of hypotheses in a weak model 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;

—Unlike the general case for rational matrices, the decision problem $\rho(\Sigma) \leq 1$ for $(0, 1)$-matrices is efficiently solvable. Specifically, given a weak model, we can decide in polynomial time whether there exists a family of sequences of observations such that the number of tracks consistent with those observations grows exponentially or whether for all families of sequences of observations the growth is at most polynomial.

These results have been previously reported by the authors [Crespi et al. 2005; Crespi et al. 2006]. The result concerning the nonexistence of intermediate rates of growth was recently proved independently by Bell [2005] on the more general domain of semigroups of complex matrices, using sophisticated algebraic techniques. Our results, based on simple graph-theoretic and combinatorial arguments, though applicable only to the restricted domain of $(0, 1)$-matrices, are stronger in that they lead to efficient algorithms to decide the questions of whether the growth is polynomial or exponential. Subsequent recent work by Blondel, Jungers, and Protassov [Jungers et al. 2006] has described an efficient algorithm for determining the actual rate of the polynomial growth, not merely the fact that the growth is polynomial.

As mentioned earlier, a weak model can be converted to a nondeterministic finite state automaton [Sheng 2006; Sheng and Cybenko 2005b]. Observation sequences of weak models are accordingly analogous to symbol sequences corresponding to state transitions in automata, which are commonly called *words* in automata theory. Furthermore, the set of words generated or accepted by an automaton is called a *language*.

A degenerate case of a weak model arises when a single observable can be associated with all states, so that there is effectively no information gained from the observations. In this degenerate case, the number of possible state sequences or hypotheses that could have generated a given observation sequence of length T is clearly the number of all possible state sequences of length T. The growth rate of state sequences of such degenerate weak models has been studied in symbolic dynamics, and it is known that the logarithm of the largest eigenvalue (equivalently the spectral radius), ρ, of an automaton's state transition matrix is the topological entropy of the set of possible state sequences of an automaton [Lind and Marcus 1995; Williams 2004]. More specifically, the entropy is

$$\lim_{k\to\infty} \frac{1}{k} \log\{\textit{number of state sequences of length } k\} = \log \rho,$$

so it is straightforward to see that the number of possible state sequences in this degenerate case is exponential if and only if $\rho > 1$, or equivalently if and only if the topological entropy is positive.

Another application in which this degenerate weak model occurs is in coding theory, where code words are generated by a finite automaton's state sequences [Lind and Marcus 1995; Marcus 1995; Williams 2004]. In that setting, the capacity of a code is the topological entropy of the automaton generating the code (a sofic shift specifically) and can be derived from the spectral radius of the automaton's state transition matrix in the way mentioned as well [Lind and Marcus 1995; Williams 2004].

The major difference between the present work and the areas we have discussed is that traditional automata theory, coding theory and symbolic dynamics postulate time-invariant state transition dynamics whereas the trackability problem we have defined involves a special kind of time-varying dynamics since the state transitions at any time are constrained by the observations made at that time but the underlying dynamics themselves do not depend on time. This is clearly not equivalent to the general case where the state transitions are themselves time-dependent in an arbitrary way, but the main results of the article do hold if the possible state transitions are drawn from a set of state transition matrices and any transition matrix is possible at any time. The reader should be able to verify that the main results do still hold in that case, but the added complexity of notation and derivation for that case was not considered worth the added partial generality.

Accordingly, the results of this article can be viewed as nontrivial extensions of well-known results in symbolic dynamics, automata theory, and coding theory to certain types of time-varying, nonstationary discrete dynamical systems. Additionally, there has been much research on the theory of ω-regular languages and Buchi automata on infinite words [Buchi 1960; Mukund 1996]. That work is primarily focused on subset construction and how to decide whether an infinite word is accepted by automata. By contrast, in this article we analyze the *number* of state transition sequences that could accept a word of finite length. We are especially interested in the worst case *growth rate* of this number.

We are not aware of similar work in automata theory with finite or infinite inputs and, in general, believe that the trackability problem as formulated here is primarily of interest to generalized sensor network applications, not automata or formal methods research community.

Extensions of these results to probabilistic models is forthcoming. For example, consider the correspondence between a weak model, W, as defined in this article and the class of Hidden Markov Models, say $\mathcal{M}$, which have nonzero transition and emission probabilities corresponding to precisely the transitions (edges) and state-to-observation relations in the weak model. We believe that h_W has polynomial growth if and only if the Shannon entropy, $H(M) = 0$, for every HMM $M \in \mathcal{M}$.

There exists much work on the design of effective tracking algorithms for sensor networks. Sensor network tracking using Multiple Hypothesis Tracking ideas originally developed by the radar tracking community has has been investigated by Oh, Sastry and others [Oh et al. 2005]. Zhao et al. [2002] have proposed an information-driven approach to determine sensor collaboration schemes in ad-hoc sensor networks by dynamically optimizing the information utility of sensor data. Yang et al. [Yang and Sikdar 2003] have developed a distributed predictive tracking protocol to track mobile targets in sensor networks. Brooks et al. [2004] used a variation of the nearest-neighbor algorithm to associate sensor detections with tracks and support mutiple hypothesis tracking with self-organizing sensor networks.

All these tracking algorithms have been proposed for specific tracking problems and sensor network architectures while this article formalizes a general concept of tracking in sensor networks and develops a quantitative theory of trackability without restriction to specific algorithms. Just as classical complexity theory makes assertions about the complexity of *problems* and not specific algorithms for solving those problems, the present work addresses properties of tracking problems without reference to specific algorithms for solving them. For example, the statement that the Traveling Salesman Problem (TSP) is NP Complete is a statement about the TSP problem and not about any specific algorithm for solving the TSP problem. Similarly, our results concern the complexity of sensor network tracking problems without reference to specific algorithmic approaches to solving them.

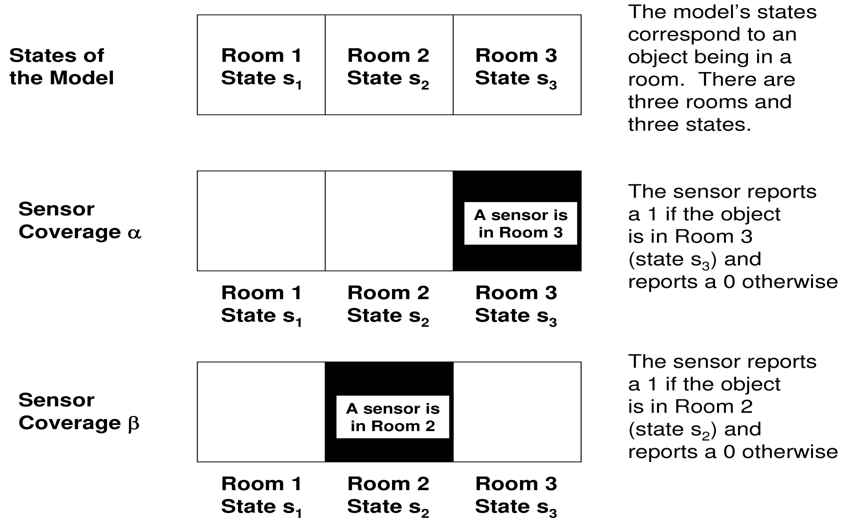


Fig. 1. The underlying state space has three states (each corresponding to occupancy of a room by an object). In Sensor Coverage α, the sensor detects presence in state s_3 (room 3) and in Sensor Coverage β, the sensor detects presence in state s_2 (room 2). The sensors report a 0 if no object is present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

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 [Berk and Fox 2005; DeSouza et al. 2006], autonomic computing [Roblee et al. 2005], dynamic social network analysis [Chung et al. 2006], and object tracking using sensor networks [Crespi et al. 2004, 2006]. An overview of process-based tracking using weak models can be found in Berk and Cybenko [2007]. Additionally, examples showing the effects of process dynamics (the underlying 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 that 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.

