Low-dimensional manifolds for exact representation of open quantum systems

Nikolas Tezak, Nina H. Amini, and Hideo Mabuchi

Low dimensional manifolds for exact representation of open quantum systems

Nikolas Tezak,[1, 2, *] Nina H. Amini,[3] and Hideo Mabuchi[1]

[1]*Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA*
[2]*Rigetti Quantum Computing, Berkeley, CA 94702, USA*
[3]*CNRS, Laboratoire des signaux et systèmes (L2S),*
CentraleSupélec, 3 rue Joliot Curie, 91192 Gif-Sur-Yvette, France

(Dated: November 20, 2017)

Weakly nonlinear degrees of freedom in dissipative quantum systems tend to localize near manifolds of quasi-classical states. We present a family of analytical and computational methods for deriving optimal unitary model transformations that reduce the complexity of representing typical states. These transformations minimize the quantum relative entropy distance between a given state and particular quasi-classical manifolds. This naturally splits the description of quantum states into transformation coordinates that specify the nearest quasi-classical state and a transformed quantum state that can be represented in fewer basis levels. We derive coupled equations of motion for the coordinates and the transformed state and demonstrate how this can be exploited for efficient numerical simulation. Our optimization objective naturally quantifies the non-classicality of states occurring in some given open system dynamics. This allows us to compare the intrinsic complexity of different open quantum systems.

I. INTRODUCTION

A given quantum mechanical system can be described in more than one way. Our choice of description is usually motivated by the insight it provides, its economy, its accuracy and, when dealing with sufficient complexity, the efficiency with which it can be numerically simulated and analyzed.

Closed quantum systems evolve unitarily, and if their Hamilton operator admits a sufficient set of individually tunable control operators one can – at least in principle – realize arbitrary unitary operations on the system's Hilbert space [1]. This enables powerful quantum computing and quantum simulation schemes that derive their advantage over classical computers from the exponential scaling of the Hilbert space dimension with system size.

We may then ask, what are the implications for open quantum systems which exhibit dissipative dynamics? As dissipation increases it becomes increasingly difficult to use them for unitary quantum computing, but there also exist applications in quantum engineering that explicitly require coupling to input and output fields, ranging from quantum limited signal amplification [2, 3], via quantum key distribution to autonomously correcting quantum memories [4]. An important class of such applications can be described in the language of quantum feedback networks [5, 6] or quantum input-output models [7].

In general, our ability to *design* quantum systems for specific tasks is severely limited by the state space complexity. This is true of closed and open systems, but as fewer guarantees exist on what dynamics are achievable with open quantum systems than in the closed system case we are even more reliant on efficient numerical

schemes. Fortunately, the dynamics of open quantum systems tend to exhibit phase space *localization* [8], which implies that there exist nonlinear sets within the system's Hilbert space that act as attractors for the quantum dynamics.

A very appealing feature of quantum network models based exclusively on nonlinear oscillators is that they allow continuously tuning between the coherent qubit regime [9] and the semi-classical weakly nonlinear limit. Several near-term applications exist for nonlinear oscillator networks: in the weakly nonlinear regime these include frameworks for photonic logic [10, 11], photonic Ising machines [12, 13] and for all-optical machine learning [14]. In the strongly nonlinear regime novel quantum error correcting schemes have been proposed [15] and implemented [16]. For all of these systems it is an extremeley interesting question to ask how their dynamics and capabilities change as the ratio of dissipation to non-linearity is increased or decreased, but no existing simulation method has allowed continuous interpolation between these regimes. Our approach remediates this by providing a framework that relies on exact quantum model transformations to exploit semi-classical localization and obtain more efficient system representations.

The key to our method is formulating the problem of finding efficient state parametrizations as an optimization problem. Specifically, we employ smoothly parametrized unitary transformations to represent states in co-moving adaptive bases. The description of, e.g., a pure quantum state $|\psi_t\rangle$ thus splits into the transformation coordinates θ_t and a residual quantum state $|\phi_t\rangle = U_{\theta_t}^\dagger |\psi_t\rangle$ that is represented in a localized basis. The diagram in Figure 1 intends to visualize this. We quantify localization in terms of complexity functionals that are shown to have precise information geometric meaning.

With this, we establish a self-consistent analytic framework that not only allows for numerical simulation but

also serves to derive coupled dynamical equations for the semi-classical group manifold coordinates and the residual quantum state. We provide expressions for different diffusive Schrödinger equations as well as (stochastic) master equations. Although *stochasticity* is not at all a necessary requirement of our approach, we expect our method to have the greatest impact for systems that are continuously weakly observed by the environment and this naturally leads us to consider stochastic dynamics.
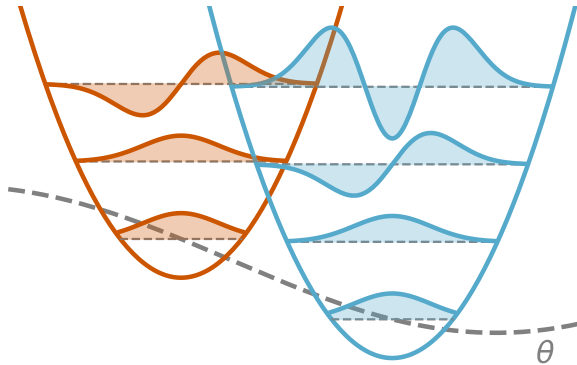

FIG. 1. Smoothly parametrized unitary maps induce a family of localized orthonormal bases parametrized by the transformation coordinates. These correspond to semi-classical phase space variables. If the group representation is irreducible and the group semi-simple, then the 'ground state' attached to each manifold point can be understood as a (generalized) coherent state [17]. For open, diffusive quantum dynamics, quantum states often localize in the vicinity of such generalized coherent states.

There already exist methods that rely on localization to analytically derive semi-classical stochastic dynamics [18] from quantum quasi-probability distributions [19–21]. These, however, require approximations that can lead to significant discrepancies with simulations based on a full quantum state description [11, 22]. Furthermore, there does not exist a general approach to incrementally increase the accuracy of these methods.

Even prior to Percival's analytic results on localization, Schack, Brun and Percival developed the MQSD simulation method [23, 24] which allowed simulating stochastic Schrödinger equations in displaced Fock bases. It can be seen as a special case of our approach, but it does not produce analytical equations of motion for the transformation coordinates. The MQSD simulation method works well for strongly dissipative oscillator systems that exhibit nearly Gaussian states with low amounts of squeezing or multi-mode correlations, but it does not generalize to other degrees of freedom. Furthermore, MQSD is based on a heuristic that has lacked a rigorous interpretation prior to our framework.

There have also been several *model-agnostic* attempts to reduce the computational complexity of simulating open quantum systems: in [25] a photonic SR-latch in the weakly quantum regime was approximated by a finite state-space Markov chain that could be re-expressed as an effective quantum network model, and in [26] a conceptually very appealing approach was proposed based on principal component analysis of simulated quantum states. One important shortcoming of both of these methods, however, is that they require at least one simulation of the full system model which can be very expensive and in many cases impossible. Furthermore, they do not preserve important qualitative features of a quantum model, such as the commutators of system observables.

The contents of this paper are as follows: First, we motivate the technical details presented in subsequent sections by presenting numerical results of our method applied to two separate systems: (a) the Kerr-nonlinearity based SR-latch as first described in [10] and discussed as an example in [25] and [26], and (b) the degenerate parametric oscillator (DPO) model which is a highly interesting system both in the semi-classical regime where it exhibits a pitchfork bifurcation and in the strongly nonlinear quantum regime, where the semi-classical fix-points limit to low decoherence cat-states that can be used to encode a qubit [15].

Next, we formally introduce the underlying 'state compression' framework which relies on positive 'penalty'-operators to quantify the complexity of quantum states. We show that their spectra can be directly related to achievable bounds on numerical truncation errors, and how this can be exploited for adapting our approach to situations in which Gaussian states poorly approximate the quantum dynamics.

We then describe how our method can be understood from an information theoretic point of view by relating our complexity functional to the quantum relative entropy. This allows to characterize the attractors near which the dynamics of a given system localize in terms of generalized Gibbs states generated by our specific choice of penalty operator which is smoothly transformed by our group.

We further present a small library of possible groups and penalty operators and finish by providing an outline of how this method can be extended to higher-dimensional Lie groups for which it is infeasible to derive the transformation differential in closed form.

There are many exciting future directions and valuable applications of our research, some of which we mention in this paper's conclusion.

II. APPLICATIONS

We first present some numerical results of our method. The technical details will be described in the following sections. Our results were obtained using a custom software package QMANIFOLD [27] that not only facilitates analytic derivations but also allows to directly carry out numerical simulations. It uses QuTiP's [28] datastructures and interfaces with our existing package QNET

[29] to automate various tedious symbolic calculations related to deriving the adjoint group representation and thus the derivation of the right generators $F_j^>(\theta_t)$. Given a particular transformation U_θ and complexity functional (cf. Section III) it can compute the minimum complexity state $|\phi_t\rangle$ and coordinates θ_t for any input state $|\psi_t\rangle$. Furthermore, given a dynamical open system model parametrized by its internal Hamiltonian H and some dissipation operators $\mathbf{L}$ it can directly derive and simulate stochastic complex quantum diffusion dynamics.

A. Toy Model: An empty cavity

Consider a simple open cavity model described by a Hamiltonian $H = \hbar\omega a^\dagger a$ and a single dissipation operator $L = \sqrt{\kappa} a$. Using the group of coherent displacements and linear excitation minimization (cf. Section VI A) of the canonical counting operator $N = a^\dagger a$ we apply the method of gradient coupled fiducial state dynamics with the fiducial state given by a displaced Gibbs state

$$\chi_{\beta,\theta} = U_\theta \underbrace{Z(\beta)^{-1} e^{-\beta N}}_{\chi_\beta} U_\theta^\dagger \quad (2.1)$$

with moving basis excitation $\langle N \rangle_{\chi_\beta} = \overline{n}_{\text{th}}$. Summarizing the manifold coordinates in a single complex amplitude $\alpha = \frac{Q+iP}{\sqrt{2}}$, this coordinate and the reduced complexity state $|\phi_t\rangle$ have the joint equations of motion (cf Section V)

$$d\alpha = -\left[i\omega + \kappa/2\right]\alpha dt + \eta \langle a \rangle_\phi dt + \sqrt{\kappa}\overline{n}_{\text{th}} dW \quad (2.2)$$

$$|d\phi_t\rangle = -\left(i\omega + \frac{\kappa}{2}\right) a^\dagger a |\phi_t\rangle dt \quad (2.3)$$
$$- \eta \langle a \rangle_\phi a^\dagger |\phi_t\rangle dt + \eta \langle a \rangle_\phi^* a |\phi_t\rangle dt$$
$$- \sqrt{\kappa}\overline{n}_{\text{th}} a^\dagger |\phi_t\rangle dW + \sqrt{\kappa}(\overline{n}_{\text{th}} + 1) a |\phi_t\rangle dW^*$$

where $|d\phi_t\rangle$ does not necessarily conserve the norm of $|\phi_t\rangle$ and where we have dropped a term $\propto |\phi_t\rangle dt$ that does not affect the evolution of any expectation values.

These dynamics can be simulated for any choice of $\overline{n}_{\text{th}}$ which can also be turned into a dynamic variable (cf. Section V F). For any finite choice of $\eta > 0$ the mode operator expectation will fluctuate around zero $\langle a \rangle_{\phi_t} \approx 0$ and in the limit $\eta \to \infty$ we recover the hard constraint $\langle a \rangle_{\phi_t}\big|_{\eta\to\infty} = 0$. We see that α only couples to the input noise process if the moving basis excitation is nonzero. On the other hand, for $n_{\text{th}} = 0 \Leftrightarrow \beta \to \infty$ we find that $|\phi_t\rangle = |0\rangle$ is a stable fix-point of the dynamics in accordance with our intuition about passive linear quantum systems. We could have equally well derived a (stochastic) master equation or a homodyne SSE using the formulas from Section V.

We can easily add a coherent displacement to this model by modifying $L \to L + \epsilon$, $H \to H + \frac{\sqrt{\kappa}}{2i}(\epsilon a^\dagger - \epsilon^* a)$. A straightforward calculation reveals that this linear displacement is fully absorbed into the dynamics of α such that relative to Equations (2.2),(2.3) we have

$$d\alpha \to d\alpha - \sqrt{\kappa}\epsilon dt \quad (2.4)$$

while $|d\phi_t\rangle$ is left unchanged.

B. Kerr-cavity based NAND latch

Mabuchi recently proposed [10] designing photonic logic gates by using nonlinear resonators in an interferometric feedback configuration. To achieve maximum power efficiency, devices could operate in a semi-classical regime where several tens of photons in a Kerr-resonator would cause the frequency to shift by one linewidth. A particularly interesting model is that of a photonic NAND-latch [10] symmetrically constructed from two Kerr-resonators with mutual coherent feedback as visualized in Figure 2. The circuit is designed such that it has two metastable states with either the first *or* the second resonator in a high photon number state while the other resonator has low internal photon number. A full quan-
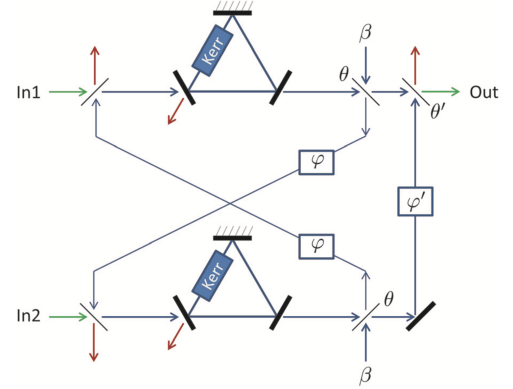

FIG. 2. A coherent NAND latch as described in [10].

tum trajectory simulation of a Stochastic Schrödinger equation with $D = d^2 = 75^2 = 5625$ levels for this model with the same parameters as in [10] and [25] is feasible on a typical workstation. In Figure 3 we present the Fock level occupation probability $\mathbb{P}(n_j = n) := \left\langle \Pi_n^{(j)} \right\rangle$ for each oscillator mode $j = 1, 2$ and each Fock-level $n = 0, 1, \ldots, d-1$ where $\Pi_n^{(j)}$ is the projection operator onto the n-th Fock state of oscillator j. We visualize this for both the original, static basis and for a coherently displaced representation, obtained by solving the linear excitation minimization problem for $N = a_1^\dagger a_1 + a_2^\dagger a_2$, i.e., the total number of excitations present in the displaced frame. We find that we can represent the dynamics of the system in a basis set that is between at least four and ten times smaller depending on whether we dynamically update the basis size during simulation *without sacrificing any precision in the simulation.* If we further allow for modest simulation error, we can achieve up to an additional order of magnitude in basis reduction. Note that
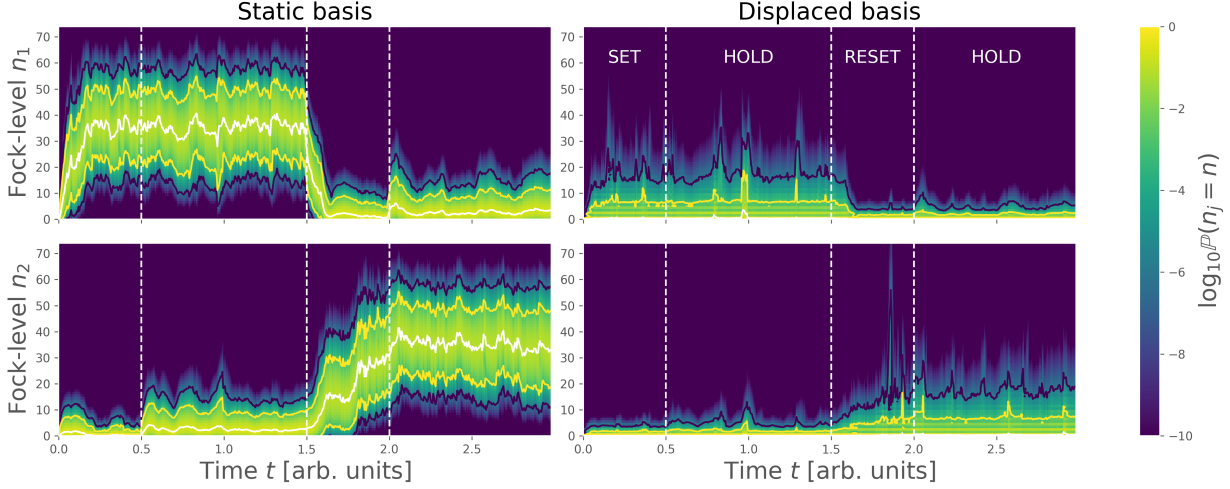
FIG. 3. Heat map of the base-10 logarithm of the occupation probability of each oscillator's Fock levels. The left panels show the results in the static basis for the first oscillator (upper left) and the second oscillator (lower left). On the right hand side we present the results obtained in a jointly displaced basis for the first oscillator (upper right) and second oscillator (lower right). The overlaid white trace gives the expected number of excitations $\left\langle a_j^\dagger a_j \right\rangle$ in each basis. The yellow (dark blue) traces show the contours at which the occupation probability reaches 10^{-3} (10^{-6}).

dimensional reduction for the same model was studied in [25, 26] and very similar physical models were studied in [11] but we claim that ours is the only approach that allows to achieve strong dimensional reduction with a controlled error and without having to simulate the full system first.

