Statistically characterizing robustness and fidelity of quantum controls and quantum control algorithms

Irtaza Khalid,[1,*] Carrie A. Weidner,[2,†] Edmond A. Jonckheere,[3,‡] Sophie G. Shermer,[4,§] and Frank C. Langbein[1,‖]

[1]*School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, United Kingdom*

[2]*Quantum Engineering Technology Laboratories, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1FD, United Kingdom*

[3]*Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*

[4]*Department of Physics, Swansea University, Swansea SA2 8PP, United Kingdom*

(Received 18 July 2022; revised 4 November 2022; accepted 2 March 2023; published 14 March 2023)

Robustness of quantum operations or controls is important to build reliable quantum devices. The *robustness-infidelity measure* (RIM_p) is introduced to statistically quantify in a single measure the robustness and fidelity of a controller as the pth order Wasserstein distance between the fidelity distribution of the controller under any uncertainty and an ideal fidelity distribution. The RIM_p is the pth root of the pth raw moment of the infidelity distribution. Using a metrization argument, we justify why RIM_1 (the average infidelity) is a good practical robustness measure. Based on the RIM_p, an algorithmic robustness-infidelity measure (ARIM) is developed to quantify the expected robustness and fidelity of controllers found by a control algorithm. The utility of the RIM and ARIM is demonstrated on energy landscape controllers of spin-$\frac{1}{2}$ networks subject to Hamiltonian uncertainty. The robustness and fidelity of individual controllers as well as the expected robustness and fidelity of controllers found by different popular quantum control algorithms are characterized. For algorithm comparisons, stochastic and nonstochastic optimization objectives are considered. Although high fidelity and robustness are often conflicting objectives, some high-fidelity, robust controllers can usually be found, irrespective of the choice of the quantum control algorithm. However, for noisy or stochastic optimization objectives, adaptive sequential decision-making approaches, such as reinforcement learning, have a cost advantage compared to standard control algorithms and, in contrast, the high infidelities obtained are more consistent with high RIM values for low noise levels.

DOI: 10.1103/PhysRevA.107.032606

I. INTRODUCTION

Fault tolerance is crucial for quantum technology and presents a particular challenge for noisy intermediate-scale quantum (NISQ) devices [1]. Broadly, there are three proposed ways to deal with noise and errors and achieve fault tolerance: (1) via error correction protocols, e.g., Shor codes [2–4] or syndrome measurements [5]; (2) using error mitigation schemes, e.g., reversing noisy dynamics [6–9], active variational noise minimization [10], or parametric modeling of architecture defects in trapped qubits [11,12]; (3) robust solutions engineering, e.g., landscape shaping of the quantum control optimization problem in search of noise-free regions [13–15], decoherence-free subspaces [16,17], or noise spectral density based filter functions [18,19]. Uncertainties that require fault-tolerance in quantum devices have two flavors: (a) interactions with the environment that lead to nonunitary dynamics and (b) inaccuracies in the control model representing a specific physical implementation that affect the evolution but do not cause nonunitary evolution.

Standard quantum control methods for steering quantum devices mostly focus on finding controls that have high fidelity using mathematical models [20–22]. However, if the operation of quantum devices is subject to noise, high fidelity itself is insufficient to gauge performance of a control scheme and extra effort is required to systematically search for solutions that are both robust against noise and have high fidelity [23,24]. This requires a notion of robustness and ideally a single measure that can capture robustness and fidelity, enabling the identification and construction of more efficient methods to find controls that satisfy both properties.

In this paper, we introduce a general statistical diagnostic based on the Wasserstein distance of order p [25] to evaluate the robustness and fidelity of quantum control solutions and the algorithms used to find them. This is applicable to any quantum control problem where the fidelity is a random variable with a probability distribution over [0,1]. The Wasserstein distance between probability distributions is a

*khalidmi@cardiff.ac.uk
†c.weidner@bristol.ac.uk
‡jonckhee@usc.edu
§lw1660@gmail.com
‖frank@langbein.org

measure of the minimal costs of probability mass transport between two distributions. In Sec. II, the pth order *robustness-infidelity measure* (RIM_p) is defined to quantify the robustness and fidelity of a quantum controller. It is the pth-order Wasserstein distance between the probability distribution for the fidelity induced by noise and the ideal distribution for a perfectly robust controller, described by a Dirac delta function at fidelity 1. We show that the RIM_p is the pth root of the pth raw moment of the infidelity distribution—a nonparametric measure independent of any particular assumption for the distribution.

Using measure-theoretic norm scaling relations between RIMs of different orders, we argue that RIM_1, the average infidelity, is sufficient as a practical measure of robustness and fidelity. As such, it meshes with the average infidelity, already used in robust [26,27] and stochastic or adaptive quantum control settings [28,29]. More generally, the RIM_p is related to the risk-tunable fidelity measure using a utility function, as introduced in Ref. [30].

The RIM has practical utility by allowing us to choose among similar, high-fidelity controllers, acting as a postselector for robust controllers, agnostic of the algorithm used to find them. This may be computationally more efficient than optimizing the RIM directly, as we see in Sec. IV C. Moreover, it can also be adapted to compare the performance of control algorithms in finding not only high-fidelity but also robust controllers. To that end, we introduce an algorithmic RIM (ARIM), averaging RIMs over multiple controllers, in Sec. II.

In Sec. IV, we illustrate the RIM and draw useful insights for robust quantum control by generating controllers for static energy landscape control of the *XX* Heisenberg model, exploiting the degree of freedom afforded by the existence of multiple optima in quantum control [31]. We analyze their robustness properties and the performance of algorithms in finding effective controllers using four optimization algorithms representing different commonly employed approaches: (1) L-BFGS: a second-order gradient-based optimization using an ordinary differential equation model of the quantum system to compute the fidelity under perfect conditions [32]; (2) proximal policy optimization (PPO): a model-free reinforcement learning algorithm, having no prior knowledge of the system [33]; (3) Nelder-Mead: a derivative-free simplex-based heuristic search method [34]; and (4) stable noisy optimization by branch and fit (SNOBFit): another derivative-free method that performs model-free learning by using regression to estimate gradients via a branch and fit method [35]. Here, (1) serves as a baseline for optimization over a noise-free fidelity objective functional under ideal conditions. (2) represents a machine-learning approach with minimal knowledge. (3) and (4) are derivative-free methods to handle stochastic objective functionals. A detailed motivation for the choice of these algorithms is presented in Sec. III B. These choices are not exhaustive but serve as a diverse set of algorithms to which we apply the RIM and ARIM, illustrating their utility and giving some indication of the performance of common control algorithms for the specific robust control problem. The algorithms were implemented in Python. Specifically, we used the SciPy library for (1) and (3) [36], and (4) is obtained from Ref. [37]. Our code and data are available at Ref. [38].

Our experimental motivation is fourfold: (A) By comparing the robustness of controllers without regard to the optimization algorithm, we wish to answer whether high fidelity implies high robustness using the RIM of the individual controllers (Sec. IV A). (B) By conducting a distributional comparison of controllers, we wish to understand how likely it is that a given algorithm produces controllers in an ideal (no-noise) setting that are robust in noisy conditions (Sec. IV A 2). (C) To study the effect of training noise of the same nature as the robustness noise model applied *during* optimization on an algorithm's ability to find robust controllers using the ARIM (Sec. IV B). For a fair comparison, we conduct (B) and (C) with a fixed number of objective function calls allotted to each algorithm. (D) In Sec. IV C, we try to understand an algorithm's asymptotic ability to find robust controllers using the ARIM through optimizing the RIM by allowing unlimited objective function calls. We consider two settings in this scenario: stochastic and nonstochastic fidelity optimization. In the latter case, we optimize over a fixed set of Hamiltonians sampled once according to a noise model, while in the former case the Hamiltonians are stochastically chosen at each objective function evaluation using the same noise model.

Our main numerical findings are:

(1) High fidelity controllers are not always robust, but the nonrobust controllers can be filtered out using the RIM.

(2) Using a consistency statistic, we show that PPO controller infidelities (RIM at no noise) are more correlated with RIM values at low noise levels compared to the other algorithms. More generally, a strong signal in the consistency statistic predicts RIM robustness while avoiding its explicit evaluation.

(3) For constrained objective function calls, there appear to be problem-dependent optimal levels of noise that produce more robust controllers for PPO in contrast to L-BFGS, SNOBFit, and Nelder-Mead.

(4) Robust controllers with respect to certain noise models in the optimization objective are obtained by all algorithms for the nonstochastic optimization objective when there are no constraints on resources.

(5) However, if the optimization objective is stochastic, the ARIM improves asymptotically only for PPO. In either case, PPO requires fewer function calls compared to the other algorithms, which highlights the potential of adaptive sequential decision-making strategies like reinforcement learning for NISQ optimization problems, where not all uncertainty can be captured by nonstochastic objective functionals (e.g., shot noise).

Lastly, from the perspective of classical control [39], it is well-known that accuracy conflicts with robustness through the $S + T = I$ formula, where S is related to tracking error and T to sensitivity of the tracking error to uncertainties. This restriction does not map directly to quantifying fidelity versus robustness in the quantum domain [24,40], although it is recovered in some cases [41]. One reason for the discrepancy is that the frequency-domain limitations of classical feedback control have limited applicability in quantifying the performance in the time domain. The RIM combines the two figures of merit into one single measure: small RIM means high fidelity and high robustness, while large RIM means poor fidelity and poor robustness.

II. MEASURING ROBUSTNESS AND FIDELITY OF QUANTUM CONTROLS

A. The general quantum control problem

The physical system we wish to control is represented by a Hamiltonian

$$H(t, \mathbf{u}) = H_0 + H_{\mathbf{u}}(t), \tag{1}$$

where the time-independent drift Hamiltonian H_0 describes the natural dynamics of the system and the control Hamiltonian $H_{\mathbf{u}}(t)$ describes the time-dependent control with the tunable, usually piecewise constant, control parameters $\mathbf{u}$. The closed-system dynamics are governed by the Schrödinger equation, which can be written in terms of the unitary evolution operator from time t_0 to time t_1,

$$U(t_0, t_1, \mathbf{u}) = \mathcal{T} \exp\left(-\frac{\imath}{\hbar}\int_{t_0}^{t_1} H(t, \mathbf{u})\, dt\right), \tag{2}$$

where $\mathcal{T}$ denotes time ordering and $\hbar$ is the reduced Planck's constant.

In general, the control problem is formulated as optimizing a fidelity $\mathcal{F}$ over a set of admissible controls. A notion of fidelity that reflects most definitions used in practice is given by $\mathcal{F} := |\langle G|K\rangle|^2$, which measures the similarity between normalized objects G and K. If we wish to prepare a state $G = |\psi_f\rangle$ from an initial state $|\psi_0\rangle$ at time t_0, then $K = U(t_0, t_1, \mathbf{u})\,|\psi_0\rangle$ and the optimization problem is given by

$$t_{\text{opt}}, \mathbf{u}_{\text{opt}} = \arg\max_{(t_1,\mathbf{u})\in\mathbb{X}} \underbrace{|\,\langle\psi_f|\,U(t_0, t_1, \mathbf{u})\,|\psi_0\rangle\,|^2}_{=\mathcal{F}(t_1,\mathbf{u})}, \tag{3}$$

where $\mathbb{X}$ is the domain of allowed controls, here including the final time t_1. A variant, up to normalization, of the state fidelity is the Hilbert-Schmidt inner product $\mathcal{F} = \text{Tr}(W^\dagger V)$ between a desired unitary transformation W and a gate achieved by control, $V = U(t_0, t_1, \mathbf{u})$. In general, we assume the fidelity is bounded, and without loss of generality we assume it lies in [0,1], where $\mathcal{F} = 1$ if and only if we have $G = e^{\imath\phi}K$, up to a global phase ϕ (and equivalently for the Hilbert-Schmidt inner product).

B. Robustness-infidelity measure

Uncertain dynamics turn the fidelity $\mathcal{F}$ into a random variable with a probability distribution $\mathbf{P}(\mathcal{F})$. Intuitively, we call a controller robust if this distribution has a low spread. While a low spread alone may indicate robustness, low fidelity means the controller does not realize the target operation well. So we also expect a fidelity close to 1. That means the perfect distribution under any uncertainties is δ_1—the Dirac delta distribution at maximum fidelity 1. In particular, we consider the delta function δ_x to be defined by an indicator cumulative distribution function (CDF):

$$C(a) = \begin{cases} 1 & \text{if } a \geqslant 0 \\ 0 & \text{if } a < 0. \end{cases} \tag{4}$$

This permits the familiar delta function property for integration with respect to a basic (rapidly diminishing) function:

$$\int_{-\infty}^{\infty} g(x)\delta_{x-a}\, dx = \int_{-\infty}^{\infty} g(x)\, dC(x-a) = g(a). \tag{5}$$

Our goal is to define a distance between probability distributions that measures closeness between the ideal and the achieved probability distribution to combine high fidelity and its robustness into a single measure.

For this, we take the Wasserstein or Earth mover's distance $\mathcal{W}$ [25,42] due to the fact that: (1) it allows us to compare two probability distributions that do not share a common support, and, in particular, compare discrete and continuous distributions; (2) its easy geometric interpretation helps with its optimization; and (3) a simplification allows it to be calculated easily, as shown next.

The dual formulation of the pth order Wasserstein distance [43] between two distributions μ, ν is given by

$$\mathcal{W}_p(\mu, \nu) = \sup_{h,g}\left[\int h(x)\, d\mu(x) - \int g(y)\, d\nu(y)\right]^{\frac{1}{p}}, \tag{6}$$

where $h(x) - g(y) \leqslant \|x - y\|^p$. Even though this form seems abstract, for one-dimensional distributions, we can analytically compute the optimal maps h, g with

Theorem 1 (Prop. 1 in Ref. [43]). The pth Wasserstein distance $\mathcal{W}_p(\mu, \nu)$ for one-dimensional probability distributions μ and ν with finite p moments can be rewritten as

$$\mathcal{W}_p(\mu, \nu) = \left(\int_0^1 |Q_\mu(z) - Q_\nu(z)|^p\, dz\right)^{\frac{1}{p}},$$

where $Q_\mu(z) = \inf\{x \in \mathbb{R} : C_\mu(x) \geqslant z\}$ denotes the quantile function and C_μ is the cumulative probability function of μ and likewise for Q_ν.

Remarkably, the optimal transport distance between one-dimensional distributions μ, ν over all possible transportation plans can be computed in terms of their quantile functions Q_μ, Q_ν. From here, following Theorem 1, it is straightforward to define the pth RIM:

$$\text{RIM}_p := \mathcal{W}_p(\mathbf{P}(\mathcal{F}), \delta_1) = \left(\int_0^1 |Q_{\mathbf{P}(\mathcal{F})}(z) - 1|^p\, dz\right)^{\frac{1}{p}}. \tag{7}$$

It can be written in terms of the raw moments (see Appendix A 1),

$$\text{RIM}_p = \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}}, \tag{8}$$

where f is a fidelity sample drawn from the distribution $\mathbf{P}(\mathcal{F})$ and $1 - f$ is the corresponding infidelity sample. We use the expectation operator defined as $\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(\cdot)] := \int(\cdot)\mathbf{P}(\mathcal{F} = f)\, df$. For $p = 1$, we recover the average infidelity:

$$\text{RIM}_1 = \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[1 - f] = 1 - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f]. \tag{9}$$

To compute the RIM_p, we estimate $\mathbf{P}(\mathcal{F})$ using n fidelity samples $f_1, f_2, \ldots, f_n$. Such samples may be obtained in practice via Monte Carlo simulation or physical experiments [44]. Hence, barring the computational or experimental expense of obtaining these samples, the RIM_p is easy to compute. In case the dynamics of the system are certain, i.e., $\mathbf{P}(\mathcal{F}) = \delta_f$ for some constant fidelity value f, the RIM_1 is

equal to the infidelity $1-f$. Moreover, the RIM_1 is small if and only if the controller is robust (in the sense of the fidelity distribution having a low spread) and is also close to the maximum fidelity.

C. The average fidelity is sufficient for robustness comparisons

We motivate why the RIM_1 is sufficient for comparing robustness and fidelity of controllers by making use of the fact that the RIMs of different orders computed on the estimated fidelity distribution are in agreement. We obtain the following bounds between the lower and higher order RIMs (see Appendix A 3):

$$\mathrm{RIM}_{p'} \leqslant n^{(\frac{1}{p}-\frac{1}{p'})}\mathrm{RIM}_p, \tag{10a}$$

$$\mathrm{RIM}_p \leqslant \mathrm{RIM}_{p'} \tag{10b}$$

for $p < p'$, where n is the number of samples used to estimate the RIM. Equation (10b) is stronger and states that RIM_p is less sensitive to outliers than $\mathrm{RIM}_{p'}$ while Eq. (10a) states that for fixed n and p, $\mathrm{RIM}_{p'}$ growth is sublinear [$\propto \exp(-1/p')$]. This can be made tighter by adding additional assumptions on the nature of $\mathbf{P}(\mathcal{F})$, but these depend on the specific control problem. The upper bound becomes loose with increasing n, but highlights the constraining nature of deviation of higher-order RIMs from RIM_1.

This means that the higher-order RIMs do not capture more useful robustness information for comparisons, with the base case in Eq. (10b) being decided by the RIM_1. RIM_1 has low sensitivity to outliers (see Appendix A 3), which makes it easier to estimate than higher-order RIMs, which, like the worst-case fidelity, are harder to accurately practically obtain (as more samples are required, which also explains the presence of n in the inequality).