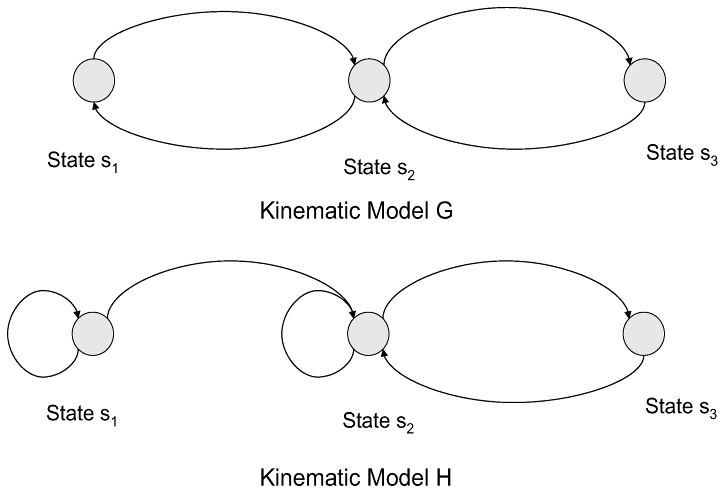


Fig. 2. There are two dynamical models, G and H, that define the allowed movement between states (rooms) of the state space.

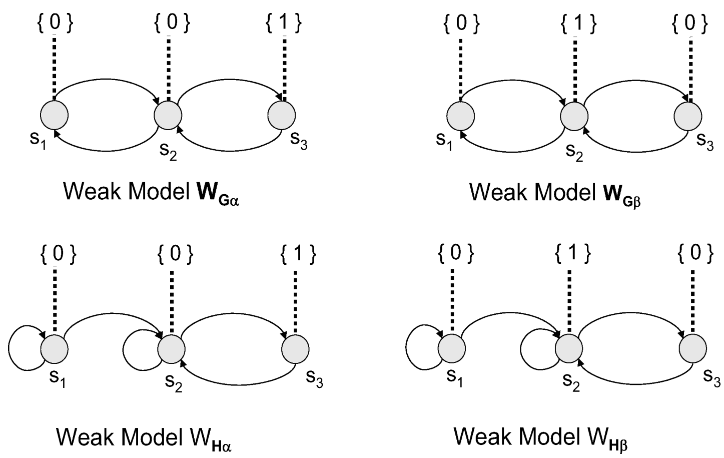


Fig. 3. There are four resulting *weak models* corresponding to the possible combinations of two different sensor coverages and two different dynamical models. This figure depicts the four models. In these graphical depictions, the possible sensor reports for each state are shown in braces connected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

Using the notation we have introduced, 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:

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

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

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

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

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

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 in this article, 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:

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

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 $\underline{1} = [111]^*$ denote the column vector of all 1's. (We will use v^* to denote the transpose of the vector v.) Having observed ξ_0, the possible states are $\underline{1}^* I_\alpha(\xi_0)$ 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

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

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

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

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 (the multiplication by A_G which propagates the states according to the model dynamics) followed by a filtering step (the multiplication by I_α which restricts the possible states according to the observation made), 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

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

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

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

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 $s_1s_1s_1$ while there are three consistent state sequences that end in state s_2, namely $s_1s_1s_2$, $s_1s_2s_2$ and $s_2s_2s_2$. There are no consistent state sequences that end in state s_3 because for that the last observation would have to be 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

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

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 to noise in the measurements, where $0 \leq k \leq m$. The value of m could be determined by a binomial distribution modelling the independent behaviors of the sensors and no coordination of distributed sensors is required to determine or use 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 is a mapping function that assigns a set of binary m-tuples to each state:

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

By $2^{\{0,1\}^m}$ we mean the set of all subsets of binary m-tuples so that $|L(v)|$ is the cardinality of the subset $L(v)$.

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 $L_k(u)$ as defined, 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 article implies that there is some noise level, $\overline{k}$, for which $h_{W_k}(t)$ grows polynomially in t when $k < \overline{k}$ while $h_{W_k}(t)$ grows exponentially in t when $k \geq \overline{k}$. It might be that $\overline{k} = 0$ or $\overline{k} = m + 1$, 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, ... 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 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, $W_{k,s} \leq W_{k',s'}$ if and only if $k \leq k'$ 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.

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 [Rota and Strang 1960] 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

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

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

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 [Lagarias and Wang 1995], and that for any finite set of matrices Σ, $\rho(\Sigma) = \overline{\rho}(\Sigma)$ [Berger and Wang 1992].

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 [2000], who reduced to it the empty word problem in Probabilistic Automata theory, which was previously known to be undecidable [Condon and Lipton 1989; Paz 1971]. (Note, real and rational problem formulations are treated separately because the complexity and decidability of some problems depends on the number system in which the problem is formulated [Blum et al. 1997].) 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 [Tsitsiklis and Blondel 1997, 2000]:

THEOREM 1 (TSITSIKLIS AND BLONDEL). *Unless $P = NP$, the Joint Spectral Radius $\rho(\{A, B\})$ of two $(0,1)$-matrices*[2] *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 bit-size of ϵ, namely $\log(1/\epsilon)$.*

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

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

where e_i is the ith standard basis vector which has all zero entries except for the ith entry which is exactly 1.

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

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

[2]By (0,1)-matrix, we mean a matrix whose entries are either 0 or 1.

where $\Sigma(\Phi) = \{A(\xi) \mid \xi \in \Phi\} \cup \{I(\xi) \mid \xi \in \Phi\}$ which we abbreviate to Σ for simplicity.

Let us now establish a useful lower bound. Let $A'(Z^T) = \prod_{i=0}^{T} A_G I(\xi_i)$.

LEMMA 2. *Let r be any n-vector with positive entries, and A be any n by n $(0, 1)$-matrix. Then $\underline{1}^* r \geq \frac{1}{n}\underline{1}^* Ar$.*

PROOF. Since A is a $(0, 1)$-matrix, $\underline{1}^* Ar = \sum_{i,j} a_{i,j} r_j \leq \sum_{i,j} r_j = n \sum_j r_j = n\underline{1}^* r$. □

LEMMA 3. *Let $A'(Z^T) = \prod_{i=0}^{T} A_G I(\xi_i)$. Then $h(Z^T) \geq \frac{1}{n} \cdot \|A'(Z^T) \cdot \underline{1}\|_1$*

PROOF. Let $Z^T = \xi_0 \xi_1 \dots \xi_T$. Then by definition

$$h(Z^T) = \underline{1}^* I(\xi_0) \prod_{i=1}^{T} A(\xi_i)\underline{1}.$$

Let $r = I(\xi_0) \prod_{i=1}^{T} A(\xi_i)\underline{1} = I(\xi_0) A'(Z^T)\underline{1}$ so that $h(Z^T) = \underline{1}^* r$. By Lemma 2,

$$\begin{aligned} h(Z^T) &= \underline{1}^* r \geq \frac{1}{n}\underline{1}^* Ar = \frac{1}{n}\underline{1}^* AI(\xi_0) \prod_{i=1}^{T} A(\xi_i)\underline{1} = \frac{1}{n}\underline{1}^* A'(Z^T)\underline{2} \\ &= \frac{1}{n} \cdot \|A'(Z^T) \cdot \underline{1}\|_1 \end{aligned}$$

□

Now, recalling the definition of the ∞-norm of a matrix:

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

the fact that in our case all matrix entries are nonnegative, and Lemma 3 it holds that

$$h(Z^T) = \|A(Z^T) \cdot \underline{1}\|_1 \geq \frac{1}{n} \cdot \|A'(Z^T) \cdot \underline{1}\|_1 = \frac{1}{n} \cdot \|A'(Z^T)\|_\infty \geq \frac{1}{n} \cdot \rho(A'(Z^T)) .$$

So we have

$$\frac{1}{n}\big(\rho(A'(Z^T)\big)^{1/(T+1)})^{T+1} \leq h(Z^T) \leq n(\overline{\rho}_{T+1}(\Sigma))^{T+1}.$$

By taking the max over Z^T and recalling that $\max_{Z^T} h(Z^T) = h_M(T)$ by definition, we obtain the inequalities

$$\frac{1}{n}\rho_{T+1}(\Sigma)^{T+1} \leq \max_{Z^T} h(Z^T) = h_M(T) \leq n(\overline{\rho}_{T+1}(\Sigma))^{T+1} \tag{5}$$

for any $T \geq 1$. (Note that Z^T contains $T+1$ observations so we take the $(T+1)$st root and so on.) The reader can verify that

$$\max_{Z^T}\{\rho(A'(Z^T))\} = \max_{A_j \in \Sigma} \left\{ \rho \left(\prod_{j=0}^{T} A_j \right) \right\}.$$

These derivations were for a fixed number of observations, T, and we now explore the implications of these relationships as $t = T$ is a variable and grows unboundedly. 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 nonnegative 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 $\rho(\Sigma) \leq 1$ 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 $L : \{1, 2\} \to 2^{\{0,1\}}$, $L(i) = \{0, 1\}$, $i = 1, 2$:

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

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

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

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

This example shows that the case $\rho(\Sigma) = 1$ is critical and can correspond to models with either bounded or unbounded hypothesis growth. In the remainder of this section we present a series of results, each proven in a separate subsection, that prove the following fundamental facts about a weak model M:

a) Either $h_M(t) = O(p(t))$, for a polynomial p, or $h_M(t) = \Omega(2^{ct})$, for $c > 0$. Moreover $h_M(t) = O(p(t))$ if and only if $\rho(\Sigma) \leq 1$.

