

**Quantum system compression: A Hamiltonian guided walk through Hilbert space**Robert L. Kosut<sup>1,2</sup>, Tak-San Ho<sup>2</sup>, and Herschel Rabitz<sup>2</sup><sup>1</sup>*SC Solutions, Sunnyvale, California 94085, USA*<sup>2</sup>*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*

(Received 2 September 2020; accepted 15 December 2020; published 12 January 2021)

We present a systematic study of quantum system compression for the evolution of generic many-body problems. The necessary numerical simulations of such systems are seriously hindered by the exponential growth of the Hilbert-space dimension with the number of particles. For a *constant* Hamiltonian system of Hilbert-space dimension  $n$  with frequencies ranging from  $f_{\min}$  to  $f_{\max}$ , we show via a proper orthogonal decomposition that for a run time  $T$  the dominant dynamics are compressed in the neighborhood of a subspace whose dimension is the smallest integer larger than the time-bandwidth product  $\Delta = (f_{\max} - f_{\min})T$ . We also show how the distribution of initial states can further compress the system dimension. Under the stated conditions, the time-bandwidth estimate reveals the *existence* of an effective compressed model whose dimension is derived solely from system properties and not dependent on the particular implementation of a variational simulator, such as a machine learning system, or quantum device, or possibly even specially adapting traditional methods of solving the time-dependent Schrödinger equation. However, finding an efficient solution procedure *is* dependent on the simulator implementation, which is not discussed in this paper. In addition, we show that the compression rendered by the proper orthogonal decomposition encoding method can be further strengthened via a multilayer autoencoder. Finally, we present numerical illustrations to affirm the compression behavior in time-varying Hamiltonian dynamics in the presence of external fields. The essential time-bandwidth product is also simply estimated for a wide class of physical systems, where typically localized high-frequency motion occurs at or around each of the many particles, and with low-frequency dynamics associated with globally distributed characteristic degrees of freedom. This estimate for the bandwidth has a generic character indicating the wide significance of expected quantum system dynamics compression. We also discuss the potential implications of the findings for machine learning tools to efficiently solve the many-body or other high-dimensional Schrödinger equations.

DOI: [10.1103/PhysRevA.103.012406](https://doi.org/10.1103/PhysRevA.103.012406)**I. INTRODUCTION**

Numerous recent studies [1–6] utilizing a flexible representation of a variational quantum state have been proposed based on artificial neural networks (ANNs) for solving many-body quantum problems. These studies have shown a favorable polynomial scaling with respect to the system’s number of particles. Such findings are consistent across many fields [7–13]; despite system complexity in the underlying physics, much of observed behavior is compressed, i.e., the dominant dynamics is manifested in significantly lower dimensions. Compression arises also in the search over the quantum control landscape as a favorable scaling of control complexity [14,15].

By compression of quantum system dynamics we mean a simulator that has these two features: (1) the simulator has a reduced number of variables that do not scale exponentially with the number of quantum particles (or, the number of simulator variables is exponentially smaller than the Hilbert-space dimension), and (2) the state error, upon using the simulator, remains satisfactory for the intended purpose. In this context, we present a *time-bandwidth product* which reveals the existence of a compressed system which satisfies the

stated features. Naturally the level of compression in point 1 above is inversely related to the degree of dynamical error tolerated in point 2. The title of the paper can be understood, since the compression occurs in the system’s Hilbert space guided by particular characteristics of the Hamiltonian involved as shown in the main body of the paper. The key time-bandwidth product is reminiscent of “Hartley’s law” [16] referenced in [17] in relation to optimal control complexity. This potential compression of quantum dynamics depends only on the system properties, including the range of the eigenvalues of Hamiltonian, the simulation run time, and the initial state. Though the compression is not dependent on the particular method of simulation, achieving a similar level of compression is expected to be dependent on the simulator implementation. Additionally we show that the compression rendered by proper orthogonal decomposition can be further reduced via a multilayer autoencoder. We remark that the compressibility analysis in this paper will mainly be carried out for constant Hamiltonians as well as the autoencoder technique for estimating the reduced dimensionality. Numerical illustrations for time-varying Hamiltonian dynamics in external (control) fields will also be presented to support the pervasive nature of system compression behavior. Evidence

further supporting this result comes from the tested Hamiltonians ranging from many-body coupled spin systems to those chosen randomly.

The paper is written in a style of introducing concepts along with the associated mathematical formulation as well as clarifying numerical illustrations throughout the text to best express the various aspects of quantum system compression as they naturally arise. The paper is organized as follows: Sec. II introduces the variational state problem, while Sec. III describes the method of proper orthogonal decomposition for obtaining linear variational states. Section IV focuses on constant Hamiltonian dynamics with an example provided in Sec. V. Sections VI and VII present the main theoretical results followed by particular numerical simulation tests in Sec. VIII. The utilization of an autoencoder to enhance compression is presented in Sec. IX. Additional numerical illustrations for time-varying dynamics in external fields are presented in Sec. X, and we present extensions on compression for unitary dynamics and for nonlinear frequency sweeps in Sec. XI. A discussion of the findings in the paper is given in Sec. XII. As remarked earlier, the time-bandwidth product is crucial to determining the expected level of compression in executing quantum system dynamics. Although the body of the paper utilizes specific knowledge of various Hamiltonians to establish the bandwidth, a general argument is presented in Sec. XII for estimating the bandwidth even in highly complex many-body systems. Finally, details of particular derivations are given in the Appendices.

## II. VARIATIONAL STATE

The goal is to simulate the quantum state  $\psi_t \in \mathbb{C}^n$  for  $t \in [0, T]$  by a *variational quantum state*  $x(\theta_t) \in \mathbb{C}^n$ , a function of a time-varying parameter  $\theta_t \in \mathbb{C}^m$  where  $m < n$ . We refer to  $\theta_t$  as the *compressed state* and to  $m$  as *compression*, respectively. The effective “small volume in Hilbert space” referred to in [8] and exposed in [9] is akin to the compressed state discussed here. In a machine learning system, such as a neural network, the “parameters” are the weights that connect all the layers. These weighting parameters are not necessarily the same as the compressed dimension or the associated state variables, though it is possible depending on the simulator implementation. In this note we leave the simulator parametrization and implementation unspecified and seek to show that generally there are fewer effective compressed states ( $m$ ) than that of the full Hilbert space ( $n$ ). Irrespective of the implementation, it is assumed that the variational state  $x(\theta_t)$  is a smooth function of the time-varying parameter vector  $\theta_t$ , and, as a result,

$$\dot{x}(\theta_t) = G(\theta_t)\dot{\theta}_t, \quad G(\theta) = \nabla_{\theta}x(\theta) \in \mathbb{C}^{n \times m}, \quad (1)$$

where  $G(\theta_t)$  as indicated is a gradient matrix.

### A. Variational formulations

In the most general case the system to be simulated is the time-varying quantum system:

$$i\dot{\psi}_t = H_t\psi_t, \quad t \in [0, T]. \quad (2)$$

The data available for simulation are the initial state  $\psi_0 \in \mathbb{C}^n$  and the time-varying Hamiltonian  $H_t \in \mathbb{C}^{n \times n}$ , a description that encompasses a quantum system under a known control or external field. The variational optimization problem that we *seek to solve* is to minimize the following *functional*:

$$E[\theta] = \frac{1}{T} \int_0^T \|\psi_t - x(\theta_t)\|_2^2 dt. \quad (3)$$

The qualifying phrase “seek to solve” is stated because the actual state flow  $\{\psi_t, t \in [0, T]\}$  is not available in the simulation context. If it were available then there would be no need for a variational version unless one seeks to *post facto* find a low-dimensional representation. In general, minimizing  $E[\theta]$  is not possible without the actual state. As in [1] for variational simulation, there is a means to query as necessary the available data: the initial state  $\psi_0$  and the time-varying Hamiltonian  $\{H_t, t \in [0, T]\}$ . The variational problem is then pragmatically posed to minimize a functional such as

$$\mathcal{E}[\theta] = \frac{1}{T} \int_0^T \|H_t x(\theta_t) - iG(\theta_t)\dot{\theta}_t\|_2^2 dt. \quad (4)$$

Depending on the context we will refer to  $E[\theta]$  or its integrand as the *state error*, and to  $\mathcal{E}[\theta]$  or its integrand as the *equation error*. To construct a variational quantum simulator in a classical device using only the available data (e.g.,  $\psi_0 \in \mathbb{C}^n$ ,  $\{H_t \in \mathbb{C}^{n \times n}, t \in [0, T]\}$ ) is equivalent to minimizing an equation error functional such as (4). This is clearly a necessary surrogate for the ideal goal: minimizing the state error (3).

### B. Compression

If the variational solution is perfect then  $x(\theta_t) = \psi_t, \forall t \in [0, T]$  from which it follows that both state and equation errors (and associated functionals) are zero. As a consequence the variational parameter  $\theta_t$  must satisfy

$$G(\theta_t)\dot{\theta}_t + iH_t\psi_t = 0, \quad x(\theta_0) = \psi_0, \quad t \in [0, T]. \quad (5)$$

These equations for finding  $\dot{\theta}_t \in \mathbb{C}^m$  with  $m < n$  are overdetermined, i.e., more equations ( $n$ ) than variables ( $m$ ). As a result, (5) will hold only if  $H_t\psi_t$  (or  $i\dot{\psi}_t$ ) is a linear combination of the columns of  $G(\theta_t)$ . In particular, a solution for  $\dot{\theta}_t$  to solve (5) will exist if and only if for  $t \in [0, T]$

$$\begin{aligned} [I_m - \Gamma_m(\theta_t)]H_t\psi_t &= 0, \\ \Gamma_m(\theta_t) &= G(\theta_t)G^\#(\theta_t) \in \mathbb{C}^{m \times m}, \end{aligned} \quad (6)$$

where  $G^\#(\theta_t) \in \mathbb{C}^{m \times n}$  is the *pseudoinverse* of  $G(\theta_t)$ . Note that properties of the pseudoinverse yield  $\Gamma_m(\theta_t)^2 = \Gamma_m(\theta_t)$ , i.e., it is idempotent. Condition (6) is necessary and sufficient for the variational state to provide a perfect simulation, i.e.,  $x(\theta_t)$  and true state  $\psi_t$  are identical.

For a perfect simulation, and clearly for an imperfect but very good simulation, especially with  $m \ll n$ , or perhaps more importantly where  $m$  does not scale exponentially with the number of particles as does  $n$ , the state flow  $\psi_t$  from (2) *must be inherently compressible*. To understand the basic foundations of quantum system compression for now we focus on a *time-invariant* quantum system, meaning that  $H_t = H$  is a constant matrix; the time-dependent extension will be returned to in Sec. X. In addition we restrict attention to a *linear* variational state, that is, the variational state gradient

is a constant matrix:  $G(\theta_t) = M \in \mathbb{C}^{n \times m}$ ,  $m < n$ . The introduction of nonlinearity into the compression process will be treated with an autoencoder in Sec. IX. Assuming that the eigenvalues sweep linearly from their lowest to highest values, we show that it is the system properties alone which allow the state and equation error measures to show a favorable scaling of compression  $m$  with system dimension  $n$ . Our analysis provides a means to assess the compression range that might be achievable from a more generally flexible variational state parametrization, i.e., with a time-varying nonlinear gradient  $G(\theta_t)$ . As in numerous studies showing favorable compression, we are led to the assertion that a variational quantum solution of a time-invariant quantum system is possible with its compression dependent solely on the system parameters: Hamiltonian eigenvalues, initial state, and simulation run time.

However, the demonstrated *existence* of system compression shown in this paper still leaves the challenge to explicitly exploit that for the performance of dynamical simulations (i.e., especially for high-dimensional many-body situations) based on (5), for example. A nascent illustration appears evident in a machine learning format utilizing artificial neural networks [1–6]. The generality of these results to wide classes of many-body quantum systems remains to be explored. Furthermore, the extension of this compression to quantum computers and especially quantum supremacy remains open for speculation. Quantum supremacy is also likely application dependent. But, in the case of simulating quantum dynamics the existence of the compression appears to automatically be discovered by utilizing the nonlinear transformation constituting ANNs, and one may speculate that for analogous quantum computer simulations of dynamics to be efficient they would need to inherently also exploit the compression.

