# Statistically Characterising 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, UK*
[2]*Quantum Engineering Technology Laboratories, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1FD, UK*
[3]*Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90007, US*
[4]*Department of Physics, Swansea University, Swansea, SA2 8PP, UK*
(Dated: July 22, 2022)

Robustness of quantum operations or controls is important to build reliable quantum devices. The *robustness-infidelity measure* ($\mathrm{RIM}_p$) is introduced to statistically quantify the robustness and fidelity of a controller as the $p$-order Wasserstein distance between the fidelity distribution of the controller under any uncertainty and an ideal fidelity distribution. The $\mathrm{RIM}_p$ is the $p$-th root of the $p$-th raw moment of the infidelity distribution. Using a metrization argument, we justify why $\mathrm{RIM}_1$ (the average infidelity) suffices as a practical robustness measure. Based on the $\mathrm{RIM}_p$, 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 by considering the problem of robust control of spin-$\frac{1}{2}$ networks using energy landscape shaping subject to Hamiltonian uncertainty. The robustness and fidelity of individual control solutions as well as the expected robustness and fidelity of controllers found by different popular quantum control algorithms are characterized. For algorithm comparisons, stochastic and non-stochastic optimization objectives are considered, with the goal of effective RIM optimization in the latter. 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 optimization objectives, adaptive sequential decision making approaches such as reinforcement learning have a cost advantage compared to standard control algorithms and, in contrast, the infidelities obtained are more consistent with higher RIM values for low noise levels.

## 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] and syndrome measurements [5]; (2) using error mitigation schemes, e.g., reversing noisy dynamics [6–9], active variational noise minimization [10], or parametric modelling 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) interaction with the environment that leads to non-unitary dynamics; (b) inaccuracies in the control model representing a specific physical implementation that affect the evolution but do not cause non-unitary 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, allowing for 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 distances 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 measure of the minimal costs of probability mass transport between two distributions. In Sec. II, the $p$-th order *Robustness-Infidelity Measure* ($\mathrm{RIM}_p$) is defined to quantify the robustness and fidelity of a quantum controller. The $\mathrm{RIM}_p$ is based on the $p$-th 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

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

$\text{RIM}_p$ is the $p$-th root of the $p$-th raw moment of the infidelity distribution – a non-parametric measure independent of any particular assumption for the distribution.

Using measure-theoretic norm scaling relations between RIMs of different order, we argue that $\text{RIM}_1$, the average infidelity, is sufficient as a practical measure of robustness and fidelity. The RIM has practical utility to choose among similar, high-fidelity controllers, acting as a post-selector for robust controllers, agnostic of the algorithm used to find them. This may be computationally more efficient than optimising the RIM directly, as we see later 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 as those benchmarked in [26]. To that end, we introduce an Algorithmic RIM (ARIM), averaging RIMs over multiple controllers, in Sec. II.

In Sec. IV, by exploiting the degree of freedom afforded by the existence of multiple optima in quantum control [27], we generate a number of controllers for energy landscape control of the XX Heisenberg model. 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 optimisation using an ordinary differential equation model of the quantum system to compute the fidelity under perfect conditions [28]; (2) Proximal Policy Optimization (PPO): a model-free reinforcement learning algorithm, having no prior knowledge of the system [29]; (3) Nelder-Mead: a derivative-free simplex-based heuristic search method [30]; 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 [31]. Here, (1) serves as a baseline for optimisation over a noise-free fidelity target functional under ideal conditions. (2) represents a machine learning approach with minimal knowledge. (3) and (4) are derivative-free methods to handle noisy target functionals. A detailed motivation for the choice of these algorithms is presented in Sec. III B. The algorithms were implemented in Python. Specifically, we used the scipy library for accessing (1) and (3) [32], and (4) is obtained from Ref. [33].

Our experimental motivation is four-fold: (a) by conducting a comparison of the robustness of controllers amongst each other without regard to the optimization algorithm, we wish to answer whether high fidelity implies high robustness using RIM at the level of the individual controller (Sec. IV A); (b) by conducting a distributional comparison of controllers we wish to ascertain the performance of control algorithms to understand how likely a given algorithm is to produce controllers that are robust in noisy conditions but were obtained in an ideal (no-noise) setting (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 algo-

rithm's ability to find robust controllers using the ARIM (Sec. IV B). To make the comparison fairly, we conduct (b) and (c) in a resource-constrained setting with a budget of a fixed number of target function calls allotted to each algorithm. (d) In Sec. IV C, we try to understand an algorithm's asymptotic ability to optimize the ARIM by removing the constraints on the number of function calls and consider two further settings in this scenario: stochastic and non-stochastic fidelity optimization. In the latter case we optimize over a fixed set of Hamiltonians sampled with respect to our noise model while in the former the Hamiltonians themselves are stochastically chosen for each target function evaluation according to the noise model.

Our main numerical findings are summarized as follows:

- High fidelity controllers are not always robust but the non-robust controllers can be filtered out using the RIM with respect to a noise model across the four aforementioned algorithms.

- Using a consistency statistic, we show that PPO controller infidelities (RIM at no model noise) are more correlated with RIM values at low levels of model noise in contrast with other algorithm infidelities. More generally, a strong signal in the consistency statistic can be used to predict RIM robustness while sidestepping its explicit evaluation.

- For constrained target 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.

- Robust controllers with respect to certain noise models in the optimization objective can be obtained for all control acquisition algorithms for the non-stochastic optimization objective when there are no constraints on resources. We demonstrate this for all control algorithms, which all optimize the ARIM asymptotically.

- However, if the optimization objective is stochastic, only PPO optimizes the ARIM asymptotically. In either case, PPO takes fewer function calls compared to the other control 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 non-stochastic target functionals, e.g., shot noise.