The bound in Eq. (10b) gets tighter for large n but also for decreasing infidelities, so in this regime, the RIMs are in agreement. Another way to see this is to note that the Wasserstein distance provides a structure-preserving geodesic between any fidelity distribution to the ideal δ_1: the distributions converge together with their RIMs of any order. So, especially when approaching the ideal distribution δ_1, i.e., in case RIM_1 is small for the high-fidelity, robust controller, there is strong agreement between RIMs of all orders. For example, the variance of distributions decreases as $\sim(1-\min\mathcal{F})^2$ as $\min\mathcal{F}\to 1$ in [0,1].

However, the fact that outliers are more influential for higher RIM orders proves useful for optimization [30] where such behavior is sought after, while our goal here is robustness and fidelity comparison. For this goal, in general, outliers are obstructive as they would hide the general distributional trend. From now on we will refer to the RIM_1 without the subscript.

D. Perturbations

Next, we define the noise in the system as perturbations of its uncertain dynamics that give rise to $\mathbf{P}(\mathcal{F})$. A perturbation to the full Hamiltonian in Eq. (1) can be expressed as $\tilde{H}(t,\mathbf{u}) = H(t,\mathbf{u}) + \gamma S \in \mathbb{C}^{n\times n}$, where $\gamma \in \mathbb{R}$ describes the strength of a perturbation and $S \in \mathbb{C}^{n\times n}$ its structure, usually normalized using some matrix norm. To induce an uncertainty into the dynamics, we treat γ and S as random variables drawn from some probability distributions. This give us a general way to represent any physically relevant uncertainties in Hamiltonian parameters.

The structure S may be fixed, e.g., describing the uncertainty in some coupling parameter for the Hamiltonian, while γ is drawn from a normal distribution. This would be consistent with a (linear) *structured perturbation* in classical robust control theory [45]. Instead, S may also be drawn from a probability distribution, describing uncertainties across multiple Hamiltonian parameters. While this generalizes structured perturbations, note that they do remain linear with respect to the strength. If S is sampled uniformly on the unit sphere, according to its normalization, we have an *unstructured perturbation*, with (uncertain) strength determined by γ. Conceptually, if γ is drawn from a normal distribution with zero mean and standard deviation σ, γS describes a "fuzzy" ball $\mathcal{B}_\sigma$ around $H(t,\mathbf{u})$. In this paper, we consider unstructured perturbations that are less idealized, in some sense, than the structured perturbations (usually considered in classical control [46]), allowing the robustness results to be interpreted generically without the need to consider specific sources of uncertainties arising from specific quantum device designs. For simplicity, we write $\mathbf{P}_\sigma(\mathcal{F})$ for a fidelity distribution obtained by unstructured perturbations drawn from $\mathcal{B}_\sigma$.

Our quantification of robustness is dependent on the choice of γS and the uncertainties in these quantities. Note that neither the choice of the noise model nor the magnitude of the noise level is restricted, as our approach is not perturbative around the optimum $(t_{\mathrm{opt}}, \mathbf{u}_{\mathrm{opt}})$, which is how noise is usually modeled in the literature [13,18,47–50]. This approach becomes relevant when confidence in an analytical physical model is low or there are missing terms that cannot be analytically or perturbatively accounted for, e.g., complicated noise sources. This is also in accordance with modern robustness theory and the μ function in classical [51] and quantum [45,52,53] settings.

To further motivate the RIM, we study how it compares with other statistical measures of robustness. The RIM generally correlates with worst-case or minimum sample fidelity, variance or higher moments, and the yield function $Y(\mathcal{F}_{\mathrm{Th}})$, which is the fraction of fidelities greater than a threshold fidelity $\mathcal{F}_{\mathrm{Th}}$. Figure 1 shows a scatter plot of RIM values versus $Y(0.95)$, $Y(0.98)$ and the worst-case fidelity for an example problem using Eq. (13) discussed in Sec. III A. The RIM has an advantage over Y in that it does not depend on an arbitrary choice of $\mathcal{F}_{\mathrm{Th}}$.

E. Measuring the performance of control algorithms

We can also apply the previous arguments to derive a measure to compare the ability of control algorithms to find high-fidelity, robust controllers. Let $\mathbf{P}(\mathrm{RIM})$ be a distribution of RIM values of controllers obtained by a particular algorithm and a particular control problem with specific uncertainties. This can be estimated by sampling L controllers produced by the algorithm. The ideal of this distribution is δ_0, so we can define the ARIM,

$$\mathrm{ARIM} := \mathcal{W}_1(\mathbf{P}(\mathrm{RIM}), \delta_0) = \mathbb{E}_{r\sim\mathbf{P}(\mathrm{RIM})}[r], \tag{11}$$

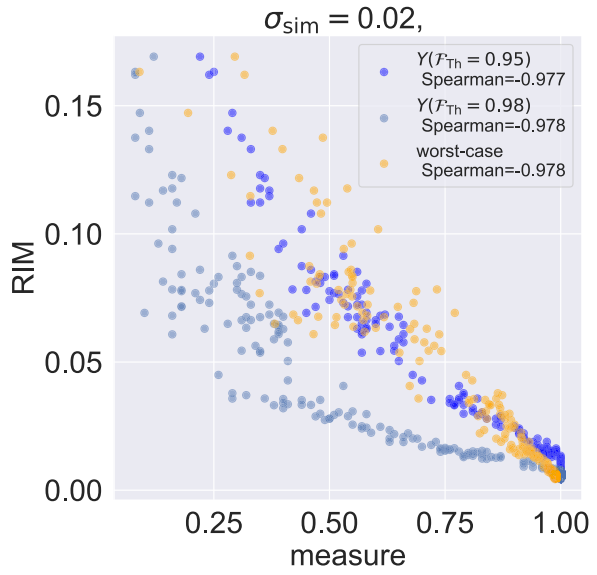

FIG. 1. RIM values generated from $\mathbf{P}_{0.02}(\mathcal{F})$ with $N = 100$ samples for 200 controllers are plotted against the yield $Y(\mathcal{F}_{\text{Th}})$ at fidelity thresholds $\mathcal{F}_{\text{Th}} = 0.95, 0.98$ and the worst case fidelity. Both measures are correlated, as encapsulated by the high negative Spearman correlation coefficients [54] and p values $<10^{-4}$.

following the same argument as before. The ARIM is small if and only if the underlying RIM distribution $\mathbf{P}(\text{RIM})$ has higher density at or near RIM $= 0$, i.e., is close to the ideal δ_0.

III. ROBUSTNESS FOR STATIC CONTROL PROBLEMS

We study the robustness of static control problems, where the controls are time independent, instead of the usual time-dependent controls. Previous work has shown that particularly robust controls can be found for these systems [23,24,40,45]. While these systems are often not fully controllable, solutions for specific operations can be found via optimization [55]. The static approach is simpler in the sense of having fewer control parameters to optimize over, which reduces computational and experimental complexity. This makes the problem suitable to demonstrate the practical usage of the RIM in a concrete example and explore the robustness properties of the control algorithms as well as the controllers they find.

A. Information transfer in the single excitation subspace of *XX* spin chains

We consider a network of M spins represented by the quantum Heisenberg model given by the Hamiltonian

$$\frac{H_{\text{Heis}}}{\hbar} = \sum_{a\in\{x,y,z\}} \sum_{j=1}^{M} J^a \sigma_j^a \sigma_{j+1}^a + \eta \sum_{j=1}^{M} \sigma_j^z, \tag{12}$$

where $\sigma_j^a = \mathbb{I}^{\otimes j-1} \otimes \sigma^a \otimes \mathbb{I}^{\otimes M-j}$ and σ^a are the usual Pauli matrices. We set $J^z = 0$ and $J^x = J^y = J$ for the XX model with uniform couplings. This model has been studied extensively, starting with Ref. [56] in 1961, and a more recent review of the system, as it relates to quantum communication, is provided in Ref. [57]. Conditions for perfect state transfer along XX chains were derived in Ref. [58] and applied to NMR systems [59]. Similar experiments have been carried out in photonic systems [60,61], and proposals for engineering similar systems with trapped ions [62] and cold atoms [63] exist.

The state space naturally decomposes into noninteracting excitation subspaces as the Hamiltonian commutes with the total excitation operator. Here we consider the first excitation subspace, the smallest space that enables transfer of one bit of information between the nodes in the network. Higher excitation subspaces may be needed for other applications, but it is desirable for information transfer to limit the space to the smallest space that is sufficient to achieve the task. This is a much smaller space and only grows as $O(M^2)$ as opposed to $O(\exp(2M))$. The Hamiltonian of the first excitation subspace is

$$\frac{(H_{XX})_{l,m}}{\hbar} = J\mathbb{1}_{l,m\pm1} + \Delta_l \mathbb{1}_{l,m}, \tag{13}$$

where $\mathbb{1}_{l,m}$ is the Kronecker delta. The static controls are local energy biases Δ_l on spin $|l\rangle$ in a diagonal matrix $H_{\boldsymbol{\Delta}} = \text{diag}(\Delta_1, \ldots, \Delta_M)$.

H_{XX} allows for transfer of single bit excitations from an initial spin state $|a\rangle$ to a final state $|b\rangle$. We define the fidelity as $\mathcal{F} = |\langle b| U(t_0, t_1, H_{XX}) |a\rangle|^2$ and the infidelity as $\mathcal{I} = 1 - \mathcal{F}$. The solution to Eq. (3) is a final time t_{opt} and a single vector of M biases $\boldsymbol{\Delta}_{\text{opt}}$. The perturbations are given by

$$(S_\sigma)_{l,m} = \sum_{k=1}^{M-1} \gamma_k^J J \mathbb{1}_{l,k}\mathbb{1}_{l,m\pm1} + \sum_{c=1}^{M} \gamma_c^C \Delta_c \mathbb{1}_{c,l}\mathbb{1}_{l,m}, \tag{14}$$

where γ_k^J and γ_c^C are the strength of the perturbation on the couplings and controls, respectively. We draw these strengths from the same normal distribution $\mathcal{N}(0, \sigma^2)$ with mean 0 and variance σ^2. A numerical example illustrating the RIM via the empirical CDF (ECDF) for two controllers is shown in Fig. 2.

Depending on the hardware platform, it is possible to consider specific practically motivated correlated noise models with correlated structured perturbations or a power-law decaying electric-field noise $(1/s)$, e.g., in trapped atomic platforms [11,65]. We have chosen to implement the simplest option of equal strength random perturbations on all nonzero entries of the Hamiltonian that is also relevant in practical settings [59–63].

B. Algorithms for static control problems

The ARIM compares algorithm performances in finding robust controllers. Here, we describe the algorithms used to find the controllers for the static control problems. For selecting algorithms, we tried to (a) investigate the performance of algorithms commonly used in the quantum control and other communities, (b) consider algorithms that do and do not require gradient information, and (c) consider reinforcement learning, more recently also used in quantum control.

L-BFGS is a common optimization algorithm used in quantum control as part of GRAPE [66] and performed well on finding high-fidelity energy landscape controllers [55]. It has not been designed for noisy optimization, but there exist smoothing modifications that attempt to address this [67–69]. For individual controller comparisons, we use standard L-BFGS with an ordinary differential equation model to

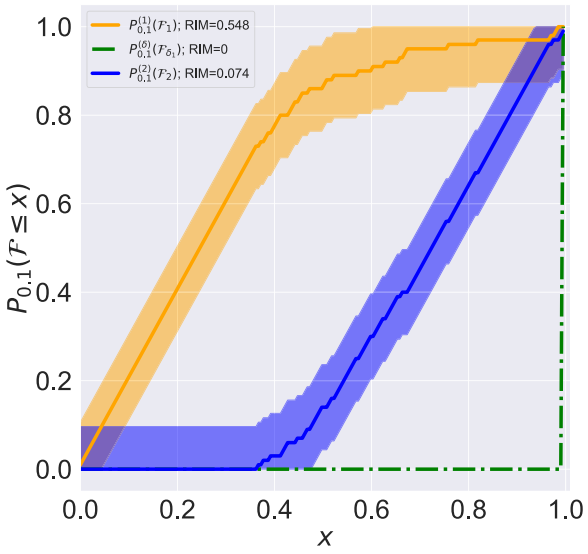

FIG. 2. To illustrate the RIM robustness measure, two static controllers for an *XX* spin chain of length five for transferring an excitation from spin $|1\rangle$ to $|3\rangle$ are compared. The empirical approximations to the CDFs for the two controllers, $l = 1, 2$, were simulated using 100 bootstrapped perturbations with $\sigma = 0.1$, giving fidelity distributions $\mathbf{P}_{0.1}(\mathcal{F}_l)$ for the fidelity random variables $\mathcal{F}_l$. The fidelity distribution $\mathbf{P}_{0.1}(\mathcal{F}_{\delta_1})$ for a perfectly robust controller with $\mathcal{F}_{\delta_1}$ is also shown. The ECDFs are estimated using 500 bootstrap repetitions. The 0.95 confidence bounds on their error are obtained using the Dvoretsky-Kiefer-Wolfowitz inequality [64]. Closeness to the perfectly robust controller can be interpreted as having a smaller area under the curve and is indicated by the RIM values.

compute the fidelity without perturbations during optimization. This serves as a baseline to understand the performance of optimizing noiseless objective functionals compared to the noisy optimization performed by all other selected algorithms and its impact on the robustness of the controllers found. We have explored stochastic gradient descent methods (e.g., ADAM [70]) and also tested a noisy version of L-BFGS that has been recently proposed that modifies the line search and lengthening procedure during the gradient update step [69] and found that our training noise scales were too large and washed away gradient information, rendering these algorithms unsuitable for our study.

Reinforcement learning has been successfully used for tackling quantum control in challenging noisy environments, resulting in similar or better performances compared to standard control methods. Promising results include the stabilization of a particle via feedback in an unstable potential [71], optimizing circuit-QED, two-qubit unitary operators under physical realization constraints [72], and optimizing multiqubit control landscapes suffering from control leakage and stochastic model errors [73], among many others.

PPO is a policy gradient method in the class of reinforcement learning algorithms [74]. It uses a discounted reward signal (e.g., the fidelity) accumulated over multiple interactions with the optimization landscape using nonparametric models: the policy function, that does the interacting by performing control operations, and the action value function, that predicts the quality of each action undertaken by the policy in terms of future payoffs in the reward signal. Both are estimated using neural networks in a control problem agnostic fashion. Doing this allows the incorporation of perturbations during training which specifically has advantages in finding robust controls for energy landscape problems [75]. Here, we use a control problem formulation of PPO for Eq. (13) as described in Ref. [75].

Nelder-Mead is a popular simplex-based control algorithm using direct search. Essentially, it keeps updating a polytope whose vertices are function evaluations toward an optimum direction. It has successfully been used in noisy experimental settings [76] due to its nonreliance on gradient information [77–79], especially when obtaining such information is resource intensive.

SNOBFit has been chosen as it has been designed to filter out quite large-scale noise in objective functionals [35]. It fits local models using objective function evaluations and implements a branching and splitting algorithm to partition the parameter space into smaller boxes with one function evaluation per box. The latter is a nonlocal search scheme that orders promising subboxes by the number of bisections required to get from the base box to that box. Subboxes with smaller bisections are worth exploring more. Like PPO, it does not rely on explicit gradient information and builds models of the optimization landscape. Thus, both algorithms should be able to cope with large amounts of noise in the form of controller and model uncertainties, environmental effects, and singularity during optimization. SNOBFit, however, differs importantly from PPO in the assumption that those models are linear. Moreover, its nonlocal optimization landscape exploration is not random and thus has comparatively a lot less variance in performance (that may or may not be poor).

IV. NUMERICAL EXPERIMENTS

To explore the robustness of controllers and corresponding control algorithms (see experimental motivation in Sec. I), we perform a Monte Carlo robustness analysis using the RIM on numerical solutions to the spin chain information transfer problem in Sec. III A for chains of length $M = 4, 5, 6, 7, 8, 9$ with $J = 1$.

We look at transitions from the start of the chain $|1\rangle$ to the end $|M\rangle$ and from $|1\rangle$ to the middle $|\lceil \frac{M}{2} \rceil\rangle$. The former transition is physically easy to control while the latter is more challenging [55], as transitions to the middle exhibit anticore behavior [80].

We collect the best 100 solutions, ranked by their fidelity, obtained by all the control algorithms. Each algorithm has a budget of 10^6 fidelity function evaluations. The budget correlates with the run time for each algorithm. It is imposed to allow for a fair comparison of the algorithm robustness performance under similar resources, while being agnostic to specific implementations and speed differences.