### III. LINEAR VARIATIONAL STATE

A common means of obtaining a *linear* variational state is the method of *proper orthogonal decomposition* (POD) or equivalently often referred to as *principal component analysis* (PCA) [18]. There are several variants of POD; here we use an unbiased version where the variational state is set to  $x(\theta_t) = M\theta_t$  with gradient matrix  $M \in \mathbb{C}^{n \times m}$  where  $m < n$  [19]. The optimization variables  $M$  and  $\theta_t$  are selected to minimize a state error functional chosen as (3):

$$\begin{aligned} \text{minimize } E_{\text{pod}} &= \frac{1}{T} \int_0^T \|\psi_t - M\theta_t\|_2^2 dt, \\ \text{subject to } M &\in \mathbb{C}^{n \times m}, \theta_t \in \mathbb{C}^m, t \in [0, T]. \end{aligned} \quad (7)$$

The well known optimal solution is

$$\theta_t = M_m^\dagger \psi_t, \quad M_m = \arg \min_{M \in \mathbb{C}^{n \times m}} \text{Tr}(I_n - MM^\dagger)C[\psi], \quad (8)$$

where  $M_m \in \mathbb{C}^{n \times m}$  is found via a singular value decomposition (SVD) of the  $n \times n$  positive semidefinite ‘‘covariance’’ matrix,

$$\begin{aligned} C[\psi] &= \frac{1}{T} \int_0^T \psi_t \psi_t^\dagger dt \\ &= [M_m \quad M_{n-m}] \begin{bmatrix} Q_m & 0 \\ 0 & Q_{n-m} \end{bmatrix} [M_m \quad M_{n-m}]^\dagger, \end{aligned} \quad (9)$$

with singular values arranged (as usual) in descending order in  $Q_m = \text{diag}(\sigma_1(C), \dots, \sigma_m(C))$  and  $Q_{n-m} = \text{diag}(\sigma_{m+1}(C), \dots, \sigma_n(C))$  and where  $[M_m \quad M_{n-m}] \in \mathbf{U}(n)$ . The resulting optimal (POD) variational state is

$$x_t = \Gamma_m \psi_t, \quad \Gamma_m = M_m M_m^\dagger \in \mathbb{C}^{n \times n}, \quad (10)$$

which produces the minimum state POD error measure: the sum of the smallest  $n - m$  singular values of the covariance matrix, i.e.,

$$E_{\text{pod}} = \text{Tr}(I_n - \Gamma_m)C[\psi] = \sum_{i=m+1}^n \sigma_i(C). \quad (11)$$

Because the matrix  $M_m$  is part of a unitary matrix, its  $m$  columns are orthonormal vectors in  $\mathbb{C}^n$  and thus  $M_m^\dagger M_m = I_m$ . The columns of  $M_m$  form a basis set for the compressed state  $\theta_t$  making  $\Gamma_m$  idempotent, i.e.,  $\Gamma_m^2 = \Gamma_m$ . Combining (10) with (4) we get the corresponding POD equation error measure:

$$\mathcal{E}_{\text{pod}} = \frac{1}{T} \int_0^T \|(H_t \Gamma_m - \Gamma_m H_t)\psi_t\|_2^2 dt. \quad (12)$$

A typical measure of model reduction error is the relative error in the cumulative sum of singular values of the covariance matrix compared to the sum of all the singular values:

$$\epsilon_m(C) = 1 - \sum_{i=1}^m \sigma_i(C) / \sum_{i=1}^n \sigma_i(C) = \sum_{i=m+1}^n \sigma_i(C). \quad (13)$$

To predict compression quantitatively for a quantum system, we make an obvious restriction as discussed next which leads to the time-bandwidth product presented in Sec. VII. Later in Sec. X we will discuss the compression effect with a time-varying Hamiltonian and suggest how a compression estimate can be obtained.

### IV. TIME-INVARIANT HAMILTONIAN DYNAMICS

Consider a *time-invariant* Hamiltonian driving the quantum dynamical system:

$$i\dot{\psi}_t = H\psi_t, \quad t \in [0, T]. \quad (14)$$

Since the Hamiltonian  $H \in \mathbb{C}^{n \times n}$  is constant, the standard eigenvalue decomposition yields

$$H = V\Omega V^\dagger \quad \begin{cases} V \in \mathbf{U}(n), \\ \Omega = \text{diag}(\omega), \quad \omega \in \mathbb{R}^n. \end{cases} \quad (15)$$

Subsequently the state flow  $\psi_t$  can be expressed variously as

$$\begin{aligned} \psi_t &= V e^{-it\Omega} V^\dagger \psi_0 = V f_t, \\ f_t &= [\alpha_1 e^{-i\omega_1 t} \quad \dots \quad \alpha_n e^{-i\omega_n t}]^T, \\ \alpha &= V^\dagger \psi_0, \end{aligned} \quad (16)$$

where  $\alpha \in \mathbb{C}^n$ ,  $\|\alpha\|_2 = 1$  is the projection of the initial state onto the natural basis of the Hamiltonian. As shown in Appendix A, the singular values of the POD covariance matrix (9) are identical with those of a positive semidefinite matrix  $R \in \mathbb{C}^{n \times n}$  which has the decomposition

$$\begin{aligned} R &= \text{diag}(\alpha) S \text{diag}(\alpha)^\dagger, \\ S &= \text{sinc}(\Lambda), \\ \Lambda_{k\ell} &= (\omega_k - \omega_\ell)T/2, \quad k, \ell = 1, \dots, n. \end{aligned} \quad (17)$$

We will refer to  $R$  as a ‘‘covariance’’ matrix and to  $S$  as the ‘‘sinc’’ matrix with respective elements

$$R_{k\ell} = \begin{cases} |\alpha_k|^2, & k = \ell, \\ (\alpha_k \alpha_\ell^*) \text{sinc} \Lambda_{k\ell}, & k \neq \ell, \end{cases}$$

$$S_{k\ell} = \begin{cases} 1, & k = \ell, \\ \text{sinc} \Lambda_{k\ell}, & k \neq \ell. \end{cases} \quad (18)$$

The corresponding normalized singular value errors in the form of (13) for  $m = 1, \dots, n$  are

$$\epsilon_m(R) = 1 - \sum_{i=1}^m \sigma_i(R),$$

$$\epsilon_m(S) = 1 - \sum_{i=1}^m \sigma_i(S)/n, \quad (19)$$

where the cumulative sum of singular value errors is normalized by their respective singular value sums:  $\text{Tr} R = \|\alpha\|_2^2 = \|\psi_0\|_2^2 = 1$  and  $\text{Tr} S = n$ . Though the calculation of the elements of  $R$  and  $S$  do *not* require any dynamical simulation, to use these expressions necessitates obtaining an eigenvalue decomposition of the system Hamiltonian, knowledge of the initial state, and calculating the singular values of  $R$  and  $S$ . We can, however, without doing any such decompositions, assert some generic properties of compression. Before discussing these, it is worthwhile to do an example.

### V. EXAMPLE: TIME-INVARIANT HAMILTONIAN SPIN SYSTEM

Consider the transverse-field Ising (TFI)  $n_{\text{spin}}$  system with Hamiltonian

$$H_{\text{TFI}} = -h \sum_{i=1}^{n_{\text{spin}}} \sigma_i^x - \sum_{i=1}^{n_{\text{spin}}-1} \sigma_i^z \sigma_{i+1}^z. \quad (20)$$

This system was studied in [1]; here we use the same model parameters and initial-state choices but only for a limited number of spins. Figure 1(a) shows the eigenvalues of  $H_{\text{TFI}}$  with TFI parameter  $h = 2$  for  $n_{\text{spin}} \in \{10, 12\}$ , ergo, space dimension  $n = 2^{n_{\text{spin}}} \in \{1024, 4096\}$ . The eigenvalues clearly sweep uniformly from maximum to minimum and sum to zero, thus forming two groups of  $n/2$  positive eigenvalues and  $n/2$  negative eigenvalues with each group having the same magnitudes:  $\omega \in \{-\omega_1, \dots, -\omega_{n/2}, \omega_1, \dots, \omega_{n/2}\}$ .

Figures 1(b) and 1(c) correspond to the two spin examples:  $n_{\text{spin}} \in \{10, 12\}$  each with run time  $T = 2$ . In each of the (b) and (c) plots, the two subplots on the right show, respectively, the root-mean-square (rms) values over  $t \in [0, T]$  of both state error and equation error, i.e.,  $\|\psi - x\|_{\text{rms}}$  and  $\|Hx - i\dot{x}\|_{\text{rms}}$ . [These are the square roots, respectively, of  $E_{\text{pod}}$  and  $\mathcal{E}_{\text{pod}}$  from (11) and (12).] The left subplots of (b) and (c) show the relative error in the cumulative sum of singular values of the covariance matrix  $R$  (blue dots) and the sinc matrix  $S$  (red dots), both calculated from (18). For comparison with the rms error measures we use the square root of the cumulative singular value errors of  $R$  and  $S$  from (19), namely,  $\sqrt{\epsilon_m(R)}$  and  $\sqrt{\epsilon_m(S)}$ . The error measures are all computed for  $H_{\text{TFI}}$  with  $h = 2$ , and for each of two initial states: one fixed at the ground state of  $H_{\text{TFI}}$  with  $h = 4$  (blue

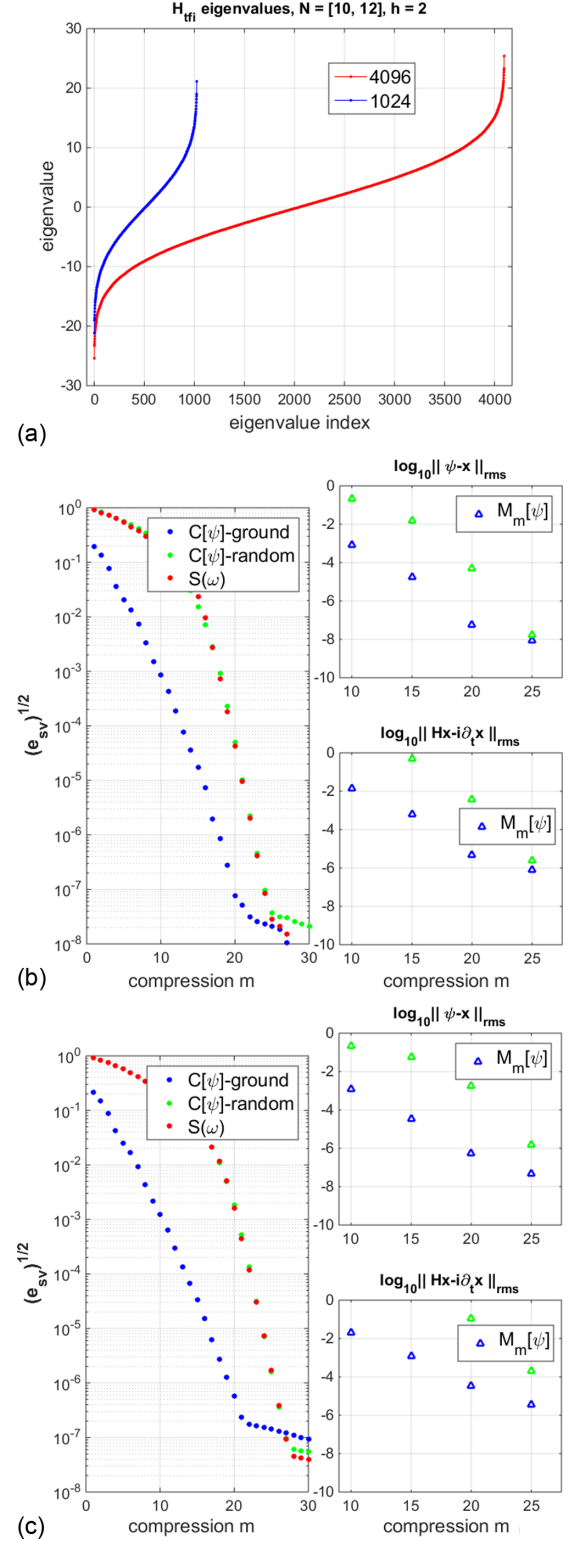


FIG. 1. TFI. All runs with  $h = 2$  (20); run time  $T = 2, 10$ , and 12 spins [eigenvalues in (a)]; and two initial states: one at the ground state of  $H_{\text{TFI}}$  with  $h = 4$  [blue dots and triangles in (b) and (c)] and one randomly selected [green dots and triangles in (b) and (c)]. The left subplots in (b) and (c) show the relative rms error [square root of (19)] of the covariance matrix  $R$  (blue dots) and the sinc matrix  $S$  (red dots) from (18) for compressions  $m = 1, \dots, 30$ . Lower right subplots in (b) and (c) show errors for  $m \in \{10, 15, 20, 25\}$ .