Lastly, from the perspective of classical control [34], it is well known that accuracy conflicts with robustness through the well known $S + T = I$ formula, where $S$ is related to tracking error and $T$ to sensitivity of the tracking error to uncertainties. This restriction does not seem to map in such an ironclad way to quantum systems in quantifying fidelity versus robustness [24]. 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 constant.

In general, the control problem is formulated as optimising a fidelity $\mathcal{F}$ over a set of admissible controls. A general notion of fidelity that reflects most definitions used in practice is given by $\mathcal{F} := |\langle G|V \rangle|^2$, which measures the similarity between normalized objects $G$ and $V$. If we wish to prepare a state $G = |\psi_f\rangle$ from an initial state $|\psi_0\rangle$ at time $t_0$, then $V = U(t_0, t_1, \mathbf{u}) |\psi_0\rangle$ and the optimisation 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 $\mathbb{X}$ represents the range of allowed controls, here including the final time $t_1$. A variant (up to normalisation) of this state fidelity is the Hilbert-Schmidt inner product $\mathcal{F} = \text{Tr}\left(G^\dagger V\right)$ between a desired unitary transformation $G$ and a gate achieved by optimal 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 $G = e^{\imath\phi}V$, up to a global phase $\phi$.

### 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 realise the target operation well. So we also expect a fidelity close to 1. That means the perfect distribution under any uncertainties is $\delta_1$ – the Dirac delta function at maximum fidelity 1. In particular, we consider the delta function $\delta_x$ to be defined by an indicator cumulative distribution function (CDF),

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

This rigorously permits the familiar delta function property during integration w.r.t. a continuous 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 in order 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, 35]. Historically, the Wasserstein distance has its inception in the Monge problem, the optimal mass transport problem of logistics; it is, on a high level, the minimum cost of transferring a probability measure $\mu$ to another one, $\nu$, by means of a measurable map $y = T(x)$ so that $\mu$ turns into $\nu$. The Kantorovich problem substitutes a more general transference plan, that is, a measure on $X \times Y$, for the Monge map $T$. (Note that the Monge infimum need not be equal to the Kantorovich minimum [36]). Taking the transport cost to be $\|x - y\|_{L^p}$, the Kantorovich cost becomes the $\mathcal{W}_p$-Wasserstein distance. We choose it here due to: (1) its weak topology and convergence properties that imply high sensitivity of convergence in terms of the closeness to the ideal; e.g., one can compare two probability distributions that do not share a common support, and in particular, one can compare a discrete and continuous distribution; (2) its easy geometric interpretation which helps with optimization of the quantity as it preserves the distributional structure on the geodesic path between two distributions, and during distributional comparisons it produces a quantification of how one distribution can be converted to another, rather than just measuring the probabilistic distance between the two; and (3) a simplification which in our case allows it to be calculated easily, as shown next.

The general dual formulation of the $p$-th order Wasserstein distance [37] between two distributions $\mu$, $\nu$ is given by the relation

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

where $h(x) - g(y) \leq \|x - y\|^p$. Even though the general form seems quite abstract, for one-dimensional distributions, we can analytically compute the optimal maps $h(\cdot), g(\cdot)$:

**Theorem 1** (*Prop. 1 in [37]*) *The p-th Wasserstein distance* $\mathcal{W}_p(\mu, \nu)$ *for one-dimensional probability distributions* $\mu$ *and* $\nu$ *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* $Q_\mu(z) = \inf\{x \in \mathbb{R} : C_\mu(x) \geq z\}$ *denotes the quantile function and* $C_\mu$ *is the cumulative probability function of* $\mu$ *and likewise for* $Q_\nu$.

Remarkably, the optimal transport distance between one-dimensional distributions $\mu, \nu$ over all possible transportation plans can be computed in terms of their quantile functions $Q_\mu$, $Q_\nu$. From here, following Thm. 1, it is straightforward to define the *p*-th *Robustness-Infidelity Measure* ($\text{RIM}_p$):

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

$\text{RIM}_p$ can be written in terms of the raw moments(see App. A 1):

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

For $p = 1$, this is the average infidelity,

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

To compute the raw moments, we can 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 [38]. Hence, barring the computational or experimental expense of obtaining these samples, the $\text{RIM}_1$ 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 $\text{RIM}_1$ is equal to the infidelity $1 - f$. Moreover, the $\text{RIM}_1$ 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

We now motivate why the $\text{RIM}_1$ is sufficient as a practical measure of robustness and infidelity. We make use of the fact that the RIMs of different orders computed on the estimated fidelity distribution are equivalent as lower and higher order RIMs bound each other (see App. A 3):

$$\text{RIM}_{p'} \leq n^{\left( \frac{1}{p} - \frac{1}{p'} \right)} \text{RIM}_p, \tag{10a}$$

$$\text{RIM}_p \leq \text{RIM}_{p'} \tag{10b}$$

for $p < p'$, where $n$ is the number of samples. This means that the higher order RIMs do not capture more information than the $\text{RIM}_1$. Moreover, given that the Wasserstein distance provides a structure-preserving geodesic

between any distribution to the ideal $\delta_1$, the distributions converge together with their RIMs of any order, especially when approaching the ideal. For example, the variance of symmetric distributions decreases as $\frac{1}{6}(1 - \min \mathcal{F})^2$ as $\min \mathcal{F} \to 1$ in $[0, 1]$.

Note that the bounds make no assumptions about independence of sample generation or the nature of the distribution $\mathbf{P}(\mathcal{F})$. However, the upper bound on $\text{RIM}_{p'}$ becomes looser with $n$ and can in principle be made much tighter by adding additional assumptions on the nature of the distribution of $\mathcal{F}$.

For higher order RIMs, the outliers are more influential (see App. A 3) and may prove useful during optimization. However, in general, the noise contributions in real quantum devices would reduce the need to weigh the outliers more [1]. From now on we will refer to the $\text{RIM}_1$ 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 normalised using some matrix norm. To induce an uncertainty into the dynamics we treat $\gamma$ 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 that one expects to be introduced during the deployment of a quantum device.

