Inferring the size of stochastic systems from partial measurements

Alain Boldini[1[0000−0003−2231−7108]] and Maurizio Porfiri[2[0000−0002−1480−3539]]

[1] Center for Urban Science and Progress and Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn NY 11201, USA

[2] Center for Urban Science and Progress, Department of Mechanical and Aerospace Engineering, and Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn NY 11201, USA, mporfiri@nyu.edu

Abstract. Inferring the size of a complex system from partial measurements of some of its units is a common problem in engineering, with significant applications in the field of structural health monitoring (SHM), where one may attempt at relating system size (number of degrees of freedom) to the integrity of the structure. Here, we demonstrate the possibility of inferring the size of a stochastic system by assembling measurements of its response into a detection matrix. In deterministic systems, the rank of the detection matrix (number of non-zero singular values) equals the size of the largest observable system component. We extend this framework to reconstruct the number of states of an unknown Markov chain, where we cannot distinguish between two or more states. In this case, we only have access to an estimate of the detection matrix, but with a larger rank, since stochasticity generates a series of non-zero singular values. We establish conditions for the correct inference of system size, relating the number of realizations and the smallest true singular value. Our work highlights connections between SHM, system identification, and control theory, paving the way for new cross-disciplinary inquiries.

Keywords: Markov chains · Networked systems · Stochastic systems.

1 Introduction

Networked systems are formed by individual entities that interact with each other [2, 6]. Because of the powerful tools available, the framework of network theory has been adopted to study a wide variety of systems, from the brain [3] to the Internet [12], as well as social networks [17]. Further, network theory has recently been applied to several complex mechanical structures, such as trusses and lattices [1, 8, 10, 11, 15, 16].

Seldom we have access to the number of units forming the system. Typically, we only have access to some of the units of the system, as accessing all of them may be impossible or not practical. The inference of the number of units in a system has significant applications in the field of structural health monitoring.

In a network representation of a mechanical structure, the nodes of the network correspond to the degrees of freedom of the system. Should we detect a change in the size of the system, and in particular a decrease in the degrees of freedom of a structure, we could assess the integrity of the structure [4].

A potential approach to solve infer the network size is to utilize the output time-series of the free-response of the system for the assembly of a so-called detection matrix [9]. Upon loose assumptions, the rank of the detection matrix is equal to the size of the largest observable component of the network system [13]. The capabilities of the detection matrix in inferring the size of mechanical systems and failures in mechanical structures has been recently demonstrated by our group [4].

Previous efforts on the use of a detection matrix have mainly focused on deterministic systems, without theoretical insights on stochastic systems. Here, we address this gap in the literature by studying the inference of the number of states of an unknown Markov chain, where we cannot distinguish between two or more states. In this case, we only have access to an estimate of the detection matrix, with a larger rank. In fact, stochasticity generates a series of "noisy" non-zero singular values, which are reduced in number by the number of realizations used in the estimation of the detection matrix. Through numerical simulations on a simple example of the problem, we identify which are tenable necessary conditions for the correct inference of the number of states of the Markov chain and study how singular values scale with the number of realizations.

2 Theory

We consider a Markov process [14] X_k with unknown finite number N of states $s_1, \ldots, s_N$. The Markov property reads

$$\Pr\left(X_k = x_k | X_0 = x_0, \ldots, X_{k-1} = x_{k-1}\right) = \Pr\left(X_k = x_k | X_{k-1} = x_{k-1}\right), \quad (1)$$

for any realization $\{x_0, \ldots, x_k\}$ of the Markov chain.

Let us define the vector of probabilities $\boldsymbol{\pi}_k = \left[\Pr\left(X_k = s_1\right), \ldots, \Pr\left(X_k = s_N\right)\right]^{\mathrm{T}}$. Then, from Eq. (1), we can write

$$\boldsymbol{\pi}_{k+1} = \mathbf{P}^{\mathrm{T}} \boldsymbol{\pi}_k = \mathbf{A} \boldsymbol{\pi}_k, \quad (2)$$

where $\mathbf{P}$ is the so-called transition matrix [14], defined element-wise as

$$P_{ij} = \Pr\left(X_k = s_j | X_{k-1} = s_i\right), \quad (3)$$

for any time instant k. Note that each row of P_{ij} must sum to one.