C. The degenerate parametric oscillator

Another simple nonlinear extension of the empty cavity is given by the degenerate parametric oscillator (DPO). DPOs can exhibit very rich dynamics and have been long been employed for amplification [30] and generation of light at tunable wavelengths [31] as well as more recently coherent optical Ising machines [32]. It can be physically realized by a resonant signal mode coupled to a strongly driven pump mode through a nonlinear parametric interaction that mediates the conversion of signal photon pairs to pump photons and vice versa. In the strongly nonlinear limit and for a low quality factor of the pump mode, the pump mode can be adiabatically eliminated yielding an open system model in a rotating frame parametrized by

$$H = i\frac{\hbar\chi}{2}\left[a^{\dagger 2} - a^2\right] \qquad (2.5)$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{\kappa}a \\ \sqrt{\beta}a^2 \end{pmatrix}. \qquad (2.6)$$

Using the same Ansatz as above but with $\overline{n}_{\rm th} = \eta = 0$ for simplicity we find for the deterministic part of the

coordinate dynamics

$$d\alpha_{\rm det} = -\left[\kappa/2 + \beta|\alpha|^2\right]\alpha dt + \chi\alpha^* dt. \qquad (2.7)$$

For positive pump phase $\chi > 0$ the dynamics are primarily captured by the evolution of the signal mode's real quadrature $Q = (\alpha + \alpha^*)/\sqrt{2}$ while the orthogonal P quadrature is suppressed. We can then fix $P = 0$ and consider the one dimensional differential equation valid in the classical limit

$$\dot{Q} \approx -\left(\frac{\kappa}{2} - \chi + \frac{\beta}{2}Q^2\right)Q. \qquad (2.8)$$

Equation (2.8) is identical to the normal form of a pitchfork bifurcation up to some re-scaling. We visualize the bifurcation diagram in Figure 4. The bifurcation exists for any non-zero two-photon loss rate $\beta > 0$, however the magnitude of β strongly affects how non-classical (which in this context we take to mean non-Gaussian) the state of the signal mode becomes. When the system is pumped only slightly above the threshold, random switching or tunneling between the two equilibria is possible. We present such a trajectory in Figure 5. For a constant steady state mode amplitude the switching rate strongly depends on the ratio of linear to two-photon loss. Additionally, in the strongly nonlinear case $\beta \geq \kappa$, the system can spontaneously evolve into cat-like states that feature exhibit significant simultaneous overlap with coherent states centered at either equilibrium.

In the limit of vanishing *linear* loss $\kappa/\beta \to 0$, the system has a decoherence-free sub-manifold spanned by the two equilibrium amplitude coherent states. In [15] Mirrahimi et al. outline a scheme to encode quantum information in such a system. A detailed study of the switching dynamics was carried out in [22].
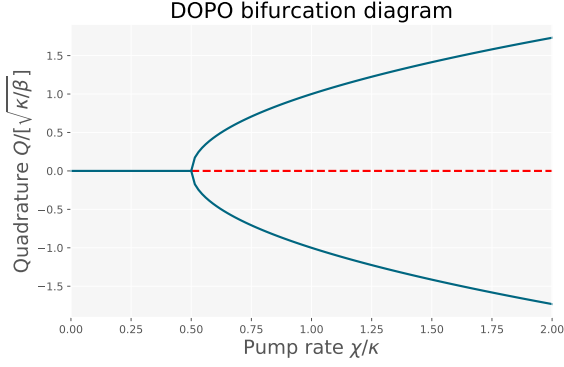
FIG. 4. When the linear loss is larger than the gain $\kappa/2 < \chi$ the real quadrature Q remains stably at 0, above a critical pump $\chi \geq \kappa/2$ this fixpoint bifurcates into two stable symmetric solutions (solid lines) and an unstable solution (dashed line) that is the continuation of the below threshold $Q = 0$ solution.

In Figure 6 we compare how each basis level contributes to a whole trajectory of states when represented in the original fixed basis to one obtained via linear excitation minimization (cf. Section VI A 1) using either a coherently displaced basis or a displaced and squeezed basis. We see that the adaptive schemes perform well in the case of strong linear dissipation but not so well in the case of strong two-photon loss. We can understand this better by inspecting typical states that occur in each evolution. In Figure 7 we present snapshots of the signal mode's Wigner function. For strong linear dissipation, the Wigner function of the signal mode typically appears quite Gaussian in shape, whereas in the strong two-photon loss case we see significant non-Gaussian features both in the transition states and when the mode is at one of the equilibria. The bad performance of the excitation minimization functional in the non-Gaussian case is much improved by the cumulant generating function (CGF) minimization approach (cf. Section III E) . In Figure 8 we compare the efficiency of the fixed basis with a coherently displaced basis where the coordinates are determined either by excitation minimization or by CGF minimization. We find that the CGF minimization (for $\lambda = 3/2$) outperforms both the fixed basis and the excitation minimization method (which is equivalent to the QSD package's approach). Here we have not even exploited the additional advantages that a displaced *and* squeezed basis combined with the CGF approach may yield. Further details of the CGF minimized state representation can be found in Appendix D.

III.   QUANTUM STATE COMPRESSION

In this section we discuss different options for quantifying the efficiency of representing a given state in a particular basis and we introduce a corresponding optimiza-

(a)$\beta/\kappa = 1/12$
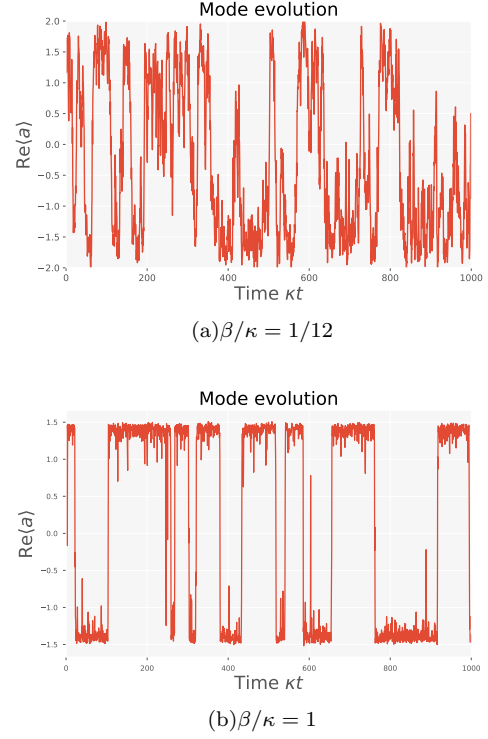

(b)$\beta/\kappa = 1$

FIG. 5.   Stochastic switching dynamics of a DPO above threshold. Figure (a) shows an example with very weak nonlinear loss $\beta \ll \kappa$ whereas (b) shows the strongly nonlinear case $\beta = \kappa$. In both cases we have chosen the parameters β, κ, χ such that the bi-stable mode amplitude equals approximately $\alpha_{r,ss} = \pm 1/\sqrt{2}$. There is a visible reduction in the switching rate and we can also see quite clearly that the magnitude of fluctuations in either bi-stable state is strongly reduced in the case of very strong nonlinearity. Specifically, the simulation parameters were $\beta = \kappa, \chi = 5\kappa/2$ and $\beta = \kappa/12, \chi = 2\kappa/3$ for the strongly and weakly nonlinear case, respectively.

tion problem. We will always assume that our adaptive Hilbert Space bases are related to the original fixed basis by means of a unitary transformation $U \in \mathcal{G}$ belonging to a connected Lie group that is locally generated by a finite dimensional Lie algebra $\mathfrak{g}$.

If our quantum state in the original, fixed basis is $|\psi_t\rangle$ we assume that it can be related to a *reduced complexity state* $|\phi_t\rangle$ via

$$|\psi_t\rangle = U_t |\phi_t\rangle \Leftrightarrow |\phi_t\rangle = U_t^\dagger |\psi_t\rangle. \qquad (3.1)$$

The generalization to mixed states is obvious, the fixed basis state ρ_t and the reduced complexity state σ_t are mutually related via

$$\rho_t = U_t \sigma_t U_t^\dagger \Leftrightarrow \sigma_t = U_t^\dagger \rho_t U_t. \qquad (3.2)$$

In the following we will work with mixed states for full generality. Special results applying only to pure states will be discussed as they arise.

(a)$\beta/\kappa = 1/12$
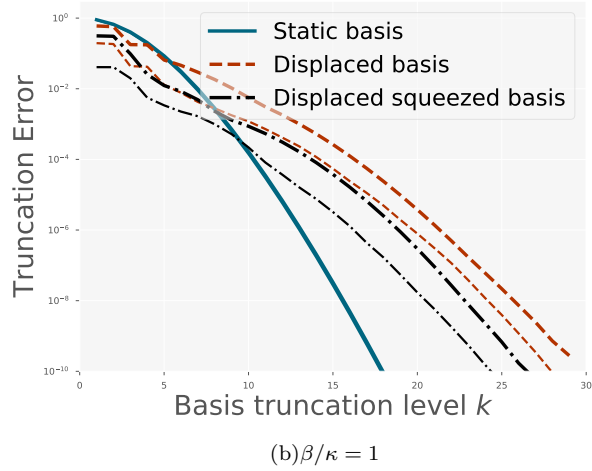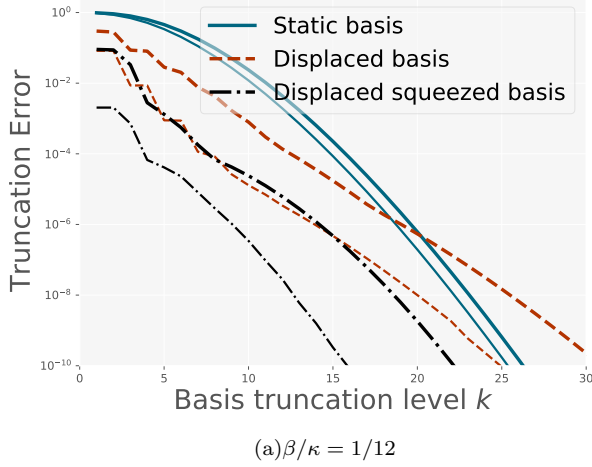

(b)$\beta/\kappa = 1$

FIG. 6. For the weakly nonlinear case (a) both the displaced basis (dashed lines) and the displaced and squeezed basis (dash-dotted lines) perform fairly well, although the displaced basis truncation error falls off less rapidly than either the static or the displaced, squeezed basis. For strongly nonlinear case (b), however, we find that the static basis (solid lines) outperforms both the displaced (dashed) and the displaced, squeezed (dash-dotted) basis. This indicates that the system dynamics depart significantly from the squeezed and displaced coherent state manifold. In both figures the thinner lines of otherwise equal style indicate the 90% level of the error, i.e., 90% of all states had lower truncation error than that.

By itself, equations (3.1) and (3.2) give an overparameterization of the original state. In the next section we outline how to remove this redundancy by deriving additional constraints on σ_t.

We will usually assume an explicit, smooth parametrization $U_t \equiv U_{\theta_t}$ by a set of coordinates $\theta_t \in D \subset \mathbb{R}^n$. Explicit parametrizations allow us to obtain analytic insight into the coupled dynamics of the

(a)$\beta/\kappa = 1/12$
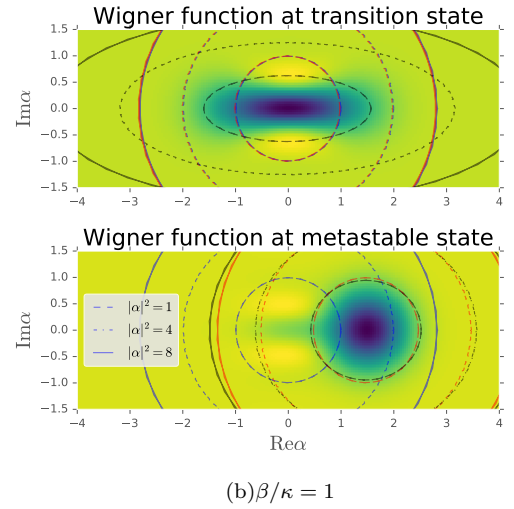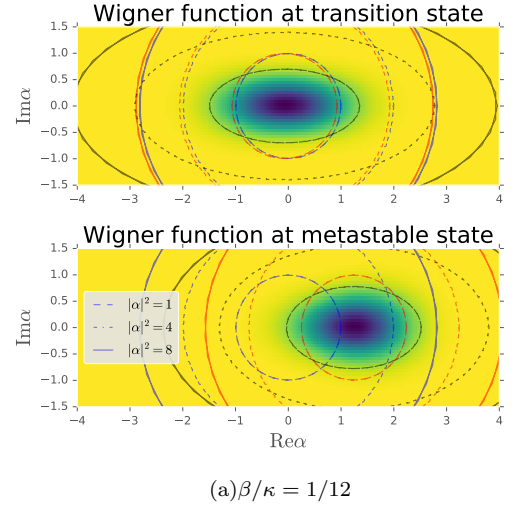

(b)$\beta/\kappa = 1$

FIG. 7. Comparing the Wigner functions of either system in typical transitions states and typical meta-stable states we see clearly that the Wigner functions of the strongly nonlinear system (b) appear much less Gaussian in shape than for the system dominated by linear dissipation. We have furthermore indicated the support set of different bases. The blue circles correspond to the fixed basis, the red circles to a coherently displaced basis and the black ellipses to a displaced and squeezed basis.

reduced complexity quantum state and the corresponding group element. We will sometimes drop the explicit time index t and write $U, \phi, \psi, \rho, \sigma, \theta$ when it is clear that they are to be taken at the same time coordinate. This, however, depends on our ability to derive explicitly [1] the partial derivatives of the transformation with respect to

---

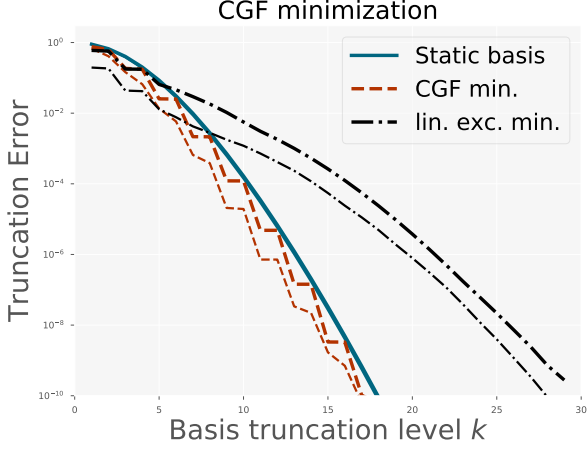[1] And in a form that can be efficiently evaluated.

FIG. 8. When changing the optimization problem to CGF minimization (dashed lines), we see that achieve higher representation efficiency than the static basis (solid lines) while so far only using a displaced, non-squeezed basis (dash-dotted lines). Again, the thinner lines of equal style indicate the 90% levels.

the coordinates

$$F_j^>(\theta) := iU_\theta^\dagger \frac{\partial U_\theta}{\partial \theta^j}. \qquad (3.3)$$

Note that these *right generators* $\{F_j^>(\theta), \; j = 1, 2, \ldots, n\}$ are necessarily Hermitian elements of the group's Lie algebra $\mathfrak{g}$.

In Appendix A we outline how to derive the right generators for some types of parametrizations. In general, however, this requires the ability to explicitly compute exponentiated matrices of the adjoint representation of $\mathfrak{g}$. For high-dimensional Lie algebras this can be fairly challenging. In Section VII we formulate a version of our method that does not require an explicit parametrization of the transformation.