The structure $S$ may be fixed, e.g. describing the uncertainty in some coupling parameter for the Hamiltonian, while $\gamma$ is drawn from a normal distribution. This would be consistent with a (linear) *structured perturbation* in classical robust control theory. Instead, $S$ may also be drawn from a probability distribution, describing uncertainties across multiple Hamiltonian parameters. While this generalises structured perturbations, note that they do remain linear w.r.t. the strength. If $S$ is sampled uniformly on the unit-sphere, according to its normalisation, we have an *unstructured perturbation*, with (uncertain) strength determined by $\gamma$. Conceptually, if $\gamma$ is drawn from a normal distribution with zero mean and standard deviation $\sigma$, $\gamma S$ 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 [39]), 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 $\gamma S$ and the uncertainties in these quantities.
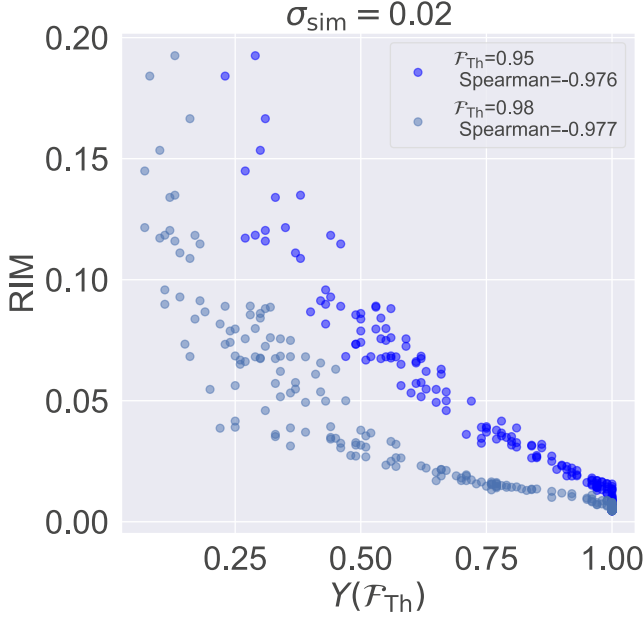
FIG. 1. RIM values generated from $\mathbf{P}_{0.02}(\mathcal{F})$ with $N = 100$ samples for 200 controllers are plotted against their yield $Y(\mathcal{F}_{\text{Th}})$ at fidelity thresholds $\mathcal{F}_{\text{Th}} = 0.95, 0.98$. Both measures are correlated, as encapsulated by the high negative Spearman correlation coefficients [48] and $p$-values $< 10^{-4}$.

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 $t_{\text{opt}}, \mathbf{u}_{\text{opt}}$ to some order, which is how noise is usually modelled in the literature [13, 18, 40–43]. 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 $\mu$ function in classical [44] and quantum [45–47] settings.

To further motivate the RIM, we study how it compares with other statistical measures of robustness. The RIM generally correlates with worst-case sample fidelity, variance or higher moments and the Yield function $Y(\mathcal{F}_{\text{Th}})$, which is the fraction of fidelities greater than a threshold fidelity $\mathcal{F}_{\text{Th}}$. Fig. 1 shows a scatter plot of RIM values versus $Y(0.95)$ and $Y(0.98)$ for an example problem using Eq. (14) discussed later. The respective measure values obtained for 200 controllers from their fidelity distribution $\mathbf{P}_{0.02}(\mathcal{F})$, indicating that $\gamma S$ has been drawn from $\mathcal{B}_{0.02}$, estimated with 100 Monte Carlo samples for each controller. The RIM has an advantage over $Y$ in that it does not depend on an arbitrary choice of $\mathcal{F}_{\text{Th}}$.

## E.  Measuring the Performance of Control Algorithms

We can also apply the previous arguments to derive a measure that summarizes the performance of a distribution of controllers generated by a control algorithm based on the RIM. This allows us, e.g., to compare algorithm performance w.r.t. their ability to find high-fidelity, robust controllers. Let the RIM values of $\mathbf{P}_{\sigma}(\mathcal{F})$ from the ideal $\delta_1$ of $M$ control solutions be drawn from a distribution $\mathbf{P}_{\sigma}(\text{RIM})$. We can peg the ideal of this distribution as $\delta_0$ and define the *Algorithmic Robustness Infidelity Measure* (ARIM)

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

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

## 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, 26, 46]: the first-order single-parameter log-sensitivity is concordant with fidelity error, meaning small log-sensitivity and small fidelity error coexist. In classical robust control discordant behaviour of the two quantities forms a fundamental limitation [34]. In Ref. [24], it was found that high robustness (in terms of single-parameter log-sensitivity) of a controller coincides with high fidelity. While often these systems are not fully controllable, solutions for specific operations can be found via optimisation, even if they are not all robust [49]. The static approach is simpler in the sense of having fewer control parameters to optimise over, which reduces computational and experimental complexity. This makes the problem suitable to demonstrate the use of the RIM and explore the robustness properties of the control algorithms as well as the controllers they find.

Eq. (1) for a static control problem is equivalent to a single time step with piecewise constant controls with Hamiltonian $H_0 + H_{\mathbf{\Delta}}$ where $\mathbf{\Delta}$ are time-independent parameters of the control Hamiltonian $H_{\mathbf{\Delta}}$. The unitary propagation operator becomes

$$U(t_0, t_1, \mathbf{\Delta}) = \exp\{-\imath(H_0 + H_{\mathbf{\Delta}})(t_1 - t_0)/\hbar\}. \quad (12)$$

Perturbations to the system are represented by $\tilde{H} = H_0 + H_{\mathbf{\Delta}} + \delta S$ where here $\delta$ is drawn from a normal distribution with variance $\sigma^2$ (see Sec. II D).