We do not have direct access to the process X_n, but only to an "output" stochastic process Y_k with $M \leq N$ output states $\bar{s}_1, \ldots, \bar{s}_M$, which in general is not even a Markov process [5]. We assume that the value of realization y_k is deterministically determined by the value of realization x_k at the same step, through a non-invertible, surjective map $f : \{s_1, \ldots, s_N\} \rightarrow \{\bar{s}_1, \ldots, \bar{s}_M\}$. We define the output vector $\boldsymbol{\phi}_k = [\Pr(Y_k = \bar{s}_1), \ldots, \Pr(Y_k = \bar{s}_M)]^{\mathrm{T}}$, and we assume that it is related to the state vector $\boldsymbol{\pi}_k$ through

$$\boldsymbol{\phi}_k = \mathbf{C}\boldsymbol{\pi}_k. \tag{4}$$

We hypothesize that the matrix $\mathbf{C}$ is a Boolean $M \times N$ matrix that embodies the f map, in which every column contains a single "1" element. This condition corresponds to a situation in which we cannot distinguish between two or more states of the Markov process X_k, which are associated to the same state $\bar{s}_i$ of Y_k. In other words, if the i-th row of $\mathbf{C}$ contains more than one "1", the states of X_k associated with the columns containing the 1 elements are mapped to the same output state $\bar{s}_i$.

Our goal is to find the number N of states of the Markov process X_k, having access only to the output stochastic process Y_k. First, we consider the case in which we know $\boldsymbol{\phi}_k$ exactly. Due to the nature of Eqs. (2) and (4), this problem is equivalent to the one of finding the size of a time-discrete system networked system from partial measurements on a subset of the nodes. Similar to Refs. [4, 9, 13], we solve this problem through the definition of a detection matrix. We consider a family of l initial probability distributions $\boldsymbol{\pi}_0^{(l)}$, $l = 1, \ldots, L$ with $L > N$, which we collect within a matrix $\boldsymbol{\Pi} = \left[\boldsymbol{\pi}_0^{(1)}, \ldots, \boldsymbol{\pi}_0^{(L)}\right]$. We can assemble the detection matrix from the corresponding output vectors $\boldsymbol{\phi}_k^{(l)}$ for $k = 0, \ldots, K-1$ and $l = 1, \ldots, L$, as follows:

$$\mathbf{T}_{(K,L)} = \begin{bmatrix} \boldsymbol{\phi}_0^{(1)} & \boldsymbol{\phi}_0^{(2)} & \ldots & \boldsymbol{\phi}_0^{(L)} \\ \boldsymbol{\phi}_1^{(1)} & \boldsymbol{\phi}_1^{(2)} & \ldots & \boldsymbol{\phi}_1^{(L)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_{K-1}^{(1)} & \boldsymbol{\phi}_{K-1}^{(2)} & \ldots & \boldsymbol{\phi}_{K-1}^{(L)} \end{bmatrix}. \tag{5}$$

The detection matrix $\mathbf{T}_{(K,L)}$ has size $MK \times L$.

By extending the claims in [4], we can prove the following statement.

Theorem 1. *Consider the system in Eqs.* (2) *and* (4)*. If* $\boldsymbol{\Pi}$ *is full-row rank* ($\mathrm{rank}(\boldsymbol{\Pi}) = L$) *and* $K \geq N$*, then the detection matrix* $\mathbf{T}_{(K,L)}$ *has the same rank of the observability matrix*

$$\boldsymbol{\mathcal{O}}_N = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{N-1} \end{bmatrix}. \tag{6}$$

Hence, if the system is observable, we can infer its size by assembling the output vectors in a detection matrix and compute its rank. However, we normally do not have access to the output vectors $\boldsymbol{\phi}_k^{(l)}$, but only to an estimate $\hat{\boldsymbol{\phi}}_k^{(l)}$ from realizations of the Markov process, from which we assemble an estimate $\hat{\mathbf{T}}_{(K,L)}$ of the detection matrix. The estimate of the detection matrix $\hat{\mathbf{T}}_{(K,L)}$ is typically full-rank, whereby a series of non-zero noisy singular values is generated due to the stochastic nature of the system. To identify necessary conditions for inference and distinguish true from noisy singular values, we explore numerically how the noisy singular values vary with the number of realizations used for the estimation of the detection matrix.