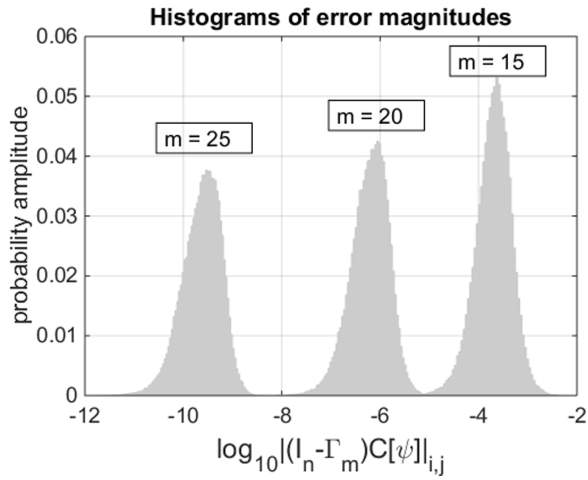


FIG. 2. Histograms of state errors. The log of the absolute values of all elements of state error matrix  $(I_n - \Gamma_m)C[\psi]$  from (11) for  $n = 2^{10}$  and for the selected compressions shown.

dots), and one with the initial state randomly selected (green dots). (Only results for the ground-state initialization were presented in [1].) As seen in Fig. 1 the selection of a random initial state (green dots in the singular value plots and green triangles in the error measure plots) increases the variational compression to achieve the same relative model error, though the increases are not dramatic. The sharp slope changes in the rms errors near full compression are the result of reaching the computational precision limit.

Figure. 2 displays the TFI data in histograms of  $\log_{10}$  of the absolute values of all elements of state error matrix  $(I_n - \Gamma_m)C[\psi]$  from (11) for  $n = 2^{10}$  and for the selected compressions shown. The decrease in error magnitudes is quite dramatic for compressions greater than 15. This value

is effectively predicted by the time-bandwidth product, i.e.,  $\Delta = (f_{\max} - f_{\min})T \approx 14$ .

Clearly there is significant dynamic compression: not only is the level small compared to the system dimension, moreover, there is little increase in compression with increasing number of spins. In both cases shown a compression of order  $m = 25$  suffices to produce very small state and equation errors.

The compression order selections ( $m$ ) are based on the initial-state and system dynamics over the specified simulation time interval  $t \in [0, T]$  with  $T = 2$ . Figure 3(a) shows what happens when the run time is extended to  $T = 4$  for compressions  $m \in \{14, 21\}$ . As expected, outside of the design range of  $t \in [0, 2]$  the state rms error as a function of time dramatically increases. In Fig. 3(b) the initial state differs from that used in the SVD of the covariance matrix, and as a result the state rms error deteriorates over the whole time interval. Two levels of initial-state perturbation are shown:  $\Delta\psi_0 \in \{0.01, 0.001\}$ . For  $m = 14$  there is no significant change whereas for  $m = 21$  the change is considerable, more than a two-order-of-magnitude increase in error. The difference in robustness to initial state may be attributed to the nominal error magnitude, i.e., the state error for  $m = 14$  is much larger than that for  $m = 21$ , and hence the latter is more sensitive to changes in the initial state.

### VI. RUN TIME AND SINGULAR VALUE BOUNDS

In the previous examples the relative singular value error of  $S$ , the sinc matrix, provides an upper bound on the errors of the covariance matrix  $R$ , and these errors are close when the initial state is random. This suggests that the efficacy of a variational quantum state of a time-invariant quantum system (14) can be obtained from the singular values of  $R$  (equivalently  $C$ ) and

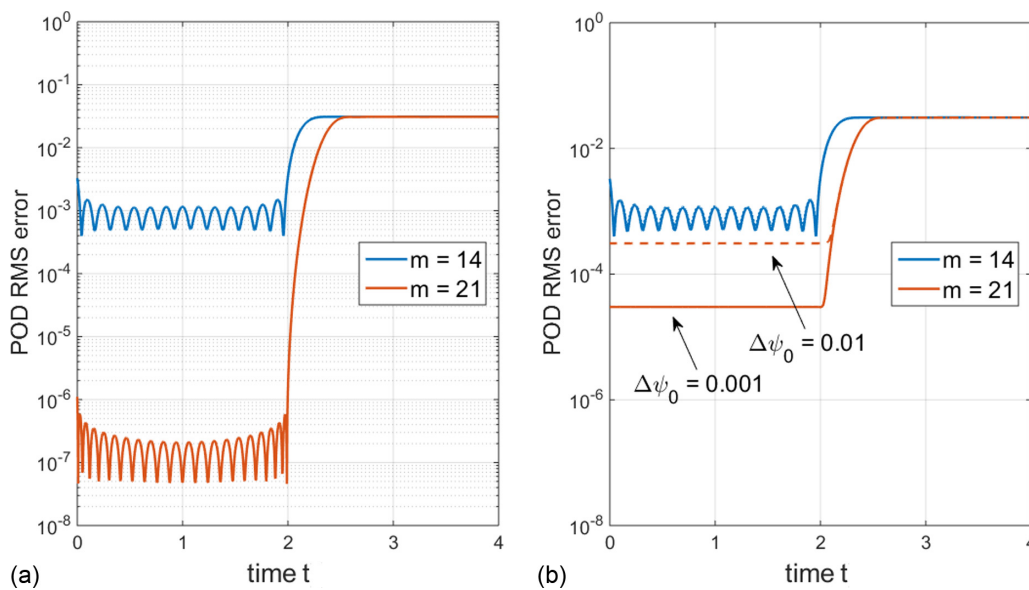


FIG. 3. Robustness of rms magnitude of POD error time responses. For the ten spin TFI system (20), (a) change to run time  $T = 4$  from variational states optimized for run time  $T = 2$  and (b) small perturbations to initial state.

the sinc matrix  $S$  (17) and (18). To see this we first examine some qualitative properties at the extremes of run time  $T$ .

### A. Run time

In the limit as the run time  $T$  goes to infinity the sinc matrix approaches the identity; hence,

$$\lim_{T \rightarrow \infty} \sigma(R) = \sigma[\text{diag}(|\alpha|^2)]. \quad (21)$$

The singular values of  $R$  become the sorted magnitudes of  $\alpha$ , the initial state expressed via the Hamiltonian eigenvectors. These are the diagonal elements of  $R$ , and all the other elements tend to zero. Compression to  $m < n$  in this case requires that the last  $n - m$  elements of  $\alpha$  are much smaller than the first  $m$  elements. At the opposite end, as the run time  $T$  goes to zero,  $R \rightarrow \alpha\alpha^\dagger$ ; thus,

$$\lim_{T \rightarrow 0} \sigma(R) = \left\{ \sum_{k=1}^n |\alpha_k|^2, 0, \dots, 0 \right\} = \{1, 0, \dots, 0\}. \quad (22)$$

Note that there is only one nonzero singular value at  $\|\alpha\|_2 = \|\psi_0\|_2 = 1$ . A reduction to a model with dimension 1 is certainly extreme, but expected. The message to take here is that the dynamics becomes further compressed as run time decreases. In effect, for small run times not very much of the space gets filled out beyond where the state started.

The extreme run-time scenarios reflect the fact that since  $\alpha = V^\dagger \psi_0$ , compression depends on how the initial state is projected onto the eigenvectors of the Hamiltonian. In the previous examples (Fig. 1) when the initial state is prepared nearby the ground state, the error measure for  $R$  is significantly smaller than for  $S$  because the initial state is localized in the Hilbert space. For a random initial state we see that the singular value errors for  $R$  and  $S$  are close since the initial state now is more spread out.

### B. Singular value bounds

More quantitative insights can be revealed using a standard singular value inequality for the product of matrices. From the relation of  $R$  and  $S$  (17), the singular values of  $R$  for each  $k = 1, \dots, n$  are bounded by

$$\sigma_k(R) \leq \min \{ \|\alpha\|_\infty^2 \sigma_k(S), \|S\| \sigma_k^2[\text{diag}(\alpha)] \}. \quad (23)$$

Another well-known inequality follows, namely,

$$\text{rank } R \leq \min \{ \text{rank } S, \text{rank } \text{diag}(\alpha) \}. \quad (24)$$

This rank inequality is only useful if  $S$  has some zero singular values and/or  $\alpha$  has some zero elements. The latter can occur when the initial state lies completely in a lower-dimensional subspace, e.g., [20]. Though this is generally unlikely, as observed in the previous spin system examples, compression is possible because many singular values of  $R$  are nearly zero, possibly driven by the localized initial state lying dominantly in a low-dimensional subspace. For example, suppose that no elements of  $\alpha$  are zero and that  $\sigma_k(S) \approx 0$  for  $k > m$  with  $m \ll n$ . Then  $\sigma_k(R) \approx 0$  for  $k > m$ , and thus  $m$  is the maximum variational compression. In this case the compression order is bounded by the number of nonzero singular (or non-small) singular values of  $S$ , the sinc matrix. This property is

clearly seen in the numerical results of Fig. 1 with the TFI example. We see this in many other cases that we have run: the sinc-matrix error bounds the covariance matrix error, i.e.,  $\epsilon_m(R) \leq \epsilon_m(S)$ , or, equivalently,

$$\sum_{i=1}^m \sigma_i(S)/n \leq \sum_{i=1}^m \sigma_i(R). \quad (25)$$

Both sides of this inequality are norms, specifically Ky Fan  $m$  norms, respectively, of  $S/n$  and  $R$ . From the *Ky Fan dominance theorem* [21] the above will hold for all  $m = 1, \dots, n$  if and only if  $\|S/n\| \leq \|R\|$  for any norm invariant unitary transformation. Only in very few cases have we seen this behavior violated and then only slightly with no particular pattern observed. Nevertheless, at the moment (25) remains a sufficient condition for a general frequency sweep. For a linear sweep approximation we can make stronger statements.

## VII. LINEAR FREQUENCY SWEEP

In many quantum systems, such as just observed for the spin system example, the eigenvalues of the Hamiltonian are almost linear from minimum to maximum, or reasonably approximated as such over a large portion of the range (see Fig. 1). Under the assumption of a linear sweep of eigenvalues, the sinc matrix  $S$  (17) becomes a symmetric (real) *Toeplitz* matrix:

$$\begin{aligned} S_{\text{lin}} &= \text{sinc} \Lambda^{\text{lin}} \in \mathbb{R}^{n \times n}, \\ \Lambda_{kl}^{\text{lin}} &= (\omega_k - \omega_\ell) T / 2 = \left( \frac{k - \ell}{n - 1} \right) \Delta \pi, \quad k, \ell = 1, \dots, n, \\ \Delta &= (\omega_{\text{max}} - \omega_{\text{min}}) T / 2\pi. \end{aligned} \quad (26)$$

The nondimensional variable  $\Delta$ , referred to here as the “time-bandwidth product,” will be seen to play the key role in establishing an approximate upper bound on the variational model order. A related measure was used in [22] in the development of a coarse-grained approach to unitary dynamics.