## A. Information Transfer in the Single Excitation Subspace of XX Spin Chains

To demonstrate the use of the RIM to characterize controls we use a simple network of $M$ spins represented by the quantum Heisenberg model given by the Hamiltonian

$$\frac{H_{\text{heis}}}{\hbar} = \sum_{a \in L} \sum_{j=1}^{M} J^a \sigma_j^a \sigma_{j+1}^a + \eta \sum_{j=1}^{M} \sigma_j^z \qquad (13)$$

where $\sigma_j^a = \mathbb{I}^{\otimes j-1} \otimes \sigma^a \otimes \mathbb{I}^{\otimes M-j}$ and $\sigma^a$ are the usual Pauli matrices indexed by $a \in L = \{x, y, z\}$. We set $J^z = 0$ and $J^x = J^y = J$ to get the XX model with uniform couplings. This model has been studied extensively, starting with Ref. [50] in 1961, with a more recent review of the system for quantum communication provided in Ref. [51]. Conditions for perfect state transfer along XX chains were derived in Ref. [52] and applied to NMR systems [53]. Similar experiments have been carried out in photonic systems [54, 55], and proposals for engineering similar systems with trapped ions [56] and cold atoms [57] exist.

The state space naturally decomposes into non-interacting 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 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 for the first excitation subspace is

$$\frac{(H_{XX})_{l,m}}{\hbar} = J \mathbb{1}_{l,m\pm1} + \Delta_l \mathbb{1}_{l,m} \qquad (14)$$

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

$H_{XX}$ 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 $t_{\text{opt}}$ and a single vector of $M$ biases $\boldsymbol{\Delta}_{\text{opt}}$. The perturbations are given by

$$\left(S^{XX}\right)_{l,m} = \sum_{k=1}^{M-1} \gamma_k^J J \mathbb{1}_{l,k} \mathbb{1}_{l,m\pm1} + \sum_{c=1}^{M} \gamma_c^C \Delta_c \mathbb{1}_{c,l} \mathbb{1}_{l,m} \qquad (15)$$

where $\gamma_k^J$ and $\gamma_c^C$ are the strength of the perturbation on the couplings and controls respectively. We draw these strengths from the same normal distribution $\mathcal{N}\left(0, \sigma^2\right)$ with mean 0 and variance $\sigma^2$. A numerical example illustrating the RIM via the empirical CDF (ECDF) for two controllers for the XX model for $M = 5$ and the transition $|1\rangle \to |3\rangle$ with $\sigma = 0.1$ is shown in Fig. 2. The
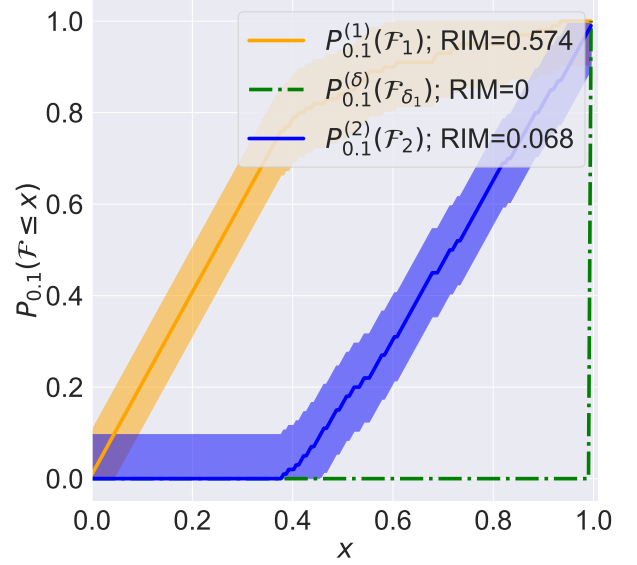


FIG. 2. To illustrate the RIM robustness measure two static controllers for an XX spin chain of length five for transferring an excitation from spin $|1\rangle$ to $|3\rangle$ are compared. The empirical approximations to the CDFs for the two controllers, $l = 1, 2$ were simulated using 100 bootstrapped 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 generated using 500 bootstrap repetitions. The 0.95 confidence bounds on their error are obtained using the Dvoretsky-Keifer-Wolfowitz inequality [58]. Closeness to the perfectly robust controller can be interpreted as having a smaller area under the curve and indicated by the RIM values.

RIM, or the average infidelity, has a well-known interpretation as the area under the cumulative fidelity distribution curve.

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, 59]. We have chosen to implement the simplest option of equal strength random perturbations on all non-zero entries of the Hamiltonian that is also relevant in practical settings [53–57].

## B. Algorithms for Static Control Problems

The ARIM can compare algorithm performance in finding robust controllers. Here, we describe the algorithms used to find the controllers for the static control problem. For selecting algorithms, we tried to (a) investigate the performance of algorithms commonly used in the quantum control community, (b) consider algorithms

that do and do not require gradient information, and (c) consider reinforcement learning, more recently also used in quantum control. 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 a specific robust control problem.

L-BFGS is a common optimisation method used in quantum control as part of GRAPE [60] and performed well on finding high-fidelity energy landscape controllers [49]. It has not been designed for noisy optimization but there exist smoothing modifications that attempt to address this [61–63]. For individual controller comparisons, we use standard L-BFGS with an ordinary differential equation model to compute the fidelity without perturbations during optimisation. This serves as a baseline to understand the performance of optimising noiseless target 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 [64]) 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 [63] 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 [65], optimizing circuit-QED, two-qubit unitary operators under physical realization constraints [66], and optimizing multi-qubit control landscapes suffering from control leakage and stochastic model errors [67], among many others.

Proximal Policy Optimization (PPO) is a policy gradient method in the class of reinforcement learning algorithms [68]. It uses a discounted reward signal (e.g., the fidelity) accumulated over multiple interactions with the optimization landscape using non-parametric 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 pay-offs 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 [69]. In this work, we use a control problem formulation of PPO for Eq. (14) as described in Ref. [69].