3 Numerics

To guide the introduction of some necessary conditions for the inference of the number of the states of the Markov chain, we rely on numerical simulations on one of the simplest possible example. Specifically, we consider a Markov process with $N = 3$ states, with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0.1 & 0.7 & 0.2 \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}. \tag{7}$$

We assume that we can only distinguish $M = 2$ of the three states, such that the second and third states cannot be distinguished. Thus, we have

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \tag{8}$$

It is easy to show that the system is fully observable.

We consider $K = 10$ time instants and $L = 10$ random initial probability distributions, such that the detection matrix is 20×10. We study how the noisy singular values vary with the number of realizations used to estimate the detection matrix. To this end, we conduct a series of simulations in which we estimate the detection matrix $\hat{\mathbf{T}}_{(K,L)}$ from a number of realizations R spanning from 10 to $100,000$ for each of the L initial probability distributions.

In Fig. 1, we show the singular values of the true detection matrix $\mathbf{T}_{(K,L)}$ and of the estimate of the detection matrix $\hat{\mathbf{T}}_{(K,L)}$ for the different numbers of realizations used to estimate it. Clearly, the true detection matrix has three non-zero singular values: we can observe a sharp gap between the group of three non-zero singular value and the set of zero (up to numerical precision) singular values (Fig. 1a). The clear separation disappeares once we consider the estimates of the detection matrix (Fig. 1b-f). Regardless of the number of realizations, all the singular values are above 10^{-3}, such that a numerical evaluation of the rank of the estimated detection matrix would always overestimate the number of states of the Markov process. While for a small number of realizations it is impossible to distinguish the second and third singular values from the noisy ones, the gap between the first three non-zero singular values and the remaining ones increases when utilizing more realizations. This observation corroborates the idea that the noisy singular values decrease in value when increasing the number of realizations used to estimate the detection matrix.

To understand how these singular values scale with the number of realizations, in Fig. 2, we show the dependence of the third and fourth singular values