The covariance matrix corresponding to  $S_{\text{lin}}$  follows from the form of (17):

$$R_{\text{lin}} = \text{diag}(\alpha) S_{\text{lin}} \text{diag}(\alpha)^\dagger, \quad (27)$$

in which  $R_{\text{lin}}$  is a Hermitian matrix but not generally of Toeplitz form. The spin system results in Fig. 4 are of sufficiently small size so that the singular values of the  $R$ ,  $S$ ,  $S_{\text{lin}}$ , and  $R_{\text{lin}}$  matrices can be directly calculated on a standard laptop. For large dimensions, this becomes infeasible. Fortunately, even for very large  $n$ , the singular values of  $S_{\text{lin}}$  can be approximated by taking the digital Fourier transform of a column of a related *circulant* matrix, for which there are a number of versions, all resulting in asymptotic approximations of the eigenvalues [23,24]. A similar procedure was utilized in [25]. In the case here, with a very specific (sinc) function forming the symmetric Toeplitz matrix elements, we can appeal to a more direct result in [26,27] on the asymptotic distribution of the Toeplitz eigenvalues. As shown in Appendix B, for large  $n$  the singular values of the Toeplitz sinc matrix  $\sigma_i(S_{\text{lin}})$ ,  $i = 1, \dots, n$  are well approximated by

$$\sigma_i(S_{\text{lin}}) \approx \begin{cases} (n - 1) / \Delta, & i < m_{\text{TBW}}, \\ 0, & i \geq m_{\text{TBW}}, \end{cases} \quad (28)$$

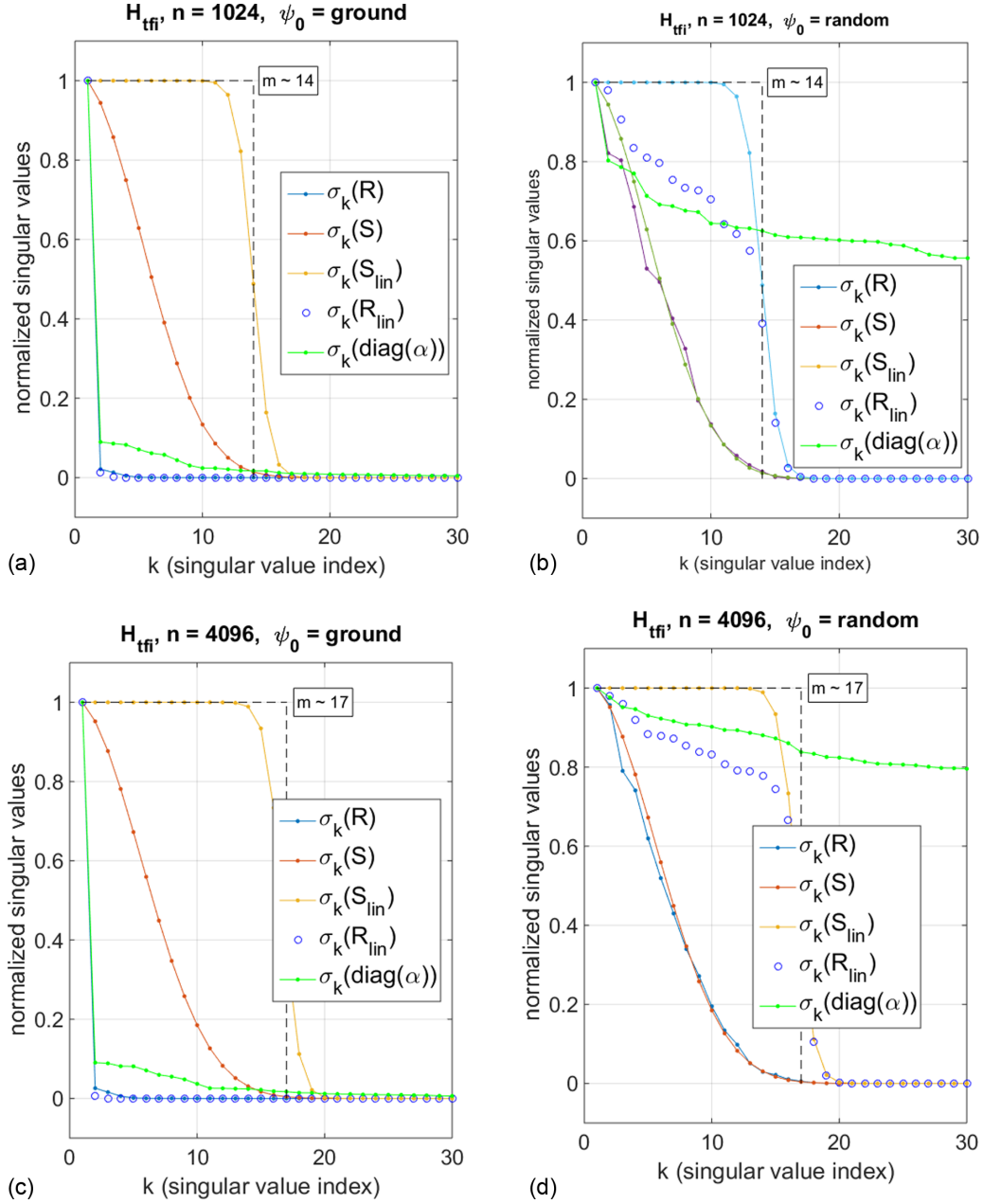


FIG. 4. Normalized singular values of  $R$ ,  $S$ ,  $S_{\text{lin}}$ ,  $R_{\text{lin}}$ , and  $\text{diag}(\alpha)$  for  $H_{\text{TFI}}$  (20) with  $h = 2$  for 10 and 12 spins and with initial states at ground for  $h = 4$  (a, c) and random (b, d). The rect function (dashed lines) is  $\bar{S}_{\text{lin}}$  from (31) with associated singular values as indicated by the text boxes  $m_{\text{TBW}} = 14$  and 17.

where the index value  $m_{\text{TBW}}$ , under the conditions of a linear frequency sweep, is the level of compression:

$$m_{\text{TBW}} = \left\lceil \frac{n}{n-1} \Delta \right\rceil, \quad \Delta = (f_{\text{max}} - f_{\text{min}})T. \quad (29)$$

The frequency spread is expressed here in Hz using  $f = 2\pi\omega$ . The corresponding covariance error (19) is

$$\begin{aligned} \epsilon_k(S_{\text{lin}}) &= 1 - \sum_{i=1}^k \sigma_i(S_{\text{lin}}) / \sum_{i=1}^n \sigma_i(S_{\text{lin}}) \\ &\approx \begin{cases} 1 - k/m & k < m_{\text{TBW}}, \\ 0 & k \geq m_{\text{TBW}}. \end{cases} \end{aligned} \quad (30)$$

Note that one could multiply the constant Hamiltonian  $H$  by  $a$  and divide the run time  $T$  by  $a$  and get the same value for  $m_{\text{TBW}}$ . This is a trivial scaling for a constant Hamiltonian; however, as we will see later (Sec. X), this simple scaling does not apply for a time-varying Hamiltonian where  $m_{\text{TBW}}$  is interpreted differently to account for the observed compression.

At the compression level  $m_{\text{TBW}}$ , the variational state errors are either very small or rapidly decreasing for  $m > m_{\text{TBW}}$ . For quantum systems with  $n \gg 1$  we can take  $m_{\text{TBW}} = \lceil \Delta \rceil$ . Let  $\bar{S}_{\text{lin}}$  denote the ideal sinc matrix with exactly  $m_{\text{TBW}} = \lceil \Delta \rceil$  nonzero constant singular values:

$$\sigma_k(\bar{S}_{\text{lin}}) = \begin{cases} n/\Delta, & k < \lceil \Delta \rceil, \\ 0, & k \geq \lceil \Delta \rceil. \end{cases} \quad (31)$$

Let  $\bar{R}_{\text{lin}}$  denote the corresponding covariance matrix:

$$\bar{R}_{\text{lin}} = \text{diag}(\alpha) \bar{S}_{\text{lin}} \text{diag}(\alpha)^\dagger. \quad (32)$$

Application of the singular value bound (23) results in

$$\sigma_k(\bar{R}_{\text{lin}}) \leq \begin{cases} (n/\Delta) \sigma_k^2[\text{diag}(\alpha)], & k < \lceil \Delta \rceil, \\ 0, & k \geq \lceil \Delta \rceil. \end{cases} \quad (33)$$

This result ensures that for a linear frequency sweep and large  $n$  the Ky Fan dominance theorem holds for the ideal pair  $(\bar{R}_{\text{lin}}, \bar{S}_{\text{lin}})$ , namely,  $\sigma_k(\bar{S}_{\text{lin}}/n) \leq \sigma_k(\bar{R}_{\text{lin}})$ , because  $\|\alpha\|_\infty^2 \geq 1/n$  since  $\|\alpha\|_2 = 1$ . Thus, for a linear sweep, the ideal relative covariance error is bounded by the relative sinc error. Further, if for all elements  $|\alpha_i|^2 = 1/n$ ,  $i = 1, \dots, n$ , then both the singular values of  $\bar{R}_{\text{lin}}$  and  $\bar{S}_{\text{lin}}$  as well as their relative errors coincide. This tendency was observed for random initial states in the spin example and is seen more pointedly in the numerical simulations presented next.

Note that the time-bandwidth product  $m_{\text{TBW}}$  is not independent of the system dimension  $n$ . For example, the spin system Hamiltonian (20) frequency range is *linear* in the number of spins  $n_{\text{spin}}$ , and since  $n = 2^{n_{\text{spin}}}$ , it follows that  $m_{\text{TBW}} = O(\log_{10} n)$ . More specifically,

$$m_{\text{TBW}} = \lceil (\log_{10} n / \log_{10} n_0) \Delta_0 \rceil \quad (34)$$

where  $\Delta_0, n_0$  are chosen nominal values for comparison with  $n \geq n_0$ . For example, with  $n_0 = 2^{10}$  and  $m_0 = \lceil \Delta_0 \rceil$ , if  $n = 2^{20}$  then  $m_{\text{TBW}} = 2m_0$ , or if  $n = 2^{80}$  then  $m_{\text{TBW}} = 8m_0$  and so on. More generally, if the frequency range scales as a polynomial function of the number of two-level particles, say  $O(N^k)$ , then  $m_{\text{TBW}} = O(\log_{10} n)^k$ . Such behavior is still a very favorable scaling with respect to the exponential scaling of the system dimension with the number of quantum particles.

## VIII. NUMERICAL RESULTS

### A. TFI system

Figure 4 shows plots of the normalized singular values of  $R, S, S_{\text{lin}}, R_{\text{lin}}$ , and  $\text{diag}(\alpha)$  for  $H_{\text{TFI}}$  (20) with  $h = 2$  for 10 and 12 spins and with two initial states: (a, c) at the ground state for  $h = 4$  and (b, d) random. Both  $R$  and  $S$  are calculated using the actual (nonlinear) eigenvalue sweep of  $H_{\text{TFI}}$  whereas  $S_{\text{lin}}$  uses a linear sweep over the same range. The text boxes with  $m_{\text{TBW}} = 14$  and 17 show the ideal predicted compression for the two spin cases assuming a linear frequency sweep. The ratio  $17/14 \approx 1.21$  follows the log scaling (34) with  $\log 4096 / \log 1024 = 1.2$  which obviously is the ratio of spins  $12/10$ . For 80 spins the compression estimate increases by a factor of 8 over ten spins to  $m = 112$ , dramatically low compared to  $n = 2^{80}$ , and so on. The dashed line rect function next to these boxes is the ideal singular value function given by (B1): a constant until index  $k = m$  and then drops to zero thereafter. The compression level where the singular values of *both*  $S$  and  $S_{\text{lin}}$  drop to essentially zero is almost identical to that predicted by (29). Additionally, the *actual* singular values of  $S_{\text{lin}}$  follow those of the ideal  $\bar{S}_{\text{lin}}$ , tending more closely as  $n$  increases from 1024 to 4096.

The singular values of the projected initial state (green curves), labeled  $\sigma_k[\text{diag}(\alpha)]$ , show that a random initial state

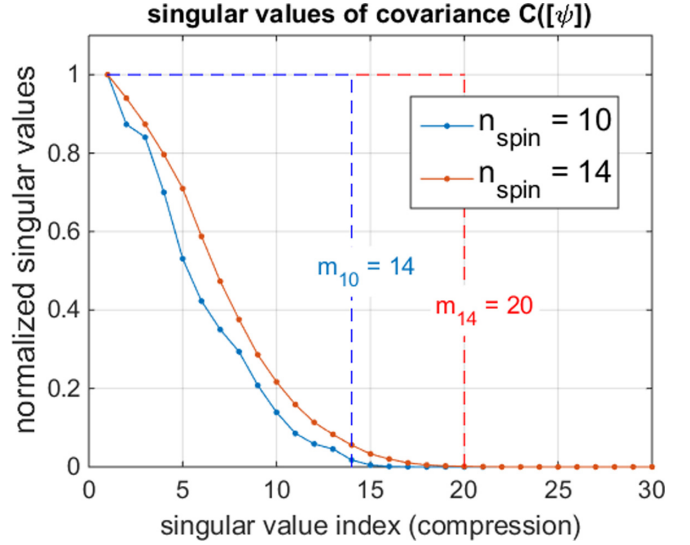


FIG. 5. Singular values of snapshot matrix (35) for the  $H_{\text{TFI}}$  system with  $n_{\text{spin}} \in \{10, 14\}$  each initialized from a random state.

is not the principle cause of compression because the state elements are spread more uniformly throughout the space than for an initial ground state which is confined.