Although our method allows for arbitrary unitary representations of Lie groups some group manifolds cannot be fully parametrized by a single coordinate patch. This can lead to additional technical difficulties which we will usually avoid by limiting ourselves to a single convex coordinate patch $\theta \in D \subset \mathbb{R}^n$ that includes the origin 0 which we always assume to map to the identity $U_0 = \mathbb{1}$. This poses no serious limitation in most cases of interest. Finally, note that any such parametrization is not unique. We can smoothly re-parametrize the coordinates and then derive the transformed generators via the chain rule. Our use of upper indices for the coordinates and lower indices for the generators is thus motivated by their covariant and contravariant transformation under such re-parametrizations.

A. The complexity functional

Consider first a canonical example: for a single bosonic degree of freedom with lowering operator a, and a transformation group given by coherent displacements $U_\theta = D(\alpha) = e^{\alpha a^\dagger - \alpha^* a}$, an intuitive constraint would be to demand that $\langle a \rangle_{\sigma_t} = 0$, or equivalently $\langle a \rangle_{\rho_t} = \theta$. This fully fixes the coordinates and removes the redundancy. This is precisely the constraint which the QSD software package implements. As we demonstrate below, this approach is equivalent to finding the coherent displacement coordinates of the nearest – in the sense of minimal quantum relative entropy distance – symmetric variance Gaussian oscillator state. This approach works well for nearly coherent states ρ_t but as we saw in Section II C it can actually *increase* the complexity when ρ_t has significant non-Gaussian features.

Below we re-derive this as the result of an optimization problem, which enables us to generalize and improve the approach. To formulate the problem, we introduce a complexity functional $\mathcal{J}(\theta; \rho)$ which, given a state ρ, attains a unique global minimum on the space of coordinates θ, we can then fix our coordinates at all times to be

$$\theta_t := \theta_*^{(\mathcal{J})}(\rho_t) = \mathrm{argmin}_\theta \, \mathcal{J}(\theta; \rho_t). \qquad (3.4)$$

B. Expectation minimization

The simplest choice of complexity functional is obtained by evaluating the expectation of a lower bounded operator M that penalizes population of undesired basis levels in the transformed basis.

$$\mathcal{J}_M(\theta; \rho) := \underbrace{\left\langle U_\theta M U_\theta^\dagger \right\rangle_\rho}_{\langle M \rangle_{U_\theta^\dagger \rho U_\theta}}. \qquad (3.5)$$

For a bosonic degree of freedom, the penalty operator M could simply be the canonical number operator $M = a^\dagger a$. As we will see below, when the transformation is given by coherent displacements $U_\theta := D(\theta^1 + i\theta^2)$ this particular choice leads to the QSD scheme $\theta_* = \left(\mathrm{Re} \left[\langle a \rangle_\rho \right], \mathrm{Im} \left[\langle a \rangle_\rho \right] \right)$.

More generally, however, *any* lower bounded operator defines a partial order relation for its eigenspaces via the ordering of its eigenvalues. According to such an order, a low complexity state would be characterized by being confined to a subspace spanned by basis states of low order. For composite systems comprising multiple degrees of freedom or even just competing measures of complexity for a single degree of freedom, we can define composite penalty operators by taking linear combinations of e mutually commuting, positive penalty operators $\{M_k \mid M_k \geq 0, [M_k, M_l] = 0, \; k, l = 1, 2, \ldots, e\}$ with

positive weights $\{\beta_k \geq 0, \ k = 1, 2, \ldots, e\}$, i.e.,

$$M(\beta) = \sum_{k=1}^{e} \beta^k M_k, \quad \beta^k > 0, \ k = 1, 2, \ldots, e. \quad (3.6)$$

We discuss how to choose the weights β below. We will generally refer to this class of optimization problems as *expectation minimization*.

C. Counting Operators

For single degrees of freedom there often exists a special class of such positive operators that has an evenly spaced and non-degenerate spectrum $\lambda_k = \lambda_0 + \lambda k$. Examples of this are the number operator for a bosonic degree of freedom and any projection to a single spatial axis of the angular momentum operator such as J_z for a system with conserved total angular momentum $\vec{J}^2 = \hbar^2 J(J+1)$. Any such operator can be normalized such that $\lambda_0 = 0$ and $\lambda = 1$ by rescaling and translation $M \to (M - \lambda_0 \mathbb{1})/\lambda$. We will refer to such normalized operators as *single degree counting operators*

$$N = \sum_{k=0}^{d} k |k\rangle \langle k|, \ d \in \mathbb{N} \cup \{\infty\}, \quad (3.7)$$

where $\{|k\rangle ; k = 0, 1, \ldots, d\}$ is the orthonormal eigenbasis of N. Such counting operators may also admit certain raising and lowering operators $A_\pm = A_\mp^\dagger$ satisfying $[N, A_\pm] = \pm A_\pm$. This implies that $A_\pm |k\rangle \propto |k \pm 1\rangle$. It is then possible to show that under the additional condition that $[A_-, A_+] = \alpha_1 \mathbb{1} + \alpha_N N$ for some real numbers α_1, α_N this reduces exactly to the above mentioned examples, i.e., in the infinite dimensional case $d = \infty$ the above conditions imply that our problem is equivalent to a harmonic oscillator with the typical raising and lowering operators $A_+ \propto a^\dagger$, while the finite dimensional case $d = 2J + 1$ is equivalent to an angular momentum space of fixed integral or half-integral J with the angular momentum ladder operators $A_\pm \propto J_\pm$.

As above we can combine individual counting operators $\{N_k, k = 1, 2, \ldots, e\}$ to define composite counting operators

$$N(\beta) := \sum_{k=1}^{e} \beta^k N_k, \quad \beta^k > 0, \ k = 1, 2, \ldots, e. \quad (3.8)$$

In most cases of interest to us the resulting operators have a unique ground state $N(\beta) |\Omega\rangle = 0$, and they always have a very simple spectrum. An important special case is realized when the individual counting operators N_k count excitations of different physical subsystems and each has its own pair of raising and lowering operators $[N_k, A_{l,\pm}] = \pm \delta_{kl} A_{l,\pm}$. It is easy to see that these must still be raising and lowering operators of the composite counting operator

$$[N(\beta), A_{l,\pm}] = \pm \beta^k A_{k,\pm}. \quad (3.9)$$

In some cases there exist additional generalized raising and lowering operators for the composite counting operator beyond the raising and lowering operators associated with individual subsystems corresponding to particle exchange or correlated particle creation. As an example, for a collection of harmonic oscillators with raising and lowering operators $\{a_1, a_1^\dagger, \ldots, a_e, a_e^\dagger\}$ we could define $N(\beta) = \sum_{k=1}^{e} \beta^k a_k^\dagger a_k$ as a composite counting operator and we would then find that the quadratic operators $\{a_j^\dagger a_k, \ j, k = 1, 2, \ldots, e\}$ which induce unitary mixing of multiple oscillator modes or $\{a_j^\dagger a_k^\dagger, \ j, k = 1, 2, \ldots, e\}$ which – along with their hermitian conjugates – induce multi-mode squeezing are also generalized raising and lowering operators:

$$\left[\sum_{k=1}^{e} \beta^k a_k^\dagger a_k, \ a_l^\dagger a_m \right] = (\beta^l - \beta^m) a_l^\dagger a_m \quad (3.10)$$

$$\left[\sum_{k=1}^{e} \beta^k a_k^\dagger a_k, \ a_l^\dagger a_m^\dagger \right] = (\beta^l + \beta^m) a_l^\dagger a_m^\dagger \quad (3.11)$$

$$\left[\sum_{k=1}^{e} \beta^k a_k^\dagger a_k, \ a_l a_m \right] = -(\beta^l + \beta^m) a_l a_m. \quad (3.12)$$

These are useful for reducing representation complexity associated with multi-degree correlations. When combining several individual counting operators, it may not be a priori known how to pick good weights $\beta := (\beta^1, \ldots, \beta^e)^T$. In Section IV A we will see that ρ itself contains all the information required to choose β.

We will generally refer to this class of optimization problems, i.e., minimizing the expectation of single or composite counting operators, as *linear excitation minimization*. The expectation value of counting operators and polynomials of counting operators can generally be expressed as a finite linear combination of operator expectations with coefficients depending on the coordinates. In some cases this allows to solve for the optimal coordinates directly and in closed form.

D. Iterative complexity reduction

While some choices of complexity functionals lead to solutions in closed form, we must usually resort to numerical optimization. For compact Lie groups represented on finite dimensional Hilbert spaces there exist some provably powerful *gradient flow* methods as described in [33, 34] but the non-compact and (at least formally) infinite dimensional case that is often of interest to us is more complicated and less well understood. Nonetheless, for specific examples of groups and parametrizations we are able to prove the convexity of some linear and even nonlinear expectation minimization schemes which allows us to employ existing schemes such as generalized Newton's method [35] for obtaining optimal coordinates. Here we derive explicit expressions for the gradient and Hessian

in terms of operator moments resulting from a second order expansion of the functional

$$\mathcal{J}(\theta + \delta\theta; \rho) = \mathcal{J}(\theta; \rho) + \sum_{j=1}^{n} y_j(\theta; \rho)\delta\theta^j \qquad (3.13)$$

$$+ \frac{1}{2} \sum_{j,k=1}^{n} h_{jk}(\theta; \rho)\delta\theta^j \delta\theta^k + O\left(\delta\theta^3\right)$$

If the complexity functional is strictly convex then the Hessian $h(\theta; \rho) = (h_{jk})_{j,k=1}^{n}$ is positive definite everywhere and an appropriate variant of Newton's method can be applied to find the optimal coordinates which are implicitly defined by requiring the gradient to vanish $y_j(\theta; \rho) = 0$, $j = 1, 2, \ldots, n$.

For the case of expectation minimization with penalty operator M, the explicit expressions for the gradient and Hessian are also given by simple expectation values

$$y_j(\theta; \rho) = \left\langle Y_j^>(\theta) \right\rangle_{U_\theta^\dagger \rho U_\theta} \qquad (3.14)$$

$$h_{jk}(\theta; \rho) = h_{kj}(\theta; \rho) = \left\langle H_{jk}^>(\theta) \right\rangle_{U_\theta^\dagger \rho U_\theta} \qquad (3.15)$$

where

$$Y_j^>(\theta) := i[M, F_j^>(\theta)] \qquad (3.16)$$

$$H_{jk}^>(\theta) := \left[F_j^>(\theta), \left[M, F_k^>(\theta)\right]\right] + i\left[M, \frac{\partial F_k^>(\theta)}{\partial\theta^j}\right]. \qquad (3.17)$$

Although it is not immediately obvious from Equation (3.17), the symmetry of the Hessian operators $H_{jk}^>(\theta) = H_{kj}^>(\theta)$ follows straightforwardly from $\frac{\partial^2 U_\theta}{\partial\theta^j \partial\theta^k} = \frac{\partial^2 U_\theta}{\partial\theta^k \partial\theta^j}$.

When ρ_t evolves with time we can now either solve the minimization problem (3.4) at each time and use this to obtain the coordinates θ_t or alternatively derive explicit (stochastic) differential equations for the coordinates. While the former will allow us to adapt our scheme to arbitrary stochastic dynamics – jump equations and diffusive dynamics – the latter method can provide us with more insight into the dynamics and open up interesting opportunities for designing control schemes.

E. Nonlinear excitation minimization

In this section we explain the importance of the level spacing of the penalty operator spectrum. Assume that we are starting with a counting operator N. Since N is positive, we can bound the probability of highly excited states using Markov's inequality

$$\mathbb{P}_\rho\left[N_\theta > N_0\right] \leq \frac{\langle N_\theta \rangle_\rho}{N_0}. \qquad (3.18)$$

Unfortunately, this bound decays only as $O\left(N_0^{-1}\right)$.

A useful alternative is then to exploit Markov's extended inequality by applying a monotonically increasing map to $N \to f(N)$. In this case we must have

$$\mathbb{P}_\rho\left[N_\theta \geq N_0\right] = \mathbb{P}_\rho\left[f(N_\theta) \geq f(N_0)\right] \leq \frac{\langle f(N_\theta) \rangle_\rho}{f(N_0)}. \qquad (3.19)$$

If $f(n)$ increases super-linearly with n, then the resulting penalty operator $f(N)$ has an increasing spacing of eigenvalues versus N and thus the bound on high excitations will decrease faster than $O\left(N_0^{-1}\right)$.

A very strong bound can be achieved with the exponential map $f(n) = \exp(\lambda n)$ parametrized by $\lambda > 0$. This gives

$$\mathbb{P}_\rho\left[N_\theta > N_0\right] \leq \left\langle e^{\lambda(N_\theta - N_0)} \right\rangle_\rho \qquad (3.20)$$

$$= e^{-\lambda N_0 + \mathcal{J}_N^{\mathrm{CGF}}}. \qquad (3.21)$$

where we have naturally been lead to introduce the *cumulant generating function* (CGF) of the penalty operator

$$\mathcal{J}_N^{\mathrm{CGF}}(\theta; \rho, \lambda) := \log\left[\underbrace{\left\langle e^{\lambda N_\theta} \right\rangle_\rho}_{\langle e^{\lambda N} \rangle_{U_\theta^\dagger \rho U_\theta}}\right], \qquad (3.22)$$

which, for very small $0 < \lambda \ll 1$ reduces to

$$\mathcal{J}_N^{\mathrm{CGF}}(\theta; \rho, \lambda) \approx \lambda \langle N_\theta \rangle_\rho + \frac{\lambda^2}{2}\mathrm{var}\left(N_\theta\right)_\rho + O(\lambda^3), \qquad (3.23)$$

showing that for small λ this optimization problem is equivalent to directly minimizing the expectation of the penalty operator while for increasing λ also penalizing large variance.

Note that when N is unbounded there may exist normalizable states ρ for which $\left\langle e^{\lambda N_\theta} \right\rangle_\rho$ diverges for any $\lambda > 0$, but this is generally true for $\langle N_\theta \rangle_\rho$ itself. We will assume here without proof that such states do not actually arise in the dynamical evolution of open quantum systems. For a constantly spaced, unbounded spectrum if there exist constants $N_0 \in \mathbb{N}$ and $\alpha \in [0, 1)$ such that $\forall n \geq N_0$ we find

$$\frac{\langle n + 1| U_\theta^\dagger \rho U_\theta |n + 1\rangle}{\langle n| U_\theta^\dagger \rho U_\theta |n\rangle} \leq \alpha,$$

then $\left\langle e^{\lambda N_\theta} \right\rangle_\rho$ exists for all $\lambda < \log 1/\alpha$. The inequality in (3.20)-(3.21) is an example of a Chernoff bound. Since the bound is satisfied for any λ we can minimize the right hand side over λ to achieve the most restrictive bound yielding

$$\log\mathbb{P}_\rho\left[N_\theta \geq N_0\right] \leq \min_\lambda \mathcal{J}_N^{\mathrm{CGF}}(\lambda) - \lambda N_0. \qquad (3.24)$$

Given N_0 the optimal λ_* leading to the lowest bound is implicitly defined via

$$\left.\frac{\partial \mathcal{J}_N^{\text{CGF}}(\lambda)}{\partial \lambda}\right|_{\lambda_*} = N_0. \qquad (3.25)$$

The negated left hand side in equation (3.24) above is proportional to the digits of relative accuracy obtained when truncating the basis at the level N_0. The above demonstrates that it is related to the cumulant generating function via a Legendre transformation:

$$-\log \mathbb{P}_\rho\left[N_\theta \geq N_0\right] \geq \mathcal{A}(N_0) := \sup_\lambda \lambda N_0 - \mathcal{J}_N^{\text{CGF}}(\lambda). \qquad (3.26)$$

We thus see that for a fixed N_0 our complexity functional guarantees a minimal accuracy with which a given quantum state can be represented in a d-dimensional subspace of the overall state space, where d is the number of eigenvalues λ_k of N such that $\lambda_k < N_0$. This is visualized in Figure 9. Conversely, assuming that $\mathcal{A}(N_0)$ is one-to-one, for any desired accuracy there exists a specific truncation level N_0 at which the Chernoff bound is tightest, making it the most efficient truncation level to achieve a certified accuracy.

We will refer to this family of nonlinear excitation minimization schemes as *CGF minimization*.