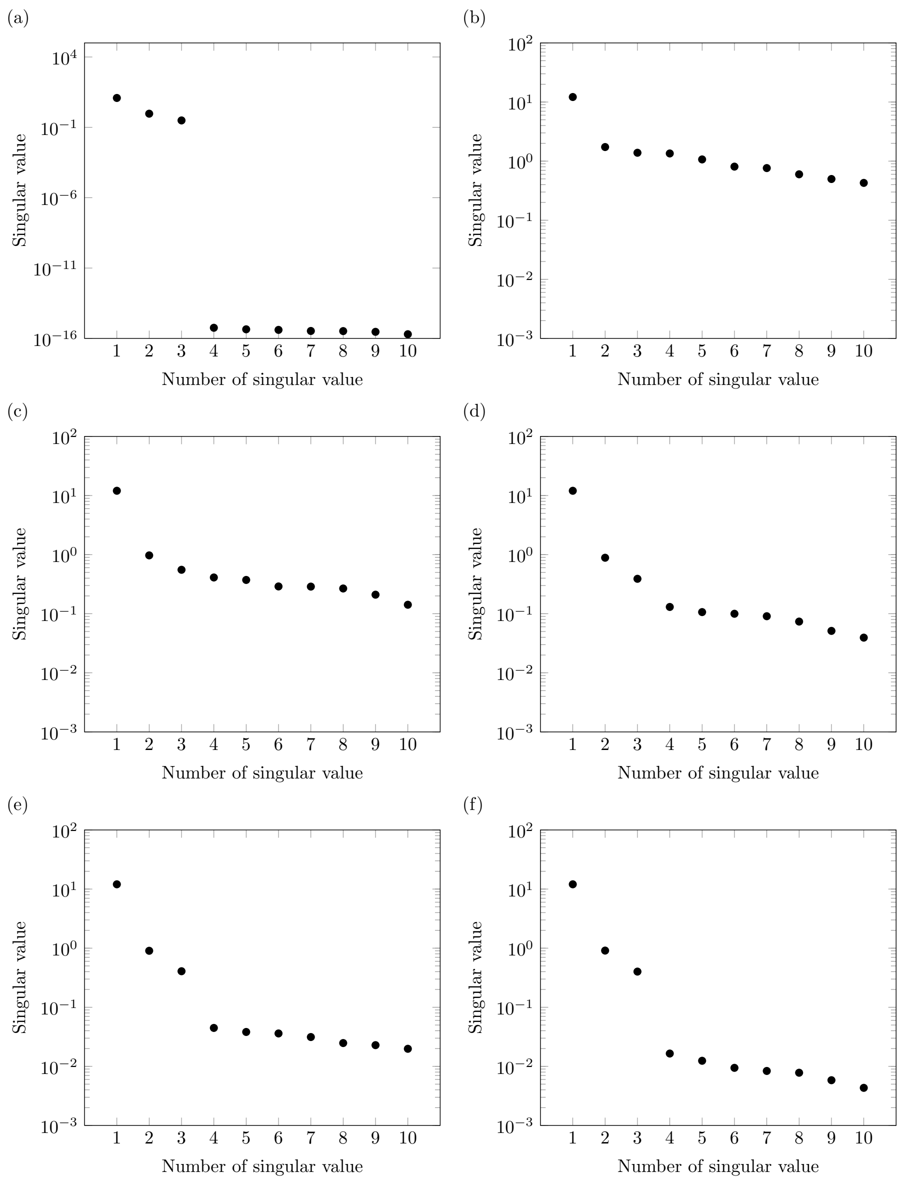

Fig. 1. Singular values of the true detection matrix $\mathbf{T}_{(K,L)}$ (a) and of the estimate of the detection matrix $\hat{\mathbf{T}}_{(K,L)}$ (b-f), for an increasing number of realizations $R = 10$ (b), $R = 100$ (c), $R = 1,000$ (d), $R = 10,000$ (e), and $R = 100,000$ (f).

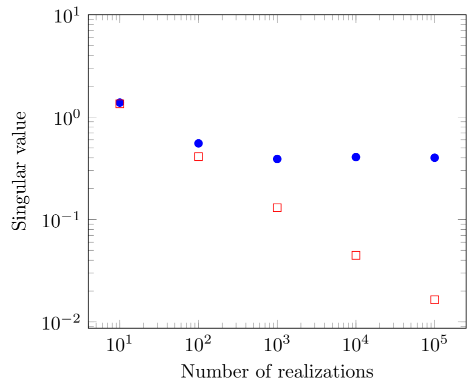

Fig. 2. Third (blue circles) and fourth (red squares) singular values of the estimate of the detection matrix $\hat{\mathbf{T}}_{(K,L)}$ as functions of the number of realizations used in the estimation.

of our system with the number of realizations used to estimate the detection matrix. For small numbers of realizations, we observe that these two singular values show a similar, almost indistinguishable trend. In fact, for a small number of realizations, the singular values are dominated by the noise associated with stochasticity, which decreases similarly for both singular values with the increase in the number of realizations. Continuing to increase the number of realizations, we observe that the true non-zero singular value stabilizes around a fixed value, while the noisy singular value continues to decrease. The largest noisy singular value scales with the inverse of the square root of the number of realizations, as confirmed by a linear fit ($R^2 = 0.999$, $p < 0.001$). This analysis provides us a condition for identifiability: the smallest non-zero singular value of the true detection matrix must be larger than the largest noisy singular value, which scales with the inverse of the square root of the number of realizations.