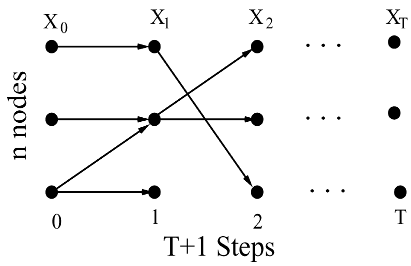


Fig. 4. Trellis structure. The n nodes in X_i are drawn along a vertical direction.

b) The question "$h_M(t) = O(p(t))$ versus $h_M(t) = \Omega(2^{ct})$" is efficiently Turing-decidable.
c) The question "$h_M(t) = \Theta(t^k)$, $k \geq 1$, versus $h_M(t) = O(1)$", in the case $\rho(\Sigma) \leq 1$, is also efficiently Turing-decidable.

4.1 a) The Number of Hypotheses Grows Either Polynomially or Exponentially

The statement in a) about the nonexistence of intermediate rates of growth was actually proved independently by Bell [2005] on the more general domain of semigroups of complex matrices and using sophisticated algebraic techniques. Our results, based on simpler graph-theoretic and combinatorial arguments, though applicable only to the restricted domain of (0, 1)-matrices, are stronger in that they lead to efficient algorithms to decide the questions in b) and c). Moreover, it has recently been shown that the exact degree of polynomial growth can be efficiently computed as well [Jungers et al. 2006].

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 4). 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 $\mathcal{T} = (\{X_i\}_{i\in[T+1]}, Y)$ 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. (We define $[k] := \{0, 1, \ldots, k - 1\}$.) 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 earlier possesses an interesting and useful property described in the following theorem. A trellis graph has the *unique path property* if there exists at most one path $\pi : [t_1, t_2] \to V$ such that $\pi(t_1) = \pi(t_2) = u$ for all nodes u and times $t_1 < t_2$.

THEOREM 4. *Let $M = (V, E, L, \Phi)$ and let $\mathcal{T}(M, Z^T) = (\{X_i\}, Y)$ be a trellis structure with the unique path property as defined above. Then*

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

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 $f_u(\pi) = l_u(\pi) = -1$ 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 $([-1, T] \times [-1, T])^n = [-1, T]^{2n}$ so there are $(T + 2)^{2n} = O(T^{2n})$ distinct hash values. 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 5).

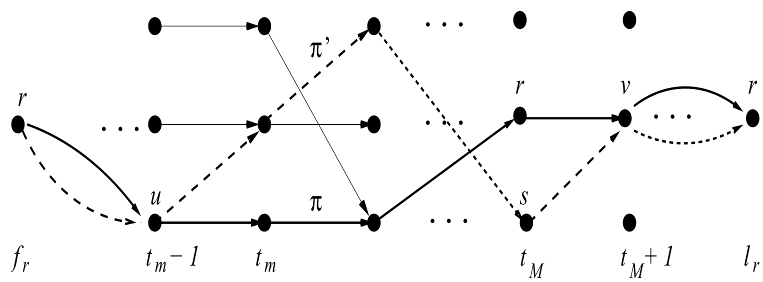


Fig. 5. At time f_r (l_r) node r is visited by both π and π' for the first (last) time.

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 this result is

Corollary 5. *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 4; that is, has the unique path property. 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 4.

Theorem 6. *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 the trellis does* not *have the unique path property so 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 $\pi_1, \pi_2 : [t_1, t_2] \rightarrow V$ 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 [Cvetkovic 1980]. 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 thus providing a proof of the fundamental fact anticipated in point a).

COROLLARY 7. *Let $M = (V, E, L, \Phi)$ and Σ be the set of matrices derived from A_M and Φ. Then*

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

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

PROOF. If $\rho(\Sigma) \leq 1$ then by Theorem 6 (specifically, its contrapositive) and Theorem 4 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 5) it must be $h_M(T) = \Omega(2^T)$. □

4.2 b). Efficient Decidability of Polynomial vs Exponential Growth

In this section 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 algorithm for answering the question. Moreover the problem is also tractable. Consequently we can efficiently decide, given a weak model, whether the number of hypotheses grows polynomially or exponentially in the worst case.

THEOREM 8. *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*

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

for $A_{j_i} \in \Sigma$ for all i. Moreover, this condition can be tested in time polynomial in $|\Sigma|$ and n.

Proof. The if-part is immediate since, as in Theorem 6, 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 $e_v^* A_{j_1} A_{j_2} \cdots A_{j_T} e_v \geq 2$. 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 $e_v^* A_{j_1} A_{j_2} \cdots A_{j_T} e_v \geq 2$ 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

$$\mathcal{T}_1 = \mathcal{T}[j_1, j_2, \ldots, j_t, j_{s+1}, \ldots, T] \text{ and } \mathcal{T}_2 = \mathcal{T}[j_{t+1}, j_{t+2}, \ldots, j_s],$$

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 $e_v^* A_{j_1} A_{j_2} \cdots A_{j_t} e_u \geq 2$ or $e_u^* A_{j_{s+1}} A_{j_{s+2}} \cdots A_{j_T} e_v \geq 2$. 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, though inefficiently, 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.

One way to improve the bound on T would be to replace V_t above with a function $g : [0, T] \to V^2$ defined as $g(t) = (\pi_1(t), \pi_2(t))$, for $\pi_1, \pi_2 \in \Pi$, and observe that $|\text{range}(g)| = O(n^2)$. This is not yet a polynomial algorithm because there would be still an exponential number of matrix sequences of polynomial length to check. However the difficulty can be circumvented as follows. Let $M = (V, E, L, \Phi)$ be a weak model and $\Sigma = \{A_1, A_2, \ldots, A_m\}$ be the set of all the (0, 1)-matrices associated with it. Let $G_k = (V, E_k)$ be the graph whose adjacency matrix is A_k. We need to check for the existence of any $(t+1)$-sequence of matrices $A_{j_0}A_{j_1}\ldots A_{j_t}$ in Σ, representing a trellis graph $\mathcal{T}(M, Z^t)$ containing two distinct paths $\pi_1, \pi_2 : [0, t] \to V$ that verify $\pi_1(0) = \pi_1(t) = \pi_2(0) = \pi_2(t) = i$. Let us define an auxiliary graph G' whose paths account for the simultaneous presence of paths π_1 and π_2 in all the possible trellises $\mathcal{T}(M, Z^t)$. Let $G' = (V', E')$ with

—$V' = \{(u, v) \mid u, v \in V\}$,
—$E' = \{((u, v), (u', v')) \mid \exists k\ (u, u'), (v, v') \in E_k\}$.

Clearly there exists a trellis graph $\mathcal{T}(M, Z^t)$ that verifies the required property for $i \neq j$ if and only if the graph G' possesses a path from (i, i) to (i, i) of length $t + 1$.