The break points (i.e., compression level) where the singular values drop significantly as predicted by the sinc matrix  $S$ , or the linear sweep matrix  $S_{\text{lin}}$ , can differ, and exceed those of the covariance matrix  $R$  (or  $R_{\text{lin}}$ ). From a ground state, the break points are generally smaller. They adhere closely to  $S$  (or  $S_{\text{lin}}$ ) for a random initial state. The former is to be expected considering that the sinc matrix,  $S$ , contains no information about any correlations with the initial state. We also see that the linear frequency sweep approximation, resulting in the symmetric Toeplitz matrix  $S_{\text{lin}}$  and the corresponding covariance  $R_{\text{lin}}$ , is in agreement with these findings.

### B. POD via snapshot

Figure 5 shows a further comparison with  $n_{\text{spin}} \in \{10, 14\}$  where the singular values are computed from the “snapshot” version of the covariance matrix (9):

$$\Psi = [\psi_{t_1} \quad \dots \quad \psi_{t_K}] \in \mathbb{C}^{n \times K} \quad (35)$$

with  $K = 200$  uniformly spaced time samples over the same simulation run time  $T = 2$  as in the previous TFI examples. This gives a sampling rate many times the maximum Hamiltonian frequency. For both spin settings the dynamics are initialized with a random state. The singular values shown are the squares of the singular values of the  $n \times 200$  snapshot matrix  $\Psi$ ; these approximate those of the covariance matrix, i.e.,  $C([\psi]) \approx (T/K) \Psi \Psi^\dagger$ . The time-bandwidth predicted compression levels  $m_{\text{TBW}} \in \{14, 20\}$  are again validated by the data and also follow the log scaling (34) for these spin systems, i.e., compression ratio  $20/14 \approx 1.43$  compared to spin ratio  $14/10 = 1.4$ .

### C. Random Hamiltonians

We generated many random Hamiltonians from a variety of distributions of elements. The results consistently confirm



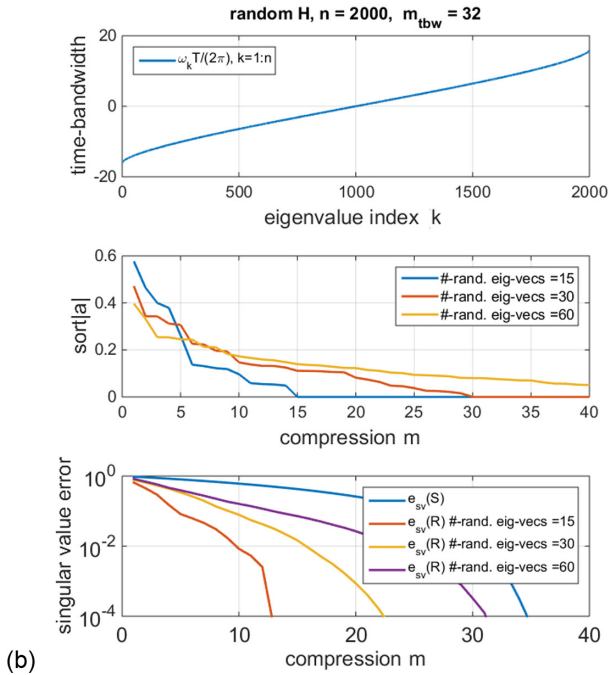
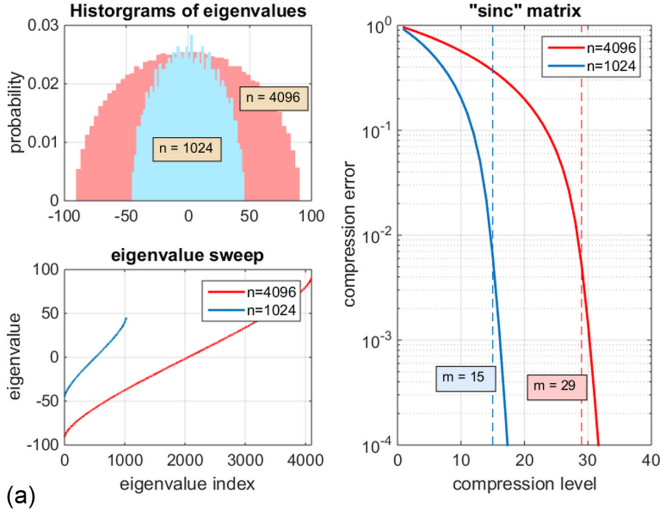


FIG. 6. Random Hamiltonians (a)  $n \in \{1024, 4096\}$  with normalized run time  $T = 1$ . The upper left plot shows the classic semicircular Wigner distribution of eigenvalues. The plot on the right shows the sinc matrix compression error (19) which matches the time-bandwidth product estimate. The lower left plot shows (nearly linear) eigenvalue sweeps. (b)  $n = 2000$ , with the initial state confined to a random selection of 15, 30, and 60 eigenvectors of the Hamiltonian.

the compression level to be in the range predicted by the time-bandwidth product. Figure 6(a) shows two instances of Gaussian random Hamiltonians for  $n \in \{1024, 4096\}$ . The upper left plot clearly displays the classic semicircular Wigner distribution of eigenvalues [28]. The sinc-matrix POD errors (19) are shown in the right plot with normalized run time  $T = 1$ . The resulting time-bandwidth compression levels are, respectively,  $m = 15$  and 29. These compression levels follow the scaling of the frequency sweep engendered by the Wigner distribution at  $O(\sqrt{n})$ : here  $29/15 = 1.93 \approx \sqrt{4096/1024} =$

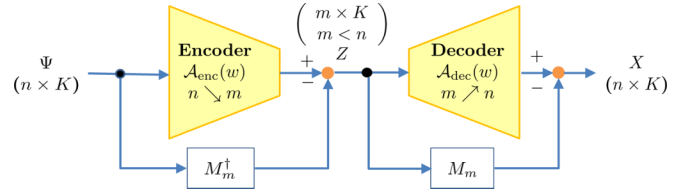


FIG. 7. Autoencoder with POD pre- and postprocessing.

2. Although not shown, the eigenvalue distribution for the spin system (20) is not semicircular; rather, it is more Gaussian in shape with small (finite) tails. These correspond to the sudden twists at the ends of the frequency sweeps as seen in Fig. 1. The Wigner distribution does not have these tails and thus only slightly twists at the end, thereby resulting in a more linear sweep.

As we have seen previously, additional compression occurs whenever the initial-state distribution is near a subspace spanned by a small number of Hamiltonian eigenvectors. Figure 6(b) depicts a typical result with a Gaussian random Hamiltonian of dimension  $n = 2000$ . For the three examples shown, we confined the random initial states to be linear combinations of 15, 30, and 60 randomly selected eigenvectors of the Hamiltonian. (Of course no initial state would be exactly so confined; this illustrates the effect.) As expected the confinement of the initial state to a further compressed subspace is seen in the middle plot where the sorted magnitudes of the unit vector (16) ( $\alpha = V^\dagger \psi_0$ ) clearly drop to zero at exactly 15, 30, and 60. The time-bandwidth product predicts a maximum compression of  $m_{TBW} = 32$  which is confirmed by the singular value errors shown in the lower plot. What is noteworthy is that for the case of the initial state being confined to a 60-dimensional subspace, the time-bandwidth product bounds the covariance level compression and the true covariance error—the same as  $e_{sv}(S)$ —begins to approach the sinc matrix error. In other words, the “sinc” matrix singular values dominate the onset of compression. A fully random initial state will cause the covariance singular values to line up with those of the sinc matrix which bounds the compression. As seen in the lower plot, as the initial-state subspace dimension increases, the errors get larger and, as expected, do not exceed the sinc matrix errors.

To emphasize this point, a variety of full length (i.e.,  $n = 2000$ ) random initial states were tested, and the compression level matched that predicted by the time-bandwidth measure. An interesting point is that the structured Hamiltonian many-body cases reported earlier in the paper, and the extreme of random Hamiltonians shown here, both displayed the same characteristic compression behavior. This situation indicates that the origin of the compression arises from the Hamiltonian spectral bandwidth and not any other special features (e.g., many-body coupling character or patterns).

IX. FURTHER COMPRESSION WITH AN AUTOENCODER

We compare our findings using POD with a variational model constructed from an autoencoder (AE) as depicted in Fig. 7. This configuration is one of several variations for seeking compression by combining an autoencoder with a

TABLE I. Comparison of rms error  $\|\Psi - X_m\|_{\text{fro}}/\sqrt{K}$  from POD only and POD with AE (Fig. 7) for  $m = 5, 10, 15$ .

$m$	POD only	POD with AE
5	0.5452	0.0101
10	0.1717	0.0070
15	0.0121	0.0030
16	0.0044	
17	0.0016	

data-driven preprocessing procedure such as POD, e.g., [29]. By disconnecting the encoder and decoder the output becomes  $X = M_m M_m^\dagger \Psi$  which is exactly the POD solution (10).

The autoencoder weights  $w \in \mathbb{R}^p$  are selected so as to minimize a weight dependent error, i.e.,

$$\begin{aligned} & \text{minimize} && E(w) = \|\Psi - X\|_{\text{fro}}^2/K, \\ & \text{subject to} && X = \mathcal{A}_{\text{dec}}(w)(Z) - M_m Z, \\ & && Z = \mathcal{A}_{\text{enc}}(w)(\Psi) - M_m^\dagger \Psi, \end{aligned} \quad (36)$$

where  $\|\cdot\|_{\text{fro}}$  is the Frobenius norm, and  $\Psi = [\psi_{t_1} \cdots \psi_{t_K}] \in \mathbb{C}^{n \times K}$  is the data [snapshot (35)] at uniform sample times  $\Delta t$ . The weight dependent interconnection structure of the encoder  $\mathcal{A}_{\text{enc}}(w) : \mathbb{C}^n \rightarrow \mathbb{C}^m$ , decoder  $\mathcal{A}_{\text{dec}}(w) : \mathbb{C}^m \rightarrow \mathbb{C}^n$ , and variational parameter dimension  $m < n$  are all specified before employing the AE. That is, the AE is seeking to achieve better predictive quality than POD for a given  $m$  value. This utilization of the AE (i.e., a special form of machine learning algorithm) can be viewed as bringing in a nonlinear feature beyond linear use of  $M$  as in POD. The flexibility inherent in the nonlinear encoding or decoding structure gives the AE the potential to produce a smaller state error for the same compression level  $m$  than that of POD which is restricted to a linear encoder or decoder structure. Finally, unlike in a variational simulator, the input state flow is available, embedded here in the snapshot matrix  $\Psi$ .

Table I and Fig. 8 compare the POD-only state errors with those from the POD with AE system. The input data comprise the snapshot matrix  $\psi \in \mathbb{C}^{1024 \times 200}$  from our previous TFI system example with a random initial state. The table shows that the POD-only rms error of 0.5452 at  $m = 5$  is reduced to 0.0101 with POD with AE, a value near to that of POD only for  $m = 15$  which has an rms error as shown of 0.0121. At  $m = 10$  and 15 the rms errors with POD with AE are approximately those with POD only at  $m = 16$  and 17, possibly indicating the limit obtainable with this multilayer autoencoder. Figure 8 highlights the large error reduction with the addition of the AE.

The POD with AE mechanism achieves an error level commensurate with the time-bandwidth product at  $m_{\text{TBW}} = 14$  for a compression  $m = 5$  and 10. Such an error reduction indicates the benefit of the inherent nonlinear variational AE state. The fundamental reason that the AE, and the particular configuration used here, can outperform POD remains an open issue. That is, given that an AE, or any neural network, can approximate most nonlinear functions, what is the character of the AE transformation that has this property, i.e., compression in this case?