Nelder-Mead is a popular simplex based control algorithm using direct search. Essentially, it keeps updating a polytope whose vertices are function evaluations towards an optimum direction. It has successfully been used in noisy experimental settings [70] due to its non-reliance on gradient information [71–73], especially when obtaining such information is resource-intensive.

Stable Noisy Optimization by Branch and Fit (SNOB-Fit) has been chosen as it has been designed to filter out quite large scale noise in target functionals [31]. It fits local models using target function evaluations and implements a branching and splitting algorithm to partition the parameter space into smaller boxes with one function evaluation point per box. The latter is a non-local search scheme that orders promising sub-boxes by the number of bisections required to get from the base box to that box. Sub-boxes 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 non-local optimization landscape exploration is not random and thus has comparatively a lot less variance in performance.

## 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 (MCRA) using 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 $\left|\lceil\frac{M}{2}\rceil\right\rangle$. The former transition is physically easy to control while the latter is more challenging [49] as transitions to the middle exhibit anti-core behaviour [74].

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

We initialise $\mathbf{\Delta}, t$ with quasi Monte Carlo sampling from the Latin Hypercube [75–77] to increase convergence rate and decrease clustering of controllers. This permits coverage of the parameter domain to be $\mathcal{O}(1/\sqrt{N})$ as opposed to $\mathcal{O}(1/N)$, where $N$ is the number of initializations and reduces the probability of duplicate controllers. Our constraints are $0 \leq t_f \leq 70$ and $-10 \leq \Delta \leq 10$ and we use 100 bootstrap samples to estimate fidelity distributions throughout. The perturbation strengths $\gamma_j^J$ and $\gamma_c^C$ are scaled by $J$ and $\Delta$ respectively as per Eq. (15). For $\sigma = 0$, $\mathbf{P}_0(\mathcal{I}) = \mathcal{I}$ is deterministic and RIM reduces to the infidelity without noise in the system.
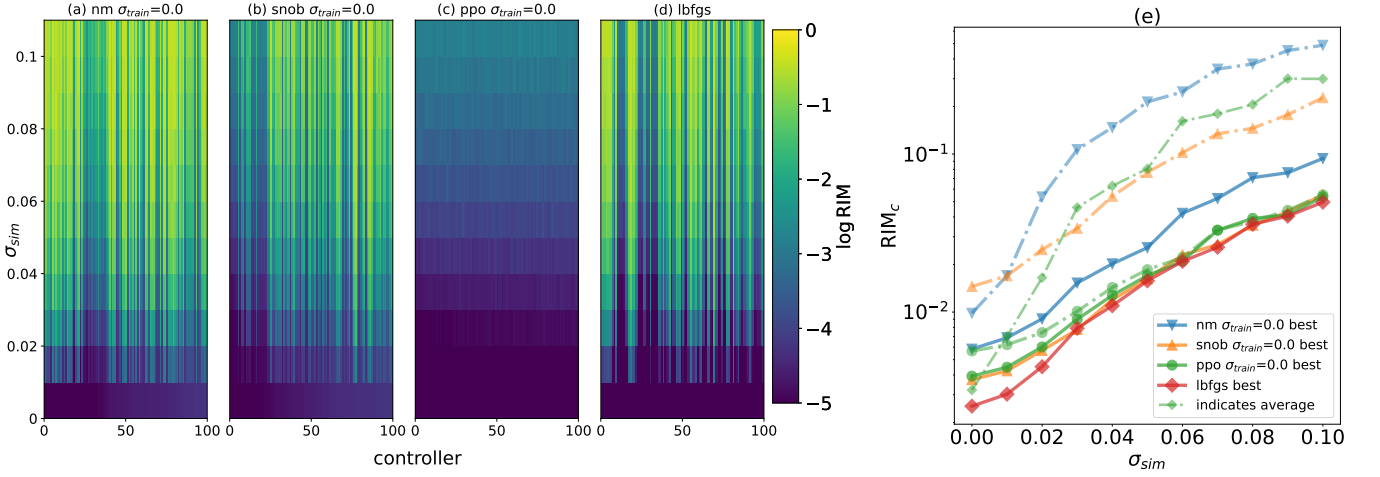
FIG. 3. (a)-(d) 100 controllers found for the XX model, Eq. (14), using Nelder-Mead, SNOBFit, PPO, and L-BFGS for $M = 5$ and the spin transition from $|1\rangle$ to $|3\rangle$ with $\sigma_{\text{train}} = 0$. The controllers are ranked in increasing order of infidelity at $\sigma_{\text{sim}} = 0$ 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 $\sigma_{\text{sim}}$ for the average and best controller (i.e., most dark over all $\sigma_{\text{sim}}$ levels) out of the 100 shown in (a)-(d) in terms of how much they preserve their corresponding RIM rank average across all $\sigma_{\text{sim}}$. Each algorithm is indicated by marker shape, and the solid and 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.

The perturbation strengths are drawn from a normal distribution with standard deviation $\sigma_{\text{train}}$ determining the strength of the noise added for the optimization. $\sigma_{\text{sim}}$ is the noise level used in the simulations to assess the robustness of the controllers found.

## A. Characterization of All Controllers under 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 non-linear and non-uniform deterioration of robustness with increasing noise which implies a tradeoff between higher fidelity at no noise and robustness at higher noise levels.