Since G' has at most n^2 nodes we can verify the presence of such a path by considering all the matrix products A'^h, for $h = 1, 2, \ldots, n^2$, and checking whether $A'^h((i, i), (i, i)) \geq 2$. The number of arithmetic operations required by this method will then depend on the cost of building A': $O(n^4|\Sigma|)$, and on the cost of computing all the first n^2 powers of the $n^2 \times n^2$ matrix A' and performing the $O(n)$ checks on the main diagonal: $O(n^{3+2\gamma})$, where $O(n^\gamma)$ is the time to multiply two $n \times n$ matrices. □

4.3 c) Efficiently Deciding Bounded vs Unbounded Polynomial Growth

Finally, in this subsection, we address the problem of efficiently deciding the question "$h_M(t) = \Theta(t^k)$, $k \geq 1$, versus $h_M(t) = O(1)$", in the case $\rho(\Sigma) \leq 1$ (see point c) at the beginning of this section). That is, we are concerned with determining whether hypothesis growth is bounded by a constant or a polynomial when $\rho(\Sigma) \leq 1$.

Recall that given a weak model $M = (V, E, L, \Phi)$ and a sequence of $t + 1$ observations, Z^t, we denote by $A = A_G$ the adjacency matrix of the kinematic specification, $G = (V, E)$, and $A(Z^t) = [a_{i,j}(Z^t)]_{i,j} = I(\xi_0)\prod_{j=1}^{t} A(\xi_j)$.

Consider the following two weak models M and M' whose state-transition kinematics are defined respectively by the following adjacency matrices:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } A' = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},$$

and whose common state-observation mapping $L : \{1, 2\} \to 2^{\{0,1\}}$ is defined as $L(i) = 0, 1$, for $i = 1, 2$. Models M and M' induce matrix sets $\Sigma = \{A, I\}$ and

$\Sigma' = \{A', I\}$ respectively. Now, observe that for all $t > 0$:

$$A^t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \text{ and } (A')^t = A'$$

and so, while $\rho(\Sigma) = \rho(\Sigma') = 1$, $h_M(t)$ grows unboundedly whereas $h_{M'}(t) \leq 2$. Moreover, observe that in the case of A^t there exist two different states i, j such that $(A^t)_{i,i} = 1$, $(A^t)_{j,j} = 1$ and $(A^t)_{i,j} = t \geq 1$. This property does not hold for M' and the results demonstrate that this property is precisely what characterizes bounded versus polynomial growth.

We generalize from these two examples a property that characterizes the growth rate of $h_M(t)$ in the case $\rho(M) \leq 1$. We also show that this property can be decided in polynomial time. In particular, given a weak model M, with adjacency matrix A and induced matrix set Σ, with $\rho(\Sigma) \leq 1$, we will establish the following two results:

(1) There exists a sequence of observations, Z^t, and two states $i, j, i \neq j$, such that $a_{i,j}(Z^t) \geq 1$, $a_{i,i}(Z^t) = 1$ and $a_{j,j}(Z^t) = 1$, if and only if $h_M(t)$ grows unboundedly.
(2) This property can be Turing-decided in polynomial time (in n and $|\Sigma|$).

We first establish the "only if" part of statement 1 above.

LEMMA 9. *Let $Z^t = (\xi_0, \xi_1, \ldots, \xi_t)$ be a sequence of $t + 1$ observations,*

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

be the transition matrix associated with the subsequence $(\xi_u, \xi_{u+1}, \ldots, \xi_v)$ and

$$A(Z^t) = [a_{i,j}(Z^t)]_{i,j} = I(\xi_0) \cdot A(Z^{1:t}).$$

If there exist i and j with $i \neq j$, such that

$$(1)\ a_{i,j}(Z^t) = c \geq 1; \quad (2)\ a_{i,i}(Z^t) \geq 1; \quad (3)\ a_{j,j}(Z^t) \geq 1$$

then $f(s) := \max\{a_{i,j}(Z^s) \mid Z^s \in \Phi^s\}$ is unbounded and so is $h_M(s) \geq f(s)$.

PROOF. Let e_i denote the $(0, 1)$-vector whose components are all zero except the i^{th} one. Thus $a_{i,j}(Z^t) = e_i^* \cdot A(Z^t) \cdot e_j$. Recall that $A(Z^t)$ is a matrix with nonnegative entries, and so

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

Now, let $A_t = A(Z^{1:t})$. We will show by induction on k that $e_i^* I(\xi_0) A_t^k e_j \geq ck$, $e_i^* I(\xi_0) A_t^k e_i \geq 1$ and $e_j^* I(\xi_0) A_t^k e_j \geq 1$. This implies that $f(s) \geq e_i^* I(\xi_0) A_t^{\lfloor s/t \rfloor} e_j \geq c \cdot \lfloor s/t \rfloor$.

—*Basis for Induction:* For $k = 1$, the statement is true by the assumptions of the lemma.

—*Inductive Step:* Assume that the statement is true for some $k \geq 1$. Recall that $I = \sum_{u=1}^{n} e_i e_i^*$ and using inequality 1 in the lemma's statement we can write:

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

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

We now establish the "if" part of statement 1, which is the harder of the necessary and sufficient assertion made. Because of the complexity of the proof, it will help the reader to have an overview of the proof structure first. Specifically, we can decompose the argument into the following two parts:

a) First, (in Theorem 10) we show that if $\rho(\Sigma) \leq 1$ and $h_M(t)$ is unbounded then there exists an observation sequence Z^t such that $A(Z^t)$ has a square submatrix of the form[3] $A(Z^t)[S|S]$, for $|S| > 1$, with the following two properties:

—$A(Z^t)[S|S]$ has at least one nonzero element in each row and in each column (note that if $\rho(\Sigma) \leq 1$ then for any Z^t no diagonal element in $A(Z^t)$ can be greater than 1 because otherwise, by definition, it would not have the unique path property and so $\rho(\Sigma) > 1$ by Theorem 6).

—The (0, 1) pattern of $A(Z^t)[S|S]$ (that is, the (0, 1)-matrix obtained by replacing each nonzero element with a 1) is not a quasi-permutation matrix, that is, a matrix obtained from a permutation matrix by zeroing out zero or more entries.

b) Secondly, (in Theorem 11) we will show that if there exists an observation sequence Z^t such that $A(Z^t)$ satisfies the conditions in a) then we can find another observation sequence $(Z')^s$ such that $A((Z')^s)$ satisfies: there exist $i, j, i \neq j$, such that $a_{i,j}((Z')^s) \geq 1$, $a_{i,i}((Z')^s) = 1$ and $a_{j,j}((Z')^s) = 1$.

These statements, when combined, are the converse of Theorem 9. With this framework thus outlined, we now proceed to establish these results.

THEOREM 10. *Let $M = (V, E, L, \Phi)$ be a weak model such that $\rho(\Sigma) \leq 1$ and $h_M(t)$ grows unboundedly. Then there exists a set of states $S \subseteq V$, $|S| > 1$, and an observation sequence Z^t such that $A(Z^t)[S|S]$ has at least one nonzero element*

[3]For sets of indices U, V, $A[U|V]$ is the submatrix of A obtained by selecting rows in U and columns in V.

in each row and in each column, its $(0, 1)$ *pattern is not a quasi-permutation matrix and for each* $k > 0$ *the diagonal elements in* $A(Z^t)^k$ *are at most* 1.

PROOF. Let $|V| = n$ be the number of states in the model. Construct the following auxiliary graph, called AG. The nodes of AG are all possible bipartite graphs involving two copies of the states of the model. There are 2^{n^2} nodes in AG since there are 2^{n^2} $(0, 1)$-matrices of order n. Since there is an unbounded number of hypotheses, pick the minimal t for which there is an observation sequence that leads to at least $N = 1 + n^{2^{n^2+1}}$ hypotheses. This is a very large number but possible by assumption.

A minimal t exists because there is only a finite number of possible observation t'-sequences with $t' \leq t$. For this minimal t and associated observation sequence, construct the corresponding trellis graph.

By construction, each edge in this trellis graph must belong to at least one path connecting a node in V at time 0 to a node in V at time t. Now each pair of adjacent copies of visited states in the trellis graph together with the trellis graph links between those visited states is a node in the auxiliary graph, AG, and so the trellis graph is equivalent to a path p through AG.