4 Conclusions

In this paper, we investigated how to infer the size of a stochastic system in the form of an unknown Markov process. We assume that we do not have direct access to the Markov process, but only to an output stochastic process in which we cannot distinguish between some of the states of the Markov process. This situation shares similarities with hidden Markov chains [5], although the problem of inferring the number of states has never been posed in this context. We seek to identify the number of states of the Markov process from the rank of a detection matrix assembled from the output measurements.

Upon assuming perfect knowledge of the output process, the rank of the detection matrix corresponds to the size of the largest observable component of the system. However, we normally have access only to an estimate of the detection matrix, which is full-rank due to the presence of non-zero noisy singular values associated with the stochastic nature of the estimate. Through numerical simulations, we show that the smallest true singular value must be larger than the largest noisy singular value for the inference to be possible. Further, we show that the largest noisy singular value scales with the inverse of the square root of the number of realizations used in the estimate of the detection matrix.

Our results pave the way for new inquiries to establish algorithms and statistical tests for the inference of the size of a stochastic system, based on estimates of the detection matrix. The algorithms may be based on our observation on the different scaling of true and noisy singular values. Statistical tests may be devised based on hard thresholding of singular values [7], based on estimates of the noisy singular values. Our work highlights important connections between SHM, system identification, and control theory.

Acknowledgments This work was supported by the National Science Foundation Grant No. CMMI-1932187 and CMMI-1953135, both awarded to MP.

References

1. Berthier, E., Porter, M.A., Daniels, K.E.: Forecasting failure locations in 2-dimensional disordered lattices. Proceedings of the National Academy of Sciences **116**(34), 16742–16749 (2019)
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U.: Complex networks: Structure and dynamics. Physics Reports **424**(4-5), 175–308 (2006)
3. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience **10**(3), 186–198 (2009)
4. Celli, P., Porfiri, M.: The detection matrix as a model-agnostic tool to estimate the number of degrees of freedom in mechanical systems and engineering structures. Chaos: An Interdisciplinary Journal of Nonlinear Science **32**(3), 033106 (2022)
5. Elliott, R.J., Aggoun, L., Moore, J.B.: Hidden Markov models: estimation and control, vol. 29. Springer Science & Business Media (2008)
6. Estrada, E., Knight, P.A.: A first course in network theory. Oxford University Press, USA (2015)
7. Gavish, M., Donoho, D.L.: The optimal hard threshold for singular values is $4/\sqrt{3}$. IEEE Transactions on Information Theory **60**(8), 5040–5053 (2014)
8. Giger, M., Ermanni, P.: Evolutionary truss topology optimization using a graph-based parameterization concept. Structural and Multidisciplinary Optimization **32**(4), 313–326 (2006)
9. Haehne, H., Casadiego, J., Peinke, J., Timme, M.: Detecting hidden units and network size from perceptible dynamics. Physical Review Letters **122**(15), 158301 (2019)
10. Kawamoto, A., Bendsøe, M.P., Sigmund, O.: Planar articulated mechanism design by graph theoretical enumeration. Structural and Multidisciplinary Optimization **27**(4), 295–299 (2004)
11. Kollmer, J.E., Daniels, K.E.: Betweenness centrality as predictor for forces in granular packings. Soft matter **15**(8), 1793–1798 (2019)
12. Pastor-Satorras, R., Vespignani, A.: Evolution and structure of the Internet: A statistical physics approach. Cambridge University Press (2004)

13. Porfiri, M.: Validity and limitations of the detection matrix to determine hidden units and network size from perceptible dynamics. Physical Review Letters **124**(16), 168301 (2020)
14. Privault, N.: Understanding Markov Chains - Examples and Applications. Springer-Verlag (2013)
15. Shai, O., Preiss, K.: Graph theory representations of engineering systems and their embedded knowledge. Artificial Intelligence in Engineering **13**(3), 273–285 (1999)
16. Smart, A., Ottino, J.M.: Granular matter and networks: three related examples. Soft Matter **4**(11), 2125–2131 (2008)
17. Wasserman, S., Faust, K., et al.: Social network analysis: Methods and applications. Cambridge University Press (1994)