We initialize $\mathbf{\Delta}, t$ with quasi Monte Carlo samples from the Latin hypercube [81–83] to increase convergence rate and decrease clustering of controllers. This permits coverage of the parameter domain with $O(1/\sqrt{N})$ samples as opposed to $O(1/N)$ for random sampling, where N is the number of initial values. Our constraints are $0 \leqslant t_f \leqslant 70$ and $-10 \leqslant \Delta \leqslant 10$. We use 100 bootstrap samples to estimate fidelity distributions

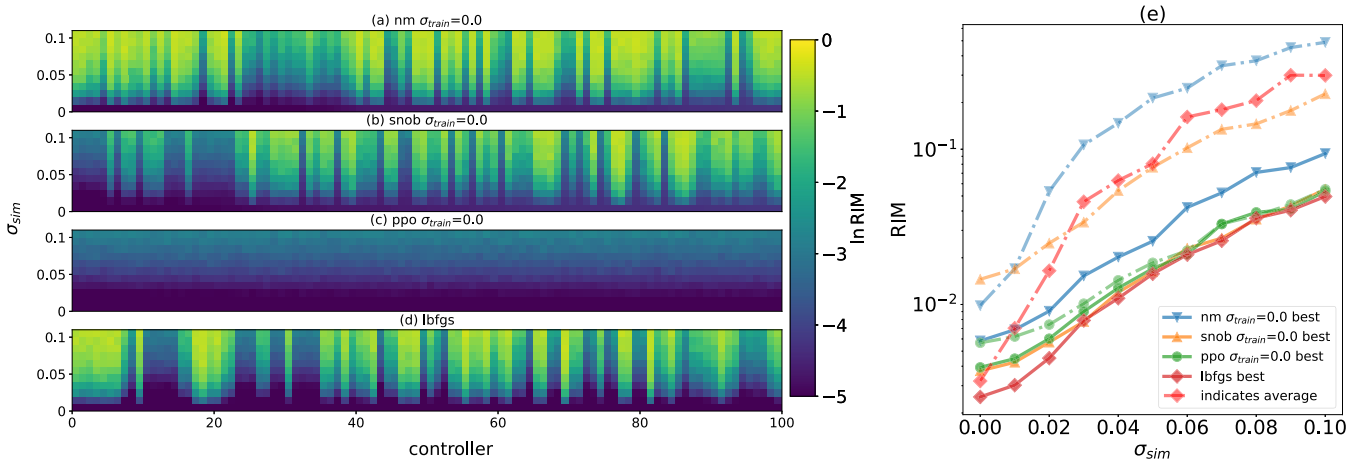

FIG. 3. (a)–(d) 100 controllers found for the *XX* spin chain model, Eq. (13), using Nelder-Mead, SNOBFit, PPO, and L-BFGS for $M = 5$ and the spin transition from $|1\rangle$ to $|3\rangle$ with $\sigma_{\text{train}} = 0$. The controllers are ranked in increasing order of infidelity at $\sigma_{\text{sim}} = 0$ from left to right. Each column represents a single controller's RIM at $\sigma_{\text{sim}} = 0, 0.01, 0.02, \ldots, 0.1$ from the bottom to the top on a log scale. Even if the infidelity or RIM at $\sigma_{\text{sim}} = 0$ is close to 0, some controllers' RIM values degrade faster than others and are hence less robust despite starting at very low infidelities. (e) RIM as a function of σ_{sim} for the average and best controller (i.e., most dark over all σ_{sim} levels) out of the 100 shown in (a)–(d) in terms of how much they preserve their corresponding RIM rank average across all σ_{sim}. Each algorithm is indicated by a marker shape, and the solid and dash-dotted lines denote the best and average controller lines, respectively. All the best controllers have very high initial fidelities and are very similar across the different control algorithms, with Nelder-Mead being only moderately worse.

throughout. The perturbation strengths γ_j^J and γ_c^C are scaled by J and Δ, respectively, as per Eq. (14). Note, for $\sigma = 0$, $\mathbf{P}_0(\mathcal{F}) = \delta_{\mathcal{F}}$ is deterministic.

The perturbation strengths are drawn from a normal distribution with standard deviation σ_{train} determining the strength of the noise added for the optimization. σ_{sim} is the noise level used in the simulations to assess the robustness of the controllers found. Implementation details are in Appendix B 1. The optimization objective is noiseless $\mathcal{F}$ for Sec. IV A, stochastic $\mathcal{F}$ with unstructured perturbation S_σ for Sec. IV B, and the RIM for the nonstochastic problem and a stochastic $\mathcal{F}$ with S_σ in Sec. IV C.

A. Characterization of all controllers found with constrained resources

1. *Ranking individual controllers*

In this section, we address our motivating question (A), whether high fidelity implies high robustness for an individual controller. We also numerically demonstrate the nonlinear and nonuniform deterioration of robustness with increasing noise which implies a trade-off between higher fidelity at no noise and robustness at higher noise levels.

To this end, we employ control algorithms to optimize an objective functional without noise, i.e., setting $\sigma_{\text{train}} = 0$ (see Sec. II D), under the general optimization conditions outlined at the start of Sec. IV. We rank these controllers by their infidelity values and then compute the RIM values for various levels of simulation noise, $\sigma_{\text{sim}} = 0.01, 0.02, \ldots, 0.1$.

For example, Figs. 3(a)–3(d) show a pseudocolor plot of the RIM values for 100 controllers found for the chosen test control problem (chain of length $M = 5$, target spin transfer $|1\rangle$ to $|3\rangle$). The lowest infidelity controllers start from the left and are indexed by columns 1 to 100, indicating their respective ranks according to their RIM at $\sigma_{\text{sim}} = 0$. The RIM values, as a function of σ_{sim}, for individual controllers grow at different rates despite starting at quite similar small values for all algorithms. The main result that applies also to all transitions (not explicitly shown here) is that the high fidelity controllers do not, in general, preserve their ranks as σ_{sim} increases, e.g., for SNOBFit [see Fig. 3(b)], the RIM for controllers $6, 8, 9, 11 - 13$ grows much more rapidly than for controllers $24 - 33$, indicated by rapid color changes from dark (low RIM) to light (high RIM) in the vertical direction. Interestingly, almost all controllers found by PPO have very low RIM across σ_{sim} values compared to the other control algorithms (color remains dark for longer). This is, however, not reflective of PPO's general behavior on the extended sample of problems we examined (see Appendix B 5). It could be limited fundamentally by the existence of robust controllers and/or the resource budget for a particular problem (see Fig. 10 in Appendix B 4 showing results for other transitions).

We further evaluate the best performing individual controller. To this end, we seek the controller that preserves its overall RIM rank average the most across the noise levels. It is computed using the reshuffled RIM ranks of each controller for all values of σ_{sim}. Likewise, we locate the controller that has the median RIM rank average across the noise levels as the averagely performing controller. Most of the RIM rank sum distributions studied were symmetric, and their median was close to their average value. So, we can try to understand average controller RIM rank order consistency in terms of how the median controller performs. We compare the RIM values of the median with the best controller in Fig. 3(e) for all algorithms, showing the RIM values for the best and median controller as a function of σ_{sim}.

For all algorithms, the best and the average controllers have similar infidelities (initial RIM value) in Fig. 3(e). Their behavior as a function of σ_{sim} is different and is generally non-

linear. Thus, the best controllers, despite being distinguishable from the others at $\sigma_{\text{sim}} = 0$, become indistinguishable for higher σ_{sim} and point at a trade-off between infidelity (at no noise) and robustness that could be leveraged when selecting a controller to be deployed for a noisy system. Moreover, the RIM curve of the best controller among all algorithms (here L-BFGS) suggests a fundamental limitation on RIM for this problem. It is likely not possible to obtain curves that are lower, but this remains theoretically unresolved.

2. *Ordinal Kendall tau for* $\mathbf{RIM}_{\sigma_{sim}}$*-rank consistency*

To address the motivating question (B), how likely a given algorithm is to produce robust controllers that were obtained in an ideal (no-noise) setting, we are interested in how consistently a controller acquisition strategy produces controllers with low RIM.

To that end, we reduce the RIM rank consistency property of the top-k controllers across two perturbation strengths $\sigma_{\text{sim}}^{(i)}$ and $\sigma_{\text{sim}}^{(j)}$ to a prediction problem by asking the following: **(Q)** *How well does the RIM rank of a controller, when ordered at strength* $\sigma_{\text{sim}}^{(i)}$*, predict the RIM rank of the controller at strength* $\sigma_{sim}^{(j)}$*?*

To answer this question, let us denote the controller RIM $\sigma_{\text{sim}}^{(i)}$-rank order by the vector $\mathbf{r}^{\sigma_{\text{sim}}^{(i)}}$, and compute an ordinal (binned or categorical) version of the Kendall-tau-B statistic $\tilde{\tau}$ [84,85], a measure of statistical dependence between $\mathbf{r}^{\sigma_{\text{sim}}^{(i)}}$ and $\mathbf{r}^{\sigma_{\text{sim}}^{(j)}}$. The ordinals are constructed only for $\mathbf{r}^{\sigma_{\text{sim}}^{(i)}}$ by binning using a discrepancy parameter α that indicates the fraction of the maximum RIM value difference within a single bin. The binned rank order $\tilde{\mathbf{r}}^{\sigma_{\text{sim}}^{(i)}}(\alpha)$ minimizes the effect of small movement in either rank due to noise. Then $\tilde{\tau}$ is computed by

$$\tilde{\tau}\left(\sigma_{\text{sim}}^{(i)}, \sigma_{\text{sim}}^{(j)}\right) = \tilde{\tau}_{i,j} = \frac{\sum_{l<m} \mathbb{I}_{l,m}^{+} + \mathbb{I}_{l,m}^{-}}{\sqrt{\left(K - t_{\text{total}}^{(i)}\right)\left(K - t_{\text{total}}^{(j)}\right)}}, \tag{15}$$

where

$$\mathbb{I}_{l,m} = \text{sgn}\left(\tilde{\mathbf{r}}_{l}^{\sigma_{\text{sim}}^{(i)}} - \tilde{\mathbf{r}}_{m}^{\sigma_{\text{sim}}^{(i)}}\right) \text{sgn}\left(\mathbf{r}_{l}^{\sigma_{\text{sim}}^{(j)}} - \mathbf{r}_{m}^{\sigma_{\text{sim}}^{(j)}}\right) \tag{16}$$

are the l, mth sign products of the rank order differences at $\sigma_{\text{sim}}^{(i)}, \sigma_{\text{sim}}^{(j)}$ with $+/-$ denoting the positive (negative) pair contributions; $K = k(k-1)/2$ is the number of total pairs being compared; $t_{\text{total}}^{(i)} = \sum_l t_l^{\sigma_{\text{sim}}^{(i)}}(t_l^{\sigma_{\text{sim}}^{(i)}} - 1)/2$ are the total numbers of ties where $\mathbb{I}_{l,m} = 0$ for $\sigma_{\text{sim}}^{(i)}$ and likewise for $t_{\text{total}}^{(j)}$. For complete positive (negative) rank order correlation $\tilde{\tau} = \pm 1$ and $\tilde{\tau} = 0$ for zero rank order correlation. For our hypothesis test, we assumed a worst case p value of 10^{-4} as an acceptance criterion on the numerical results that will follow and also that the controllers generating these rank orders are independent of each other. In this case, this constraint is satisfied by the i.i.d. noise model for a given set of unique controllers corresponding to different points in a static optimization landscape. The independence over the choice of controllers is not necessary as all the consistency comparisons are for this fixed choice of controllers.

For our earlier spin chain example ($M = 5$ spins, transfer from $|1\rangle$ to $|3\rangle$), we focus on $\tilde{\tau}$ for $\sigma_{\text{sim}}^{(i)} = 0, \sigma_{\text{sim}}^{(j)}$ pairs that is sufficient to answer **(Q)**. More specifically, we aim to understand how well the no-noise RIM (i.e., the average infidelity) ranks correlate with the general RIM ranks. This is shown in Fig. 4 for each optimization algorithm for $\alpha = 0.05$. For the $\sigma_{\text{sim}}^{(i)} \geqslant 0.03$, the RIM rank order is the most consistent with $\tilde{\tau} \gtrsim 0.6$ for PPO excluding other algorithms. But there is larger shuffling of the ranks of PPO controllers as σ_{sim} increases with deteriorating $\tilde{\tau}$ and SNOBFit takes over. This may be due to small numerical differences in RIM [see Fig. 3(c)] observed, and thus a stronger consistency for $\sigma_{\text{sim}} \leqslant 0.03$ is captured. We highlight in Appendix B 2 that the reason why PPO infidelities correlate more with RIM values at higher σ_{sim} is because it optimizes a discounted RIM ($\sum_i \gamma^i \text{RIM}^{(i)}$ for $0 \leqslant \gamma \leqslant 1$) as its reward function. We extend this analysis for $\sigma_{\text{train}} > 0$ in Appendix B 3 to further corroborate that the infidelity rank order for PPO correlates most with higher order RIMs.

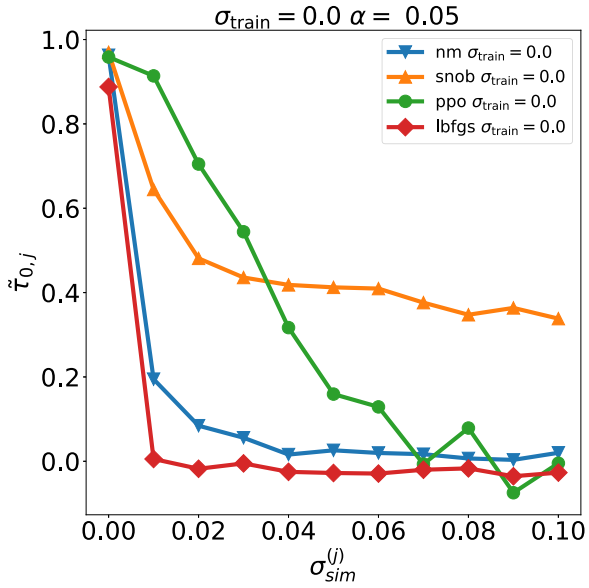

FIG. 4. RIM rank order consistency statistic $\tilde{\tau}$ for the 100 controllers found for the problem $M = 5$, $|1\rangle$ to $|3\rangle$ between the two levels: no simulation noise, $\sigma_{\text{sim}}^{(i)} = 0$ and $\sigma_{\text{sim}}^{(j)}$ from $\{0.0, 0.01, \ldots, 0.1\}$ for (a) Nelder-Mead, (b) SNOBFit, (c) PPO, and (d) L-BFGS without training noise. In other words, this is the correlation of infidelity rank order with the general RIM ranks. The $\tilde{\tau}_{0,j}$ values decline the slowest for PPO until $\sigma_{\text{sim}}^{(j)} = 0.04$ and then SNOBFit takes over compared to the rest. This shows, for this case, that the PPO infidelity rank order correlates the most with RIM rank order for $\sigma_{\text{sim}} \leqslant 0.03$.

The other algorithms typically have a sharper drop at $\sigma_{\text{sim}}^{(j)} = 0, 0.01$ step where the infidelity rank order for L-BFGS and, to a lesser extent, Nelder-Mead is completely noninformative (due to very high fidelity values without noise) and is not consistent with the orders at larger σ_{sim}. This is most likely because the controllers found are the result of second order, gradient-based, or similarly successful search methods for finding optima precisely. Since PPO and SNOBFit are gradient-free, for $\sigma_{\text{sim}} \geqslant 0.03$, their controllers are more consistent in comparison. In this case, the infidelity rank order is more informative of the RIM rank order than, e.g., L-BFGS, as fidelities are not being fully maximized due to the absence of a strong gradient direction. Note that a viable link between the consistency statistic and a generic gradient-based

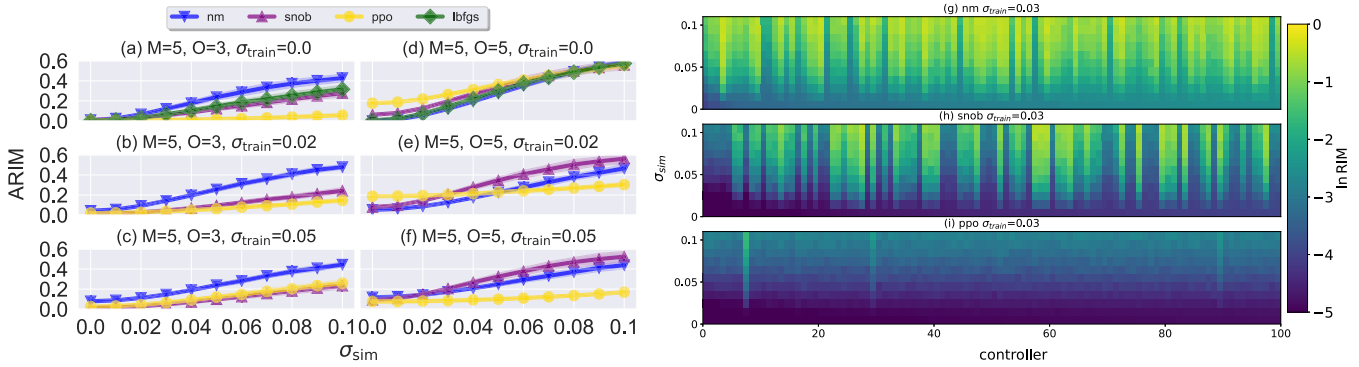

FIG. 5. ARIM as a function of σ_{sim} for $M = 5$ where (a)–(c) are the end-to-middle ($|1\rangle$ to $|3\rangle$) and (d)–(f) are the end-to-end ($|1\rangle$ to $|5\rangle$) transitions (end denoted by O). The ARIM is computed from a distribution of RIM values for 100 controllers for each σ_{sim} for SNOBFit, Nelder-Mead, PPO, and L-BFGS. We identify each algorithm with a unique marker and/or color. Both PPO and SNOBFit are run multiple times at $\sigma_{\text{train}} = 0, 0.02, 0.05$. PPO has higher variance with respect to σ_{train} than SNOBFit and Nelder-Mead, whose performance curves are more in line with the L-BFGS curve for $\sigma_{\text{sim}} \geqslant 0.05$ and mostly worse for $\sigma_{\text{sim}} \leqslant 0.05$. 95% confidence intervals (shading) are computed using nonparametric bootstrap resampling [86] with 100 resamples. (g)–(i) show individual-controller comparisons for $\sigma_{\text{train}} = 0.03$ ranked by fidelity (leftmost is highest).

algorithm is hard to establish, so this does not preclude the existence of algorithms that are $\tilde{\tau}$-wise better.