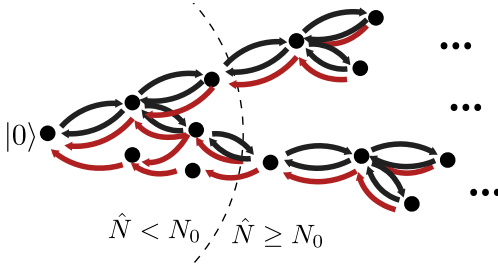

$$\hat{N} < N_0 \,, \, \hat{N} \geq N_0$$

FIG. 9. Many physically relevant Lie groups admit unitary representations in which there exists a natural ordering of the basis states according to the spectrum of some generalized energy or counting operator. Here we visualize such basis levels by black dots and suggest that they are (partially) ordered from left to right. The black arrows represent coherent transitions induced by a Hamilton operator, while the red arrows indicate dissipation induced transitions. Transforming the dynamics to a parametrized basis induces a mapping from this graph to one with potentially more transitions, e.g., under a squeezing transformation $a \to \cosh r a + \sinh r a^\dagger$ which generally increases the terms of the Hamilton operator, but if the transformation coordinates are chosen wisely, the system state can be kept close to the left side of this graph, i.e., it is trapped in a low-dimensional subspace of the overall transformed basis.

IV. INFORMATION THEORETIC INTERPRETATION

In this section we demonstrate that expectation minimization can also be interpreted as minimizing the quantum relative entropy between ρ and a transformed canonical Gibbs state $\chi_{\theta,\beta} := U_\theta \chi_\beta U_\theta^\dagger$ where $\chi_\beta := Z(\beta)^{-1} e^{-\beta M}$ and $Z(\beta) = \text{Tr}\left(e^{-\beta M}\right)$, which is itself the canonical Gibbs state associated with the transformed penalty operator

$$\chi_{\theta,\beta} = Z(\beta)^{-1} \exp\left(-\beta \underbrace{U_\theta M U_\theta^\dagger}_{=:M_\theta}\right). \qquad (4.1)$$

To see this, note that the quantum relative entropy between ρ and $\chi_{\theta,\beta}$ is given by

$$S\left(\rho\|\chi_{\theta,\beta}\right) = \text{Tr}\left(\rho\left[\log \rho - \log \chi_{\theta,\beta}\right]\right) \qquad (4.2)$$
$$= -H(\rho) - \left\langle\log \chi_{\theta,\beta}\right\rangle_\rho \qquad (4.3)$$
$$= -H(\rho) + \log Z(\beta) + \beta \left\langle M_\theta\right\rangle_\rho. \qquad (4.4)$$

Here the first and second terms $-H(\rho) + \log Z(\beta)$ do not depend on θ and therefore minimizing the quantum relative entropy over all possible coordinates θ is equivalent to excitation minimization as discussed above. We may also minimize (4.4) over β to derive an optimal Gibbs weight β_* for a given quantum state ρ and the associated optimal coordinates θ_*. By the construction of the Gibbs state, this is equivalent to ρ and χ_{θ,β_*} having equal expectation values of the penalty operator, i.e.,

$$\left\langle M\right\rangle_\sigma = -\partial_\beta \log Z(\beta)|_{\beta_*} = \left\langle M\right\rangle_{\chi_{\beta_*}} \qquad (4.5)$$
$$\Leftrightarrow \left\langle M_{\theta_*}\right\rangle_\rho = \left\langle M_{\theta_*}\right\rangle_{\chi_{\theta_*,\beta_*}}. \qquad (4.6)$$

This then affords a very nice interpretation of expectation minimization: It finds the transformation coordinates that minimize the quantum relative entropy between the actual state and a manifold of generalized Gibbs states. These form the set of maximum Von-Neumann entropy states constrained to have a specific expectation of the penalty operator. In this sense, given a penalty operator M, they form the *the least biased manifold of states to compare a given state with*.

When $\beta = \beta_*$ the minimum relative entropy is given by the difference in entropies of the actual state and the nearest thermal state

$$S(\rho\|\chi_{\theta_*,\beta_*}) = H(\chi_{\theta_*,\beta_*}) - H(\rho) \qquad (4.7)$$
$$= H(\chi_{\beta_*}) - H(\sigma). \qquad (4.8)$$

Note that for a pure state $\rho_t = |\psi_t\rangle\langle\psi_t|$ the second term vanishes, whereas the positivity of $S(\rho\|\chi) \geq 0$ (with equality if and only if $\rho = \chi$) implies that χ_{θ_*,β_*} really is the maximum (over all states) Von-Neumann entropy state with minimum (over all parameters) expected penalty $\left\langle M_{\theta_*}\right\rangle_\rho = \left\langle M_{\theta_*}\right\rangle_{\chi_{\theta_*,\beta_*}}$. By its definition, the Von-Neumann entropy only depends on the spectrum of the density matrix, which is invariant under unitary transforms. This allows us to compute it in either the original or the moving basis $H(\rho_t) = H(\sigma_t)$.

In the specific case of collective oscillator states with $M = N(\beta) = \beta^1 N_1 + \cdots + \beta^e N_e$, i.e., multi-mode

counting operators, the minimal quantum relative entropy between a given state and a generalized Gaussian state has been employed as a measure of non-Gaussianity [36, 37]. If we take U_θ to allow for all Gaussianity preserving transformation (coherent displacements, unitary and more generally symplectic transformations) then excitation minimization corresponds exactly to finding the nearest Gaussian state. Often, however, we may restrict the transformation to a sub-manifold of the full set of Gaussian transformations to avoid a large decrease of the sparsity of the operators generating the moving basis dynamics. More generally, for any composite penalty operator $M = M(\beta) = \beta^1 M_1 + \cdots + \beta^e M_e$ the quantum relative entropy splits into a sum

$$S(\rho\|\chi_{\theta,\beta}) = -H(\rho) + \log Z(\beta) + \sum_{k=1}^{e} \beta^k \langle M_{k,\theta}\rangle_\rho, \tag{4.9}$$

where $Z(\beta) = \mathrm{Tr}\left(\exp(-\sum_{k=1}^{e}\beta^k M_k)\right)$ and $M_{k,\theta} = U_\theta M_k U_\theta^\dagger$. This information geometric interpretation then affords us a method for optimally composing penalty operators: We demand that the relative entropy be minimized not only over the transformation coordinates θ but also over the complexity weights β.

A special situation arises when the penalty operators act on different degrees of freedom, in which case we can decompose the partition function

$$\log Z(\beta) = \sum_{k=1}^{e} \log \underbrace{\mathrm{Tr}_k\left(\exp(-\beta^k M_k)\right)}_{Z_k(\beta^k)}. \tag{4.10}$$

Additionally, in certain cases the group transformation may be decomposable into single degree transformations

$$U_\theta = U_{1,\theta_1} \otimes U_{2,\theta_2} \otimes \cdots \otimes U_{e,\theta_e} \tag{4.11}$$

such that each factor U_{k,θ_k} is a particular unitary operator parametrized by its own disjoint set of coordinates $\theta_k = (\theta_k^1, \ldots, \theta_k^{n_k})$ and acting non-trivially only on the degree of freedom labeled k. In this case we have

$$M_{k,\theta} = U_\theta M_k U_\theta^\dagger = U_{k,\theta_k} M_k U_{k,\theta_k}^\dagger \tag{4.12}$$

which implies that the optimization can be carried out independently for the coordinates θ_k for each $k = 1, 2, \ldots, e$ and also prior to optimizing over the complexity weights β^k.

A. Quadratic expansion of the relative entropy

Any choice of parametrized quantum state family combined with a quantitative measure of distances between quantum states induces a local distance metric for the parameters θ and β. For small coordinate displacements $\delta\theta$ and $\delta\beta$ the quantum relative entropy can be expanded to second order which yields

$$\delta S(\rho\|\chi_{\theta,\beta}) = \sum_{k=1}^{e} z_k \delta\beta^k + \sum_{j=1}^{n} y_j \delta\theta^j \tag{4.13}$$
$$+ \frac{1}{2}\sum_{j=1}^{n}\sum_{l=1}^{n} h_{jl}\delta\theta^j\delta\theta^l$$
$$+ \frac{1}{2}\sum_{k=1}^{e}\sum_{j=1}^{n} y_{kj}\delta\beta^k\delta\theta^j$$
$$+ \frac{1}{2}\sum_{k=1}^{e}\sum_{m=1}^{e} g_{km}\delta\beta^k\delta\beta^m + O(\delta^3)$$

where the coefficients are

$$z_k := \underbrace{Z(\beta)^{-1}\frac{\partial Z(\beta)}{\partial\beta^k}}_{-\langle M_{k,\theta}\rangle_{\chi_{\theta,\beta}}} + \langle M_{k,\theta}\rangle_\rho, \tag{4.14}$$

$$y_{kj} := \left\langle \frac{\partial M_{k,\theta}}{\partial\theta^j}\right\rangle_\rho, \tag{4.15}$$

$$y_j := \sum_{k=1}^{e}\beta^k y_{kj}, \tag{4.16}$$

$$h_{jl} := \sum_{k=1}^{e}\beta^k \left\langle \frac{\partial^2 M_{k,\theta}}{\partial\theta^j\partial\theta^l}\right\rangle_\rho, \tag{4.17}$$

$$g_{km} := \frac{\partial^2 \log Z(\beta)}{\partial\beta^k\partial\beta^m} = \mathrm{Cov}\left[M_{k,\theta}, M_{m,\theta}\right]_{\chi_{\theta,\beta}}. \tag{4.18}$$

Near a minimum the linear contributions vanish and in the vicinity of such minima we thus have a quadratic form in the coordinate displacements which can be interpreted as a local distance measure. We will use this expansion later on when we derive effective dynamics for the complexity weights. When ρ is itself a member of the parametrized manifold $\rho = \chi_{\theta_*,\beta_*}$, then for $(\theta,\beta) = (\theta_* + \delta\theta, \beta_* + \delta\beta)$ the linear contribution to the above expansion clearly vanishes and the quadratic coefficients are given by the Kubo-Mori quantum Fisher information metric [38]. The operator derivatives appearing in these definitions can be transformed to the moving bases where they can be expressed in terms of the transformation generators as

$$\left\langle \frac{\partial M_{k,\theta}}{\partial\theta^j}\right\rangle_\rho = \left\langle i[M_k, F_j^>(\theta)]\right\rangle_\sigma \tag{4.19}$$

$$\left\langle \frac{\partial^2 M_{k,\theta}}{\partial\theta^j\partial\theta^l}\right\rangle_\rho = \left\langle \left[F_j^>(\theta), [M_k, F_l^>(\theta)]\right]\right\rangle_\sigma \tag{4.20}$$
$$+ \left\langle i\left[M_k, \frac{\partial F_l^>(\theta)}{\partial\theta^j}\right]\right\rangle_\sigma.$$

The derivatives of the partition function will generally depend on the particular choice of penalty operators, but we point out that the gradient of the relative entropy with

respect to β vanishes iff

$$0 = Z(\beta)^{-1}\frac{\partial Z(\beta)}{\partial \beta^k} + \langle M_{k,\theta}\rangle_\rho, \quad k = 1, 2, \ldots, e. \quad (4.21)$$

As in the case of a single penalty operator these conditions are equivalent to fixing each β^k such that the expected penalty agrees between the Gibbs state and the actual state ρ:

$$\langle M_{k,\theta}\rangle_\rho = \langle M_{k,\theta}\rangle_{\chi_{\theta,\beta_*}}, \quad k = 1, 2, \ldots, e. \quad (4.22)$$

This then answers our previous question from Section III B of how to optimally choose the weights β. The joint minimization of the quantum relative entropy over both β and θ can also be understood as *fitting* a generalized Gibbs state parametrized by (β, θ) to ρ.

V.   DYNAMICS IN A MOVING BASIS

In this section we describe how to derive coupled equations of motion for the transformation coordinates and the low complexity quantum state such that the coordinates rapidly converge either exactly or approximately to an optimal value.

A.   Pure state dynamics

We will start by discussing the case of pure state evolution, induced by a (stochastic) Schrödinger equation

$$|d\psi_t\rangle = -idG_t |\psi_t\rangle. \quad (5.1)$$

The generator for a closed system and deterministic dynamics is simply $dG = \hbar^{-1}Hdt$, whereas for an open system evolving according to an unnormalized stochastic Schrödinger equation (SSE) we might have

$$dG_t^{(\text{het})} = \left[\hbar^{-1}H - \frac{i}{2}\sum_{k=1}^{f} L_k^\dagger L_k\right]dt + i\sum_{k=1}^{f} dM_t^{k*}L_k, \quad (5.2)$$

where the $\{L_k, k = 1, 2, \ldots, f\}$ are the Lindblad collapse operators. This particular complex diffusive stochastic unraveling of the open system dynamics can be interpreted as a system whose output modes are all measured using heterodyne detection with perfect fidelity. The associated complex heterodyne measurement processes are given by $dM_t^k = \langle L_k\rangle_{\psi_t} dt + dW_{k,t}$, with complex Wiener processes satisfying

$$\mathbb{E}[\Delta W_k(t_0, t_1)\Delta W_l(t_0, t_1)] = 0, \quad (5.3)$$
$$\mathbb{E}[\Delta W_k^*(t_0, t_1)\Delta W_l(t_0, t_1)] = \delta_{kl}|t_1 - t_0| \quad (5.4)$$

where

$$W_k(t) := \int_0^t dW_{k,t'}, \quad (5.5)$$

$$\Delta W_k(t_0, t_1) := W_k(t_1) - W_k(t_0) = \int_{t_0}^{t_1} dW_{k,t'}. \quad (5.6)$$

Instead of the complex valued measurement process we could also consider single quadrature homodyne measurements. This yields

$$dG_t^{(\text{hom})} = \left[\hbar^{-1}H - \frac{i}{2}\sum_{k=1}^{f}\left(L_k^\dagger L_k + L_k^2\right)\right]dt$$
$$+ i\sum_{k=1}^{f} dM_t^{(\text{hom}),k}L_k. \quad (5.7)$$

Here the measurement processes are real valued $dM_t^{(\text{hom}),k} = \left\langle L_k + L_k^\dagger\right\rangle_{\psi_t} dt + dY_{k,t}$ with real Wiener increments satisfying

$$\mathbb{E}[\Delta Y_k(t_0, t_1)\Delta Y_l(t_0, t_1)] = \delta_{kl}|t_1 - t_0|, \quad (5.8)$$

where analogously we have introduced

$$Y_k(t) := \int_0^t dY_{k,t'}, \quad (5.9)$$

$$\Delta Y_k(t_0, t_1) := Y_k(t_1) - Y_k(t_0) = \int_{t_0}^{t_1} dY_{k,t'}. \quad (5.10)$$

In either case, it is essential to represent the SSE in the Stratonovich picture as this enables us to carry out stochastic projections [39]. Contrary to the usual notation we will always take XdY to indicate a Stratonovich stochastic differential as opposed to an Ito differential.

We could also consider situations where some output channels are measured via homodyne and some via heterodyne measurements and we can even generalize to dynamics with discrete jumps such as the quantum jump trajectories encountered when modeling direct photon detection.

As outlined above the state vector $|\psi_t\rangle$ in the fixed basis is related to the reduced complexity state vector $|\phi_t\rangle$ via

$$|\psi_t\rangle = U_{\theta_t} |\phi_t\rangle \Leftrightarrow |\phi_t\rangle = U_{\theta_t}^\dagger |\psi_t\rangle. \quad (5.11)$$

It then follows that the transformed state evolves according to a modified SSE

$$|d\phi_t\rangle = -idK_{t,\theta_t} |\phi_t\rangle, \quad (5.12)$$

with effective generator

$$dK_{t,\theta_t} := \underbrace{U_{\theta_t}^\dagger dG_t U_{\theta_t}}_{=:dG_{t,\theta_t}} - \sum_{j=1}^{n} F_j^>(\theta_t)d\theta_t^j. \quad (5.13)$$

We see that the transformed state has dynamics generated not only by the transformed SSE generator dG_{t,θ_t} but also by the explicit time dependence of the unitary mapping $-\sum_{j=1}^{n} F_j^>(\theta_t)d\theta_t^j$. Note that the coordinate dynamics will in general be stochastic. If we were working in the Ito picture, dK_{t,θ_t} would have to include terms induced by second order stochastic differentials $d\theta^j d\theta^k$

(and even $d\theta^k dM^{j*}$) for which the Ito table is a priori unknown. Thus, working in the Stratonovich picture is essential as it allows us to manipulate the processes with the product and chain rules of ordinary calculus. As of yet, we have not specified $d\theta_t^j$. In Section V C we will derive a family of dynamics that constrain the moving basis state exactly or approximately to a minimum of a complexity functional as discussed in Section III.