Note that between two consecutive stages in the trellis graph the number of hypotheses increases by at most a factor of n^2 (corresponding to the node in AG which is the maximally connected bipartite graph). Consequently, in order to achieve $N = 1 + n^{2^{n^2+1}}$ hypotheses it must be that $t \geq 2^{n^2}$. Moreover, when the path p goes through a node in AG that corresponds to a quasi-permutation (partial matching) on the states, there is no growth in the number of hypotheses and so p must in fact contain at least 2^{n^2} nodes that are not quasi-permutations. Call those nodes "increasing" nodes.

Now AG has 2^{n^2} nodes, so fewer than 2^{n^2} increasing nodes, and as we said the trellis graph represents a path $p : B_1, B_2, \ldots, B_t$ in AG containing at least 2^{n^2} increasing nodes. Therefore, by the pigeonhole principle, there must be at least one increasing node B occurring at least twice in p; that is, there exist $u < v$ such that $B_u = B_v = B = (V, V, E_{VV})$. This means that there is a cycle C in AG starting and ending with an increasing node B. Back in the trellis graph, this cycle corresponds to a sequence of stages $X_{t_1}, X_{t_1+1}, \ldots, X_{t_2}, X_{t_2+1}$ and associated observations, where the pattern of states and links between X_{t_1} and X_{t_1+1} is exactly the same as between X_{t_2} and X_{t_2+1}. Moreover, that pattern is a bipartite graph B that is not a perfect or partial matching because it is an increasing node of AG. This construct is illustrated in Figure 6.

Define the class of trellis paths $\Pi(X, Y) = \{\pi : [t_1, t_2] \to V \mid \pi(t_1) \in X, \pi(t_2) \in Y\}$ and let $S = \{i \in V \mid \exists \pi \in \Pi(V, V), \pi(t_1) = i\}$. Then the following are true:

(1) For any node $s \in S$ there exists a path $\pi \in \Pi(S, S)$ such that $\pi(t_2) = s$. Namely any node $s \in S$ is reachable at time t_2 by some trellis path starting from some node in S at time t_1. In fact assume $s \in S$. Since the link pattern repeats at time t_2, if a trellis path leaves $s \in S$ at time t_1 through an edge $(s, s') \in E_{VV}$ then there must be a trellis path that passes through $s \in S$ at time t_2 and leaves it along the same edge. And this implies that there

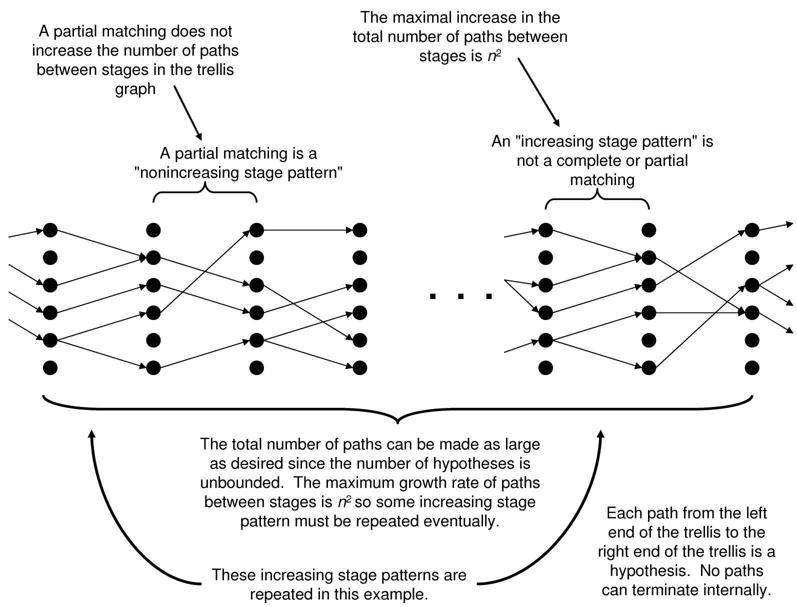


Fig. 6. Because there is an unbounded number of hypotheses, there is an unbounded number of paths in the associated worst case trellis graph. Therefore, some patterns between stages of the trellis graph must be repeated and this is used in the proof of Theorem 10.

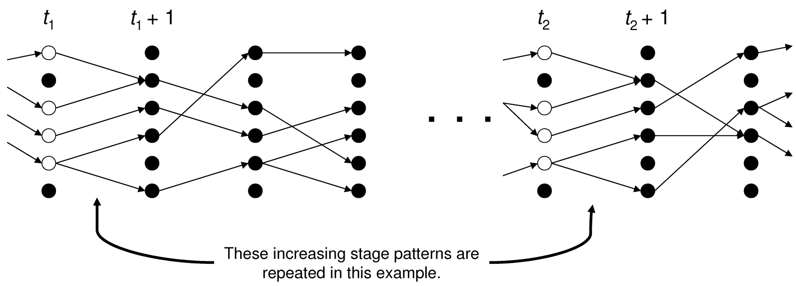
exists a trellis path that starts from some node in S at time t_1 and ends in s at time t_2.

(2) For any node $s \in S$ there exists a path $\pi \in \Pi(S, S)$ such that $\pi(t_1) = s$. Namely from any $s \in S$ at time t_1 we can reach some other node in S at time t_2. In fact assume $s \in S$. Then by definition there exists a path $\pi \in \Pi(S, V)$ with $\pi(t_1) = s$ and $\pi(t_2) = v \in V$. Now π can be prolonged until some node in V at time t. Thus there exists an edge $(v, v') \in E_{VV}$ and so there exists also a trellis path $\pi \in \Pi(S, S)$ that leaves v at time t_1 which gives $v \in S$.

(3) The class of paths $\Pi(S, S)$ cannot form a matching between the nodes in S. If this were not the case we would be able to remove the piece of trellis graph from t_1 to t_2 thus obtaining a shorter observation sequence achieving the same number of hypotheses which would contradict the assumption of the minimality of t.

(4) For each $s \in S$ there is at most one path $\pi \in \Pi(S, S)$ starting and ending in s. In fact we have already observed that otherwise we would have $\rho(\Sigma) > 1$, contradicting the hypothesis. As a consequence it must be that $|S| > 1$.

Consider the matrix $A(Z^{t_1:t_2})[S|S]$. Conditions 1, 2, and 3 imply that its (0, 1) pattern cannot be a quasi-permutation matrix and moreover each column and each row has at least one nonzero element. Condition 4 implies that the diagonal elements are at most 1. This last condition must hold for any integer power of $A(Z^{t_1:t_2})$. □



S is the set of states that paths pass through at times t_1 and t_2. They are not solid in the trellis graph above to highlight them.

For *u* and *v* in *S*, define the matrix *P* by *P(u,v)* = 1 if there is a path from *u* at time t_1 to *v* at time t_2 and *P(u,v)* = 0 otherwise. Note that there can be at most one path from any state *s* at time t_1 to the same *s* at time t_2 because otherwise there would be an exponential growth in hypotheses contradicting the assumptions of the theorem.

On the other hand, there is a 1 in every row and every column of *P* because there are edges leaving and entering every state in *S*.

Fig. 7. Using special repeated stages, it is possible to construct an auxiliary matrix P that has special properties used in the proof of Theorem 11.

THEOREM 11. *Let $M = (V, E, L, \Phi)$ be a weak model and assume there exist a set of states $S \subseteq V$, $|S| > 1$, and an observation sequence Z^t such that $A(Z^t)[S|S]$ has at least one nonzero element in each column and in each column, its $(0, 1)$ pattern is not a quasi-permutation matrix, and, for each $k > 0$, the diagonal elements in $A(Z^t)^k$ are at most 1. Then there exist an observation sequence $Z^T = \xi_0, \xi_1, \ldots, \xi_T$ and two different states $i, j \in S$ such that $a_{i,j}(Z^T) \geq 1$, $a_{i,i}(Z^T) = 1$ and $a_{j,j}(Z^T) = 1$.*

PROOF. Let $|S| = n$ and P be the $(0, 1)$ pattern matrix of $A(Z^t)[S|S]$. We prove that there exist u, v, $u \neq v$, such that $P^{n!}(u, u) = 1$, $P^{n!}(v, v) = 1$ and $P^{n!}(u, v) > 0$. And this will be enough as for any u, v and for any power k, $P(u, v)^k \leq a_{u,v}(Z^t)$.