Finally, note that $\tilde{\tau}$ should be thought of as a proxy of reliability of an algorithm's capability to generate numerical control solutions whose infidelity values are more consistent and predictive of their RIM values at higher σ_{sim}. If strong correlation is obtained, this circumvents (or at least increases confidence for circumventing the latter's) computation.

However, high RIM rank order consistency does not imply that the RIM values remain low at higher noise. Rather, it indicates how much the RIM of a controller is predictive of the controller's relative robustness performance at a higher noise level. The nonparametric nature of $\tilde{\tau}$ removes information about the fidelity value range and should be viewed in conjunction with Figs. 3(a)–3(e). If the correlation signal is strong, it could be used to sidestep the evaluation of the RIM at nonzero noise in favor of using the infidelity instead, eliminating the need for expensive sampling.

B. Comparison of control algorithms with constrained resources

We address our motivating question (C): What is the effect of training noise on a control algorithm's ability to find robust controllers? The overall picture is complex in terms of algorithm rankings. We numerically confirm that there is a problem-dependent optimal noise level that best smooths the optimization landscape for algorithms to more consistently find robust controllers.

We collect 100 controllers at training perturbations $S_{\sigma_{\text{train}}}$ with training noise level $\sigma_{\text{train}} \in \{0, 0.01, \ldots, 0.05\}$ for PPO, SNOBFit, and Nelder-Mead. We do not consider any training noise for L-BFGS, since only the former algorithms are designed to perform optimization with noisy perturbations. This involves using a stochastic fidelity (objective) function call evaluated under the single structured perturbation $S_{\sigma_{\text{train}}}$ (exactly analogous to $S_{\sigma_{\text{sim}}}$).

We select $\sigma_{\text{sim}} \in \{0, 0.01, \ldots, 0.1\}$ to evaluate the RIM of the controllers found at different noise levels with a budget of 10^6 objective function calls per run. Each run corresponds to 100 controllers found under this budget constraint. The ARIM is then used to quantify an algorithm's performance, with respect to robustness and fidelity, based on the 100 controllers that it found during the run.

We only show the representative end-to-middle and end-to-end transition for the state-preparation problem for $M = 5$ at $\sigma_{\text{train}} = 0, 0.02, 0.05$ in Fig. 5. Results for other spin-transitions and training noises are presented in Appendix B 5.

Recall from Eq. (11) that the ARIM is a measure of how far the distribution $\mathbf{P}(\text{RIM})$ is from its ideal δ_0. The ARIM curves at different training noises in Figs. 5(a)–5(f) increase at different rates σ_{sim}, starting from similar base ARIM values at $\sigma_{\text{sim}} = 0$ for each algorithm. Note that the base ARIM value coincides with the average infidelity over controllers, in the absence of training noise.

A spread in ARIM curves indicates that the probabilistic distance of RIM values with respect to the ideal for all controllers increases at different rates. So, the algorithm represented by the slowest growing curve is the best to find robust controllers.

Overall, SNOBFit's and Nelder-Mead's ARIM curves at various training noises perform similarly to L-BFGS across all problems. However, there are distinctions in the region of $\sigma_{\text{sim}} \leqslant 0.05$, where L-BFGS curves start at lower ARIM values and grow more quickly compared to SNOBFit curves at various noise levels. In the region of $\sigma_{\text{sim}} \geqslant 0.05$, the SNOBFit curves comparatively grow more slowly, possibly because the fidelity has degraded so much that further deterioration is less likely across all 100 controllers. The Nelder-Mead curves exhibit similar behavior to the SNOBFit curves in that there is less variance with respect to the σ_{train} levels, both when overall performance is good and when it is poor.

Compared to other algorithms, there is more variance in the PPO ARIM curves across training noises for a particular spin transfer problem, with some curves overlapping each other. The best performing ARIM curve is PPO at $\sigma_{\text{train}} = 0.05$ for the end-to-end transition shown in Fig. 5(f) (and for six of eight cases in Appendix B 5). This indicates that PPO is often capable of finding robust solutions, but the

optimal value of training noise varies across the transition problems.

We also present an extended RIM analysis (like in Sec. IV A) for the controllers found for the same transition problem at training noises for the derivative-free approaches. The RIMs at $\sigma_{\text{sim}} \in 0, \ldots, 0.1$ are plotted in Figs. 5(g)–5(i) for PPO, SNOBFit, and Nelder-Mead at $\sigma_{\text{train}} = 0.03$. On an individual level, SNOBFit and Nelder-Mead controllers share more algorithmic robustness and fidelity characteristics with each other across σ_{train} than with PPO controllers: i.e., they have high RIM variance within distribution per σ_{train}. This performance is also comparable to the L-BFGS controllers shown in Fig. 3(d). On the other hand, individually, the controllers found by PPO differ significantly across σ_{train} where notably the RIM and ARIM values stay uniformly very low for the case $\sigma_{\text{train}} = 0, 0.03$ and the controllers are generally distinctly robust compared to SNOBFit and Nelder-Mead controllers.

Finally, we suggest possible explanations for these differences in behavior between algorithms. Since SNOBFit constructs local quadratic models to estimate gradients, it effectively filters out the perturbations $S_{\sigma_{\text{train}}}$. The manifestation of this effect is that the controllers at one training noise react similarly with respect to the RIM, compared to controllers at other training noises (including the case of no training noise) as well as controllers found by L-BFGS. For Nelder-Mead, there are fewer noise-adaptation mechanisms compared to PPO and SNOBFit for large noise perturbations that might affect the quality of the estimated gradient direction and hence the rate of growth of the ARIM with respect to simulation noise at higher training noise levels is unavoidable.

In contrast, PPO does not filter out the perturbations under $S_{\sigma_{\text{train}}}$ and forms its policy gradient estimates from stochastic fidelity function evaluations, which likely differentiates it from SNOBFit. PPO also effectively estimates the fidelity landscape nonlinearly using a fixed two-layer linear (100×100 dimensional) neural network, which may lead to generally better ARIM performance.

C. Comparison of control algorithms with unconstrained resources

We consider the behavior of the aforementioned control algorithms with an unconstrained number of objective function calls to address our motivating question (D), that is, we seek to understand an algorithm's ability to find robust controllers via the ARIM—without the function call constraint. Furthermore, we wish to ascertain what the effect of the training noise level σ_{train} is on ARIM optimization.

We consider two objective function settings: (i) *stochastic objective*: for each evaluation, a new Hamiltonian is drawn according to the noise model, which corresponds to 1 $S_{\sigma_{\text{train}}}$ perturbation in $\mathcal{F}$ during a single evaluation and (ii) *nonstochastic objective*: where the evaluation is over k perturbed, but fixed, Hamiltonians, predrawn from the noise model such that optimization objective is a deterministic RIM computed from k fixed training perturbations $\{S^{(i)}_{\sigma_{\text{train}}}\}_1^k$. In this case, the function calls are counted as k as they amount to k different fidelity function evaluations per one optimization objective call. Furthermore, since we cannot compute the analytical gradient of both objective functions to use L-BFGS [32] in both settings, we use a version of L-BFGS that approximates the Hessian using forward differences.

We fix the control problem to be the end-to-middle M=5 transition. We consider the change in average ARIM over $\sigma_{\text{sim}} \in \{0, 0.01, \ldots, 0.1\}$ for the top 100 controllers with respect to function calls. Three training noises $\sigma_{\text{train}} = 0, 0.05, 0.1$ are considered. The controller rankings are maintained with respect to the objective function and are updated in steps of 10^6 function calls up to 4×10^7. For the stochastic setting (i), we maintain the controller ranking via the stochastic fidelity function evaluation. For the nonstochastic setting (ii), we maintain the ranking through the deterministic RIM obtained using $k = 100$ pre-drawn training perturbations $\{S^{(i)}_{\sigma_{\text{train}}}\}_1^{100}$. The choice of the hyperparameter k was obtained using cross validation. Specifically, for a particular training noise and control algorithm, we picked a k from $\{10, 100, 10\,000\}$ to compute a RIM in the objective function and then compared it to a RIM computed using $k' = 10\,000$ different training perturbations $\{S^{(i)}_{\sigma_{\text{train}}}\}_1^{10^4}$ that comprise a large validation set during the optimization run for all 100 controllers. We found no significant empirical difference in error between the objective function RIMs for $k = 100$ and $k = 10\,000$. Note that the variance in the RIM decreases as $O(1/k)$ by the law of large numbers.

For the nonstochastic setting (ii), it can be seen from Figs. 6(a)–6(c) (solid lines) that the average ARIM of all algorithms reduces asymptotically with the number of function calls. PPO attains the lowest final average ARIM values at 4×10^7 function calls for each σ_{train} but its final ARIM is not markedly better than the other algorithms considered here. However, this setting is quite expensive in terms of the total number of function calls.

For the stochastic setting (i) [dashed lines in Figs. 6(a)–6(c)], increasing the training noise reduces the ability of the control algorithm to find robust controllers for all the σ_{train}, and the average ARIM is not minimized to the same extent as in setting (ii). This makes sense, since the stochastic objective (i) is a noisy fidelity with a reduced focus on robustness. The average ARIM is no longer reliably improved by any control algorithm in setting (i).

Within setting (i), we note that PPO converges to the lowest average ARIM value compared to the rest of the algorithms. This is theoretically justifiable as PPO optimizes a discounted RIM by design (see Appendix B 2). Another thing to note is that the lowest average ARIM values obtained by PPO in all three settings are similar, even though they are not attained at the same number of function calls. This suggests that PPO's ARIM performance can be made independent of the training noise levels, given an unconstrained number of objective function calls. However, this might be difficult to achieve completely since, even for PPO, there is a selection bias for low infidelity, but not low RIM. This manifests itself in the fact that the average ARIM starts increasing, albeit slowly, with respect to the number of function calls after the lowest average ARIM value is reached. Furthermore, we note that sharp transitions like the stochastic PPO curves depicted in Fig. 6(b) are also typically reported in classical reinforcement learning contexts

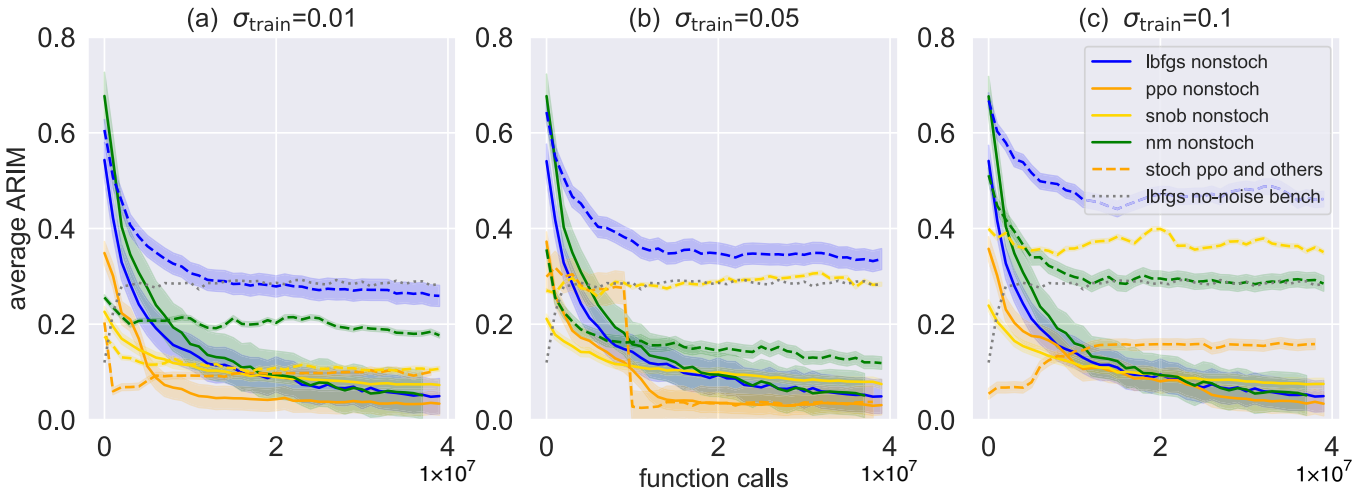

FIG. 6. Asymptotic ARIM performance when the number of objective function calls is unconstrained for the $M = 5$ end-to-middle spin transfer problem. The ARIMs are averaged over the σ_{sim} set $\{0, 0.01, \ldots, 0.1\}$. The stochastic objective setting (i) is shown with dashed lines and deterministic RIM objective setting (ii) with solid lines; different algorithms correspond to different colors. An L-BFGS no-noise benchmark is shown with a dotted line. The target RIM is computed using 100 nonstochastic fidelity evaluations. The ARIM is computed and averaged over a $\sigma_{sim} = 0.0, 0.01, \ldots, 0.1$ set. (a)–(c) correspond to training noise $\sigma_{train} = 0, 0.05, 0.1$, where the curves are for 100 controllers ranked by the corresponding objective function evaluation (i), (ii) and are updated every 10^6 function calls. For setting (ii), all control algorithms asymptotically reduce the average ARIM, but this is not cost competitive with the stochastic setting (i) where PPO performance reaches the local minimum for all noise levels with fewer function calls. We see that the training noise level can help the landscape exploration process; this positively affects PPO in (a), (c) and Nelder-Mead in (b). For setting (i), L-BFGS, then SNOBFit, then Nelder-Mead, then PPO is the most prone to performance deterioration with respect to σ_{train} due to the differences in their reliance on (estimated) gradient information. In these plots, the shading indicates 95% confidence intervals, determined by using bootstrap resampling.

and are linked to sharp improvements in the reward by the algorithm [87].

To compare these results with the more standard noiseless fidelity maximization as a benchmark, we also plot the average ARIM for L-BFGS with a noiseless fidelity objective function and analytical gradient information. This version of L-BFGS accumulates sharp peaks in the fidelity landscape with more function calls since it is gradient based and is effectively climbing to the sharpest peak in the fidelity landscape. Hence its average ARIM flatlines quickly with respect to function calls to a higher value compared to the other control algorithms in setting (i), with the exception of the forward differencing L-BFGS.

Contrasting settings (i) and (ii) for a single control algorithm, the point at which there is an advantage for nonstochastic optimization via setting (ii) is around 10^7 function calls for the algorithms, excluding L-BFGS with noise. For the regime below 10^7 function calls, setting (i) has a clear advantage over (ii) for PPO and SNOBFit.

V. CONCLUSION

We have presented the RIM_p, a statistical generalization of the infidelity in the robustness sense, defined with respect to perturbations of arbitrary noise level S_σ in the fidelity function. The RIM_p is the pth order Wasserstein distance of the infidelity distribution induced by S_σ from some ideal distribution that is impervious to S_σ. We show that it is the pth root of the pth raw moment of the infidelity distribution and can be evaluated using perturbed fidelity function evaluations in physical experiments or Monte Carlo simulations. For $p = 1$, the infidelity measure RIM_1 reduces to the average infidelity. Using a metrization argument, we justify why RIM_1 is a practical robustness measure for quantum control problems due to the convergence of RIM_p values, for all p, given highly robust and high fidelity controllers. The RIM_1 also has a nice interpretation as the area under the curve of the cumulative distribution of the infidelity. We leave further analyses of the utility of the RIM_p for, e.g., the optimization of robustness, for future work.

The RIM is further generalized to define an ARIM to compare the performance of control algorithms in terms of their ability to find robust high-fidelity controllers. Even though the RIM and ARIM are illustrated for static controls, they can be computed in any situation that generates a fidelity distribution over [0,1], including time-dependent controls and open quantum systems, enabling their use and further study for a wide range of practical quantum control problems.

We have used the RIM under model and controller noise to quantify the performance, in terms of the robustness and fidelity, of individual controllers for excitation transfer in spin chains by energy landscape shaping. The controllers were obtained by four control algorithms (PPO, SNOBFit, Nelder-Mead, L-BFGS) at simulation noise scales of up to 10%. Using the RIM, we found that high-fidelity controllers can vary widely in robustness to noise across all algorithms that we studied, although there are notable differences in algorithmic

efficacy with respect to robustness, as indicated by the ARIM. We also demonstrate a consistency statistic that can be used to differentiate control algorithms by how correlated their controller infidelities are with the RIM_1. This provides a method to predict robustness via the RIM_1 without its explicit evaluation.

To compare the control algorithms, we studied their ARIM performance for multiple spin transfer problems. Under constrained function calls of a stochastic objective function (noisy infidelity), PPO performed better than SNOBFit and Nelder-Mead at certain, problem-specific training noise levels. SNOBFit performance at different training noise levels was similar, regardless of whether it was good or bad, suggesting that it is filtering out the noise. Nelder-Mead exhibits similarly consistent behavior across training noise levels with less than optimal performance for all but one problem. With unconstrained stochastic function calls, PPO showed excellent performance compared to the other algorithms, independent of the training noise level, since its reward accumulation strategy implicitly optimizes a discounted RIM.