For pure states the expectation of any hermitian observable $X = X^\dagger$ evaluated according to the moving basis state ϕ_t evolves as

$$d \langle X \rangle_{\phi_t} = -2 \operatorname{Im} \left[\sigma \left(dK_{t,\theta_t}, X \right)_{\phi_t} \right] \tag{5.14}$$

$$= -2 \operatorname{Im} \left[\sigma \left(dG_{t,\theta_t}, X \right)_{\phi_t} \right] \tag{5.15}$$

$$- i \sum_{j=1}^{n} \left\langle \left[F_j^> (\theta_t), X \right] \right\rangle_{\phi_t} d\theta^j.$$

In deriving this, it was taken into account that (5.12) does not preserve the norm. Note that when X explicitly depends on time or θ_t one needs to add additional partial derivatives accordingly. We have used Percival's notation [8] for the *quantum correlation*

$$\sigma \left(A, B \right)_{\phi_t} := \left\langle A^\dagger B \right\rangle_{\phi_t} - \left\langle A \right\rangle_{\phi_t}^* \left\langle B \right\rangle_{\phi_t},$$

which defines a semi-definite inner product on the space of operators (cf. Appendix B).

B. Mixed state dynamics

Here we discuss the case of mixed quantum states and deterministic or stochastic Lindblad master equations which are necessary when some of the output channels are unobserved or some measurements have non-ideal fidelity. Note that usually imperfect measurement can be modeled by splitting the output channel on a beamsplitter with transmissivity equal to the fidelity and then observing the transmitted output with perfect fidelity and the reflected output not at all.

The generalization to stochastic master equations is straightforward. In general the fixed basis state ρ_t may evolve according to

$$d\rho_t = -idG_t\rho_t + i\rho_t dG_t^\dagger \tag{5.16}$$

$$+ \sum_{j=1}^{f'} \left[c_j \rho_t c_j^\dagger - \frac{1}{2} \{c_j^\dagger c_j, \rho_t\} \right] dt$$

where dG_t is defined exactly as in the pure state case and the additional Lindblad collapse operators $\{c_1, c_2, \ldots, c_{f'}\}$ correspond to unobserved output channels. The moving basis dynamics follow from a straight-

forward extension of the pure state case

$$d\sigma_t = -idG_{t,\theta_t}\sigma_t + i\sigma_t dG_{t,\theta_t}^\dagger \tag{5.17}$$

$$+ i \left[\sum_{j=1}^{n} F_j^> (\theta_t) d\theta_t^j, \sigma_t \right] dt$$

$$+ \sum_{j=1}^{f'} \left[c_{j,\theta_t} \sigma_t c_{j,\theta_t}^\dagger - \frac{1}{2} \{c_{j,\theta_t}^\dagger c_{j,\theta_t}, \sigma_t\} \right] dt \tag{5.18}$$

where $c_{j,\theta_t} := U_{\theta_t}^\dagger c_j U_{\theta_t}$. In the mixed state case, the expectations of Hermitian observables evolve as

$$d \langle X \rangle_{\sigma_t} = -2 \operatorname{Im} \left[\sigma \left(dG_{t,\theta_t}, X \right)_{\sigma_t} \right] \tag{5.19}$$

$$- i \sum_{j=1}^{n} \left\langle \left[F_j^> (\theta_t), X \right] \right\rangle_{\sigma_t} d\theta^j$$

$$+ \sum_{j=1}^{f'} \operatorname{Re} \left[\left\langle [c_{j,\theta_t}^\dagger, X] c_{j,\theta_t} \right\rangle_{\sigma_t} \right] dt.$$

Here we have overloaded the notation for the quantum correlation [8] for mixed states in the natural way, i.e.,

$$\sigma \left(A, B \right)_{\sigma_t} := \left\langle A^\dagger B \right\rangle_{\sigma_t} - \left\langle A \right\rangle_{\sigma_t}^* \left\langle B \right\rangle_{\sigma_t},$$

We thus find almost the same result as for pure states except for an additional contribution $\operatorname{Re} \left[\left\langle [c_{j,\theta_t}^\dagger, X] c_{j,\theta_t} \right\rangle_{\sigma_t} \right]$ for each unobserved output channel.

C. Gradient flow coupled dynamics

Assume now that we have fixed a complexity functional with Hessian $(h_{jk}(\theta_t))$ and gradient $(y_j(\theta_t))$ and that at a given time t we are starting at optimal coordinates, i.e., we have already solved for θ_t such that the complexity gradient $y_j(\theta_t) = 0$, $j = 1, 2, \ldots, n$. Then we can determine the coordinate increments $d\theta_t^j$ by requiring $y_j(\theta_{t+dt}) = y_j(\theta_t) + dy_j(\theta_t) = 0$. We may then derive the coordinate dynamics by computing the differential change of the gradient coefficients $dy_j(\theta_t)$ as a function of $|d\psi_t\rangle$ and $d\theta_t$ solve for $d\theta_t$ such that $dy_j(\theta_t) = 0$.

More generally, if we assume that we are not starting exactly at optimal coordinates but close to the optimum, then we can instead choose a decay parameter $\eta > 0$ and solve for $d\theta$ such that

$$dy_j(\theta_t) \overset{!}{=} -\eta \, y_j(\theta_t) \, dt, \quad j = 1, 2, \ldots, n. \tag{5.20}$$

This reduces to the above case when we are already at the optimum coordinates, but for a good choice of η it leads to increased robustness to slight deviations as they are exponentially damped over time.

Inserting $X = Y_j^>(\theta)$ into (5.19) while accounting for its explicit dependence on θ we obtain

$$dy_j(\theta_t) = \sum_{k=1}^n h_{jk}(\theta_t)d\theta_t^k - dq_j(\theta_t), \qquad (5.21)$$

where we have defined the *bias flow*

$$dq_j(\theta_t) := 2\,\mathrm{Im}\left[\sigma\left(dG_{t,\theta_t}, Y_j^>(\theta_t)\right)_{\sigma_t}\right] \qquad (5.22)$$

$$-\sum_{l=1}^{f'} \mathrm{Re}\left[\left\langle [c_{l,\theta_t}^\dagger, Y_j^>(\theta_t)]c_{l,\theta_t}\right\rangle_{\sigma_t}\right] dt.$$

The second contribution to the bias flow only arises for mixed state dynamics with unobserved output channels.

Combining Equations (5.20) and (5.22) we find for $j = 1, 2, \ldots, n$

$$\sum_{k=1}^n h_{jk}(\theta_t)d\theta_t^k = dq_j - \eta\, y_j(\theta_t)\, dt. \qquad (5.23)$$

Assuming a strictly convex complexity functional and thus a positive definite Hessian $h(\theta_t) = (h_{jk}(\theta_t))_{j,k=1}^n$ this relationship can be solved for the coordinate differentials

$$d\theta_t^k = \sum_{j=1}^n \left[h(\theta_t)^{-1}\right]^{kj} [dq_j - \eta\, y_j(\theta_t)\, dt], \qquad (5.24)$$

for $k = 1, 2, \ldots, n$. We point out that the bias flow is linear in dG_{t,θ_t} and has a single contribution per c_{j,θ_t} which simplifies its derivation.

We will briefly review the independent contributions of Hamiltonian and dissipative deterministic contributions as well as stochastic terms in dG_{t,θ_t}. For a Hamiltonian and thus hermitian contribution $\hbar^{-1}H dt = \hbar^{-1}H^\dagger dt$ we find

$$dq_{H,j} = -i\left\langle [\hbar^{-1}H, Y_j^>(\theta_t)]\right\rangle_{\sigma_t} dt. \qquad (5.25)$$

For an anti-hermitian dissipative drift term $V dt = \frac{1}{2i}L^\dagger L dt$ we find

$$dq_{V,j} = 2\,\mathrm{Im}\left[\sigma\left(\frac{1}{2i}L^\dagger L, Y_j^>(\theta_t)\right)_{\sigma_t}\right] dt \qquad (5.26)$$

$$= \mathrm{Re}\left[\sigma\left(L^\dagger L, Y_j^>(\theta_t)\right)_{\sigma_t}\right] dt \qquad (5.27)$$

$$= \mathrm{cov}\left(L^\dagger L, Y_j^>(\theta_t)\right)_{\sigma_t} dt. \qquad (5.28)$$

Finally, each complex diffusion term $dQ = i dM^* L$ contributes

$$dq_{dQ,j} = 2\,\mathrm{Im}\left[\sigma\left(i dM^* L, Y_j^>(\theta_t)\right)_{\sigma_t}\right] \qquad (5.29)$$

$$= -2\,\mathrm{Re}\left[\sigma\left(dM^* L, Y_j^>(\theta_t)\right)_{\sigma_t}\right] \qquad (5.30)$$

Considered as a function of σ_t and θ_t this can be used to derive the corresponding Ito SDEs which, however, are in general quite complicated.

D. Gradient coupled fiducial state dynamics

If the system state remains localized near a semi-classical manifold of generalized Gibbs states $U_\theta \Omega U_\theta^\dagger$, then it may be advantageous to evaluate the Hessian $(h_{jk}(\theta_t))$ and the bias flow (dq_j) not in the actual moving basis state σ_t but in a reference state $\Omega = Z(\beta)^{-1}e^{-\sum_{l=1}^e \beta^l M_l}$ instead. In this case, Equation (5.24) becomes

$$d\theta^k = \sum_{j=1}^n [h^{(\Omega)}(\theta_t)^{-1}]^{kj}\left[dq_j^{(\Omega)} - \eta\, y_j(\theta_t)\, dt\right], \quad (5.31)$$

for $k = 1, 2, \ldots, n$. The advantage of this is that $[h^{(\Omega)}(\theta_t)^{-1}]^{kj}$ and $dq_j^{(\Omega)}$ are purely functions of θ and other scalar model parameters. Thus, (5.31) can be understood as semi-classical equations of motion for the coordinates coupled to the true quantum state via the gradient $\eta\, y_j(\theta_t)\, dt$.

E. Information projected fiducial state dynamics

It turns out that for an expectation minimization functional the quantum state independent contributions to the dynamics $d\theta^{k,\mathrm{sc}} := \sum_{j=1}^n [m^{(\Omega)}(\theta_t)^{-1}]^{kj} dq_j^{(\Omega)}$ are fully equivalent to the manifold projection method proposed in [40] except that the metric on the tangent space is taken to be the Kubo-Mori metric [41] that naturally arises as a generalized quantum Fisher information metric associated with differential increases in quantum relative entropy [38]. In fact, in this setting one may assume the complexity weights β to be time dependent and derive coupled dynamics for them, as well. We discuss a special case of this in the following.

F. Gibbs manifold projection

Here we briefly sketch how to use our framework to obtain approximate low dimensional dynamics that arise when the state ρ_t is constrained to the Gibbs-manifold parametrized by the group transformation coordinates θ_t and the complexity weights β_t at all times. For a Gibbs state $\chi_{\theta,\beta} = Z(\beta)^{-1}\exp\left(-\sum_{k=1}^e \beta^k M_{k,\theta}\right)$ both the quantum relative entropy as well as its linear variation $\sum_{k=1}^e z_k \delta\beta^k + \sum_{k=1}^n y_k \delta\theta^k$ (cf. Equations (4.14) and (4.16)) with respect to θ and β vanish because it is already on the Gibbs manifold we project to. We can now apply the full procedure outlined in [40] to also project the stochastic dynamics of ρ_t onto the manifold. Generally this amounts to choosing $d\theta$ and $d\beta$ such that $dy = (dy_1, \ldots, dy_n)$ and $(dz_1, \ldots, dz_e)$ are orthogonal to the tangent space spanned by $d\theta$ and $d\beta$ with respect to the Kubo-Mori associated with the second order variation of the quantum relative entropy.

In the interest of brevity we simply present the result of that derivation. The coordinate increments $d\theta$ and $d\beta$ are the implicit solutions of the following linear system

$$\sum_{l=1}^{n} h_{jl} d\theta_t^l + \sum_{k=1}^{e} y_{kj} d\beta^k = dq_j \qquad (5.32)$$

$$\sum_{j=1}^{n} y_{kj} d\theta^j + \sum_{k=1}^{e} g_{km} d\beta_t^m = dm_k \qquad (5.33)$$

with h_{jl}, y_{kj} and g_{km} defined as in Equations (4.15) - (4.18) for $\rho \equiv \chi_{\theta,\beta}$. The flow vectors $dq = (dq_1, \ldots, dq_n)$ and $dm = (dm_1, \ldots, dm_e)$ are given by

$$dq_j := 2 \operatorname{Im} \left[\sigma \left(dG_{t,\theta_t}, Y_j^{>}(\theta_t) \right)_{\chi_{\beta_t}} \right] \qquad (5.34)$$

$$- \sum_{l=1}^{f'} \operatorname{Re} \left[\left\langle [c_{l,\theta_t}^\dagger, Y_j^{>}(\theta_t)] c_{l,\theta_t} \right\rangle_{\chi_{\beta_t}} \right] dt$$

$$dm_k := 2 \operatorname{Im} \left[\sigma \left(dG_{t,\theta_t}, M_k \right)_{\chi_{\beta_t}} \right] \qquad (5.35)$$

$$- \sum_{l=1}^{f'} \operatorname{Re} \left[\left\langle [c_{l,\theta_t}^\dagger, M_k] c_{l,\theta_t} \right\rangle_{\chi_{\beta_t}} \right] dt,$$

both of which can be evaluated in the moving basis via the diagonal Gibbs state χ_β for which all off-diagonal operator moments conveniently vanish. In many cases of interest the above expressions can be evaluated analytically as all operator moments become explicit functions of only β and θ, yielding coupled, low dimensional ODEs or SDEs that approximately describe the original quantum dynamics. We note that the transformation coordinates θ can often be directly associated with classical or semi-classical quantities such as generalized canonical position and momentum variables, but the complexity weights β have a purely statistical interpretation as they encapsulate the uncertainty originally encoded in the exact quantum state.

VI. EXAMPLES OF MANIFOLDS

Here we present some examples of transformation groups, penalty operators and the corresponding generators and sensitivity variables. A given transformation U_θ is characterized by its differential form $U_{\theta_t}^\dagger dU_{\theta_t} = -i \sum_{j=1}^{n} F_j^{>}(\theta_t) d\theta_t^j$, its adjoint action on its own Lie-Algebra $U_{\theta_t}^\dagger X U_{\theta_t}$ and the sensitivity operators whose expectations form the gradient and Hessian of a particular expectation minimization problem. The sensitivity operators can be derived for any given complexity functional.

A. Coherent displacement

The simplest example is that of coherent displacements $\theta = (Q, P)$. Working in this representation we have

$$U_\theta = e^{-iQp + iPq}, \qquad U_\theta^\dagger a U_\theta = a + \frac{Q + iP}{\sqrt{2}}, \qquad (6.1)$$

where the generators and right generators are given by

$$q = \frac{a + a^\dagger}{\sqrt{2}}, \qquad p = \frac{a - a^\dagger}{\sqrt{2}i}, \qquad (6.2)$$

$$F_1^{>}(\theta_t) = p + P/2, \qquad F_2^{>}(\theta_t) = -q - Q/2. \qquad (6.3)$$

Coherent displacements of different modes commute and therefore the extension to multiple oscillator modes follows trivially.

1. Linear excitation minimization

For the canonical penalty operator given simply by the photon number operator $M = N := a^\dagger a$ the gradient and Hessian operators are

$$Y_1^{>}(\theta_t) = -q, \quad Y_2^{>}(\theta_t) = -p \qquad (6.4)$$

$$\left(H_{jk}^{>}(\theta_t) \right) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad (6.5)$$

We see that the Hessian is constant and strictly positive definite. Since the domain of the coordinates is $\mathbb{R}^2$ and thus clearly convex and closed, we always have a unique optimum.