Let $G = (V', E')$ be the graph defined as $V' = \{1, 2, \ldots, n\}$, and $(u, v) \in E'$ if and only if $P(u, v) = 1$. Various properties of P are evident from a sample depiction in Figure 7. Then the following facts are true:

(1) Each node in G has indegree and outdegree at least 1, and so there are no isolated nodes.
(2) G is not a cycle cover or a partial cycle cover since P is not a quasi-permutation matrix (each permutation matrix corresponds uniquely to a cycle cover in G).
(3) For any integer $k > 0$, $P^k(u, v)$ is the number of directed paths (not necessarily simple) from node u to node v of length k.
(4) Each node belongs to at most one cycle. So no path can connect a cycle to itself and each (weakly) connected component of the graph consists of

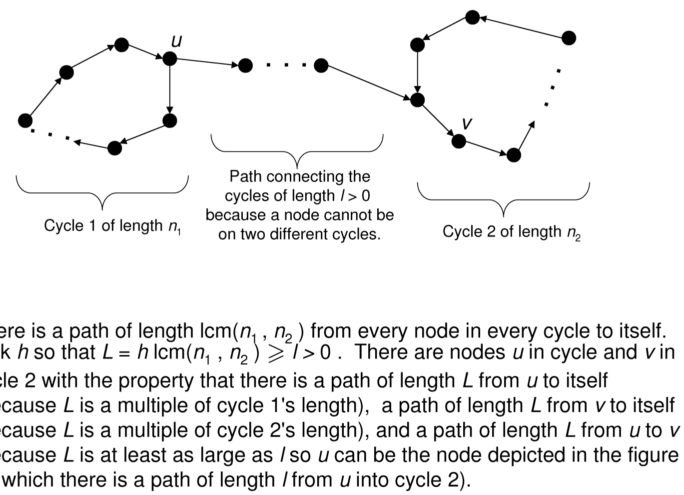


Fig. 8. Using the properties of the derived P, it is possible to identify nodes with the properties specified by Theorem 11.

simple cycles possibly connected one-way by simple paths. Otherwise we would be able to find a node u such that $P^k(u, u) > 1$ for some k.

(5) There must exist at least two different cycles connected by at least a simple path (or an edge) in G. In fact
  (a) If all the cycles were node disjoint they would form a cycle cover.
  (b) If we had just simple ”open” paths then their start node would have indegree zero and their ending node outdegree zero.
  (c) Finally, if we had only one cycle connected to a simple path then the simple path would either start from a node with indegree 0 or end at a node of outdegree 0.

At this point let C_1, C_2 be two disjoint cycles of n_1 and n_2 nodes respectively.

Let $\pi = (v_{j_1}, v_{j_2}, \ldots, v_{j_{l+1}})$ be the shortest path connecting one way C_1 to C_2. So $v_{j_1} \in C_1$, $v_{j_{l+1}} \in C_2$ and no other node in π belongs to either C_1 or C_2. Note that it might be $l = 1$, and also $n_1 = 1$ and/or $n_2 = 1$.

Let $h > 0$ such that $L = h \cdot \text{lcm}(n_1, n_2) \geq l$. Then $P^L(u, u) = 1$ for any node $u \in C_1 \cup C_2$. In fact there exists exactly a path of length L from any node in either C_1 or C_2 to itself. Pick node $v_{j_1} \in C_1$. Then $P^L(v_{j_1}, v_{j_1}) = 1$ because v_{j_1} belongs to cycle C_1. But v_{j_1} is also the node in C_1 closest to C_2 so follow path π towards C_2. After $L \geq l$ edges we will enter C_2. Let $v \in C_2$ be the node in C_2 at distance L from v_{j_1}. Again it has to be $P^L(v, v) = 1$ but also $P^L(u_{j_1}, v) \geq 1$ (and not just 1 because C_1 and C_2 may be connected by more than a simple path). Finally, observe that $\text{lcm}(n_1, n_2) | n!$ and $n! \geq l$ and so the claim follows for $L = n!$. The reasoning behind this argument is depicted in Figure 8. □

Polynomial time decidability of unbounded polynomial growth. Let $M = (V, E, L, \Phi)$ be a weak model with $\rho(\Sigma) \leq 1$. We have just established that $h_M(t)$ grows unboundedly if and only if there exists an observation sequence $Z^T = \xi_0, \xi_1, \ldots, \xi_T$ and two different states i, j such that $a_{i,i}(Z^T) = 1$, $a_{j,j}(Z^T) = 1$

and $a_{i,j}(Z^T) \geq 1$. In the following we show how to decide this condition in polynomial time.

Let $\Sigma = \{A_1, A_2, \ldots, A_m\}$ be the set of all the (0, 1)-matrices associated with model M. Let also $G_k = (V, E_k)$ be the graph whose adjacency matrix is A_k.

We need to check for the existence of a $(T+1)$-sequence of matrices $A_{j_0}A_{j_1}\ldots A_{j_T}$ in Σ, representing a trellis graph $\mathcal{T}(M, Z^T)$ that contains paths $\pi_1, \pi_2, \pi_3 : [0, T] \to V$ verifying

(1) $\pi_1(0) = \pi_1(T) = i$
(2) $\pi_2(0) = \pi_2(T) = j$
(3) $\pi_3(0) = \pi_1(0)$ and $\pi_3(T) = \pi_2(T)$.

Let us define an auxiliary graph $\hat{G}$ whose paths account for the simultaneous presence of paths π_1, π_2, π_3. Let $\hat{G} = (\hat{V}, \hat{E})$ with

—$\hat{V} = \{(u, v, w) \mid u, v, w \in V\}$,
—$\hat{E} = \{((u, v, w), (u', v', w')) \mid \exists k\ (u, u'), (v, v'), (w, w') \in E_k\}$.

Clearly there exists a trellis graph $\mathcal{T}(M, Z^T)$ that verifies the required property for $i \neq j$ if and only if graph $\hat{G}$ possesses a path from (i, j, i) to (i, j, j) of length $T + 1$.

Since $\hat{G}$ has at most n^3 nodes we can verify the presence of such a path by considering all the matrix products $\hat{A}^h$, for $h = 1, 2, \ldots, n^3$, and checking whether $\hat{A}^h((i, j, i), (i, j, j)) \geq 1$. The number of arithmetic operations required by this method depends on the cost of building $\hat{A}$: $O(n^6|\Sigma|)$, and on the cost of computing the first n^3 powers of A' and checking $O(n^2)$ entries: $O(n^{5+3\gamma})$, where $O(n^\gamma)$ is the cost of multiplying two $n \times n$ matrices.

Here is an explicit description of the algorithm:

Input: $M = (V, E, L, \Phi)$
Output: **true** if and only if $h_M(t) = O(1)$

```
Σ := {A(ξ) | ξ ∈ Φ} ∪ {I(ξ) | ξ ∈ Φ}
if ρ(Σ) > 1 then
   return false
Ĝ := (V̂, Ê) with V̂ and Ê defined as above.
Â := A_Ĝ (adjacency matrix of Ĝ)
B := Â
for k := 1 to |V|^3
   for i, j := 1 to |V|
      if i ≠ j and B((i, j, i), (i, j, j)) ≥ 1 then
         return false
   end
   B := B × Â
end
return true
```

---

5. APPLICATIONS TO SENSOR NETWORKS

Many defense, security, and surveillance applications involve the detection of dynamic processes using large amounts of observation data gather by sensors. Hidden Markov Models (HMM) and Multiple Hypotheses Tracking (MHT) algorithms are widely used in this area to infer the states of objects from the sensor observation sequences. For example, Warrender and Forrest [Warrender and Forrest 1999] employed HMM's to characterize normal system call sequences in executing computer programs and then used these HMM's as normal signatures to detect abnormal software behavior. Brook et al. [2004] used a variation of the nearest-neighbor algorithm to associate sensor detections with tracks and supported multiple hypothesis tracking with self-organizing distributed sensors in military applications. Those works, like others in this area, do not address underlying theoretical performance issues such as we do in this article with the trackability framework that we have introduced.