To this end, we employ control algorithms to optimise a target functional without noise, i.e., setting $\sigma_{\text{train}} = 0$ (see Sec. II D), under the general optimisation 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)-(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 $\sigma_{\text{sim}}$ for individual controllers grow at different rates despite starting at quite similar small values for all algorithms. The main takeaway that applies also to all transitions (not explicitly shown here) is that the high fidelity controllers do not, in general, preserve their ranks as $\sigma_{\text{sim}}$ increases, e.g., for SNOBFit (see Fig. 3(b)), the RIM for controllers $5, 7, 8, 10 - 12$ grows much more rapidly than for controllers $17 - 22$, 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 $\sigma_{\text{sim}}$ 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 examined in Sec. IV B. It could be limited fundamentally by the existence of robust controllers and/or the resource budget for a particular problem (see Fig. 11 in App. B 3 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 can be computed using the reshuffled RIM ranks of each controller for all values of $\sigma_{\text{sim}}$. Likewise, we locate the controller that has the median RIM rank average as the average 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 per-

forms. 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 $\sigma_{\rm sim}$. At all levels, the RIM-wise best controller always has a better RIM than the median controller with its RIM value growing comparatively more slowly.

For all algorithms, the best and the average controllers have similar infidelities (initial RIM value) in Fig. 3(e), but their behaviour as a function of $\sigma_{\rm sim}$ is different and is generally non-linear. The best controllers for PPO, SNOBFit and L-BFGS algorithms also behave similarly in RIM at higher $\sigma_{\rm sim}$. Nelder-Mead performs slightly worse with the RIM of the best controller being consistently higher for all noise levels. Thus, the best controllers, despite being distinguishable at $\sigma_{\rm sim} = 0$, become indistinguishable for higher $\sigma_{\rm sim}$ 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 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 can approximately reduce the RIM rank consistency property of the top-$k$ controllers across two perturbation strengths $\sigma_{\rm sim}^{(i)}$ and $\sigma_{\rm sim}^{(j)}$ to a prediction problem by asking the following: *How well does the RIM rank of a controller, when ordered at strength $\sigma_{sim}^{(i)}$, predict the RIM rank of the controller at strength $\sigma_{sim}^{(j)}$?* Let us denote the controller RIM $\sigma_{\rm sim}^{(i)}$-rank order by the vector $\mathbf{r}^{\sigma_{\rm sim}^{(i)}}$.

We answer this question by computing an ordinal (binned/categorical) version of the Kendall-tau-B statistic $\tilde{\tau}$ [78, 79], a measure of statistical dependence between $\mathbf{r}^{\sigma_{\rm sim}^{(i)}}$ and $\mathbf{r}^{\sigma_{\rm sim}^{(j)}}$. The ordinals are constructed only for $\mathbf{r}^{\sigma_{\rm sim}^{(i)}}$ by binning using a discrepancy parameter $\alpha$ that indicates the fraction of the maximum RIM value difference within a single bin. The binned rank order $\tilde{\mathbf{r}}^{\sigma_{\rm sim}^{(i)}}(\alpha)$ minimizes the effect of small movement in either rank. $\tilde{\tau}$ is computed by

$$\tilde{\tau}(\sigma_{\rm sim}^{(i)}, \sigma_{\rm sim}^{(j)}) = \tilde{\tau}_{i,j} = \frac{\sum_{l<m} \mathbb{I}_{l,m}^+ + \mathbb{I}_{l,m}^-}{\sqrt{\left(K - t_{\rm total}^{(i)}\right)\left(K - t_{\rm total}^{(j)}\right)}} \quad (16)$$
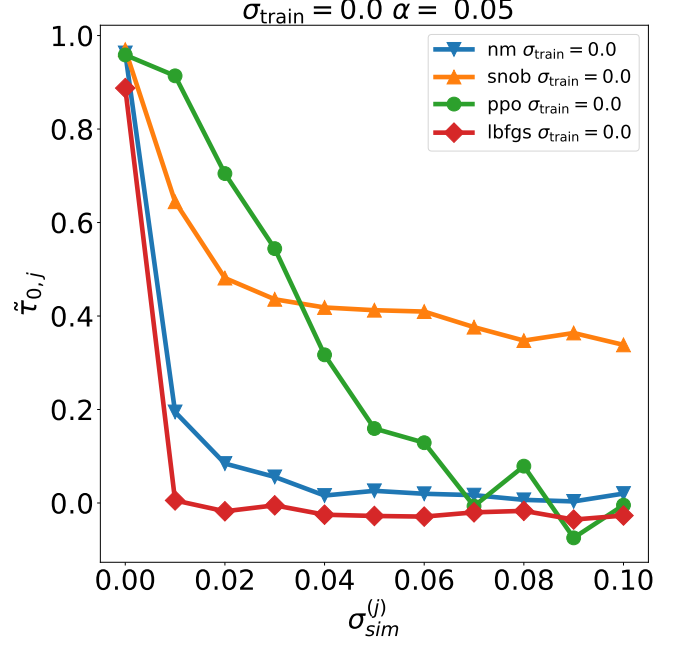


FIG. 4. RIM rank order consistency statistic $\tilde{\tau}$ for the 100 controllers found for the problem $M = 5, |1\rangle$ to $|3\rangle$ between the two levels: no simulation noise, $\sigma_{\rm sim}^{(i)} = 0$ and $\sigma_{\rm sim}^{(j)}$ from $\{0.0, 0.01, \ldots, 0.1\}$ for (a) Nelder-Mead, (b) 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 discrepancy parameter $\alpha = 0.05$ is used to smooth the $\sigma_{\rm sim}^{(i)}$-ranks to reduce noise due to small movements in the rank ordering. The $\tilde{\tau}_{0,j}$ values decline the slowest for PPO until $\sigma_{\rm sim}^{(j)} = 0.04$ and then SNOBFit takes over compared to the rest. This shows, in this case, that the PPO infidelity rank order correlates the most with RIM rank order for $\sigma_{\rm sim} \leq 0.03$.