2. Nonlinear excitation minimization

Assuming a nonlinearly transformed penalty operator $M = f(N)$ the gradient operators are

$$Y_1^{>}(\theta_t) = i[f(N), p] = \frac{[a^\dagger, f(N)] - [a, f(N)]}{\sqrt{2}} \qquad (6.6)$$

$$= -\frac{a^\dagger \Delta_f(N) + \Delta_f(N) a}{\sqrt{2}}, \qquad (6.7)$$

$$Y_2^{>}(\theta_t) = i[q, f(N)] = -i\frac{a^\dagger \Delta_f(N) - \Delta_f(N) a}{\sqrt{2}} \qquad (6.8)$$

where we have introduced

$$\Delta_f(N) := f(N+1) - f(N), \qquad (6.9)$$

which allows us to write down two useful (and equivalent) rules

$$[a^\dagger, f(N)] = -a^\dagger \Delta_f(N) \Leftrightarrow [a, f(N)] = \Delta_f(N) a. \qquad (6.10)$$

The Hessian operators are then given by

$$H_{11}^>(\theta_t) = \Delta_f(N) + a^\dagger \Delta_f^{(2)}(N)a \qquad (6.11)$$
$$+ \frac{a^{\dagger 2}\Delta_f^{(2)}(N) + \Delta_f^{(2)}(N)a^2}{2},$$

$$H_{12}^>(\theta_t) = H_{21}^>(\theta_t) = -\frac{a^{\dagger 2}\Delta_f^{(2)}(N) - \Delta_f^{(2)}(N)a^2}{2i}, \qquad (6.12)$$

$$H_{22}^>(\theta_t) = \Delta_f(N) + a^\dagger \Delta_f^{(2)}(N)a \qquad (6.13)$$
$$- \frac{a^{\dagger 2}\Delta_f^{(2)}(N) + \Delta_f^{(2)}(N)a^2}{2}.$$

Here the second order differences have been introduced:

$$\Delta_f^{(2)}(N) := \Delta_f(N+1) - \Delta_f(N) = \Delta_{\Delta_f}(N). \qquad (6.14)$$

To analyze under what conditions the Hessian may be positive, observe first that the Hessian operator matrix $H^> := (H_{jk}^>(\theta_t))$ factors as

$$H^> = V^\dagger \tilde{H}^> V, \qquad (6.15)$$

where we have defined

$$\tilde{H}^> := \begin{pmatrix} \Gamma & \Sigma \\ \Sigma^\dagger & \Gamma \end{pmatrix}, \qquad (6.16)$$

$$V := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} = (V^\dagger)^{-1} \qquad (6.17)$$

via

$$\Gamma := \Delta_f(N) + a^\dagger \Delta_f^{(2)}(N)a \qquad (6.18)$$
$$\Sigma := \Delta_f^{(2)}(N)a^2. \qquad (6.19)$$

V may be interpreted as a linear transformation from a real valued quadrature representation of the displacement $\delta\theta = (\delta Q, \delta P)^T$ to complex amplitudes $(\alpha, \alpha^*)^T = V\delta\theta$. Since V contains only scalars, the matrices $h, \tilde{h}$ of element-wise expectations, i.e.,

$$h_{jk} = \left\langle H_{jk}^>(\theta_t)\right\rangle_\sigma, \quad \tilde{h}_{jk} = \left\langle \tilde{H}_{jk}^>(\theta_t)\right\rangle_\sigma, \qquad (6.20)$$

are related by a similarity transform $h = V^\dagger \tilde{h}V$ and therefore have the same eigenvalues. Equipped with the above definitions we are thus able to formulate and prove the following

Theorem: A sufficient condition for the positive semi-definiteness of the Hessian h is that the following operators are non-negative

$$0 \le \Delta_f^{(2)}(N-1) \qquad (6.21)$$
$$0 \le \Gamma - \Xi \qquad (6.22)$$
$$0 \le \Gamma \qquad (6.23)$$

where Γ is defined as above and

$$\Xi := \frac{a^\dagger\left[\Delta_f^{(2)}(N+1) + \Delta_f^{(2)}(N-1)\right]a + \Delta_f^{(2)}(N)}{2}. \qquad (6.24)$$

Proof: See Appendix C.

This theorem is helpful because the operators appearing in its conditions are all diagonal in the eigenbasis of N, which implies that we can verify the positivity simply by evaluating the inequalities for each eigenstate of N. Note furthermore that the non-negativity of Γ is also a necessary condition and that when $\Gamma - \Xi > 0$ is strictly positive, the Hessian is strictly positive.

Sometimes we only wish to determine whether the Hessian is positive when the states $\sigma = \Pi\sigma\Pi$ are restricted to a subspace of the full state space characterized by a projection operator Π. In this case the operator inequalities must only hold for the projected operators $\Pi\Sigma\Pi, \Gamma\Pi, \Xi\Pi$.

Applying the above theorem to some interesting test cases we find that $f(n) = n^2$ leads to a convex objective, as this leads to $\Delta_f^{(2)}(N-1) \ge 0, \Gamma = 1 + 4N, \Xi = 1 + 2n$.

For $f(n) = e^{\lambda n}$ with $\lambda > 0$, which is relevant for CGF minimization, we find that $\Gamma - \Xi$ is not positive everywhere, but for small enough λ there exists an $n_* = \left\lceil (3x^2 - x^3)/(-1+x)^3\right\rceil\big|_{x=e^\lambda}$ such that at least on the subspace of number states lower than n_* the condition holds: $\Pi_{N \le n_*}(\Gamma - \Xi) \ge 0$. This upper bound on N becomes arbitrarily large as λ goes to zero.

B. Squeezing and displacement

Our next example will be a mixture of squeezing and displacement albeit each only parametrized by a single variable $\theta = (Q, R)$, which does not allow to realize the most general pure Gaussian state.

$$U_\theta = e^{-iQp}e^{iRs}, \quad U_\theta^\dagger a U_\theta = \frac{Q}{\sqrt{2}} + \cosh Ra - \sinh Ra^\dagger \qquad (6.25)$$

with generators

$$p = \frac{a - a^\dagger}{\sqrt{2}i}, \qquad s = \frac{a^2 - a^{\dagger 2}}{2i}, \qquad (6.26)$$
$$F_1^>(\theta_t) = e^R p, \qquad F_2^>(\theta_t) = -s. \qquad (6.27)$$

For the canonical complexity functional $N = a^\dagger a$ the sensitivity operators are

$$Y_1^>(\theta_t) = -e^R q, \quad Y_2^>(\theta_t) = 2r = a^2 + a^{\dagger 2} \qquad (6.28)$$

$$\left(H_{jk}^>(\theta_t)\right) = \begin{pmatrix} e^{2R} & -e^R q \\ -e^R q & 2 + 4a^\dagger a \end{pmatrix}. \qquad (6.29)$$

This still leads to a positive definite Hessian, but it is now dependent on the coordinates and the state. It is straightforward to derive the more general case of arbitrary displacements and squeezing, but the expressions become more tedious. They are implemented in our software package QMANIFOLD [27].

C. Spin coherent displacement

Our final example here is for angular momentum states and a single irreducible representation labeled by J such that $\mathbf{J}^2 = J(J+1)$ and the spectrum of J_z is given by $-J, -J+1, \ldots, J-1, J$. The commutator relationships are $[J_z, J_\pm] = \pm J_\pm$ and $[J_+, J_-] = 2J_z$. The ladder operators can also be represented in terms of the x-y generators $J_\pm = J_x \pm iJ_y$. We can then define a unitary Lie group explicitly via

$$U_\theta = e^{-\mu J_+} e^{-\log(1+|\mu|^2)J_z} e^{\mu^* J_-}, \tag{6.30}$$

$$U_\theta^\dagger J_- U_\theta = \frac{J_- - \mu^2 J_+ - 2\mu J_z}{1+|\mu|^2}, \tag{6.31}$$

$$U_\theta^\dagger J_z U_\theta = \frac{\mu^* J_- + \mu J_+ + (1-|\mu|^2)J_z}{1+|\mu|^2} \tag{6.32}$$

where $\mu = \theta^1 + i\theta^2$ (6.33)

$$F_1^>(\theta_t) = \frac{-iJ_- + iJ_+ - 2\theta^2 J_z}{1+|\mu|^2} \tag{6.34}$$

$$F_2^>(\theta_t) = \frac{-J_- - J_+ + 2\theta^1 J_z}{1+|\mu|^2}. \tag{6.35}$$

The action of this unitary on a fixed reference state, e.g., $U_\theta |j_z = -J\rangle$ can be identified with the spin coherent states as introduced by Radcliffe [42]. For the canonical spin complexity functional $N = J_z$ the sensitivity operators are

$$Y_1^>(\theta_t) = \frac{-J_- - J_+}{1+|\mu|^2}, \quad Y_2^>(\theta_t) = \frac{iJ_- - iJ_+}{1+|\mu|^2} \tag{6.36}$$

$$\left(H_{jk}^>(\theta_t)\right) = \frac{4}{1+|\mu|^2} \begin{pmatrix} \mathrm{Re}\,[\mu J_-] - J_z & \mathrm{Im}\,[\mu J_-] \\ \mathrm{Im}\,[\mu J_-] & -\mathrm{Re}\,[\mu J_-] - J_z \end{pmatrix}, \tag{6.37}$$

where $\mathrm{Re}\,[\mu J_-] = (\mu J_- + \mu^* J_+)/2$ and $\mathrm{Im}\,[\mu J_-] = (\mu J_- - \mu^* J_+)/2i$. This does not generally lead to a positive definite Hessian, e.g., consider the $j_z = +J$ eigenstate of J_z for which the expected Hessian is clearly negative definite. For any given state the eigenvalues of $\left(\left\langle H_{jk}^>(\theta_t)\right\rangle_\sigma\right)$ are given by

$$\lambda_\pm = \frac{4}{1+|\mu|^2}\left[-\langle J_z\rangle_\sigma \pm |\mu|\sqrt{\langle J_x\rangle_\sigma^2 + \langle J_y\rangle_\sigma^2}\right] \tag{6.38}$$

Assuming that $\langle J_z\rangle_\sigma < 0$, these are non-negative iff

$$|\mu|^2\left[\langle J_x\rangle_\sigma^2 + \langle J_y\rangle_\sigma^2\right] \leq \langle J_z\rangle_\sigma^2. \tag{6.39}$$

Two sufficient but not necessary conditions for this are given by

$$|\mu|^2 < \frac{\langle J_z\rangle_\sigma^2}{J(J+1) - \langle J_z^2\rangle_\sigma}, \tag{6.40}$$

and a more restrictive version that depends only on $\langle J_z\rangle_\sigma^2$.

$$|\mu|^2 < \frac{\langle J_z\rangle_\sigma^2}{J(J+1) - \langle J_z\rangle_\sigma^2}. \tag{6.41}$$

This follows directly from $\langle J_k\rangle_\sigma^2 \leq \langle J_k^2\rangle_\sigma$ for $k = x, y, z$ and from $\sum_{k\in\{x,y,z\}} J_k^2 = \mathbf{J}^2 = J(J+1)$.

VII. COORDINATE FREE METHOD

The methods we have introduced in the preceding section are quite appealing in that they can allow us to simulate significantly larger open quantum systems than is possible with static bases. They can also yield analytic insight into semi-classical dynamics and provide more intuition. With the analytic results presented in Section VI one can construct commuting product transformations for fairly complex systems, but they are unlikely to work well when considering non-factoring group manifolds of coordinate dimensions beyond $O(10)$ as it becomes very difficult to obtain a fully exponentiated parametrized transform from which we can derive the right generators $\{F_j^>(\theta_t)\}$. These, however, are essential to the derivation of most other important quantities. Moreover, having an explicit coordinate based representation of the unitary U_θ is useful for transforming states from the static to the moving basis. There are, however, examples of groups that can at least in principle be employed without ever having to derive an explicit parametrization of U_θ. Here we sketch out how this can be achieved in principle.

Consider a Lie group and a state space and dynamics for which

1. the model operators that generate the dynamics H, L and c can be represented as polynomials of elements of a group's finite dimensional Lie algebra, and

2. a full Hilbert space basis can be generated by repeated action of some raising operators $\{A_{k,+}\}$ that are inside the Lie algebra starting from a unique reference state $|\Omega\rangle$, which is itself fully characterized as the unique zero-eigenvalue eigenstate of an operator $M|\Omega\rangle = 0$, where M is also either inside the group's Lie algebra or a polynomial of Lie algebra elements.

In this case we may use the adjoint representation of the group transform itself as the parametrization, i.e., if $\{Y_1, Y_2, \ldots, Y_q\} \subset \mathfrak{g}$ is a hermitian basis of the Lie algebra, then if a group element g is unitarily represented on the Hilbert space as $U(g)$ there also exists a corresponding element in the adjoint representation of the group $S(g)$ such that $U(g)Y_k U(g)^\dagger = \sum_{j=1}^q S_k^j(g)Y_j$. This adjoint representation will typically require only a polynomial number of parameters in the system size. In some cases where a part of the Lie algebra can be expressed as polynomials of some other basis elements further reduction is possible, e.g., consider two bosonic modes a, b and a Lie algebra that also contains $a^\dagger b$ or ab. In this case, if we know how a and b transform under U then we also know how $a^\dagger b$ or ab transform.

If we start at $t = 0$ with the identity $g = e$ and $S_k^j(e) = \delta_k^j$, then at each time step we can consider arbitrary elements from the Lie algebra and write $U(g)^\dagger \delta U(g) = -i \sum_{j=1}^q \delta \mu^j Y_j$. This implies that

$$\delta[U(g)Y_k U(g)^\dagger] = U(g)U(g)^\dagger \delta U(g) Y_k U(g)^\dagger \quad (7.1)$$
$$+ U(g)Y_k \delta U(g)^\dagger U(g) U(g)^\dagger$$

$$= -i \sum_{j=1}^q \delta \mu^j U(g)[Y_j, Y_k]U(g)^\dagger \quad (7.2)$$

$$= -i \sum_{j,l=1}^q \delta \mu^j c_{jk}^l U(g) Y_l U(g)^\dagger \quad (7.3)$$

$$= -i \sum_{j,l,m=1}^q \delta \mu^j c_{jk}^l S_l^m(g) Y_m \quad (7.4)$$

$$= \sum_{m=1}^q \delta S_k^m(g) Y_m. \quad (7.5)$$

Here we have used the convention for the structure constants given in Equation A3 and the final two lines allow us to read off the differential change of the adjoint transformation

$$\delta S_k^m(g) = -i \sum_{j,l=1}^q \delta \mu^j c_{jk}^l S_l^m(g). \quad (7.6)$$

Armed with this, we can now reformulate the complexity reduction problem. Assume that we wish to find a transform U_* such that it minimizes a function $\mathcal{J}_M(U) = \langle UMU^\dagger \rangle_\rho$, i.e.,

$$U_* = \mathrm{argmin}_U \mathcal{J}_M(U). \quad (7.7)$$

Expanding this to second order in a a perturbation $\delta U = -i \delta \mu^j Y_j$ we find

$$\mathcal{J}_M(U + \delta U) = \langle UMU^\dagger \rangle_\rho + \sum_{j=1}^q \underbrace{\left\langle U \frac{[Y_j, M]}{i} U^\dagger \right\rangle_\rho}_{y_j} \delta \mu^j$$

$$(7.8)$$

$$+ \frac{1}{2} \sum_{j,k=1}^q \underbrace{\langle U[Y_{\{k}, [M, Y_{j\}}]]U^\dagger \rangle_\rho}_{h_{kj}} \delta \mu^j \delta \mu^k$$

$$+ O(\delta \mu^3).$$

Here we have used the notation $[Y_{\{k}, [M, Y_{j\}}]]$ to indicate that the expression is symmetrized over k and j, i.e., $[Y_{\{k}, [M, Y_{j\}}]] \equiv \frac{1}{2} \{[Y_k, [M, Y_j]] + [Y_j, [M, Y_k]]\}$. As above, the gradient and Hessian coefficients can be more efficiently evaluated in the current lowest complexity state $\sigma = U^\dagger \rho U$. Assuming that (h_{jk}) is positive semi-definite at all steps, we can use Newton's method or a similar technique to identify good coordinate updates $\delta \mu$ and update our adjoint representation $S \to S + \delta S$

as specified in (7.6) as well as the transformed state $\sigma = U^\dagger \rho U$ according to

$$\sigma \to \sigma + i \sum_{j=1}^q [Y_j, \sigma] \delta \mu. \quad (7.9)$$

As above, we may also combine this complexity reduction with dynamics of ρ or σ, respectively. The results of our previous sections carry over with little modification. Although the lack of a fixed parametrization of U may cause us to worry about what the specific meaning of σ really is, our initial assumptions guarantee that the adjoint representation matrix S uniquely fixes how our Hamiltonian and all other relevant operators are transformed to the low complexity basis and furthermore guarantees that a representation in terms of the original basis states can at least in principle be recovered in the new basis by transforming $M_U := UMU^\dagger$ using S and then solving for the unique eigenvector $|\Omega_U\rangle = U|\Omega\rangle$ with zero eigenvalue of M_U. All other basis states can then be recovered by acting on $|\Omega_U\rangle$ with the transformed raising operators $\{UA_{k,+}U^\dagger\}$. This prescription should work straightforwardly for more complicated groups such as the $SU(n)$ symmetry groups of multiple bosonic or fermionic modes or the symplectic group $Sp(2n)$ for an ensemble of oscillators or distinguishable particles.