As discussed in Section 1, weak models have the structural ingredients of HMM's but do not specify the probabilities associated with state transitions and observations. Therefore, weak models can characterize a wider class of dynamic processes in those applications, such as nonstationary processes and processes with no statistical models, which motivated us to analyze trackability and other properties of such models. Our interest in this problem also arose from the study of effectively using weak models for detecting and tracking processes in a variety of applications such as computer security [Berk and Fox 2005; DeSouza et al. 2006], autonomic computing [Roblee et al. 2005], dynamic social network analysis [Chung et al. 2006] and object tracking using sensor networks Crespi et al. [2004, 2006]. Simple examples showing the effects of process dynamics (the underlying state machine model) and sensor coverage on hypothesis growth are developed in the Appendix.

The theoretical results developed in this article can be used to evaluate sensor placement and process modelling to improve the effectiveness of tracking algorithms. As discussed in Section 2, given an observation sequence $Z^t = \xi_0\xi_1 \ldots \xi_t$, the number of possible consistent hypotheses is given by

$$h(Z^t) = I(\xi_0)\prod_{i=1}^{i=t} A(\xi_i) = I(\xi_0)\prod_{i=1}^{i=t} AI(\xi_i), \tag{14}$$

where $A(\xi_i) = AI(\xi_i)$.

A is the underlying state transition matrix, which is determined by the inherent physical or logical constraints of the tracked target (for example, a kinematic model for a ground vehicle). Although the dynamics of a tracked object may not be changed by the model, the particular model used for tracking it can affect the performance of the tracking algorithm considerably. For example, states in a kinematic model may be the position of a target or the position and velocity of a target which would lead to different state spaces and state transition possibilities. Moreover, target states in a model can include heat signatures and other properties in addition to position and/or velocity. $I(\xi_i)$ describes the association matrices between states and their sensor observations, which can be modified by changing sensor placements. There are similarities between the role of $I(\xi_i)$

in this work and the concept of *gating* in classical tracking algorithms [Wang et al. 2002].

This discussion makes the point that models for tracking can involve specifying both the dynamics and sensor phenomenology even when the target is an adversary whose actual behavior we cannot change. We can still develop different kinds of models for an adversarial target that we wish to track.

Therefore, as shown in Equation (14), we can change the growth rate of tracking hypotheses by adding sensors or changing sensor placements. If each $I(\xi_i)$ only has a single 1 on its diagonal (that is, all states are directly distinguishable from their sensor observations), we can map any observation sequence into a unique state sequence directly and we only have one hypothesis—the real state sequence itself. In practice, we may need many sensors to achieve this and some states may not be directly observable by any sensor. Now the challenging problem is: given the state transition matrix A of a tracked target, what is an effective sensor placements to reduce the hypotheses' growth rate.

In this article we proved that, given a state transition matrix A and its stateobservation association matrices $I(\xi_i)$, it is polynomial-time decidable (in the size of the state space, n, and the number of association matrices $I(\xi_i)$) whether the worst-case hypothesis growth is exponential, or polynomial, or bounded. In fact, our proofs also include outlines of how to design these polynomial algorithms. Therefore, given a state transition matrix A, it is possible to compare several sensor deployment plans and choose the best one according to the hypothesis growth rates.

For example, with a given number of sensors, a sensor placement resulting in bounded hypothesis growth may be better than other placements that could lead to exponential hypothesis growth. Moreover, based on results developed in this article, we have polynomial complexity algorithms to decide whether a planned sensor placement results in bounded or polynomial growth of hypotheses. Therefore, we believe that results such as presented in this article are useful for the effective design and efficient use of sensor networks and the associated tracking algorithms.

An important problem beyond those discussed here involves tracking multiple targets simultaneously [Bar-Shalom and Li 1995; Reid 1979b; Yu and Wu 2005]. Multitarget tracking of a fixed number of targets can be reduced to single target tracking in some circumstances by combining the individual target kinematic models into a single product model. The penalty for this is exponential growth of the model's state space so this is rarely done. The results of this article apply to tracking performance of individual models nonetheless.

6. SUMMARY AND FUTURE WORK

This article 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 modeling and tracking problem. We have shown that the decision problem $\rho(\Sigma) \leq 1$ for Σ a finite set of $n \times n$ $(0, 1)$-matrices is in fact decidable in a Turing equivalent model of computation in time polynomial in Σ and n, unlike the general problem for sets of rational matrices which is known to be even 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 low-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 low-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 article. The various examples presented in this article 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 Sheng and Cybenko [2005a].

ACKNOWLEDGMENTS

We sincerely thank Vincent Blondel, Raphaël Jungers and Leonid Gurvitz for many helpful discussions, pointers to related work and suggestions for extensions of our earlier results, specifically the structure of the efficient algorithms for deciding between polynomial and exponential growth. We also thank the anonymous reviewers for providing valuable suggestions that significantly improved the original submitted manuscript.

APPENDIX

The results of this article 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—Exponential growth in the number of hypotheses.

—Example 2—Simple constraints on the kinematics of Example 1 result in bounded hypothesis growth.

—Example 3—Modified kinematics from Example 1 result in initial exponential growth which eventually stops and the overall hypothesis number in the limit is bounded.

—Example 4—Same kinematics as Example 3 but with modified sensor coverage yields unbounded polynomial growth.

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: $(x, y) \rightarrow (x \pm 1, y)$ or $(x, y) \rightarrow (x, y \pm 1)$. 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 9, 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.

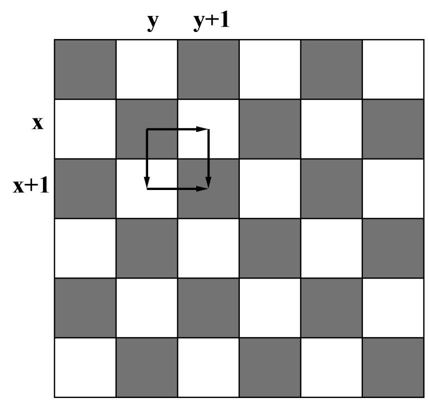


Fig. 9. The black cells are monitored by binary sensors.

(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 Figure 9):

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

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 observation between any pair of sensor observations, this sequence of 3 sensor observations is technically a sequence of 5 observations.) According to Theorem 6 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.

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:

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

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 Figure 9) 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 10. 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

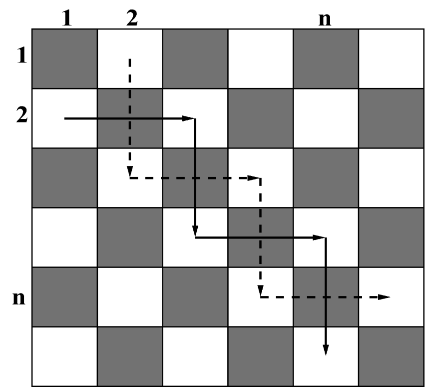


Fig. 10. The only two paths consistent with observations $X, (2, 2), X, (3, 3), X, \ldots, (n, n), X$. Character X stands for "no-observation." When the vehicle deviates from this diagonal pattern, the ambiguity disappears completely.

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).

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.

The following facts are true for this example:

(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) 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 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

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

The number of solutions, $h(T)$, can be determined using standard methods from combinatorics [Balakrishnan 1995]:

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

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 of arbitrary degree and this growth is over all time since the vehicle can stay in the same horizontal line indefinitely.

GLOSSARY OF SYMBOLS

This glossary lists notation in the order it is used together with a brief description and the page on which it is first introduced and defined.