where

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

are the $l, m$-th sign products of the rank order differences at $\sigma_{\rm sim}^{(i)}, \sigma_{\rm sim}^{(j)}$ with the $+/-$ denoting the positive/negative pair contributions; $K = k(k-1)/2$ is the number of total pairs being compared; $t_{\rm total}^{(i)} = \sum_l t_l^{\sigma_{\rm sim}^{(i)}}(t_l^{\sigma_{\rm sim}^{(i)}} - 1)/2$ are the total number of ties where $\mathbb{I}_{l,m} = 0$ for $\sigma_{\rm sim}^{(i)}$ and likewise for $t_{\rm total}^{(j)}$. 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 have 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 $\sigma_{\mathrm{sim}}^{(i)} = 0, \sigma_{\mathrm{sim}}^{(j)}$ pairs that is sufficient to answer the question raised at the start. 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 $\sigma_{\mathrm{sim}}^{(i)} \geq 0.03$, 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 $\sigma_{\mathrm{sim}}$ increases with deteriorating $\tilde{\tau}$ and SNOB-Fit takes over. This may be due to the minor numerical differences in RIM (see Fig. 3(c)) observed, and thus a stronger consistency for $\sigma_{\mathrm{sim}} \leq 0.03$ is captured. We highlight in App. B 1 that the reason why PPO infidelities correlate more with RIM values at higher $\sigma_{\mathrm{sim}}$ is because it optimizes a discounted RIM ($\sum_i \gamma^i \mathrm{RIM}^{(i)}$ for $0 \leq \gamma \leq 1$) as its reward function.

The other algorithms typically have a sharper drop at $\sigma_{\mathrm{sim}}^{j)} = 0, 0.01$ step where the infidelity rank order for L-BFGS and, to a lesser extent, Nelder-Mead is completely non-informative (due to very high fidelity values) and is not consistent with the orders at later $\sigma_{\mathrm{sim}}$. This is most likely because the controllers found are the result of second order, gradient-based or similarly successful search methods for finding local optima.

On the other hand, PPO ($\sigma_{\mathrm{sim}} \geq 0.03$) and SNOBFit are gradient-free so their controllers are more consistent in comparison. For these cases, 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.

These behaviors resemble a general trend that was observed for these algorithms for all spin transitions studied, which will be discussed in the subsequent sections. 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 $\sigma_{\mathrm{sim}}$. 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 infidelity (or higher RIM value) of a controller is predictive of its RIM value (even for a higher RIM value) at a different noise level. The non-parametric nature of $\tilde{\tau}$ removes information about the range $[\min F, \max F]$ and should be viewed in conjunction $3(a) - (f)$. If the correlation signal is strong, it could be used to sidestep the evaluation of RIM at non-zero noise in favour of using the infidelity instead, eliminating the need for expensive sampling.

A consistency statistic analysis for PPO trained under noise is presented in App. B 2 that corroborates our findings that infidelity rank order for PPO correlates more with higher RIM in contrast with the other studied algo-

rithms.

### B. Comparison of Control Algorithms with Constrained Resources

We now address our motivating question (c): what is the effect of training noise on a control algorithm's ability to find robust controllers? We find the overall picture complex in terms of algorithm rankings and numerically confirm a spin-transition-dependent presence of optimal noise levels for the ARIM that best smooths the optimization landscape.

#### 1. Algorithm Performance Comparison Using the ARIM

Only PPO, Nelder-Mead and SNOBFit are designed to perform noisy optimization with perturbations, although L-BFGS can be applied to optimize the average fidelity over an ensemble of systems, which will be considered later. We optimize for $\sigma_{\mathrm{train}} \in \{0, 0.01, \ldots, 0.1\}$ and select $\sigma_{\mathrm{sim}} \in \{0, 0.01, \ldots, 0.1\}$ to evaluate the RIM of the controllers found at different noise levels with a budget of $10^6$ target function call per instance. For XX spin chains, we use the ARIM to quantify the performance of 100 controllers found by all algorithms at training perturbation strengths $\sigma_{\mathrm{train}} = 0, 0.01, \ldots, 0.05$ for PPO, SNOBFit, L-BFGS and Nelder-Mead. Fig. 5 shows the ARIM curves as a function of $\sigma_{\mathrm{sim}}$ where an optimal curve is flat at ARIM $= 0$.

We identify each algorithm with a unique marker and each $\sigma_{\mathrm{sim}}$ with a unique color. For the end-to-middle control problem with $M = 5$, the ARIM curves without training noise in Fig. 5(b) show a divergence at higher $\sigma_{\mathrm{sim}}$ from similar base ARIM values of $\sigma_{\mathrm{sim}} = 0$ for each algorithm. This indicates that the probabilistic distance of all controller RIM values from the ideal increases at different rates, making the algorithm's controller represented by the slowest growing curve (here, PPO, although different $\sigma_{\mathrm{train}}$ values do better than others for different problems) the most robust.

On average, most SNOBFit ARIM curves at various training noises perform similarly to L-BFGS across all problems. However, there are distinctions in the region $\sigma_{\mathrm{sim}} \leq 0.05$ where L-BFGS curves start at lower ARIM values and usually degrade more quickly compared to SNOBFit at various noise levels while in the region $\sigma_{\mathrm{sim}} \geq 0.05$ the latter curves grow comparatively more slowly, possibly because the performance determinant, i.e., the fidelity, has degraded so much that further deterioration is less likely across all 100 controllers.