VIII. SUMMARY OF OUR METHODS

In this work we have presented analytical and computational methods that allow to simulate closed and open quantum systems assuming deterministic – unitary or ensemble averaged – or stochastic dynamics affecting either pure or mixed states.

Given a particular dynamical quantum system to be simulated, we suggest the following steps to apply our methods. We assume as the most general case of mixed states and stochastic evolution as described in Equation (V B) defined by a generator dG and a set of unobserved dissipation operators $\{c_j, j = 1, 2, \ldots, f'\}$.

1. Based on intuition or prior simulations, identify the degrees of freedom that are most likely to localize.

2. Define a joint penalty operator $M = \sum_{k=1}^e \beta^k M_k \geq 0$ and choose a set of transformation generators $\{X_j, j = 1, 2, \ldots, n\}$ that are elements of a finite dimensional Lie algebra and use the methods outlined in the Supplementary Material to obtain an exponentiated form of the parametrization U_θ and consequently the right generators $\{F_j^>(\theta_t) = iU_{\theta_t}^\dagger \frac{\partial U_{\theta_t}}{\partial \theta^j}, j = 1, 2, \ldots, n' \leq n\}$. This generally requires explicit exponentiation of matrices in the adjoint representation and can be done using a computer algebra system. In Section VI we have provided a number of examples that can be combined to handle composite systems.

3. Derive the explicitly transformed generators dK_{t,θ_t} as defined in Equation (5.13) and the transformed dissipation operators for the unobserved channels $c_{j,\theta_t} = U_{\theta_t}^\dagger c_j U_{\theta_t}$. This can most easily be done by writing these operators as polynomials of the transformation generators and then applying the explicitly obtained adjoint representation to transform them.

4. Simulate the joint equations of motion for the reduced complexity state σ_t and the coordinates θ_t defined as

$$d\sigma_t = -i dG_{t,\theta_t}\sigma_t + i\sigma_t dG_{t,\theta_t}^\dagger \qquad (8.1)$$

$$+ i\left[\sum_{j=1}^n F_j^>(\theta_t)d\theta_t^j, \sigma_t\right] dt$$

$$+ \sum_{j=1}^{f'}\left[c_{j,\theta_t}\sigma_t c_{j,\theta_t}^\dagger - \frac{1}{2}\{c_{j,\theta_t}^\dagger c_{j,\theta_t}, \sigma_t\}\right] dt \qquad (8.2)$$

$$d\theta_t = h_t^{-1}\left[dq_t - \eta\, y_t dt\right], \qquad (8.3)$$

where the expected gradient y_t and Hessian h_t are defined in Equation (3.14) and where the bias flow dq_t has been defined in Equation (5.22). The gain parameter $\eta \geq 0$ can be freely chosen. A reasonable heuristic to picking a good value is to linearize the equation of motion for as a function of θ_t and choose $\eta \geq \min\{-\operatorname{Re}[\lambda], \lambda \in \operatorname{spec}(J)\}$ where J is the Jacobian of that linearization. The special case of pure state simulations is handled equivalently.

5. The resulting simulated trajectory for the reduced complexity state σ_t can be used to test whether further improvements can be achieved by allowing more general transformations. On the other hand, if an inspection of the coordinate trajectory θ_t reveals that some parameters are nearly constant or more generally the coordinate trajectory is itself confined to some lower dimensional set then it may be possible to remove some degrees of freedom from the transformation via an embedding transformation $\tilde{\theta} \mapsto \theta$ where $\tilde{\theta}$ has fewer coordinates. In general it can be useful to iterate the above steps a few times to identify a good parametrization.

IX. CONCLUSION AND OUTLOOK

As stated at the outset, we have presented a family of methods that allow one to investigate the inherent complexity of quantum states by attempting to reduce the total number of variables required for their description. We provided an information theoretic interpretation of our method and several alternate prescriptions for deriving *exact* coupled dynamics of the semi-classical group coordinates θ and the reduced complexity quantum state σ. Our method includes earlier work by Schack, Brun

and Percival [23] as a special case but provides a substantially larger, analytic framework that can also be applied in cases where their method does not work. Further, we have shown how our method connects to earlier model-reduction work on nonlinear projection of quantum models [40]. Our approach is extremely flexible and suited for arbitrary Lie groups as long as their associated Lie algebras have finite dimension. In practice the explicit coordinate representations will work best for low dimensional Lie-algebras, and we have also outlined how to implement a method that does not rely on an explicit coordinate parameterization.

This work has many promising future directions, some key examples are the application of our simulation scheme to quantum measurement and control problems. In particular, our method is very nicely suited to model noisy dispersive or high power qubit readout of superconducting quantum circuits [43] and it could be used for a more principled approach to the quantum feedback model for autonomous state preparation considered in [44]. It could also provide more rigorous simulations of the quantum effects in coherent optical Ising machines or coherent machine learning devices than existing techniques based on quasi-probabilities [14, 45, 46].

By completely projecting the reduced complexity state to the nearest Gibbs state, our method can also be used to derive further reduced order projected models such as the Maxwell-Bloch type model considered in [40] or a semi-classical coupled mode theory for nonlinear resonators similar to the Wigner method proposed in [11]. It would be very appealing to work out in more detail how our approach could improve our understanding of quantum feedback networks [5, 6] by conceptually separating both the node-systems and the interconnecting fields into quasi-classical and quantum components. A very promising framework for this avenue is given by the recent work of John Gough [47] on controlled flows.

It is also possible to extend the formalism to quantized fields, in particular it may be very useful to apply it to traveling wave fields inside quantum feedback networks with time delays. In this scenario the reduced complexity state for the bath modes could be modeled using Matrix Product States as in [48] or by an approximate delay model [49]. Recent and exciting work by Sarovar et al. [50] addresses a slightly different question, i.e., that of which quantum models are robustly solvable on an analog quantum simulator, but using similar information geometric techniques as our approach. Ultimately, we expect the question of *efficient* representation to be closely related to the question of *robust* representation. Finally, we believe that it is worth investigating whether the inherent complexity of the quantum states a system evolves through can be related to its computational power. A formal framework for classical dynamical input/output systems has been introduced in [51].

Appendix A: General construction of the coordinate transformation

In this secection we outline how to construct complex transformations and derive the right generators. The simplest construction for the transformation is by chaining single parameter transformations

$$U_\eta := V_{\eta^1}^{(1)} V_{\eta^2}^{(2)} \cdots V_{\eta^n}^{(n)}, \tag{A1}$$

$$\text{where } V_{\eta^j}^{(j)} := \exp(-\eta^j X_j), \; j = 1, 2, \ldots, n \tag{A2}$$

Thus far we allow for complex coordinates $\{\eta^j\}$ and arbitrary, i.e., not necessarily hermitian, generators $\{X_j\}$. We assume that the generators X_j are elements of a finite dimensional Lie algebra $\mathfrak{g}$. Given a basis $\{Y_1, Y_2, \ldots, Y_q\} \subset \mathfrak{g}$ for the Lie algebra with structure constants c_{jk}^l implicitly defined via

$$[Y_j, Y_k] = \sum_{l=1}^q c_{jk}^l Y_l, \tag{A3}$$

we represent each transformation generator in this basis as $X_j = \sum_{k=1}^q R_j^k Y_k$. Using the structure constants, it is straightforward to compute the conjugation of any basis element by a single parameter transformation [52] to be

$$V_{\eta^j}^{(j)-1} Y_k V_{\eta^j}^{(j)} = \sum_{l=1}^q \underbrace{[\exp(A_{\eta^j}^{(j)})]_k^l}_{=:S_{\eta^j}^{(j)}} Y_l \tag{A4}$$

$$\text{with } [A_{\eta^j}^{(j)}]_k^l = \eta^j \sum_{h=1}^q R_j^h c_{hk}^l. \tag{A5}$$

The matrices $A_{\eta^j}^{(j)}$ are typically very sparse and can be exponentiated symbolically using a tool such as Mathematica [53] the SymPy package [54]. The resulting matrices $S_{\eta^j}^{(j)} \in \mathbb{C}^{n \times n}$ are elements of the adjoint representation of the transformation group and can be used to directly transform the generators. As they each depend on only a single coordinate, they satisfy $S_{-\eta^j}^{(j)} = \left(S_{\eta^j}^{(j)}\right)^{-1}$. With this, it is straightforward to see that

$$dU_\eta = U_\eta \sum_{j=1}^n X_j^>(\eta) d\eta^j \tag{A6}$$

$$\text{with } X_j^>(\eta) := V_{\eta^n}^{(n)-1} \cdots V_{\eta^{j+1}}^{(j+1)-1} X_j V_{\eta^{j+1}}^{(j+1)} \cdots V_{\eta^n}^{(n)} \tag{A7}$$

$$= \sum_{l=1}^n \sum_{k=1}^q R_j^l [S_{\eta^{j+1}}^{(j+1)} \cdots S_{\eta^n}^{(n)}]_l^k Y_k. \tag{A8}$$

We see that by requiring this particular differential form of U_η, i.e., with all generators on the right hand side, each generator is additionally transformed $X_j \to X_j^>(\eta)$ by all single parameter transformations that appear to its right.

The differential transformation can be equivalently expressed with the differential generators on the left side of U_{η_t}:

$$dU_\eta = \left[\sum_{j=1}^n X_j^<(\eta) d\eta^j\right] U_\eta \tag{A9}$$

$$\text{with } X_j^<(\eta) = U_\eta X_j^>(\eta) U_\eta^{-1} \tag{A10}$$

$$= V_{\eta^1}^{(1)} \cdots V_{\eta^{j-1}}^{(j-1)} X_j V_{\eta^{j-1}}^{(j-1)-1} \cdots V_{\eta^1}^{(1)-1} \tag{A11}$$

$$= \sum_{k=1}^n [S_{-\eta^{j-1}}^{(j-1)} \cdots S_{-\eta^1}^{(1)}]_j^k X_k. \tag{A12}$$

If we now assume that each generator is antihermitian $X_j = iF_j$ and if we restrict the coordinates to real values $\eta^j = \theta^j \in \mathbb{R}$, then the resulting transformation is unitary. This is desirable because observables evaluated in the transformed frame $\langle \phi_t | M' | \phi_t \rangle = \langle \psi_t | U_{\theta_t}^\dagger M U_{\theta_t} | \psi_t \rangle$ are actually simply the conjugated observables $M' = U_{\theta_t}^{-1} M U_{\theta_t}$.

To transform any operator M from the basis associated with $|\psi_t\rangle$ to the moving basis $|\phi_t\rangle$ one first needs to express it exclusively in terms of functions of the generators $M = f(X_1, X_2, \ldots, X_n)$ where $f : \mathbb{C}^n \to \mathbb{C}$ is analytic in all variables, typically f is a polynomial. We then have

$$M'_\theta := U_\theta^\dagger M U_\theta$$
$$= f(X_1'(\theta), X_2'(\theta), \ldots, X_n'(\theta)), \tag{A13}$$

$$\text{with } X_j'(\theta) := \sum_k \underbrace{[S_{\theta^1}^{(1)} \cdots S_{\theta^n}^{(n)}]_j^k}_{=:S_\theta} X_k. \tag{A14}$$

We see that having the adjoint representation single parameter transformation matrices $\{S_{\theta^j}^{(j)}, j = 1, 2, \ldots, n\}$ allows us to do all necessary computations. We remark that representing $M = f(X_1, X_2, \ldots, X_n)$ generally does not imply a unique function f as some generators may themselves be polynomials of the other generators.

Finally, note that there exist alternate ways [55] of parameterizing groups and deriving partial derivatives that may come in useful in more complex cases. Our rules for chained single parameter transformations derived here can straightforwardly generalized to chained multi-parameter transformations.

For very complex parameterizations, analytical/symbolic methods may fail but in that case it should still be possible to work in a purely numerical representation that stores and integrates both θ and elements of the adjoint representation of U_θ.

Appendix B: Properties of the quantum correlation

The quantum correlation is not a strictly positive definite inner product because the quantum self-correlation

of an operator A vanishes in any eigenstate

$$A \left| \phi_t \right\rangle = \lambda_a \left| \phi_t \right\rangle \Leftrightarrow \sigma \left(A, A \right)_{\phi_t} = 0. \qquad \text{(B1)}$$

The sufficiency "$\Rightarrow$" of this condition is obvious, the necessity "$\Leftarrow$" follows from the Cauchy-Schwarz inequality for the regular Hilbert space inner product.

Restricted to hermitian operators $A^\dagger = A, B^\dagger = B$ the quantum correlation can be decomposed into its real and imaginary part as

$$\sigma \left(A, B \right)_{\phi_t} = \text{cov} \left(A, B \right)_{\phi_t} + i \left\langle \frac{[A,B]}{2i} \right\rangle_{\phi_t},$$

with the symmetrized covariance function

$$\text{cov} \left(A, B \right)_{\phi_t} = \frac{1}{2} \left\langle \{A, B\} \right\rangle_{\phi_t} - \left\langle A \right\rangle_{\phi_t} \left\langle B \right\rangle_{\phi_t}.$$

Appendix C: Proof of our theorem

Since Γ is non-negative, the trace of $\tilde{h}$ is non-negative: $\text{Tr}\left(\tilde{h} \right) = 2 \left\langle \Gamma \right\rangle_\sigma \geq 0$. Due to the hermiticity of $\tilde{h}$ its eigenvalues must be real and it thus suffices to prove that its determinant is non-negative. For any non-negative operator $A \geq 0$ we must have $0 \leq \left\langle (a + a^\dagger) A (a + a^\dagger) \right\rangle_\sigma$ for any state σ. Inserting $A = \Delta_f^{(2)}(N - 1)$ – which is positive by the first condition of our theorem – this yields

$$0 \leq 2 \, \text{Re} \left[\left\langle \underbrace{\Delta_f^{(2)}(N) a^2}_{\Sigma} \right\rangle_\sigma \right] \qquad \text{(C1)}$$
$$+ \left\langle \underbrace{a^\dagger \left[\Delta_f^{(2)}(N+1) + \Delta_f^{(2)}(N-1) \right] a + \Delta_f^{(2)}(N)}_{\Xi} \right\rangle_\sigma.$$

This inequality is satisfied not only for σ but also for any unitarily rotated state $\sigma(\phi) := e^{i\phi N} \sigma e^{-i\phi N}$. The second expectation value in (C1) is invariant under any such transformation, whereas it imparts a complex phase on the first expectation value.

$$\left\langle \underbrace{\Delta_f^{(2)}(N) a^2}_{\Sigma} \right\rangle_{\sigma(\phi)} = e^{2i\phi} \left\langle \underbrace{\Delta_f^{(2)}(N) a^2}_{\Sigma} \right\rangle_\sigma. \qquad \text{(C2)}$$

We can therefore always find a ϕ_* such that the first expectation is real valued and non-positive

$$2 \, \text{Re} \left[\left\langle \underbrace{\Delta_f^{(2)}(N) a^2}_{\Sigma} \right\rangle_{\sigma(\phi_*)} \right] = -2 \left| \left\langle \underbrace{\Delta_f^{(2)}(N) a^2}_{\Sigma} \right\rangle_\sigma \right|. \quad \text{We}$$

have thus shown that

$$\left| \left\langle \Sigma \right\rangle_\sigma \right| \leq \left\langle Xi \right\rangle_\sigma. \qquad \text{(C3)}$$

Under the second condition to our theorem we have $\Gamma - \Xi \geq 0$ which implies $\left\langle \Gamma \right\rangle_\sigma \geq \left\langle \Xi \right\rangle_\sigma$. We thus know that the determinant of $\tilde{h}$ can be bounded from below as

$$\text{Det}(\tilde{h}) = \left\langle \Gamma \right\rangle_\sigma^2 - \left| \left\langle \Sigma \right\rangle \right|^2 \geq \left\langle \Gamma \right\rangle_\sigma^2 - \left\langle \Xi \right\rangle_\sigma^2 \geq 0. \qquad \text{(C4)}$$