In contrast, when optimizing the RIM_1 (average infidelity) over a fixed ensemble of perturbations, all algorithms were capable of asymptotically finding an optimum. However, this approach is expensive in terms of the number of function calls compared to the aforementioned stochastic optimization setting with a noisy fidelity function as the objective. Our results also show that for stochastic settings, e.g., shot noise, PPO (or more generally reinforcement learning) is a promising approach to obtain robust controllers.

A limitation of this work is that we require the computation of multiple controllers per control problem. In simulation, this further involves numerous time-consuming matrix exponential evaluations to generate a large number of samples per controller to approximate the RIM measure. More work is necessary to elucidate the fundamental limitations of the optimization landscape. Nevertheless, our statistical robustness approach is a useful tool that can be applied in a wide range of quantum control scenarios where analytic approximations with small and/or uncorrelated noise are unsuitable.

ACKNOWLEDGMENTS

I.K. acknowledges support for this work by a PhD scholarship from the School of Computer Science and Informatics, Cardiff University. The authors acknowledge the support of the Supercomputing Wales project to obtain the computational results presented herein, which is partly funded by the European Regional Development Fund (ERDF) via the Welsh Government.

APPENDIX A: RIM_p CALCULATIONS

1. *p*th order RIM

In the subsequent argument, recall that the quantile function is the inverse of the CDF function. Following Theorem 1, we can write the RIM_p as

$$\mathrm{RIM}_p = \left(\int_0^1 |Q_{\mathbf{P}(\mathcal{F})}(z) - Q_{\delta_1}(z)|^p \, dz\right)^{\frac{1}{p}}. \tag{A1}$$

Note that both terms in the integrand are 0 at $z = 0$. Then, for $z \in (0, 1]$, by definition,

$$\begin{aligned} Q_{\delta_1}(z) &= \inf\{x \in \mathbb{R} : C_{\delta_1}(x) \geqslant z > 0\} \\ &= \inf\{x \in \mathbb{R} : C(x = 1) \geqslant z > 0\} \\ &\quad \text{using the CDF of } \delta_1 \text{ in Eq. (4)} \\ &= 1. \end{aligned} \tag{A2}$$

Another way to see this is to use the inverse property $Q_{\delta_1}(\cdot) = C_{\delta_1}^{-1}(\cdot)$. The CDF is 0 in the interval $[-\infty, 1)$ and 1 in $[1, \infty]$. Next, we perform a change of variable $z = C_{\mathbf{P}(\mathcal{F})}(f)$. The differential is given by $dz = \frac{dC_{\mathbf{P}(\mathcal{F})}(f)}{df} df = \mathbf{P}(\mathcal{F} = f) df$ as the derivative of the CDF with respect to the random variable is the probability distribution function. Substituting the terms, we get

$$\mathrm{RIM}_p = \left(\int_0^1 |Q_{\mathbf{P}(\mathcal{F})}(C_{\mathbf{P}(\mathcal{F})}(f)) - 1|^p \mathbf{P}(\mathcal{F} = f) \, df\right)^{\frac{1}{p}}. \tag{A3}$$

Now we use the fact that $Q_{\mathbf{P}(\mathcal{F})}(C_{\mathbf{P}(\mathcal{F})}(f)) = f$ (inverse property) to obtain

$$\mathrm{RIM}_p = \left(\int_0^1 \mathbf{P}(\mathcal{F} = f)|f - 1|^p \, df\right)^{\frac{1}{p}}. \tag{A4}$$

Since the domain of integration remains invariant, for fidelity measures with support in [0,1], it can be extended to $[-\infty, \infty]$. We obtain

$$\begin{aligned} \mathrm{RIM}_p &= \left(\int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F} = f)|f - 1|^p \, df\right)^{\frac{1}{p}} \\ &= \left(\int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F} = f)(1 - f)^p \, df\right)^{\frac{1}{p}} \\ &\quad \text{as } f \leqslant 1, \text{ switch the order and drop } |\cdot| \\ &= \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1 - f)^p]. \end{aligned} \tag{A5}$$

We obtain the last line using the expectation operator $\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(\cdot)] := \int(\cdot)\mathbf{P}(\mathcal{F} = f) \, df$. For $p = 1$, using Eq. (A5), $\mathrm{RIM}_1 = 1 - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f]$.

We can further decompose the RIM_p as a sum of expectations of various powers of the fidelity:

$$\begin{aligned} \mathrm{RIM}_p &= \left(\sum_{k=0}^{p} \binom{p}{k} (-1)^k \int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F} = f) f^k \, df\right)^{\frac{1}{p}} \\ &\quad \text{using the binomial theorem} \\ &= \left(\sum_{k=0}^{p} \binom{p}{k} (-1)^k \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f^k]\right)^{\frac{1}{p}}. \end{aligned} \tag{A6}$$

For example, using Eq. (A6) for $p = 2$, we obtain

$$\begin{aligned} \mathrm{RIM}_2 &= \sqrt{1 - 2\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f] + \mathrm{Var}(f) + \mathbb{E}^2{}_{f\sim\mathbf{P}(\mathcal{F})}[f]} \\ &\quad \text{using Eq. (A6)} \\ &\quad \mathrm{Var}(X) = \mathbb{E}_{X\sim\mathbf{P}}[X^2] - \mathbb{E}^2_{X\sim\mathbf{P}}[x] \\ &= \sqrt{\mathrm{RIM}_1 + \mathrm{Var}(f) - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f]\mathrm{RIM}_1} \\ &= \sqrt{\mathrm{Var}(f) + \mathrm{RIM}_1^2} \\ &\quad \text{expanding } \mathrm{RIM}_1 \text{ and simplifying.} \end{aligned} \tag{A7}$$

Likewise, we get

$$\mathrm{RIM}_3 = \left(\mathrm{RIM}_1^3 + 3\,\mathrm{Var}(f) + \mathbb{E}^3{}_{f\sim\mathbf{P}(\mathcal{F})}[f] - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[f^3]\right)^{\frac{1}{3}}. \tag{A8}$$

The degree of distinguishability of the fidelity distribution from the ideal becomes better for higher p at the cost of the outliers becoming more influential.

2. Error bound on the RIM_p and ARIM_p estimators

Here we propose a probably approximately correct alternative error bound for an estimation $\widehat{\mathrm{RIM}}_p$ of RIM_p in Eq. (A5) based on an empirical estimate $\widehat{\mathbf{P}}(\mathcal{F})$ of its generating probability distribution $\mathbf{P}(\mathcal{F})$. With probability at least $1-\delta/2$,

$$\begin{aligned}
&|\widehat{\mathrm{RIM}}_p - \mathrm{RIM}_p| \\
&\quad = |\mathbb{E}_{f\sim\widehat{\mathbf{P}}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}} - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}}| \\
&\quad \leqslant |\mathbb{E}_{f\sim\widehat{\mathbf{P}}(\mathcal{F})}[(1-f)^p] - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]|^{\frac{1}{p}} \\
&\quad = [t]\left|\int_0^1 \widehat{\mathbf{P}}(\mathcal{F}=f)(1-f)^p\,df \right. \\
&\qquad \left. - \int_0^1 \mathbf{P}(\mathcal{F}=f)(1-f)^p\,df\right|^{\frac{1}{p}} \\
&\quad \leqslant \left(\int_0^1 |\widehat{\mathbf{P}}(\mathcal{F}=f) - \mathbf{P}(\mathcal{F}=f)|(1-f)^p\,df\right)^{\frac{1}{p}} \\
&\quad = \left(\int_0^1 \left|\widehat{\mathbf{P}}(\mathcal{F}=f) - \mathbb{E}_{\widehat{\mathbf{P}}\sim\mathbf{D}}[\widehat{\mathbf{P}}(\mathcal{F}=f)]\right|(1-f)^p\,df\right)^{\frac{1}{p}} \\
&\quad \leqslant \frac{C^{\frac{1}{p}}}{p+1} = \frac{1}{p+1}\left(\frac{\ln\frac{4}{\delta}}{2n}\right)^{\frac{1}{2p}},
\end{aligned} \tag{A9}$$

where the second line and the fourth line come from using the reverse triangle inequality and in the fifth line we rewrite the true distribution $\mathbf{P}(\mathcal{F})$ as $\mathbb{E}_{\widehat{\mathbf{P}}\sim\mathbf{D}}[\widehat{\mathbf{P}}(\mathcal{F})]$, which is true for any unbiased empirical estimator. We use McDiarmid's inequality to obtain the bounding constant C using the fact that the probability distribution $\mathbf{D}$ generates a family of random variable empirical distributional estimators $\widehat{\mathbf{P}}_j = \frac{1}{n}\sum_{i=1}^n \delta_{f_i}$, where we have the differences occurring only on the kth coordinate,

$$\left|\widehat{\mathbf{P}}(f_1,\ldots,f_k,\ldots,f_n) - \widehat{\mathbf{P}}(f_1,\ldots,f_{k'},\ldots,f_n)\right| \leqslant \frac{1}{n}, \tag{A10}$$

where n is the number of samples. A similar bound can also be derived for the $\widehat{\mathrm{ARIM}}$ estimator. This error bound is similar to the DKW (Dvoretzky-Kiefer-Wolfowitz) bound for the ECDF and would suffice in generating the 95% confidence intervals for Fig. 5 without the need to do bootstrap resampling.

3. Relative order of RIM_p

Using Lyapunov's inequality, stating that $\mathbb{E}[|X|^q]^{1/q} - \mathbb{E}[|X|^p]^{1/p} \geqslant 0$ for $q \geqslant p > 0$ for some $\mathbb{E}[|X|^t] < \infty$, we show that

$$\begin{aligned}
&\mathrm{RIM}_q - \mathrm{RIM}_p \\
&\quad = \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^q]^{\frac{1}{q}} - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}} \\
&\quad = \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[|(1-f)|^q]^{\frac{1}{q}} - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[|(1-f)|^p]^{\frac{1}{p}} \\
&\quad \geqslant 0.
\end{aligned} \tag{A11}$$

For any $q \geqslant p \geqslant s > 0$, it follows that $\mathrm{RIM}_q \geqslant \mathrm{RIM}_p \geqslant \mathrm{RIM}_s$. The converse is true without the pth roots. The linearity of expectations implies that $\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p - (1-f)^q] \geqslant 0 \iff 0 < p \leqslant q$.

We can also derive a lower bound on RIM_p. For some $p' \geqslant p$, we have

$$\begin{aligned}
\mathrm{RIM}_{p'} \leqslant \mathrm{RIM}_p^{\frac{p}{p'}} &= \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p'}} \\
&= \frac{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}}}{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p]^{\frac{1}{p}-\frac{1}{p'}}} \\
&\leqslant \frac{\mathrm{RIM}_p}{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)]^{1-\frac{p}{p'}}} \\
&\leqslant \frac{\mathrm{RIM}_p}{(\min_f(1-f))^{1-\frac{p}{p'}}},
\end{aligned} \tag{A12}$$

where the relation in the second to last line is obtained by applying Jensen's inequality and the final line is obtained from the observation that $\min_f(1-f) < \mathbb{E}[1-f]\ \forall f$. Note that this result still depends on the data. Higher orders p and p' of the RIM are related to each other in a concave sense and when $p, p' \to \infty$, the RIMs become more equivalent. Conversely, near perfect fidelity, all the RIMs are converging to 0, but the presence of an outlier fidelity sample strongly governs how much discrepancy there still is between a higher-order RIM and a lower order RIM. This discrepancy is still concavely dependent on p and p'.

We arrive at equivalence relations for RIMs of different orders by noting that $\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}[(1-f)^p] \geqslant m \sup_f(1-f) = m$ for the smallest positive finite measure $m > 0$ on the domain set on which we define the probability distribution $\mathbf{P}(f)$. This follows from the continuity of f and the continuity of $\mathbf{P}(f)$. If f already has an ideal distribution, then this is trivially true. Equation (A12) yields

$$\mathrm{RIM}_{p'} \leqslant m^{(\frac{1}{p'}-\frac{1}{p})}\mathrm{RIM}_p. \tag{A13}$$

In practical settings, e.g., when using the ECDF, $m \geqslant \frac{1}{n}$. Intuitively, this follows from the observation that for any $\mathbb{E}_{\mathbf{P}(X)}[X^p] = \int P(X)X^p\,dX \approx \frac{1}{n}\sum_{i=1}^n X_i^p$ using samples $X_1,\ldots,X_n$. For the estimated $\widehat{\mathrm{RIM}}_p$:

$$\widehat{\mathrm{RIM}}_{p'} \leqslant n^{(\frac{1}{p}-\frac{1}{p'})}\widehat{\mathrm{RIM}}_p. \tag{A14}$$

This implies that RIMs of different orders are equivalent (in the convergence sense) when the bound holds.

Instead, motivated by the convergence in the Wasserstein distance, for fixed n, RIM_1 can effectively constrain any RIM_p with $p > 1$, since growth in p is sublinear. This implies that the RIMs converge when they tend to 0, as seen in Fig. 7. We also note that the higher order RIMs increase the measure's sensitivity to outliers greatly, even though growth in the RIM is sublinear in p. For most practical purposes, the first-order

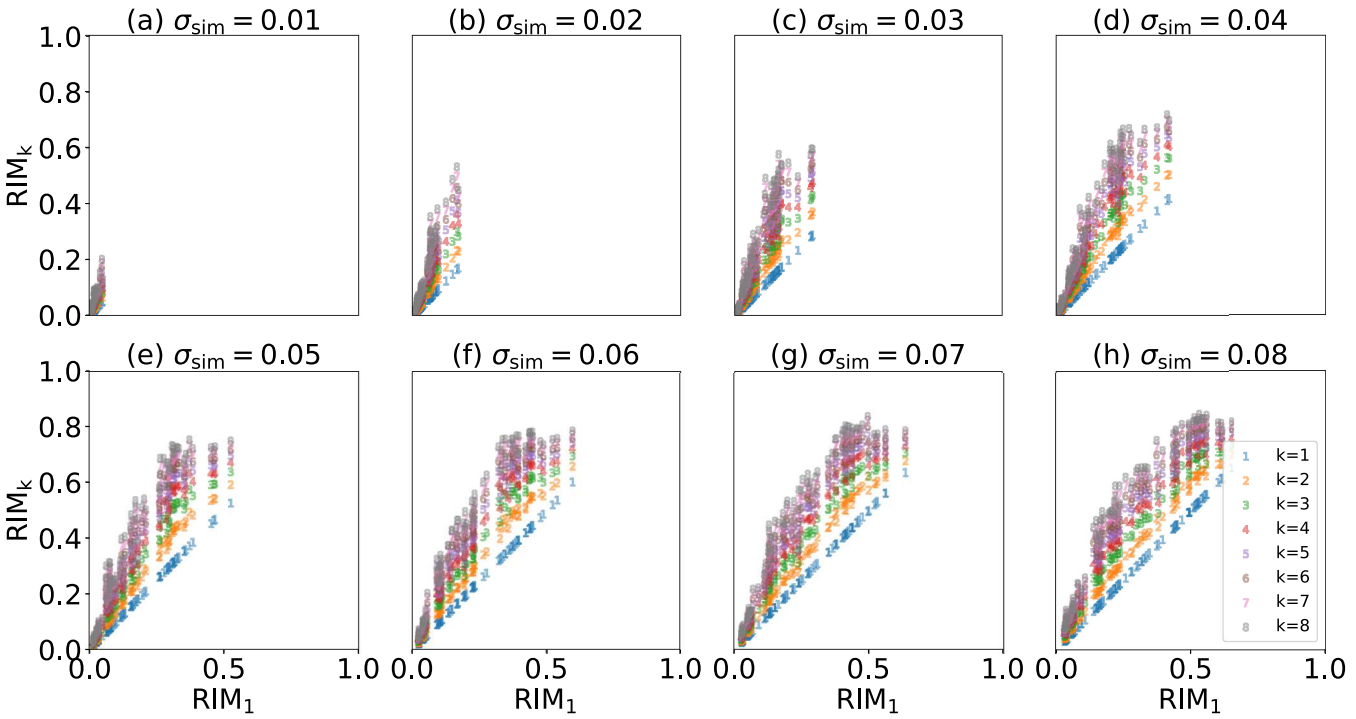

FIG. 7. RIM_k scaling as a function of RIM_1 for 100 controllers for $M = 5$ and the transition from $|1\rangle$ to $|3\rangle$ and $N = 100$ samples per controller for RIM evaluation. Each subplot (a)–(h) corresponds to a noise level σ_{sim} indexing the fidelity probability distribution $\mathbf{P}_{\sigma_{sim}}(\mathcal{F})$. There is convergence and thus more agreement in RIM_k for small values.

RIM measure should be sufficient for performance measurements, especially in the quantum technologies setting.

APPENDIX B: OPTIMIZATION ALGORITHMS

1. Implementation details of the optimization objectives

Details about the optimization objectives for the numerical results in Sec. IV are given in Table I. In every section, for every σ_{sim}, the RIM is evaluated using $N = 100$ Monte Carlo $S_{\sigma_{sim}}$ perturbations to the fidelity function. The RIM itself is only optimized in Sec. IV C for the nonstochastic case (ii) where 100 $S_{\sigma_{sim}}$ are sampled at the start and are reused for every function call. Note, however, that we count these as 100 function calls as these amount to 100 fidelity function evaluations. Also, for better performance, in Sec. IV C for the stochastic case (i), instead of using the analytical form for the

TABLE I. Implementation details for various optimization settings in the paper. For Sec. IV C, the asymptotic setting, (i) refers to the stochastic scenario and (ii) refers to the nonstochastic scenario where the RIM is optimized using the same 100 fixed set of perturbations $\{S_{\sigma_{train}}\}$ per function call.