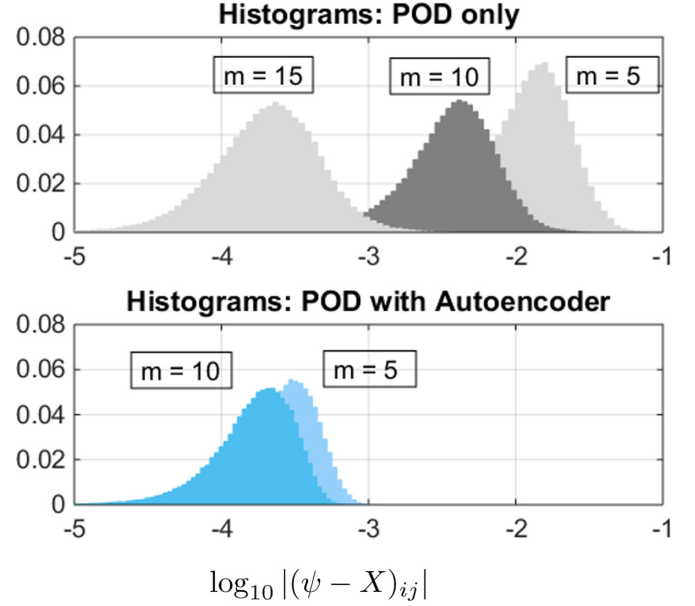


FIG. 8. Histograms of  $\log_{10}$  of the absolute value of all elements of the state error matrix  $\Psi - X$  (36). Upper: POD only for  $m = 5, 10, 15$ . Lower: POD with autoencoder for  $m = 5, 10$ .

## X. COMPRESSION OF TIME-VARYING HAMILTONIAN DYNAMICS

Consider the time-varying quantum system

$$\dot{\psi}_t = (H_0 + c_t H_1) \psi_t, \quad \psi_0 \in \mathbb{C}^n, \quad t \in [0, T], \quad (37)$$

where the external field  $c_t$  is independently, identically, and uniformly distributed in  $[-c_{\text{mag}}, c_{\text{mag}}]$  at each of the  $K$  time intervals in  $[0, T]$ . The resulting state samples are stored in the “snapshot” matrix:

$$\Psi = [\psi_{t_1} \cdots \psi_{t_K}] \in \mathbb{C}^{n \times K}, \quad (38)$$

which is used to compute the sampled-data version of the covariance matrix (9) via  $C = \Psi \Psi^\dagger / K$ . The dark thick (blue) curves in Fig. 9 are the singular value errors  $\varepsilon_m(C)$  vs compression level  $m$  from (13) for  $n = 128$  with 100 trials at  $K = 200$  time samples for  $T = 2$  and with each field magnitude  $c_{\text{mag}} = \{0.1, 1, 10\}$  with both  $H_0, H_1 \in \mathbb{C}^{128 \times 128}$  randomly generated, normalized to  $\|H_0\| = \|H_1\| = 10$ , and held fixed throughout the 100 trials. The two light thin (pink) curves are the singular value errors for one trial with  $n = 1024$  at the two field magnitudes  $c_{\text{mag}} = \{1, 10\}$  and again randomly generated  $H_0, H_1 \in \mathbb{C}^{1024 \times 1024}$  and normalized also to 10. We ran many cases at  $n = 1024$ ; so as not to crowd the figure we show two representative examples, and the rest fell in the same range as the blue curves. What is interesting to note is that there is compression in these cases and the levels do not depend very much on the Hilbert-space dimension. They do, however, depend significantly on the external field magnitude and certainly the relative magnitudes of the Hamiltonians. Fixing these at 10 for both  $n = 128$  and 1024 shows this effect. The results depicted are qualitatively similar to what we expect for the time-invariant Hamiltonians and the compression level predicted by the time-bandwidth product.

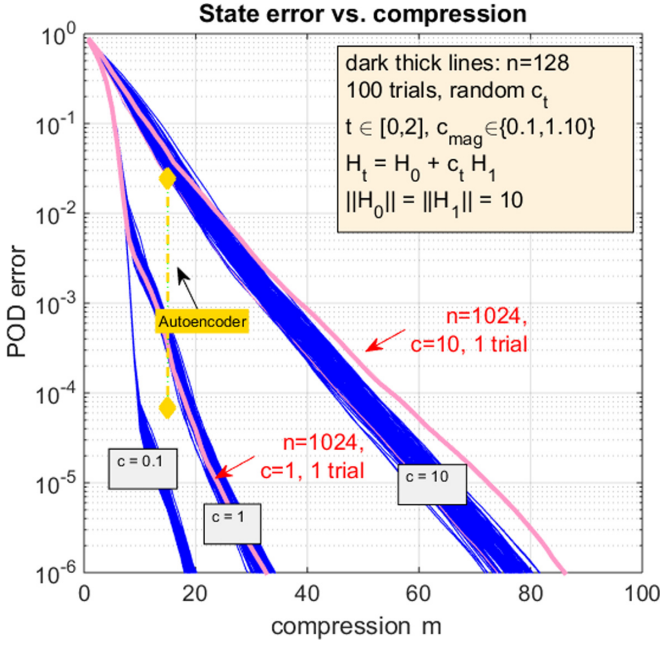


FIG. 9. Linear-time-varying random field.  $c_t, t \in [0, 2]$ . Dark thick (blue) curves are POD errors (13) for  $n = 128$  from 100 trials each at uniform random control with magnitude ranges  $\{\pm 0.1, \pm 1, \pm 10\}$  from fixed random Hamiltonians normalized as indicated. The two light thin (pink) curves are for  $n = 1024$  from one trial each with  $c = 10$  and  $1$  and random Hamiltonians as previously normalized. Applying the autoencoder (Fig. 7) with POD at compression level  $m = 15$ ,  $\varepsilon_{15} = 0.0245$  (upper diamond) results in an error approximately  $7 \times 10^{-5}$  (lower diamond).

Though compression is clearly revealed by the POD procedure, it is based on a *linear* variational state. A nonlinear variational state, such as one obtained from a neural network, would have the potential for improvement. Applying the autoencoder of Fig. 7 with POD initiated at the compression level  $m = 15$ ,  $\varepsilon_{15} = 0.0245$  (upper diamond) results in an error approximately  $7 \times 10^{-5}$  (lower diamond), almost a two-order-of-magnitude error reduction for the same compression level.

Although a detailed mathematical analysis of the time-dependent compression behavior remains to be determined, we speculate here on how a time-bandwidth measure can be developed for a time-varying quantum system such as (37). Following the exposition and notation in [30], consider the *norm-bounded linear differential inclusion* (NLDI):

$$i\dot{\psi}_t^N \in \mathcal{H}_{\text{NLDI}} \psi_t^N, \quad \psi_0^N = \psi_0, \\ \mathcal{H}_{\text{NLDI}} = \{H_\delta = H_0 + \delta H_1 \mid |\delta| \leq c_{\text{mag}}\}, \quad (39)$$

where the set  $\mathcal{H}_{\text{NLDI}} \subset \mathbb{C}^{n \times n}$ . Any solution of (37) is also a solution of the NLDI and in most cases the converse also holds. As a result, many characteristics of all solutions of the NLDI are inherited by all solutions of (37). Our *speculation* is that if compression is one of these characteristics, then the time-bandwidth product can be applied to the NLDI to establish the onset level of compression for (37). Table II shows the results of using the worst-case frequency spread from  $\mathcal{H}_{\text{NLDI}}$  to predict the onset of compression for (37). Though the

TABLE II. Comparison of POD compression error levels at  $10^{-4}$  and  $10^{-5}$  with predicted compression onset from the worst-case spread of the eigenvalues of  $\mathcal{H}_{\text{NLDI}}$  (39) for system (37) with respect to  $c_{\text{mag}} \in \{0.1, 1, 10\}$  for 100 trials each for systems evolving from (37) with dimension  $n = 128$  and  $\|H_{0,1}\| = 10$ .

$c_{\text{mag}}$	Compression level		
	NLDI prediction max $\Delta \text{eig}\{\mathcal{H}_{\text{NLDI}}\}$	POD error $10^{-4}$	POD error $10^{-5}$
0.1	7	10	13–15
1	10	17–19	22–26
10	65	44–52	58–67

predicted NLDI onset of compression is in the neighborhood produced by POD, absent a more rigorous theoretical analysis we leave this objective for a future study. One possible path we will explore is using approaches based on robust control theory for multiple uncertainties as applied in [31] for finding the set of uncertain eigenvalues as might be characterized by the time-varying nature of a Hamiltonian with multiple time-varying fields.

## XI. EXTENSIONS OF THE ANALYSIS

In this section we briefly discuss potential extensions of the time-bandwidth product theory to (A) unitary dynamics and (B) nonlinear frequency sweeps for constant Hamiltonians.

### A. Unitary dynamics

The time-bandwidth product (29) also predicts the approximate size of the compression of a variational simulation of unitary dynamics:

$$i\dot{U}_t = H U_t, \quad U_0 = I_n, \quad t \in [0, T]. \quad (40)$$

For  $\nu = 1, \dots, n$  let  $u_{t,\nu} \in \mathbb{C}^n$  denote the columns of  $U_t$ . Each column of the unitary evolves according to the *same* Hamiltonian system, i.e.,

$$i\dot{u}_{t,\nu} = H u_{t,\nu}, \quad u_{0,\nu} = \varepsilon_\nu, \quad t \in [0, T]. \quad (41)$$

Since the initial unitary is the identity matrix, it follows that the initial value of the  $\nu$ th column is  $\varepsilon_\nu \in \mathbb{R}^n$ , a vector with single nonzero element equal to one in the  $\nu$ th place. Let  $X_t$  denote the variational unitary approximation with columns  $x_{t,\nu} \in \mathbb{C}^n$ ,  $\nu = 1, \dots, n$ . Using the Frobenius norm the state (unitary) error is

$$E = \frac{1}{T} \int_0^T \|X_t - U_t\|_{\text{fro}}^2 dt = \sum_{\nu=1}^n \frac{1}{T} \int_0^T \|x_{t,\nu} - u_{t,\nu}\|_2^2 dt. \quad (42)$$

As shown in Appendix C, under a linear sweep of Hamiltonian eigenvalues, the previously defined sinc matrix bounds all the singular values of each subsystem covariance, resulting in the total error bound:

$$E \approx E_{\text{lin}} \leq n \left( 1 - \sum_{i=1}^m \sigma_i(S_{\text{lin}}/n) \right). \quad (43)$$

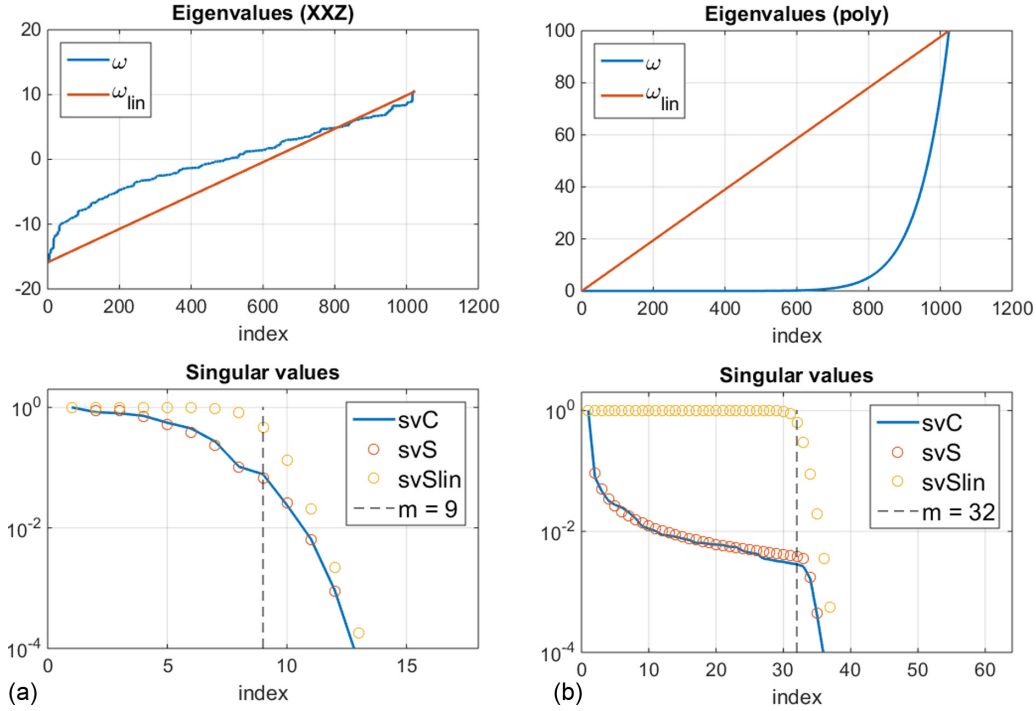


FIG. 10. Two types of eigenvalue distributions with system dimension  $n = 1024$ : (a) from an  $XXZ$  spin chain exhibiting characteristic piecewise constant variations, and (b) from a fictitious system with an exaggerated polynomial distribution.