This proves our claim. $\qquad \square$

Appendix D: Further results for simulating under CGF minimization

Here we present some additional results obtained in simulating the DOPO system in a basis that minimizes the CGF functional. In particular, we investigate the relationship between the manifold coordinates and the expectation of the oscillators's mode operators. While excitation minimization will always enforce $\left\langle a \right\rangle_{\phi_t} \equiv 0 \Leftrightarrow \left\langle a \right\rangle_{\psi_t} = \alpha = \frac{Q + iP}{\sqrt{2}}$ CGF optimization generally does not lead to such a linear relationship as can be seen in Figure 10. We can also see that different regions in phase space
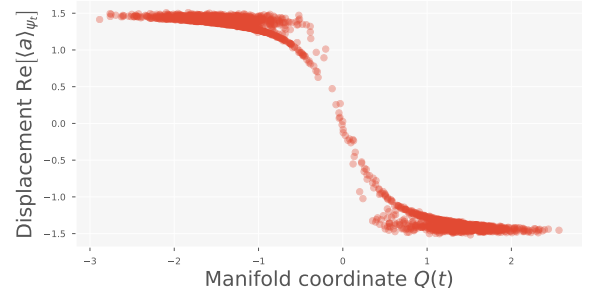
FIG. 10. The optimal manifold coordinate $Q(t)$ under CGF minimization appears mostly monotonically but not linearly related to the mode expectation $\text{Re}\left[\left\langle a \right\rangle_{\psi_t} \right]$.

lead to different complexity as measured by the CGF (cf Figure 11). This motivates using a simulation method in which even the basis size is adapted to the inherent complexity of the current dynamics. Finally, we note that
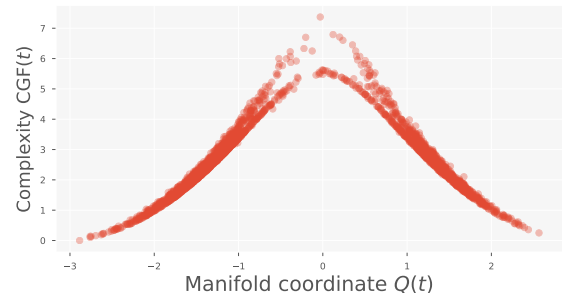
FIG. 11. When the system state localizes near the 'origin' i.e., $Q = 0$, the complexity increases, i.e., more basis levels are necessary for accurate representation.

even within the displaced basis there appears to be yet lower dimensional attractors for the reduced complexity state $|\phi_t\rangle$. In Figure 12 we have visualized the distribution of the first three moving basis level populations when transforming to the CGF optimal basis. In Figure 13 we have simulated the gradient coupled fiducial state dynamics for the DPO system above threshold for different values of the coupling gain η. All stochastic simulations were carried out with the same random seed and thus the same realization of the innovation process. Figure 13 (c) shows some discrepancies between states reconstructed from simulations with lower η and $\eta_0 = 200$. Specifically we compute $d_{\mathrm{FS}}^2(\psi^{(\eta)}, \psi^{(\eta_0)})$, where the fixed basis representation states are obtained from the respective moving bases representation states via $\left|\psi^{(\eta)}(t)\right\rangle = U_{\theta^{(\eta)}} \left|\phi^{(\eta)}(t)\right\rangle$ and the Fubini-Study distance is defined as:

$$d_{\mathrm{FS}}(\phi, \phi') := \arccos\left(|\langle\phi|\phi'\rangle|\right) \tag{D1}$$

The errors decrease with increasing η suggesting that a strong gradient coupling gain $\eta = O(100)$ is preferable. We intend to investigate this more systematically in future work. We note that the errors appear to stay constant over time, which is encouraging. Furthermore, we have simulated our system with different sizes of the basis (cf Figure 13 (d)) and evaluated the error relative to the most accurate simulation. Surprisingly, we find that the truncation error remains roughly constant, i.e., even for this randomly switching system, the low-dimensional approximations to the system state track the actual system state very well.

[1] S. G. Schirmer, H. Fu, and A. I. Solomon, Physical Review A 63, 063410 (2001), arXiv:0010031v2 [arXiv:quant-ph].

[2] C. Caves, Physical Review D, Physical Review D 26 (1982).

[3] B. R. Mollow and R. J. Glauber, Physical Review, Physical Review 160, 1076 (1967).

[4] J. Kerckhoff, H. Nurdin, D. Pavlichin, and H. Mabuchi, Physical Review Letters, Physical Review Letters 105, 040502 (2010).

[5] J. Gough and M. R. James, Communications in Mathematical Physics, Communications in Mathematical Physics 287, 1109 (2008).

[6] J. Gough and M. R. James, IEEE Transactions on Automatic Control, IEEE Transactions on Automatic Control 54, 2530 (2009).

[7] J. Combes, J. Kerckhoff, and M. Sarovar, arXiv , 1 (2016), arXiv:1611.00375.

[8] I. C. Percival, Journal of Physics A: Mathematical and General, Journal of Physics A: Mathematical and General 27, 1003 (1999).

[9] H. Mabuchi, Physical Review A, Physical Review A 85, 015806 (2012).

[10] H. Mabuchi, Applied Physics Letters, Applied Physics Letters 99, 153103 (2011).

[11] C. Santori, J. S. Pelc, R. G. Beausoleil, N. Tezak, R. Hamerly, and H. Mabuchi, Physical Review Applied, Physical Review Applied 1, 054005 (2014).

[12] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, Physical Review A, Physical Review A 88, 063853 (2013).

[13] T. Inagaki, K. Inaba, R. Hamerly, K. Inoue, Y. Yamamoto, and H. Takesue, Nature Photonics, Nature Photonics 2, 1 (2016).

[14] N. Tezak and H. Mabuchi, EPJ Quantum Technology, EPJ Quantum Technology 2, 10 (2015).

[15] M. Mirrahimi, Z. Leghtas, V. V. Albert, S. Touzard, R. J. Schoelkopf, L. Jiang, and M. H. Devoret, New Journal of Physics, New Journal of Physics 16, 045014 (2014).

[16] Z. Leghtas, S. Touzard, I. M. Pop, A. Kou, B. Vlastakis, A. Petrenko, K. M. Sliwa, A. Narla, S. Shankar, M. J. Hatridge, M. Reagor, L. Frunzio, R. J. Schoelkopf, M. Mirrahimi, and M. H. Devoret, Science 347, 853 (2015).

[17] A. Perelomov, "Generalized coherent states and some of their applications," (1977).

[18] S. Carter, Physical Review A, Physical Review A 51, 3274 (1995).

[19] E. Wigner, Physical Review, Physical Review 40, 749 (1932).

[20] K. Husimi, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series 22, 264 (1940).

[21] E. C. G. Sudarshan, Physical Review Letters, Physical Review Letters 10, 277 (1963).

[22] P. Kinsler and P. D. Drummond, Physical Review A, Physical Review A 43, 6194 (1991).

[23] R. Schack, T. a. Brun, and I. C. Percival, J. Phys. A: Math. Gen., J. Phys. A: Math. Gen. 28, 16 (1995).

[24] T. Steimle, G. Alber, and I. C. Percival, Journal of Physics A: Mathematical and General, Journal of Physics A: Mathematical and General 28, L491 (1999).
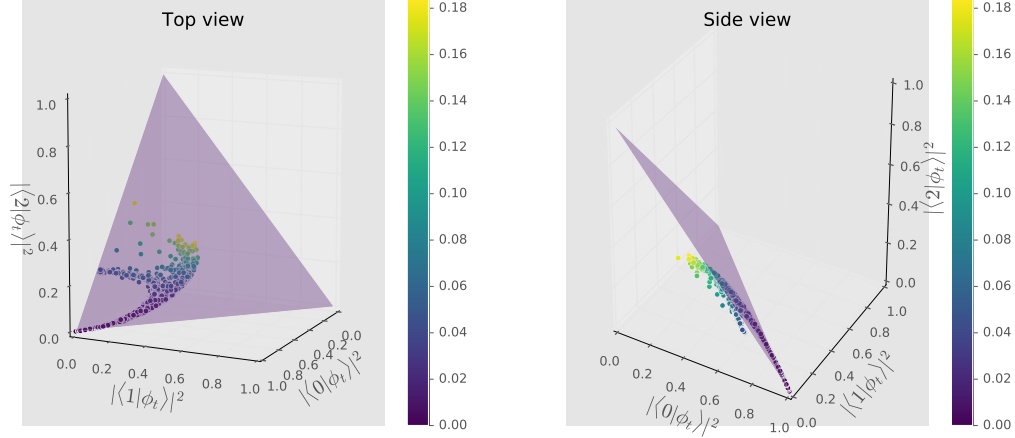
FIG. 12. The probability simplex spanned by the excitation probability of the first three basis levels. For the optimal CGF trajectory the basis level populations remain nearly confined to this simplex, but diverge slightly from it, especially when the first and second excited basis levels are non-trivially populated. The points are color coded according to their 'missing probability'-distance p' from the simplex, i.e. $p' := 1 - \langle \Pi_0 + \Pi_1 + \Pi_2 \rangle_{\phi_t}$.

[25] N. Tezak, A. Niederberger, D. S. Pavlichin, G. Sarma, and H. Mabuchi, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **370**, 5270 (2012).

[26] Z. Shi and H. I. Nurdin, *Journal of Physics D: Applied Physics*, Journal of Physics D: Applied Physics **49**, 465501 (2016).

[27] N. Tezak, "QMANIFOLD," (2016), access to the code is currently available on request and it will likely be publicly released in the future.

[28] J. Johansson, P. Nation, and F. Nori, *Computer Physics Communications*, Computer Physics Communications **183**, 1760 (2012).

[29] N. Tezak and M. Goerz, "Qnet," (2012).

[30] S. Akhmanov, A. Kovrigin, A. Piskarskas, V. Fadeev, and R. Khokhlov, Jetp Lett **2**, 191 (1965).

[31] J. A. Giordmaine and R. C. Miller, Phys. Rev. Lett. **14**, 973 (1965).

[32] S. Utsunomiya, K. Takata, and Y. Yamamoto, *Optics Express*, Optics Express **19**, 18091 (2011).

[33] C. K. Li, Y. T. Poon, and T. Schulte-Herbrueggen, *Math Comp*, Math Comp **80**, 1601 (2008).

[34] T. Schulte-Herbrueggen, S. J. Glaser, G. Dirr, and U. Helmke, *Reviews in Mathematical Physics*, Reviews in Mathematical Physics **22**, 597 (2010).

[35] R. Mahony and J. H. Manton, *Journal of Global Optimization*, Journal of Global Optimization **23**, 309 (2002).

[36] M. G. Genoni, M. G. A. Paris, and K. Banaszek, *Physical Review A - Atomic, Molecular, and Optical Physics*, Physical Review A - Atomic, Molecular, and Optical Physics **78**, 4 (2008).

[37] M. G. Genoni and M. G. A. Paris, *Physical Review A - Atomic, Molecular, and Optical Physics*, Physical Review A - Atomic, Molecular, and Optical Physics **82**, 1 (2010).

[38] D. Petz and C. Ghinea, *Quantum Probability and Related Topics*, Quantum Probability and Related Topics **27**, 261 (2011).

[39] R. v. Handel and H. Mabuchi, *Journal of Optics B: Quantum and Semiclassical Optics*, Journal of Optics B: Quantum and Semiclassical Optics **7**, S226 (2005).

[40] H. Mabuchi, *Physical Review A*, Physical Review A **78**, 15 (2008).

[41] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States* (Cambridge University Press, Cambridge, 2006) p. 76.

[42] J. M. Radcliffe, *Journal of Physics A: General Physics*, Journal of Physics A: General Physics **4**, 313 (2001).

[43] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Reviews of Modern Physics **82**, 1155 (2010), arXiv:0810.4729.

[44] C. K. Andersen, J. Kerckhoff, K. W. Lehnert, B. J. Chapman, and K. Mølmer, Physical Review A **93**, 012346 (2016), arXiv:1508.07780.

[45] K. Takata, A. Marandi, and Y. Yamamoto, *Physical Review A*, Physical Review A **92**, 043821 (2015).

[46] D. Maruo, S. Utsunomiya, and Y. Yamamoto, *Physica Scripta*, Physica Scripta **91**, 083010 (2016).

[47] J. E. Gough, arXiv (2017), arXiv:1701.06355.

[48] H. Pichler and P. Zoller, , 1 (2015), arXiv:1510.04646.

[49] G. Tabak and H. Mabuchi, EPJ Quantum Technology **3**, 3 (2016), arXiv:1510.08942.

[50] M. Sarovar, J. Zhang, and L. Zeng, EPJ quantum technology **4**, 1 (2017).

[51] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Scientific reports **2**, 514 (2012).

[52] R. R. Puri, *Mathematical methods of quantum optics* (2001) p. 285.

[53] I. Wolfram Research, "Mathematica," (1988), http://www.wolfram.com/mathematica.

(a)Manifold coordinate $Q(t)$

(b)Mode expectation $\langle a \rangle_{\psi_t}$

(c)Error distance d_{FS}^2
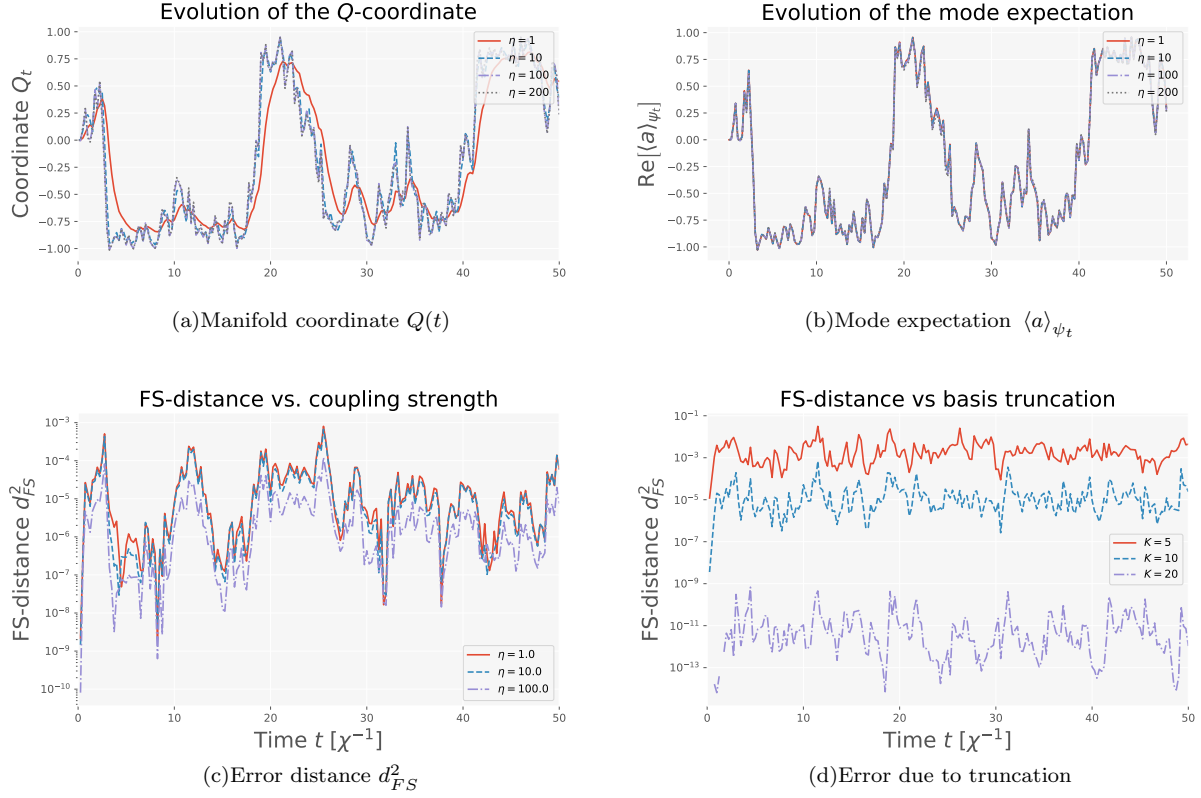
(d)Error due to truncation

FIG. 13. Gradient coupled fiducial state dynamics (simulated with $K = 30$ basis levels) for varying coupling strength $\eta \in [1, 10, 100, 200]$. In (a) we compare the Q coordinate trajectories and find that they agree very well for large η and appear somewhat 'low-passed' for $\eta = 1$. The mode expectation values (shown in (b)) agree very well for all values of η. Finally, in (c) we present the Fubini-Study distance between each trajectory and the $\eta = 200$ trajectory. In (d) we compare the Fubini-Study distance between a trajectory simulated with $K = 40$ moving basis levels to trajectories carried out with $K = 5, 10, 20$. Unsurprisingly, a larger basis decreases the error, but the errors do not appear to accumulate over time. All simulations were carried out with $\eta = 10$.

[54] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz, "Sympy: symbolic computing in python," (2017).

[55] R. M. Wilcox, *Journal of Mathematical Physics*, Journal of Mathematical Physics **8**, 962 (1967).