Section	Objective function (OF) and args.	Train (OF) noise	Algorithm	Total OF calls	Single call cost
IV A 1	$\mathcal{F}$	No	L-BFGS	10^6	1
IV A 1	$\mathcal{F}$	No	PPO	10^6	1
IV A 1	$\mathcal{F}$	No	SNOBFit	10^6	1
IV A 1	$\mathcal{F}$	No	Nelder-Mead	10^6	1
IV B	$\mathcal{F}$	No	L-BFGS	10^6	1
IV B	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	PPO	10^6	1
IV B	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	SNOBFit	10^6	1
IV B	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	Nelder-Mead	10^6	1
IV C(i)	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	L-BFGS	∞	1
IV C(i)	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	PPO	∞	1
IV C(i)	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	SNOBFit	∞	1
IV C(i)	$\mathcal{F}$ and 1 $S_{\sigma_{train}}$	Yes	Nelder-Mead	∞	1
IV C(ii)	RIM and 100 fixed $S_{\sigma_{train}}$	No	L-BFGS	∞	100
IV C(ii)	RIM and 100 fixed $S_{\sigma_{train}}$	No	PPO	∞	100
IV C(ii)	RIM and 100 fixed $S_{\sigma_{train}}$	No	SNOBFit	∞	100
IV C(ii)	RIM and 100 fixed $S_{\sigma_{train}}$	No	Nelder-Mead	∞	100

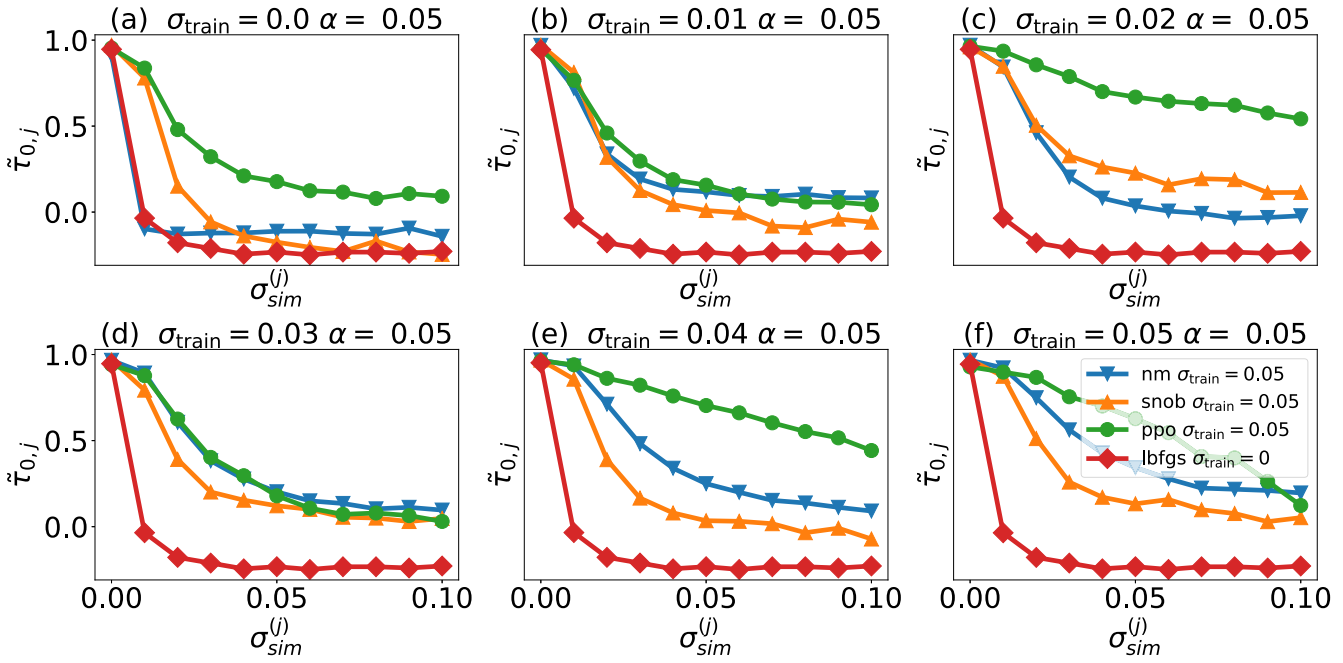

FIG. 8. Consistency statistic $\tilde{\tau}_{0,j}$ for all algorithms at $\sigma_{\text{train}} = 0.0, \ldots, 0.05$ for discrepancy parameter $\alpha = 0.05$ for $M = 5$ and the transition from $|1\rangle$ to $|3\rangle$. Case (a) was presented in the main text. For (f), $\sigma_{\text{train}} = 0.04$, PPO is actually more robust in terms of ARIM growth compared with (e) as seen from their positions in Fig. 5(b). Characteristically, most low ARIMPPO controllers show high rank consistency in the region $0 \leqslant \sigma_{\text{sim}} \leqslant 0.04$. Nelder-Mead is similar to L-BFGS in all plots except (e) and (f), where it shows slightly more consistency than PPO and SNOBFit controllers.

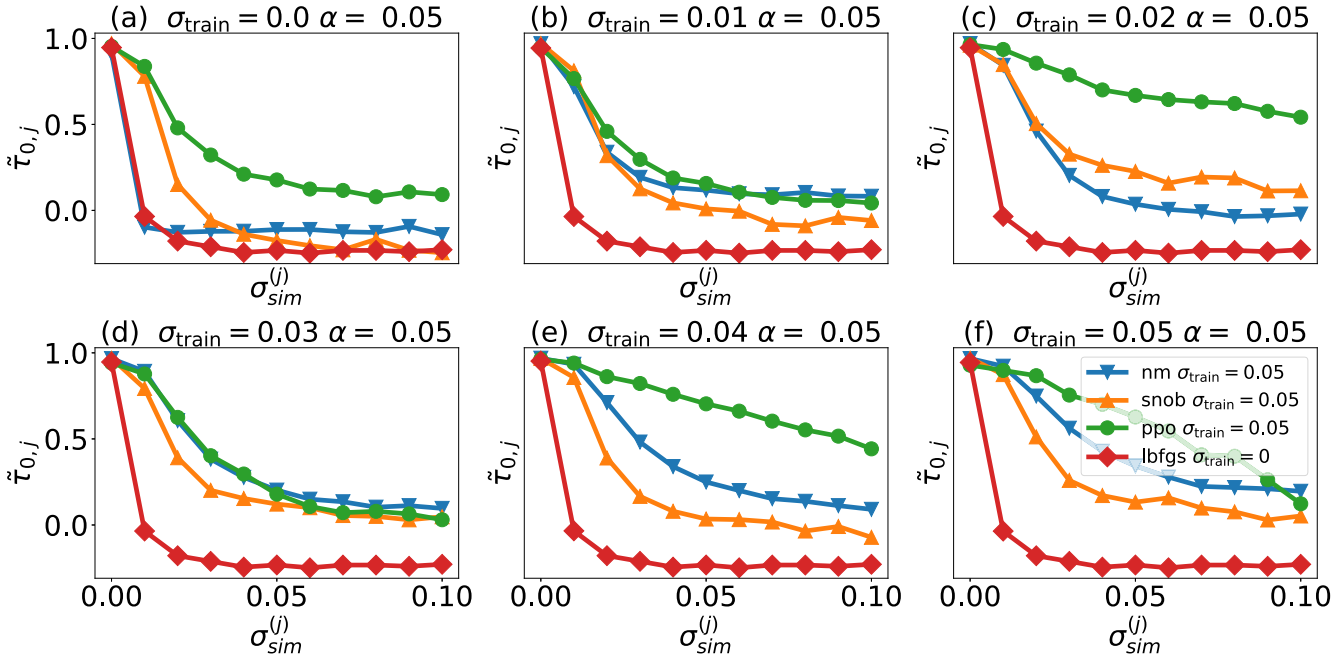

FIG. 9. Consistency statistic $\tilde{\tau}_{0,j}$ for all algorithms at $\sigma_{\text{train}} = 0.0, \ldots, 0.05$ for discrepancy parameter $\alpha = 0.05$ for $M = 5$ and the transition from $|1\rangle$ to $|4\rangle$. Again, the PPO curves are the most consistent.

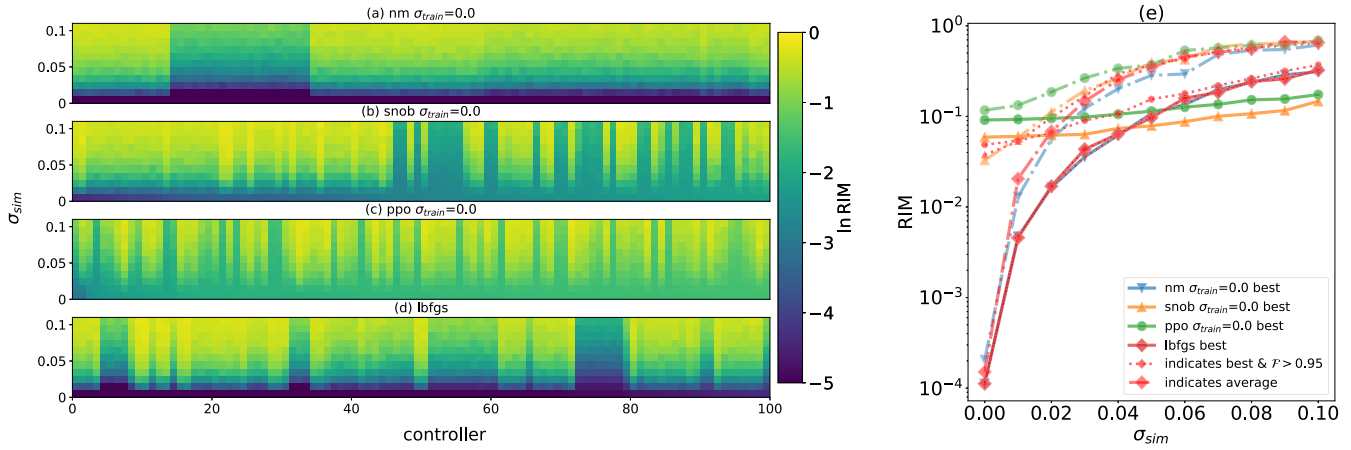

FIG. 10. (a)–(d) 100 controllers found for the *XX* spin chain model, Eq. (13), using Nelder-Mead, SNOBFit, PPO ($\sigma_{\text{train}} = 0$), and L-BFGS for $M = 5$ and the spin transition from $|1\rangle$ to $|5\rangle$. All algorithms find controllers that are not very robust as indicated by the RIM. PPO has notably worse initial infidelities for all controller compared to Fig. 3(c) but their degradation is slow, as seen from (e). This is only the case for this noise level and Fig. 12(r) indicates the existence of a much better controller set at $\sigma_{\text{sim}} = 0.05$ that is similar in performance to Fig. 3(c). From (e), we can see that Nelder-Mead and L-BFGS optimize the infidelity to $<10^{-4}$. However, these best controllers decay in robustness very quickly as well.

gradient of fidelity $\mathcal{F}$, we use finite differences to approximate the gradients $\nabla_{\Delta}\mathcal{F}$ (where Δ are the controls).

2. PPO optimizes a discounted RIM_1

We follow the standard finite-horizon Markov decision process formulation for the reinforcement learning setting for states, actions and one-step state transition rewards (s_t, a_t, r_t) that are sampled in trajectories $\tau = \{(s_t, a_t, r_t) : t = 1, \ldots, T\}$ stored in the buffer $\mathbf{D}$. The PPO algorithm uses a clip objective to update the policy π_θ parameters θ with first-order constraints that minimize policy distributional divergence. The policy objective is

$$\theta_{k+1} \propto \arg\max_{\theta} \sum_{\tau\in\mathbf{D}} \sum_{a_t,s_t,r_t\in\tau}^{T} \min\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A_{\pi_{\theta_k}} \times (s_t, a_t)\,\text{clip}\left(\epsilon, A_{\pi_{\theta_k}}(s_t, a_t)\right)\right], \tag{B1}$$

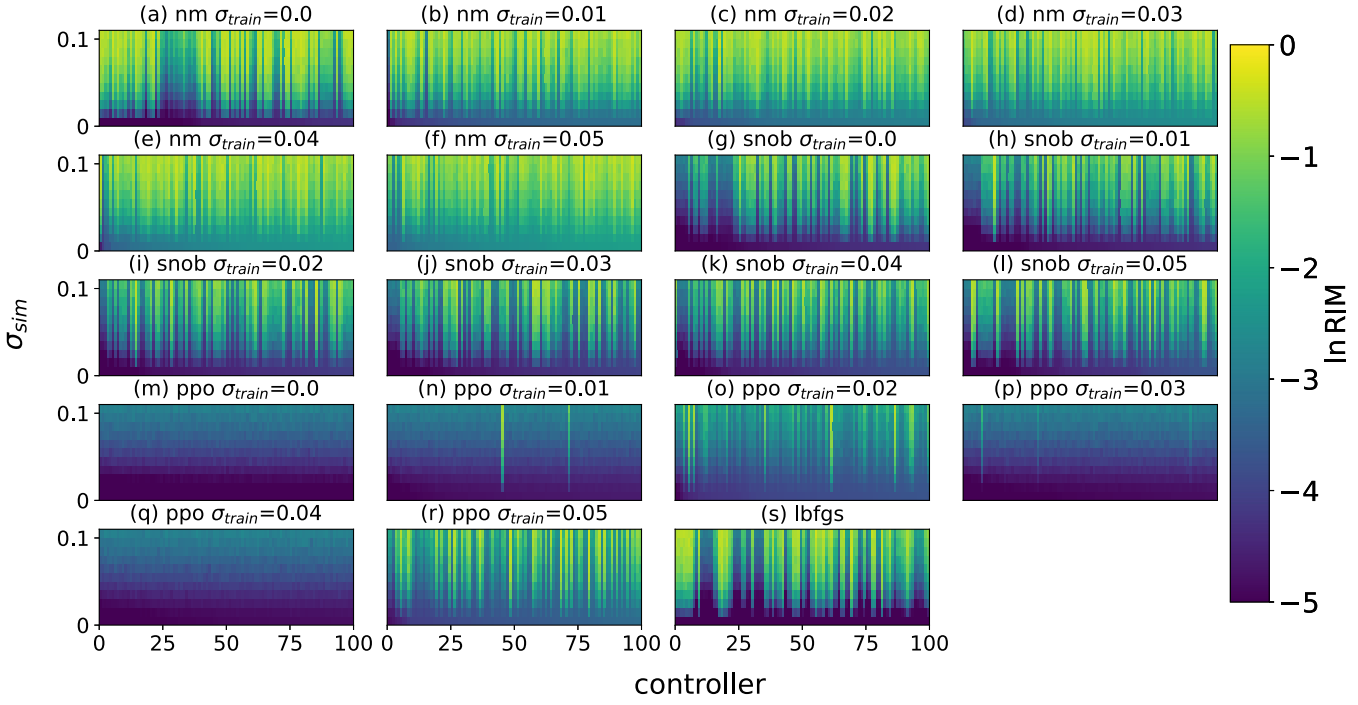

FIG. 11. Individual-controller comparison between (a)–(f) Nelder-Mead, (g)–(k) SNOBFit, and (m)–(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity (left) for the case $M = 5$ and the spin transition from $|1\rangle$ to $|3\rangle$. (s) shows the L-BFGS results for the same spin transition problem.