As we have shown for compression of the quantum state dynamics with a constant Hamiltonian, we see the same here, i.e., for large  $n$  the singular values of  $S_{\text{lin}}$  approach those of  $\bar{S}_{\text{lin}}$ , and as in (30) the error tends asymptotically to zero, thereby ensuring compression for a value of  $m$  not dependent on the exponential growth of the Hilbert-space dimension with the number of particles.

### B. Nonlinear frequency sweeps

The more common case to expect from an arbitrary Hamiltonian is a nonlinear eigenvalue sweep (e.g., see Fig. 1). Figure 10 presents two other types of eigenvalue spreads and the corresponding singular value plots of the covariance matrix  $C$ , the sinc matrix  $S$ , and the linear sinc matrix  $S_{\text{lin}}$ . The upper plot in (a) shows the eigenvalues from an  $XXZ$  spin chain with Hamiltonian  $H = \sum_i \sigma_x^i \sigma_x^{i+1} + \sigma_y^i \sigma_y^{i+1} + \delta \sigma_z^i \sigma_z^{i+1}$ . The characteristic feature of these eigenvalues is that they are piecewise constant over varying intervals. With  $\delta = 1.5$  and with a random initial state, the lower plot in (a) shows that the linear approximation gives a compression that is a comparable range with that obtained from POD of a snapshot matrix. On the right the upper plot in (b) shows a fictitious eigenvalue distribution made from an exaggerated polynomial where the eigenvectors are selected randomly to generate the Hamiltonian. Though the linear approximation is clearly not very good, the compression estimate remains in the range predicted by the time-bandwidth product.

Though these circumstances violate the assumption leading to the time-bandwidth product analysis, the examples presented earlier in the paper show that a nonlinear eigenvalue

sweep does not result in a serious problem: the predicted compression from the linear eigenvalue sweep is in the range computed from the actual system. To further support this finding we propose to stretch the nonlinear eigenvalue sweep engendered by an  $n$ -dimensional Hamiltonian into a straightened out linear sweep over the same range. This is accomplished by equating the path lengths of the two sweeps.

The path length of  $n$  nonlinear monotonically increasing eigenvalues from  $\omega_1$  to  $\omega_n$  is

$$d(\omega) = \sum_{k=1}^{n-1} [(\omega_{k+1} - \omega_k)^2 + 1]^{1/2}. \quad (44)$$

If the sweep were linear then  $d(\omega_{\text{lin}}) = \sqrt{(\Delta\omega)^2 + (n-1)^2}$  with  $\Delta\omega = \omega_n - \omega_1$ . To keep the same eigenvalue range with the stretched linear sweep requires interpolating between the stretched gaps in successive eigenvalues with an additional  $n' - n$  eigenvalues where

$$n' = \lceil 1 + [d(\omega)^2 - (\Delta\omega)^2]^{1/2} \rceil. \quad (45)$$

For the two TFI frequency sweeps shown in Fig. 1 for  $n \in \{1024, 4096\}$  the stretched system dimension increases modestly to  $n' \in \{1029, 4101\}$ . Since the compression  $m$  scales logarithmically with dimension (34), the relative time-bandwidth compression estimate is  $\log_{10} n' / \log_{10} n - 1 \in \{1.47 \times 10^{-4}, 7.03 \times 10^{-4}\}$ . These changes are not visible in Fig. 4 where the predicted compressions under the linear sweep assumption are compared with the order where the actual singular values become insignificant. Similarly for the  $XXZ$  and polynomial eigenvalues shown in Fig. 10:  $n'_{\text{xxz}} = 1046$  and  $n'_{\text{poly}} = 1047$ .

## XII. SUMMARY

The main result reported here is the introduction of the time-bandwidth product:

$$\Delta = (f_{\max} - f_{\min})T, \quad (46)$$

which, for a quantum state evolving for time  $T$  with a constant Hamiltonian with frequencies ranging from  $f_{\min}$  to  $f_{\max}$ , provides an estimate of compression to a reduced system whose dimension is the nearest integer. We show that the predicted compressed dimension is exact (in a well-defined asymptotic sense) if the Hamiltonian frequencies (eigenvalues) range linearly when ordered from minimum to maximum. Since the time-bandwidth product does not depend on the initial state, it is, in effect, predicting the range of the worst-case level of compression, or more precisely, where compression begins as defined in the Introduction. Though every real system has a nonlinear frequency sweep, numerous simulations of systems with random (constant) Hamiltonians shown here do not violate the predicted approximate level of compression, i.e., no order-of-magnitude changes. In general, the time-bandwidth product is consistent with the level or onset of observed compression.

The time-bandwidth product is of course not independent of the Hilbert-space dimension. For spin systems composed of two-level particles, the frequency range typically scales linearly with the number of particles. As a result the time-bandwidth estimated compression level scales logarithmically. More generally a polynomial scaling of frequency range will still result in a compression that beats exponential scaling.

A lower value of the predicted level of compression, and sometimes significantly lower, is possible dependent upon how the initial state is distributed amongst the Hamiltonian eigenvectors. In particular, for a system with a random initial state, the system dimension has little impact on the variational compression, the main driver being the product of the frequency sweep range and the simulation run time. In contrast, for a specific initial state, such as one close to the system ground state, the interaction with the Hamiltonian eigenvectors plays a significant role in further reducing the variational compression. We showed how this comes into play using a related measure: if the initial state is predominantly in or near a subspace spanned by a small number of Hamiltonian eigenvectors, compression will be in the range of the initial subspace dimension, lower than that indicated by the time-bandwidth product. With enough run time, and with the initial state not predominantly confined to a subspace, all of the Hilbert space will eventually be populated. The good news is that this upper dimension can only grow linearly with run time: with a huge number of states the run time would have to be very long to make such an impact.

A limitation of the time-bandwidth product is that it is derived from a *linear* variational model. In contrast, a machine learning system is built to implement a *nonlinear* transformation, and thereby potentially delivering a lower level of compression with the same or smaller error. We demonstrated this effect using an autoencoder. On the other hand, the time-bandwidth product does reveal the *existence* of a useful level of compression, indicating that with a machine learning sys-

tem to actually solve for the dynamics is an evolving area of challenging research.

For a time-varying Hamiltonian, e.g., a system affected by time-varying external fields, the time-bandwidth product is not strictly applicable. Nevertheless, our preliminary analysis based on differential inclusions in Sec. X as well as numerous simulations show that compression still holds, though with an expected dependence on the field strength. It also seems reasonable to expect compression with a time-varying Hamiltonian to be worse than the time-invariant case, as the external field can be thought of as moving the initial state around through some portion of the specified Hilbert space. For a reasonably posed control problem, even for an exponentially large many-body system, it would normally not be expected to define the goal for the control to move from one end of Hilbert space to the other.

Various means were discussed for determining the time-bandwidth product to determine the degree of compression for any particular system. The essential procedure relied on either the system's dynamics appropriately sampled or, more particularly, full knowledge of the Hamiltonian's eigenvalues. Although the latter numerical determination would be forbidding for a truly high-dimensional many-body quantum system, a ready estimate of the bandwidth may be obtained in the typical scenario of the physical system having localized high-frequency components at or around the particles involved, and low-frequency components corresponding to wave-function states dispersed over the entire many-body system. This scenario is typical of many realistic systems. For example, in the case of the Ising  $n_{\text{spin}}$  system widely used throughout the paper, each spin has locally high-energy states in the presence of an applied field and only weak coupling between neighboring spins. In this circumstance, the highest energy should be approximately proportional to  $n_{\text{spin}}$  (i.e., actually  $\pm n_{\text{spin}}$ ) with the frequency bandwidth expected to scale as  $O(2n_{\text{spin}})$ . The latter result of the bandwidth scaling is very close to that found numerically in the paper, varying to some degree, depending upon the nature of the initial state of the system and the desired level of accuracy of the dynamics; the difference is inconsequential given that the Hilbert-space dimension is exponentially large,  $n = 2^{n_{\text{spin}}}$ .

More generally, consider a large molecule, nanoparticle, or arbitrary agglomeration of chemically bonded atoms. In such circumstances, the highest-frequency motion in the Born-Oppenheimer approximation (neglecting electronic excitation, which can easily be included as well) typically involves the vibrations between the nearest-neighbor chemically bonded atoms, which include bound states and the energy continuum corresponding to dissociation of one or more chemical bonds. Any such physical situation with a time-independent Hamiltonian would start with an initial state corresponding to some finite energy of the system. In the limiting case, the energy associated with the initial state is sufficient to break every chemical bond in the sample. For simplicity, assuming that the highest continuum frequency limit reached by each dissociated bond has approximately the same value  $O(\omega_{\max})$ , along with there being  $N$  pairs of such strongly coupled bonds, the maximum frequency is estimated to be  $O(N\omega_{\max})$ . Similarly the lowest frequency may simply be taken as zero or a small value [i.e., compared to

$O(N\omega_{\max})$ ] corresponding to various types of low-frequency motion spread over the entire sample of atoms. Interestingly in this case, considering the continuum being accessible, the Hilbert space is infinite dimensional, yet the time-bandwidth product specifying the compressed dynamics is estimated to only be of dimension  $O(N\omega_{\max}T)$ . The compression found here is dramatic (i.e., the same qualitative conclusion would arise even if only one bond broke) and certainly essential for considering any practical implementation of the dynamics, whether that be by machine learning, a quantum device, or possibly a clever use of more traditional methods of solving the Schrödinger equation.

To summarize, an arbitrary large quantum many-body coupled system with  $N'$  atoms [i.e., the number of bonds  $N$  is expected to be  $O(\alpha N')$  when  $\alpha \in [2, 4]$ ] is expected to have an exponentially large Hilbert space (even possibly infinite in size if the continuum is accessible), while the level of compression is expected to approximately scale as  $O(N\omega_{\max}T)$ . If the system was categorized as having a random structure, and thus a random Hamiltonian, then  $\omega_{\max}$  should be replaced by a suitable statistical average over the local high-frequency values at or around each atom. Similar arguments should also apply to time-dependent Hamiltonians with reasonably bounded controls, as discussed in Sec. X. Thus, we may generally conclude that compression is not only expected, but also essential, for the practical consideration of implementing any form of complex many-body quantum dynamics simulation.

The time-bandwidth predicted compression in some ways reveals that the Schrödinger equation is a giant variational minimization machine within which is found a “discovery”: *compression*. Returning to the second paragraph of the paper with respect to points 1 and 2 there, one can also view the system compression as an asymptotic result, where the formulation presented in the paper and the numerical evidence clearly indicate that a tolerable level of the onset of compression appears to typically set in at very low values of  $m$  with  $m \ll n$ . That possibility is of course buried in the data that go into any machine learning system. It would seem, then, that a neural net quantum simulator would have to find a compressed system, or else how could it simulate quantum dynamics (with usefully small errors) without using a number of parameters equivalent to the exponential size of the quantum state? Though we have no proof at this time, the reports of successful simulations of quantum many-body dynamics with nonexponential scaling of neural network parameters lend support to the time-bandwidth prediction of the existence of compression.

#### ACKNOWLEDGMENTS

All of the authors acknowledge support by the Data X project at Princeton University (Grant No. DataX-19-003). R.L.K. was partly supported under the Defense Advanced Research Projects Agency Physics of Artificial Intelligence program (Contract No. HR00111890031). R.L.K. thanks Shaowu Pan for alerting us to the POD modified autoencoder structure and Jun Kyu Lee and Kamal Nayal for the implementation and data assembly thereof.