Notation	*Description*
$G = (V, E)$	a finite directed graph or, equivalently, an automaton with states V and state transitions E
A, A_G	vertex/node adjacency matrix of G
Φ	the set of all possible observables by sensors
$L(v)$	the set of observables corresponding to state v
weak model	a finite automata whose dynamics are determined by traditional state machine and which emits a set of observable events when occupying a state with no probabilistic assumptions or models
$W = (V, E, L, \Phi)$	a weak model
ξ_t	an observation at time t
$Z^T = \xi_0\xi_1 \ldots \xi_{T-1}\xi_T$	an observation sequence of length $T+1$
$\mathcal{H}_W(Z^T)$	the set of hypotheses or state sequences that could have generated the observation sequence Z^T
$h_W(Z^T)$	the cardinality of $\mathcal{H}_W(Z^T)$
$h(Z^T)$	equal to $h_W(Z^T)$ when the weak model W is understood by context
$h(T)$	equal to $\max_{Z^T} h(Z^T)$
(0,1)-matrix	a matrix whose entries are either 0 or 1
Σ	a finite set of (0,1)-matrices
$\rho(\Sigma)$	the Joint Spectral Radius of Σ

$I(\xi), I_W(\xi)$	a diagonal matrix which specifies which states, in weak model W, correspond to observation ξ
$\underline{1}$	vector whose entries are all 1's
v^*	transpose of v
$z_W(\xi_0\xi_1\ldots\xi_k)$	row vector whose ith coordinate is the number of hypotheses in weak model W starting in any state and ending in state i at time k consistent with the observations $\xi_0\xi_1\ldots\xi_k$
$A(\xi_j), A_G(\xi_j)$	$A(\xi_j) = AI(\xi_j)$, $A_G(\xi_j) = A_G I(\xi_j)$ are the weak model dynamics constrained by the observation ξ_j
$\{0,1\}^m$	the set of all binary vectors of length m
$2^{\{0,1\}^m}$	the set of all subsets of binary vectors of length m
$d_H(x, y)$	the Hamming distance between binary vectors x and y
$\sigma(A)$	spectrum of matrix A; that is, the set of eigenvalues of A
$\rho(A)$	spectral radius of A
$\overline{\rho}(\Sigma), \rho(\Sigma)$	Joint Spectral Radius and Generalized Joint Spectral Radius of the set Σ of matrices
$\Sigma(\Phi)$	the set of matrices $A(\xi)$ for $\xi \in \Phi$
$[k], [t_1, t_2]$	$[k] = \{0, \ldots, k\}$, $[t_1, t_2] = \{t_1, t_1 + 1, \ldots, t_2\}$
$\mathcal{T}(M, Z^T)$	trellis graph for the model M and observations Z^T
π	a path through a trellis graph

REFERENCES

BALAKRISHNAN, V. K. 1995. *Combinatorics, Including Concepts of Graph Theory*. McGraw-Hill, New York.

BAR-SHALOM, Y. AND LI, X.-R. 1995. *Multitarget-Multisensor Tracking: Principles & Techniques*. Yaakov Bar-Shalom, Storrs, CT.

BELL, J. P. 2005. A gap result for the norms of semigroups of matrices. *Linear Algebra Appl. 402*, 101–110.

BERGER, M. AND WANG, Y. 1992. Bounded semigroups of matrices. *Linear Algebra Appl. 166*, 21–27.

BERK, V. AND CYBENKO, G. 2007. Process query systems. *IEEE Computer*, 62–70.

BERK, V. H. AND FOX, N. 2005. Process query systems for network security monitoring. In *Proceedings of the SPIE*, vol. 5403. Orlando, FL.

BLONDEL, V. AND TSITSIKLIS, J. 2000. The boundedness of all products of a pair of matrices is undecidable. *Syst. Control Lett. 41,* 2.

BLUM, L., CUCKER, F., SHUB, M., AND SMALE, S. 1997. *Complexity and Real Computation*. Springer, New York.

BROOKS, R., FRIEDLANDER, D., KOCH, J., AND PHOHA, S. 2004. Tracking multiple targets with self-organizing distributed ground sensors. *J. Parall. Distrib. Comput. 64,* 7, 874–884.

BUCHI, J. 1960. On a decision method in restricted second order arithmetic. *Z. Math. Logik Grundlag. Math 6*, 66–92.

CHUNG, W. ET AL. April 2006. Dynamics of process-based social networks. In *Proceedings of the SPIE* (Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V). Orlando, FL.

CONDON, A. AND LIPTON, R. 1989. On the complexity of space bounded interactive proofs. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*. 462–467.

CRESPI, V., CHUNG, W., AND JORDAN, A. B. 2004. 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.

Crespi, V., Cybenko, G., and Jiang, G. 2006. What is trackable? In *Proceedings of the SPIE* (Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V). Orlando, FL.

Crespi, V., Cybenko, G., and Jiang, G. 2005. The theory of trackability with applications to sensor networks. Tech. rep., TR2005-555, Department of Computer Science, Dartmouth College.

Crespi, V., Cybenko, G., and Sheng, Y. 2006. Tracking in complex situations and environments. In *Proceedings of the SPIE* (Unattended Ground, Sea, and Air Sensor Technologies and Applications VIII Conference). Orlando, FL.

Cvetkovic, D. 1980. *Spectra of Graphs: Theory and Application*. Academic Press.

Cybenko, G., Berk, V. H., Gray, R. S., and Jiang, G. 2004. 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.

Daubechies, I. and Lagarias, J. 1992. Sets of matrices all infinite products of which converge. *Linear Algebra Appl. 162*, 227–263.

DeSouza, I. et al. 2006. Detection of complex cyber attacks. In *Proceedings of SPIE* (Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V). Orlando, FL.

Jungers, R., Protasov, V., and Blondel, V. D. 2006. Efficient algorithms for deciding the type of growth of products of integer matrices. Preprint.

Lagarias, J. and Wang, Y. 1995. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra Appl. 214*, 17–42.

Lind, D. and Marcus, B. 1995. *An Introduction to Symbolic Dynamics*. Cambridge University Press, Cambridge, UK.

Marcus. 1995. Symbolic dynamics and connections to coding theory, automata theory and system theory. In *Proceedings of the 50th Symposium in Applied Mathematics, American Mathematical Society*. (PSAM).

Mukund, M. 1996. Finite-state automata on infinite inputs. In *the 6th National Seminar on Theorectical Computer Science*. Banasthali, Rajasthan, India.

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

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

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

Reid, D. 1979a. An algorithm for tracking multiple targets. *IEEE Trans. Automat. Control 24,* 6, 843–854.

Reid, D. B. 1979b. An Algorithm for Tracking Multiple Targets. *IEEE Trans. Automat. Control 24,* 6.

Roblee, C., Berk, V., and Cybenko, G. 2005. Large-scale autonomic server monitoring using process query systems. In *Proceedings of the IEEE International Conference on Autonomic Computing, Seattle, WA*.

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

Sheng, Y. August 2006. The theory of trackability and robustness for process detection. Ph.D. Thesis, Thayer School of Engineering, Dartmouth College.

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

Sheng, Y. and Cybenko, G. 2005b. Distance measures for nonparameteric weak process models. In *Proceedings of IEEE Systems, Man and Cybernetics Conference*, To appear.

Tsitsiklis, J. and Blondel, V. 1997. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate. *Math. Control, Signals, and Syst. 10*, 31–40. (Correction in 10, 381).

Tsitsiklis, J. and Blondel, V. 2000. A survey of computational complexity results in systems and control. *Automatica 36*, 1249–1274.

Wang, M.-H., You, Z.-S., and Peng, Y.-N. 2002. A novel tracking gate algorithm. In *Proceedings of the 4th World Congress on Intelligent Control and Automation*, Vol. 2, 2002.

WARRENDER, B. AND FORREST, S. 1999. Detecting intrusion using system calls: alternative data models. In *Proceedings of the IEEE Symposium on Security and Privacy*. Oakland, CA, 133–145.

WILLIAMS, S. G. 2004. *Symbolic Dynamics and its Applications*. American Mathematical Society, Providence, RI.

YANG, H. AND SIKDAR, B. 2003. A protocol for tracking mobile targets using sensor networks. In *Proceedings of the 1st IEEE Workshop on Sensor Network Protocols and Applications*. Anchorage, AL. 71–81.

YU, T. AND WU, Y. 2005. Decentralized multiple target tracking using netted collaborative autonomous trackers. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

ZHAO, F., SHIN, J., AND REICH, J. 2002. Information-driven dynamic sensor collaboration. *IEEE Signal Process. 19,* 2, 61–72.

Received July 2006; revised April 2007, July 2007; accepted August 2007