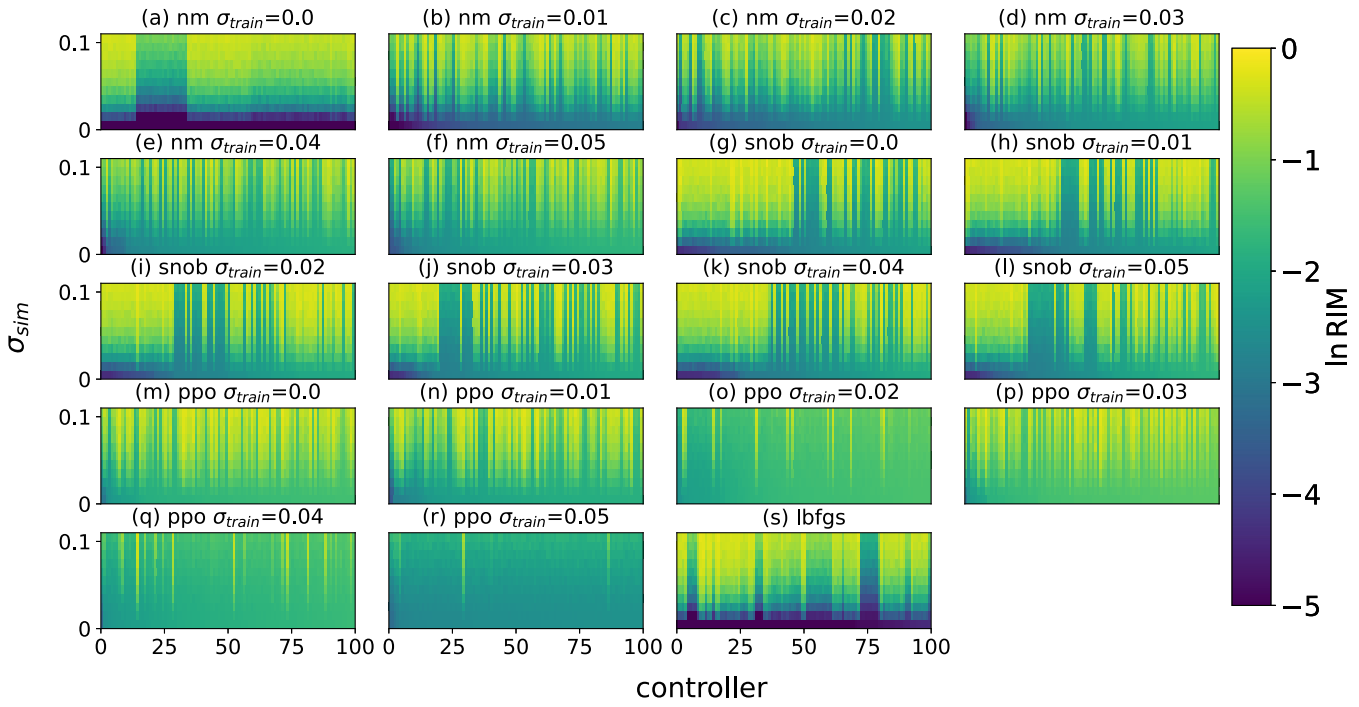

FIG. 12. Individual-controller comparison between (a)–(f) Nelder-Mead, (g)–(k) SNOBFit, (m)–(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity for the case $M = 5$ and the spin transition from $|1\rangle$ to $|5\rangle$. (s) shows the L-BFGS result for $\sigma_{\text{train}} = 0$.

where $\pi(\cdot)$ is the policy probability distribution. The advantage estimates are

$$A_{\pi_{\theta_k}}(s_t, a_t) = \sum_{i=t}^{T-1} (\gamma\lambda)^{i-t}\big(r_t + \gamma V_{\phi_k}(s_{t+1}) - V_{\phi_k}(s_t)\big), \quad \text{(B2)}$$

with value function $V_\phi(s_t) = \mathbb{E}_\pi[\sum_{i=0}^{T-1} \gamma^i r_{t+i+1} | s = s_t]$, where ϕ are the value function parameters. The value function is regressed onto discounted rewards sampled according to $\pi(\cdot)$. The clip function truncates the advantages to be between $(1 \pm \epsilon)A_{\pi_{\theta_k}}$. The value function's optimization objective is

$$\phi_{k+1} \propto \arg\max_\phi \sum_{\tau\in\mathbf{D}} \sum_{t=0}^{T} \left(V_\phi(s_t) - \sum_{i=t}^{T} \gamma^i r_i\big(s_t^\tau\big) \right)^2. \quad \text{(B3)}$$

The algorithm tries to maximize this expression. In the case of flat rewards and advantages $\lambda = \gamma = 1$, the advantage estimates are

$$A_{\pi_{\theta_k}}(s_t, a_t) = V_{\phi_k}(s_t) - \left(V_{\phi_k}(s_T) + \sum_{i=t}^{T-1} r_t \right) = V_{\phi_k}(s_t) - \widehat{V_{\phi_k}}(s_t). \quad \text{(B4)}$$

The value function can be written in terms of an expectation under the policy, as an average reward: $V_\phi(s_t) = T\mathbb{E}_\pi[\frac{1}{T}\sum_{i=t}^{T} r_i | s = s_t]$. The optimal value function is defined by $V_*(s_t) = \max_\pi V_\phi(s_t)$, which is maximized if the policy is optimal, i.e., $\pi_\theta = \pi_{\theta^*}$ at $\theta = \theta^*$. Near optimality, the advantages are approximately 0 as there should be no advantages conferred to the optimal policy π_{θ^*}, which also has an optimal value function. Thus, $\widehat{V_{\phi^*}}(s_t) \to V_{\phi^*}(s_t)$ as $A_{\pi_{\theta^*}} \to 0$. The sample rewards minus the predicted rewards by the value function go to 0 in Eq. (B1). The same argument applies with discounts $\gamma, \lambda < 1$ and, hence, it can be shown that the algorithm optimizes a discounted RIM_1 estimator as its value function. Most reinforcement learning algorithms effectively optimize the average or cumulative reward $\hat{J} \propto \sum_i r_i$ due to the one-step heuristic application of the Bellman principle of optimality [88].

3. More consistency statistic plots

This section expands the discussion on the consistency statistic in the main text in Sec. IV A 2. We plot the consistency statistic $\tilde{\tau}_{0,j}$ for all algorithms for $\alpha = 0.05$ for the case $M = 5$ and the transition $|1\rangle$ to $|3\rangle$ in Figs. 8(a)–8(f) [8(a) is Fig. 4] and $|1\rangle$ to $|4\rangle$ in Figs. 9(a)–9(f) for multiple training noise levels. Note that for each subplot the L-BFGS curve is always the same at $\sigma_{\text{train}} = 0$. The controllers found by PPO at $\sigma_{\text{train}} = 0.05$ are less consistent for some noise levels than others, e.g., $\sigma_{\text{sim}} \geqslant 0.04$ compared with the controllers found at $\sigma_{\text{train}} = 0.04$. This is also true for SNOBFit and Nelder-Mead. Moreover, the decline in the correlation values is smoothest for PPO compared to the rest for nearly all twelve instances shown in both figures. With more training noise, Nelder-Mead is sometimes closer in consistency to the controllers found to L-BFGS, e.g., Figs. 8(a) and 8(b). But it produces more consistent controllers with increasing training noise likely due to diminishing returns of the gradient direction, makes its behavior more like SNOBFit and PPO.

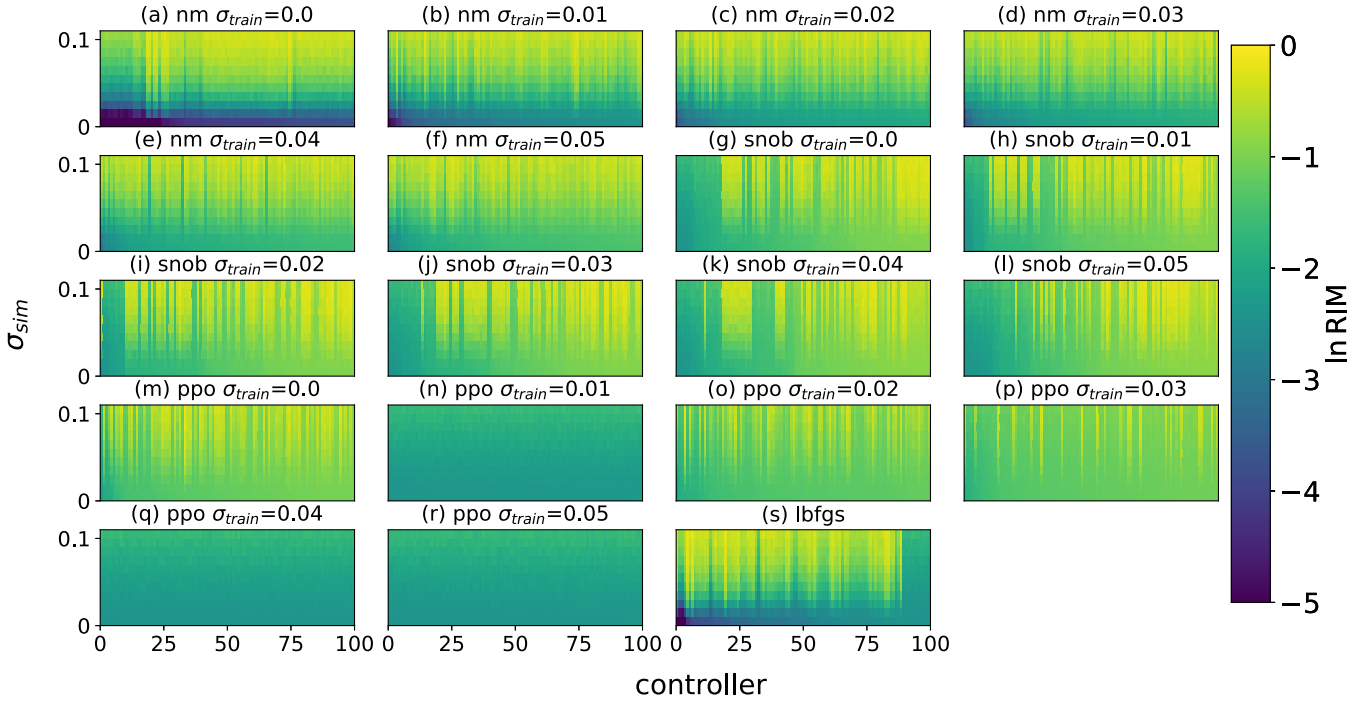

FIG. 13. Individual-controller comparison between (a)–(f) Nelder-Mead, (g)–(k) SNOBFit, (m)–(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity for the case $M = 6$ and the spin transition from $|1\rangle$ to $|6\rangle$. (s) shows the L-BFGS result for $\sigma_{\text{train}} = 0$.

For most PPO runs, the consistency statistic is highest for $\sigma_{\text{sim}} \leqslant 0.04$ and thus the infidelity rank order is a good predictor of RIM rank order for higher σ_{sim}, which was not observed for any of the other algorithms. Also note that this analysis does not reveal anything about how high the RIM values are for the controllers (a drawback of the nonparametric test) and

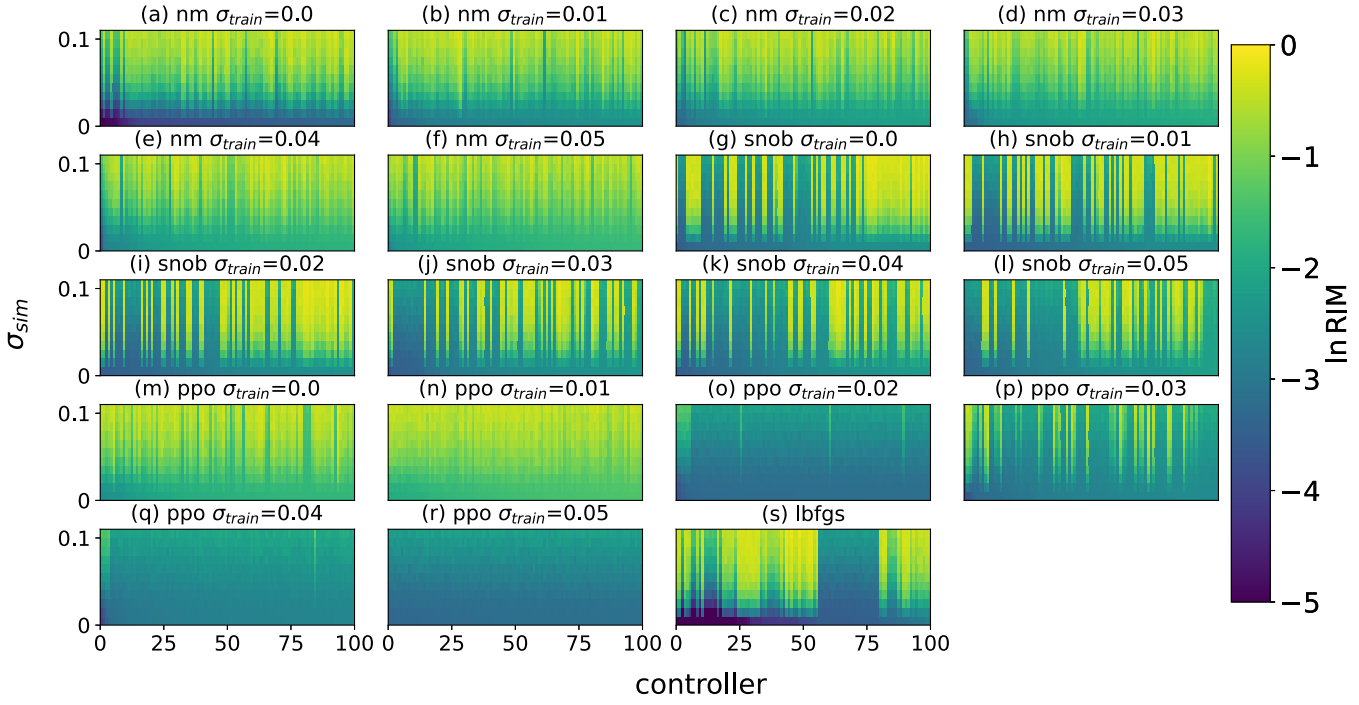

FIG. 14. Individual-controller comparison between (a)–(f) Nelder-Mead, (g)–(k) SNOBFit, (m)–(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity for the case $M = 6$ and the spin transition from $|1\rangle$ to $|4\rangle$. (s) shows the L-BFGS result for $\sigma_{\text{train}} = 0$.

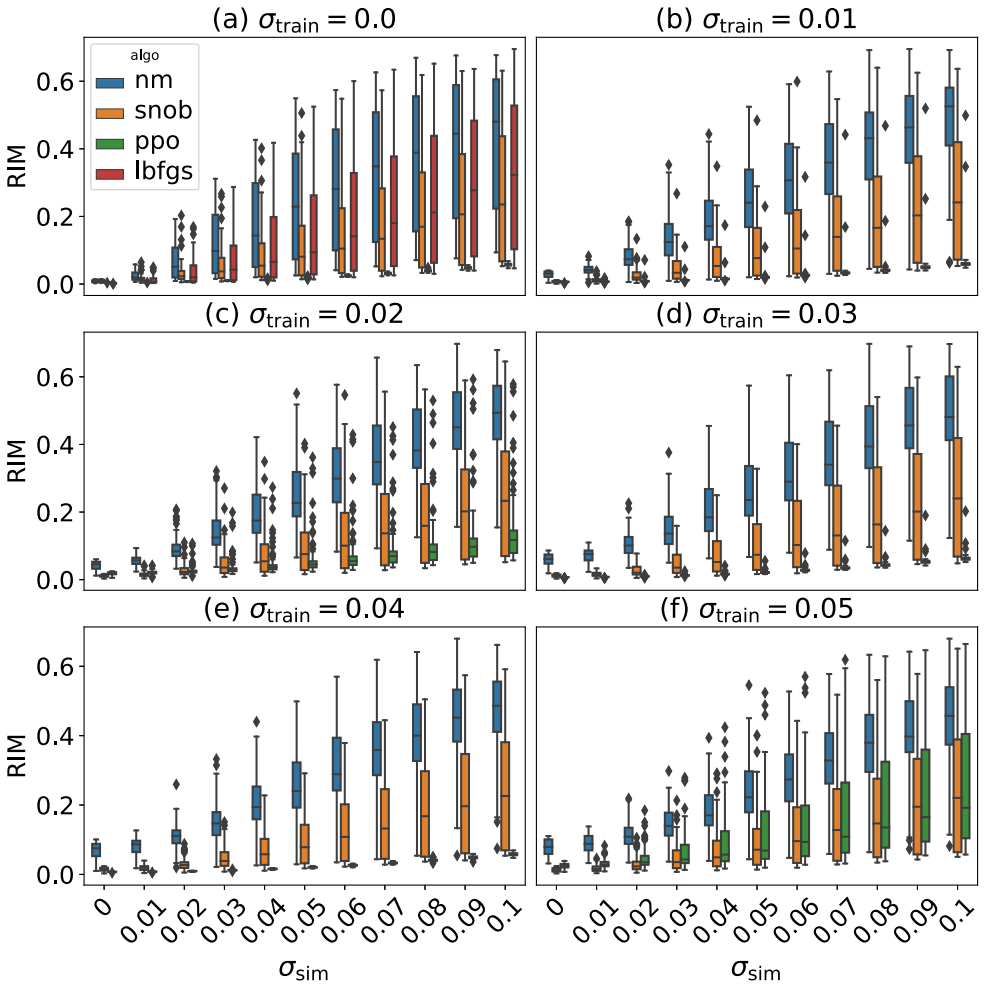

FIG. 15. Box plots of the RIM for the 100 controllers for $M = 5$, $|1\rangle$ to $|3\rangle$ shown in Fig. 11 in the main text found by Nelder-Mead, SNOBFit, and PPO for various σ_{train} (a)–(f). For the case $\sigma_{\text{train}} = 0$ in (a), we also show L-BFGS box plots as a reference. On the distributional level, PPO controllers are generally the more robust of the three with respect to the RIM, but there is high variance across σ_{train} compared to the SNOBFit and Nelder-Mead controllers. The median SNOBFit RIM value per σ_{sim} is higher than L-BFGS, so it has a longer left tail. The Nelder-Mead controllers have the most weight on their right tails and are comparatively the worst.

should be processed as companion plots to the figures where these explicit values are shown.