#### APPENDIX A: SINGULAR VALUES OF $R$

Using the eigenvalue decomposition of the Hamiltonian and state flow (15) and (16), the covariance matrix (9) can be expressed as

$$C[\psi] = V F V^\dagger, \quad F = \frac{1}{T} \int_0^T f_t f_t^\dagger dt \in \mathbb{C}^{n \times n}. \quad (\text{A1})$$

Because  $V$  is unitary, the singular values of the state covariance matrix  $C$  (we drop the  $C[\psi]$  notation for clarity) are identical to those of  $F$ :

$$\left( \begin{array}{l} F = W \begin{bmatrix} Q_m & 0 \\ 0 & Q_{n-m} \end{bmatrix} W^\dagger, \\ W = [W_m \quad W_{n-m}], \quad W_m \in \mathbb{C}^{n \times m} \end{array} \right), \\ \times \left( \begin{array}{l} [M_m \quad M_{n-m}] = V W \\ M_m = V W_m \in \mathbb{C}^{n \times m} \end{array} \right). \quad (\text{A2})$$

The matrices  $Q_m$  and  $Q_{n-m}$  are diagonal and contain the singular values of  $C$  (or  $F$ ) in descending order [same as in (9)]. Since  $V, W \in \mathbf{U}(n)$ , the singular vectors of  $C$  are the columns of the unitary product  $VW \in \mathbf{U}(n)$  from which the first  $m$  columns provide the basis for the variational model reduction corresponding to the model error  $\text{Tr } Q_{n-m}$  (11). An equivalent expression for  $F$  is

$$F = \text{diag}(e^{-i\omega T/2}) R \text{diag}(e^{-i\omega T/2})^\dagger, \\ R = \text{diag}(\alpha) S \text{diag}(\alpha)^\dagger, \quad (\text{A3})$$

with  $R$  and  $S$  as defined in (17) and (18). Since  $\text{diag}(e^{-i\omega T/2})$  is a unitary, the singular values of  $F$  are the same as those of  $R$  and also  $C$ .

#### APPENDIX B: SINGULAR VALUES OF $S_{\text{lin}}$

Let  $T_L \in \mathbb{R}^{L \times L}$  be a real symmetric Toeplitz matrix whose first column is generated by the series  $\tilde{f}_L = \{f_\ell, \ell \in [0, L-1]\}$ . From [26], if the Fourier transform  $F(\omega) = \mathcal{F}(\tilde{f}_\infty)$  satisfies certain monotonicity conditions, then as  $L \rightarrow \infty$  the eigenvalues of  $T_L$  approach those of the Fourier transform  $F(\omega)$ . The first column of the sinc matrix with a linear sweep  $S_{\text{lin}}$  (26) consists of the first  $n$  terms of the of the  $L$ -length series  $\tilde{s}_L = \{\text{sinc}(\frac{\pi \Delta \ell}{n-1}), \ell \in [0, L-1]\}$ . With  $n$  fixed, the Fourier transform of  $\tilde{s}_\infty$  satisfies the aforementioned conditions; i.e., from standard tables,

$$F(\omega) = \sum_{\ell=-\infty}^{\infty} e^{-i\omega \ell} \text{sinc}\left(\frac{\Delta \pi \ell}{n-1}\right) \\ = \begin{cases} (n-1)/\Delta, & |\omega| \leq \pi \Delta / (n-1), \\ 0, & \pi \Delta / (n-1) < |\omega| < \pi. \end{cases} \quad (\text{B1})$$

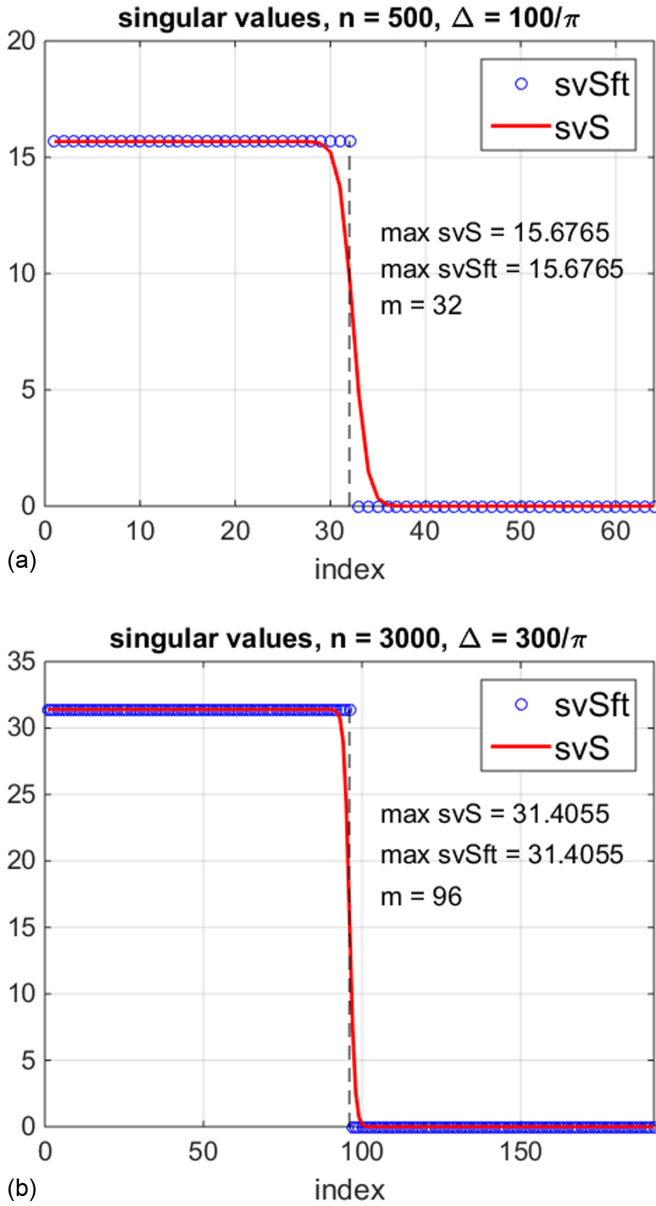


FIG. 11. Singular values of ideal (28) and  $S_{\text{lin}}$ : (a)  $\{n = 500, \Delta = 100/\pi\}$  and (b)  $\{n = 3000, \Delta = 300/\pi\}$  with compression estimate  $m_{\text{TBW}} = \lceil (\frac{n}{n-1})\Delta \rceil$  from (29).

Discretizing  $\omega = 2\pi k/n$ ,  $k = 0, \dots, n-1$ , and then sorting and grouping the absolute values of  $|F(\omega_k)|$ , we get the singular value (asymptotic in  $n$ ) approximation (28). (In [27] this type of asymptotic approximation was used for Toeplitz eigenvalues.) Figure 11 compares the ideal (Fourier transform) singular values (28) labeled svSft with those from the linear sweep matrix  $S_{\text{lin}} \in \mathbb{R}^{n \times n}$  labeled svS for two instances: (a)  $\{n = 500, \Delta = 100/\pi\}$  and (b)  $\{n = 3000, \Delta = 300/\pi\}$  with compression estimate  $m_{\text{TBW}} = \lceil (\frac{n}{n-1})\Delta \rceil$  from (29).

### APPENDIX C: UNITARY COMPRESSION

Applying the POD method, *mutatis mutandis*, to each of the  $n$  systems (41), with the variational compression fixed for all at  $m$ , results in the optimal *linear* variational unitary as

$$X_t = [\Gamma_1 u_{t,1} \quad \dots \quad \Gamma_n u_{t,n}], \quad (\text{C1})$$

where the rank- $m$  matrices  $\Gamma_v = M_v M_v^\dagger \in \mathbb{C}^{n \times n}$  with each  $M_v \in \mathbb{C}^{n \times m}$  formed as in (9) from the  $m$  singular vectors corresponding to the  $m$  largest singular values of each covariance matrix:

$$C_v = \frac{1}{T} \int_0^T u_{t,v} u_{t,v}^\dagger dt. \quad (\text{C2})$$

The optimal (POD) state (unitary) error and the corresponding  $R$  and  $S$  matrices are

$$\begin{aligned} E &= \sum_{v=1}^n \text{Tr}(I_n - \Gamma_v) C_v, \\ C_v &= V \text{diag}(e^{-i\omega T/2}) R_v \text{diag}(e^{i\omega T/2}) V^\dagger, \\ R_v &= \text{diag}(\alpha_v) S \text{diag}(\alpha_v)^\dagger, \quad \alpha_v = V^\dagger \varepsilon_v. \end{aligned} \quad (\text{C3})$$

Here  $S$  is *exactly* the previously defined sinc matrix (17). The error bound (43) follows directly.

- [1] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [2] S. Czischek, M. Gärtner, and T. Gasenzer, Quenches near Ising quantum criticality as a challenge for artificial neural networks, *Phys. Rev. B* **98**, 024311 (2018).
- [3] M. Schmitt and M. Heyl, Quantum dynamics in transverse-field Ising models from classical networks, *Sci. Post. Phys.* **4**, 013 (2018).
- [4] G. Fabiani and J. H. Mentink, Investigating ultrafast quantum magnetism with machine learning, *Sci. Post. Phys.* **7**, 4 (2019).
- [5] S. Das Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, *Phys. Today* **72**, 48 (2019).
- [6] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and Lenka Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [7] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [8] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, Quantum Simulation of Time-Dependent Hamiltonians and the Convenient Illusion of Hilbert Space, *Phys. Rev. Lett.* **106**, 170501 (2011).
- [9] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Parameter space compression underlies emergent theories and predictive models, *Science* **342**, 604 (2013).

- [10] J. K. Freericks, B. K. Nikolić, and O. Frieder, The nonequilibrium quantum many-body problem as a paradigm for extreme data science, *Int. J. Mod. Phys. B* **28**, 1430021 (2014).
- [11] H. W. Lin, M. Tegmark, and D. Rolnick, Why does deep and cheap learning work so well, *J. Stat. Phys.* **168**, 1223 (2017).
- [12] X. Han and S. A. Hartnoll, Deep Quantum Geometry of Matrices, *Phys. Rev. X* **10**, 011069 (2020).
- [13] Z. Huang and A. V. Balatsky, Complexity and geometry of quantum state manifolds, [arXiv:1711.10471](https://arxiv.org/abs/1711.10471).
- [14] B. Russell, H. Rabitz, and R.-B. Wu, Control landscapes are almost always trap free: A geometric assessment, *J. Phys. A: Math. Theor.* **50**, 205302 (2017).
- [15] R. L. Kosut, C. Arenz, and H. Rabitz, Quantum control landscape of bipartite systems, *J. Phys. A: Math. Theor.* **52**, 165305 (2019).
- [16] See R. V. L. Hartley, Transmission of information, *Bell Syst. Tech. J.* **7**, 535 (1928), for Hartley's law: "It is shown that when the storage of energy is used to restrict the steady state transmission to a limited range of frequencies the amount of information that can be transmitted is proportional to the product of the width of the frequency-range by the time it is available."
- [17] S. Lloyd and S. Montangero, Information Theoretical Analysis of Quantum Optimal Control, *Phys. Rev. Lett.* **113**, 010502 (2014).
- [18] J. N. Kutz, *Data-Driven Modeling and Scientific Computation: Methods for Complex Systems and Big Data* (Oxford University, New York, 2013).
- [19] In POD/PCA often a bias term is included:  $x(\theta_t) = M\theta_t + b$ .
- [20] A. Kumar and M. Sarovar, On model reduction for quantum dynamics: Symmetries and invariant subspaces, *J. Phys. A: Math. Theor.* **48**, 015301 (2014).
- [21] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University, Cambridge, England, 1990).
- [22] P. Xie and E. Weinan, Coarse-grained spectral projection (CGSP): A scalable and parallelizable deep learning-based approach to quantum unitary dynamics, [arXiv:2007.09788](https://arxiv.org/abs/2007.09788) (July 2020).
- [23] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*, 2nd ed. (Chelsea, New York, 1984).
- [24] Robert M. Gray, On the asymptotic eigenvalue distribution of Toeplitz matrices, *IEEE Trans. Inf. Theory* **18**, 725 (1972).
- [25] G. Fisanick-Englot and H. Rabitz, Studies of inelastic molecular collisions using impact parameter methods. I. Model calculations, *J. Chem. Phys.* **62**, 1409 (1975).
- [26] A. Bottcher, J. M. Bogoya, S. M. Grudsky, and E. A. Maximenko, Asymptotics of eigenvalues and eigenvectors of Toeplitz matrices, *Sbornik: Math.* **208**, 1578 (2017).
- [27] S.-E. Ekström, C. Garoni, and S. Serra-Capizzano, Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form, *Experimental Mathematics*, **27**, 478 (2018).
- [28] E. P. Wigner, On the distribution of the roots of certain symmetric matrices, *Ann. Math.* **67**, 325 (1958).
- [29] S. Pan and K. Duraisamy, Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability, *SIAM J. Appl. Dynamical Systems* **19**, 480 (2020).
- [30] S. Boyd, L. E. Ghaoui, Eric Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory* (SIAM, Philadelphia, PA, 1994).
- [31] M. Kishida and R. D. Braatz, On the analysis of the eigenvalues of uncertain matrices by  $\mu$  and  $\nu$ : Applications to bifurcation avoidance and convergence rates, *IEEE Transactions on Automatic Control* **61**, 748 (2016).