4. More individual controller plots

The results presented in Fig. 3 ($M = 5$ and transition $|1\rangle$ to $|3\rangle$) are not reflective of PPO's general behavior on the extended sample of problems examined in Sec. IV B. Figure 10 shows the case ($M = 5$ and transition $|1\rangle$ to $|4\rangle$), where all the controllers found are not very robust. This is likely either due to unlucky sampling of the space of possible controllers or their nonexistence. Note that SNOBFit and PPO are similar in their RIM degradation, as observed from Fig. 10(e). We also provide some more cases ($M = 5$ and transition $|1\rangle$ to $|5\rangle$ in Fig. 12) and ($M = 6$ and transitions: $|1\rangle$ to $|4\rangle$ in Fig. 14, $|1\rangle$ to $|6\rangle$ in Fig. 13) for algorithm comparison of controllers under noisy training in Sec. IV B to highlight some of the variation of controller quality for different regimes of noise and spin chain transitions observed in the main ARIM comparison presented in Fig. 16. Each individual subplot is the result of an independent run of each algorithm with a stochastic fidelity function evaluated under the unstructured perturbations using the same approach as earlier with σ_{sim}. These are also plotted for a more distributional comparison as pairwise box plots in Fig. 15. For both Figs. 11 and 15, we also show L-BFGS results for comparison.

5. Full ARIM comparisons

For the cases $M = 6, 7$, both types of transitions appear to be challenging for PPO, SNOBFit and Nelder-Mead at most, if not all, training perturbation strengths; especially the end-to-end $M = 7$ transition [Fig. 16(d)], where PPO at $\sigma_{\text{train}} = 0.05$ is only marginally better than the rest of the algorithm runs, excluding Nelder-Mead. A pertinent question is whether this is genuinely reflective of the landscape or if, for PPO, our budget constraint of 10^6 target functional calls is insufficient for larger system sizes, as the control

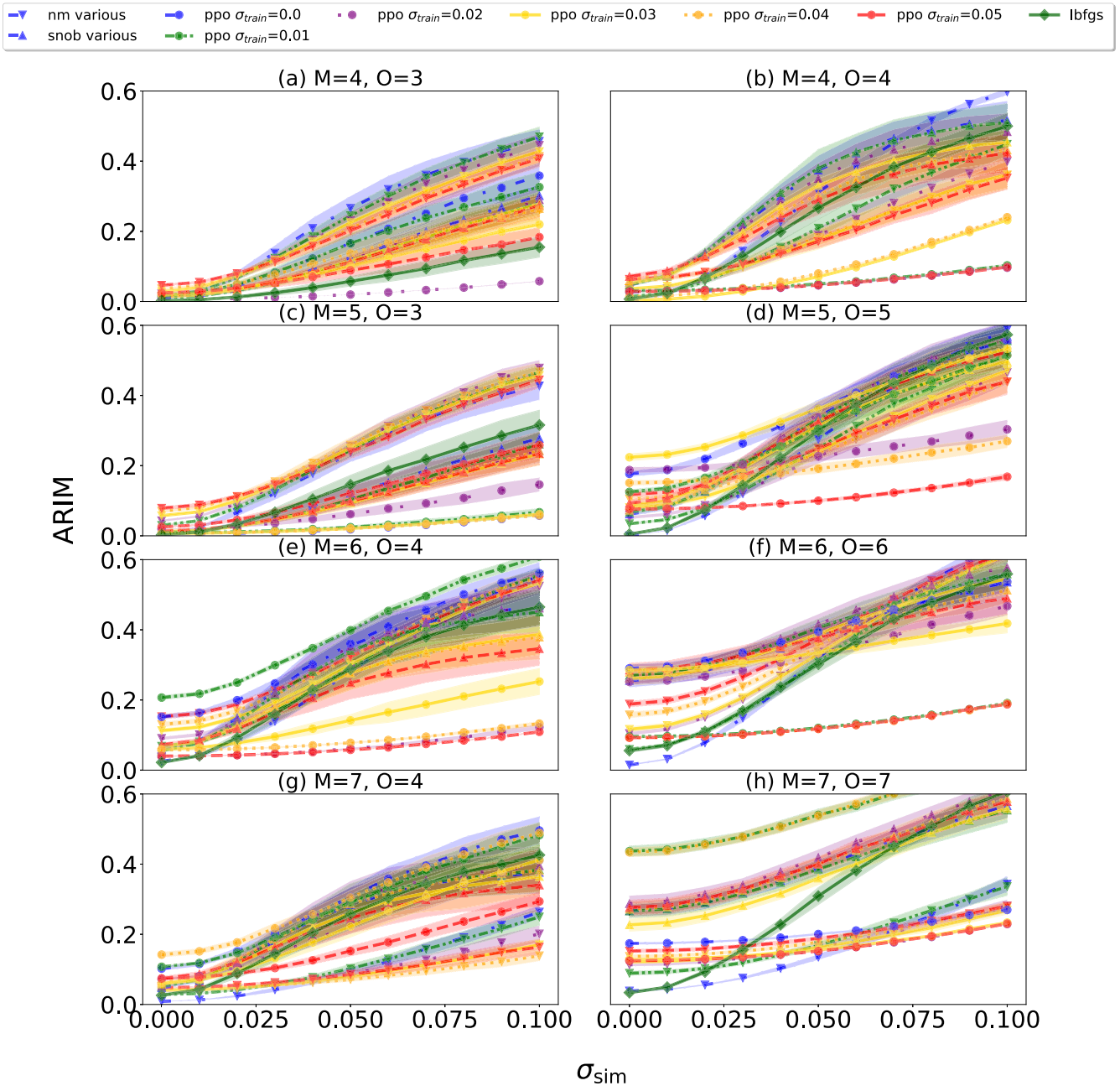

FIG. 16. ARIM as a function of σ_{sim} for $M = 4, 5, 6, 7$, where the left column contains end-to-middle transitions and the right column contains end-to-end transitions. The final state is denoted by O. The ARIM is computed from a distribution of RIM values for 100 controllers for each σ_{sim} for SNOBFit, Nelder-Mead, PPO and L-BFGS indicated by their marker shapes and line styles. Both PPO and SNOBFit are run multiple times at $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, which is indicated by the color of the ARIM curve. For all problems, PPO has higher variance with respect to σ_{train} than SNOBFit and Nelder-Mead. The latter pair's performance curves are more in line with the L-BFGS curve for $\sigma_{\text{sim}} \geqslant 0.05$ and mostly worse for $\sigma_{\text{sim}} \leqslant 0.05$. For most of the problems the best performing (lowest) curve across all problems is PPO at $\sigma_{\text{train}} = 0.05$ (brown) except in (a), where it is PPO at $\sigma_{\text{train}} = 0.02$ and in (g) where it is Nelder-Mead at $\sigma_{\text{train}} \geqslant 0.04$. 95% confidence intervals (shading) are computed using nonparametric bootstrap resampling [86] with 100 resamples.

problem is exponentially dependent on the number of control degrees of freedom. The former hypothesis might hint at a fundamental limitation on robustness of this particular control landscape. The fact that most noisy Nelder-Mead curves for these problems are clustering together suggests that noise could also help in reaching robust areas in the control landscape faster by regularizing or smoothing the landscape by an appropriate degree. We investigate asymptotic algorithm behavior with respect to the training noise in Sec. IV C to illustrate this and show that there is convergence in PPO performance for all the noise levels at sufficiently many function calls.

[1] J. Preskill, Quantum **2**, 79 (2018).
[2] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).
[3] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).
[4] A. M. Steane, Proc. R. Soc. London A **452**, 2551 (1996).
[5] D. Gottesman, arXiv:0904.2557.
[6] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).
[7] K. Temme, S. Bravyi, and J. M. Gambetta, Phys. Rev. Lett. **119**, 180509 (2017).
[8] S. J. Beale, J. J. Wallman, M. Gutiérrez, K. R. Brown, and R. Laflamme, Phys. Rev. Lett. **121**, 190501 (2018).
[9] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X **8**, 031027 (2018).
[10] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).
[11] M. Cetina, L. N. Egan, C. A. Noel, M. L. Goldman, A. R. Risinger, D. Zhu, D. Biswas, and C. Monroe, arXiv:2007.06768.
[12] L. N. Egan, Scaling quantum computers with long chains of trapped ions, Ph.D. thesis, University of Maryland, 2021.
[13] D. Hocker, C. Brif, M. D. Grace, A. Donovan, T.-S. Ho, K. M. Tibbetts, R. Wu, and H. Rabitz, Phys. Rev. A **90**, 062309 (2014).
[14] L. Viola, E. Knill, and S. Lloyd, Phys. Rev. Lett. **82**, 2417 (1999).
[15] P. Solinas, P. Zanardi, and N. Zanghì, Phys. Rev. A **70**, 042316 (2004).
[16] D. A. Lidar, I. L. Chuang, and K. B. Whaley, Phys. Rev. Lett. **81**, 2594 (1998).
[17] E. Jonckheere, A. Shabani, and A. T. Rezakhani, in *Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, USA* (IEEE, Piscataway, NJ, 2014), pp. 5794–5801.
[18] C. Kabytayev, T. J. Green, K. Khodjasteh, M. J. Biercuk, L. Viola, and K. R. Brown, Phys. Rev. A **90**, 012316 (2014).
[19] T. J. Green, J. Sastrawan, H. Uys, and M. J. Biercuk, New J. Phys. **15**, 095004 (2013).
[20] T. Chen and Z.-Y. Xue, Phys. Rev. Appl. **14**, 064009 (2020).
[21] R. Srinivas, S. Burd, H. Knaack, R. Sutherland, A. Kwiatkowski, S. Glancy, E. Knill, D. Wineland, D. Leibfried, A. C. Wilson *et al.*, Nature (London) **597**, 209 (2021).
[22] R. Blumel, N. Grzesiak, N. Pisenti, K. Wright, and Y. Nam, arXiv:1905.09292.
[23] Q. Ansel, S. J. Glaser, and D. Sugny, J. Phys. A: Math. Theor. **54**, 085204 (2021).
[24] E. Jonckheere, S. Schirmer, and F. Langbein, Int. J. Robust Nonlinear Control **28**, 2383 (2018).
[25] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, Heidelberg, 2009), Vol. 338.
[26] J.-S. Li and N. Khaneja, IEEE Trans. Autom. Control **54**, 528 (2009).
[27] R.-B. Wu, H. Ding, D. Dong, and X. Wang, Phys. Rev. A **99**, 042327 (2019).
[28] G. Turinici, Phys. Rev. A **100**, 053403 (2019).
[29] C. Chen, D. Dong, R. Long, I. R. Petersen, and H. A. Rabitz, Phys. Rev. A **89**, 023402 (2014).
[30] X. Ge and R.-B. Wu, Phys. Rev. A **104**, 012422 (2021).
[31] C. Brif, R. Chakrabarti, and H. Rabitz, New J. Phys. **12**, 075008 (2010).
[32] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, ACM Trans. Math. Softw. **23**, 550 (1997).
[33] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, arXiv:1707.06347.
[34] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965), https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf.
[35] W. Huyer and A. Neumaier, ACM Trans. Math. Softw. **35**, 1 (2008).
[36] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, Nat Methods **17**, 261 (2020).
[37] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. De Jong, in *Proceedings of the IEEE Int. Conf. Quantum Computing and Engineering, Denver, CO, USA* (IEEE, Piscataway, NJ, 2020), pp. 267–277.
[38] I. Khalid, C. A. Weidner, E. A. Jonckheere, S. G. Shermer, and F. C. Langbein, *RobChar: Robust Characterisation of Quantum Controls and Control Algorithms (v1.0.1)* (Zenodo, 2022), https://doi.org/10.5281/zenodo.6891152.
[39] M. G. Safonov, A. J. Laub, and G. L. Hartmann, IEEE Trans. Automat. Contr. **26**, 47 (1981).
[40] S. G. Schirmer, E. A. Jonckheere, and F. C. Langbein, IEEE Trans. Automat. Contr. **63**, 2523 (2018).
[41] S. O'Neil, S. G. Schirmer, F. C. Langbein, C. A. Weidner, and E. Jonckheere, arXiv:2210.15783.
[42] M. Arjovsky, S. Chintala, and L. Bottou, in *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh (PMLR, 2017), Vol. 70, pp. 214–223.
[43] A. Ramdas, N. G. Trillos, and M. Cuturi, Entropy **19**, 47 (2017).
[44] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).
[45] E. Jonckheere, S. Schirmer, and F. Langbein, IEEE Trans. Automat. Contr. **62**, 6568 (2017).
[46] J. Doyle, in *IEE Proceedings D Control Theory and Applications* (American Physical Society, College Park, MD, 1982), Vol. 129, pp. 242–250.
[47] S. C. Hou, M. A. Khan, X. X. Yi, D. Dong, and I. R. Petersen, Phys. Rev. A **86**, 022321 (2012).
[48] Z. Shen, M. Hsieh, and H. Rabitz, J. Chem. Phys. **124**, 204106 (2006).
[49] E. Barnes, X. Wang, and S. D. Sarma, Sci. Rep. **5**, 12685 (2015).
[50] M. Demiralp and H. Rabitz, Phys. Rev. A **57**, 2420 (1998).
[51] K. Zhou and J. C. Doyle, *Essentials of Robust Control* (Prentice Hall, Upper Saddle River, NJ, 1998), Vol. 104.
[52] S. G. Schirmer, F. C. Langbein, C. A. Weidner, and E. Jonckheere, IEEE Trans. Autom. Control **67**, 6012 (2022).
[53] S. G. Schirmer, F. C. Langbein, C. A. Weidner, and E. A. Jonckheere, in *Proceedings of the IEEE Conf. Decision and Control, Norwalk, Connecticut, USA* (IEEE, Piscataway, NJ, 2021), pp. 4158–4163.
[54] C. Spearman, Am. J. Psychol. **15**, 72 (1904).
[55] F. C. Langbein, S. Schirmer, and E. Jonckheere, in *Proceedings of the IEEE Conf. Decision and Control, Osaka, Japan* (IEEE, Piscataway, NJ, 2015), p. 6454.
[56] E. Lieb, T. Schultz, and D. Mattis, Annal. Phys. **16**, 407 (1961).
[57] S. Bose, Contemp. Phys. **48**, 13 (2007).
[58] M. Christandl, N. Datta, A. Ekert, and A. J. Landahl, Phys. Rev. Lett. **92**, 187902 (2004).
[59] J. Zhang, G. L. Long, W. Zhang, Z. Deng, W. Liu, and Z. Lu, Phys. Rev. A **72**, 012331 (2005).
[60] M. Bellec, G. M. Nikolopoulos, and S. Tzortzakis, Opt. Lett. **37**, 4504 (2012).

[61] A. Perez-Leija, R. Keil, A. Kay, H. Moya-Cessa, S. Nolte, L.-C. Kwek, B. M. Rodríguez-Lara, A. Szameit, and D. N. Christodoulides, Phys. Rev. A **87**, 012309 (2013).
[62] T. Graß and M. Lewenstein, EPJ Quantum Technol. **1**, 1 (2014).
[63] R. E. Barfknecht, T. Mendes-Santos, and L. Fallani, Phys. Rev. Res. **3**, 013112 (2021).
[64] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, Ann. Math. Statist. **27**, 642 (1956).
[65] M. Brownnutt, M. Kumph, P. Rabl, and R. Blatt, Rev. Mod. Phys. **87**, 1419 (2015).
[66] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).
[67] A. Asl and M. L. Overton, arXiv:2006.11336.
[68] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Phys. Rev. A **84**, 022305 (2011).
[69] H.-J. M. Shi, Y. Xie, R. Byrd, and J. Nocedal, arXiv:2010.04352.
[70] D. P. Kingma and J. Ba, arXiv:1412.6980.
[71] Z. T. Wang, Y. Ashida, and M. Ueda, Phys. Rev. Lett. **125**, 100401 (2020).
[72] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, npj Quantum Inf. **6**, 6 (2020).
[73] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, npj Quantum Inf. **5**, 33 (2019).
[74] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 2018).
[75] I. Khalid, C. A. Weidner, E. A. Jonckheere, S. G. Schirmer, and F. C. Langbein, in *Proceedings of the IEEE Conference on Decision and Control (CDC), Austin, Texas, USA* (IEEE, Piscataway, NJ, 2021), p. 4133.
[76] C. A. Weidner and D. Z. Anderson, Phys. Rev. Lett. **120**, 263201 (2018).
[77] E. Zahedinejad, S. Schirmer, and B. C. Sanders, Phys. Rev. A **90**, 032310 (2014).
[78] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Phys. Rev. Lett. **112**, 240504 (2014).
[79] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011).
[80] E. Jonckheere, F. C. Langbein, and S. Schirmer, Quantum Inf. Proc. **13**, 1607 (2014).
[81] W.-L. Loh, Ann. Stat. **24**, 2058 (1996).
[82] M. Stein, Technometrics **29**, 143 (1987).
[83] A. B. Owen, *Monte Carlo Book: The Quasi-Monte Carlo Parts* (2019).
[84] M. G. Kendall, *Rank Correlation Methods*, 3rd ed. (C. Griffin, London, 1962).
[85] A. Agresti, *Analysis of Ordinal Categorical Data* (John Wiley & Sons, Hoboken, New Jersey, 2010), Vol. 656.
[86] B. Efron, J. Am. Stat. Assoc. **82**, 171 (1987).
[87] E. J. Rafols, M. B. Ring, R. S. Sutton, and B. Tanner, in *IJCAI'05: Proceedings of the 19th International Joint Conference on Artificial Intelligence, San Francisco, CA, United States* (Morgan Kaufmann, 2005), pp. 835–840.
[88] P. Thomas, in *Proceedings of the 31st International Conference on Machine Learning, Bejing, China* (PMLR, 2014), Vol. 32, pp. 441–448.