The Nelder-Mead curves exhibit similar behaviour to the SNOBFit curves in that there is less variance w.r.t. the $\sigma_{\mathrm{train}}$ levels. Only for both $M = 7$ problems is performance close to optimal for most curves, where the performance is comparable with PPO. Otherwise, the controllers found by Nelder-Mead are, on average, similar in
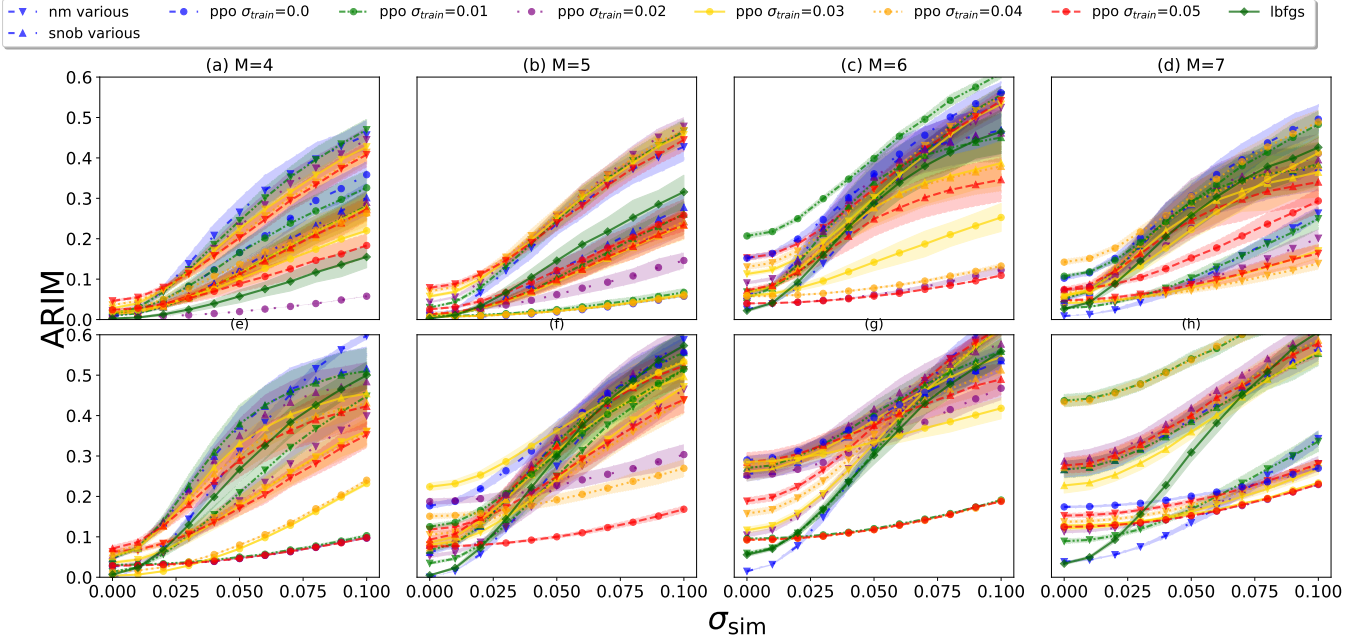
FIG. 5. ARIM as a function of $\sigma_{\mathrm{sim}}$ for $M = 4, 5, 6, 7$ where the top row (a)-(d) are end-to-middle transitions and the bottom row (e)-(h) are end-to-end transitions. The ARIM is computed from a distribution of RIM values for 100 controllers for each $\sigma_{\mathrm{sim}}$ for SNOBFit, Nelder-Mead, PPO and L-BFGS indicated by their marker shapes. Both PPO and SNOBFit are run multiple times at $\sigma_{\mathrm{train}} = 0, 0.01, \ldots, 0.05$ which is indicated by the color of the ARIM curve. For all problems, PPO has higher variance with respect to $\sigma_{\mathrm{train}}$ than SNOBFit and Nelder-Mead. The latter pair's performance curves are more in line with the L-BFGS curve for $\sigma_{\mathrm{sim}} \geq 0.05$ and mostly worse for $\sigma_{\mathrm{sim}} \leq 0.05$. For most of the problems the best performing (lowest) curve across all problems is PPO at $\sigma_{\mathrm{train}} = 0.05$ (brown) except in (a) where it is PPO at $\sigma_{\mathrm{train}} = 0.02$ and in (d) where it is Nelder-Mead at $\sigma_{\mathrm{train}} \geq 0.04$. The error shading is computed using non-parametric bootstrap resampling [80] with 100 resamples (this can also be done using a DKW-like inequality shown in App. A 2).

ARIM behaviour to L-BFGS and SNOBFit.

There is more variance in the PPO ARIM curves across training noises with some curves overlapping each other. The best performing ARIM curve is PPO at $\sigma_{\mathrm{train}} = 0.05$ for 6 of 8 cases shown in Fig. 5 for these problems, with the exception of the end-to-middle $M = 4$ transition (see Fig. 5(a)), where PPO at $\sigma_{\mathrm{train}} = 0.05$ has an ARIM curve that is slightly worse than than the L-BFGS curve and the $M = 7$ end-to-middle transition, where the noisy Nelder-Mead curves dominate for $\sigma_{\mathrm{train}} > 0$. This indicates that PPO is often capable of finding robust solutions, but the optimal value of training noise varies across the transition problems.

### 2. Noisy Learning vs Noisy Search

For the cases $M = 6, 7$, both types of transitions appear to be 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. 5(d)) where PPO at $\sigma_{\mathrm{train}} = 0.05$ is only marginally better than the rest of the algorithm instances 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 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 scale. We investigate asymptotic algorithm behavior w.r.t. 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 an appropriate number of function calls.

We now conduct a comparison between the derivative-free approaches considered in this paper. We extend the RIM analysis of 100 controllers found at training noises $\sigma_{\mathrm{train}} = 0.01, 0.02, \ldots, 0.05$ for the same case ($M = 5$, end-to-middle) considered (without training noise) in Sec. IV A. The RIM values found by Nelder-Mead, PPO and SNOBFit at each $\sigma_{\mathrm{train}}$ are plotted as a function of $\sigma_{\mathrm{sim}}$ in Fig. 6. Each individual subplot is the result of an independent run of each algorithm with a noisy fidelity function evaluated under the structured perturbation $S_{\sigma_{\mathrm{train}}}$ using the same approach as earlier with $\sigma_{\mathrm{sim}}$. These are also plotted for a more distributional compar-
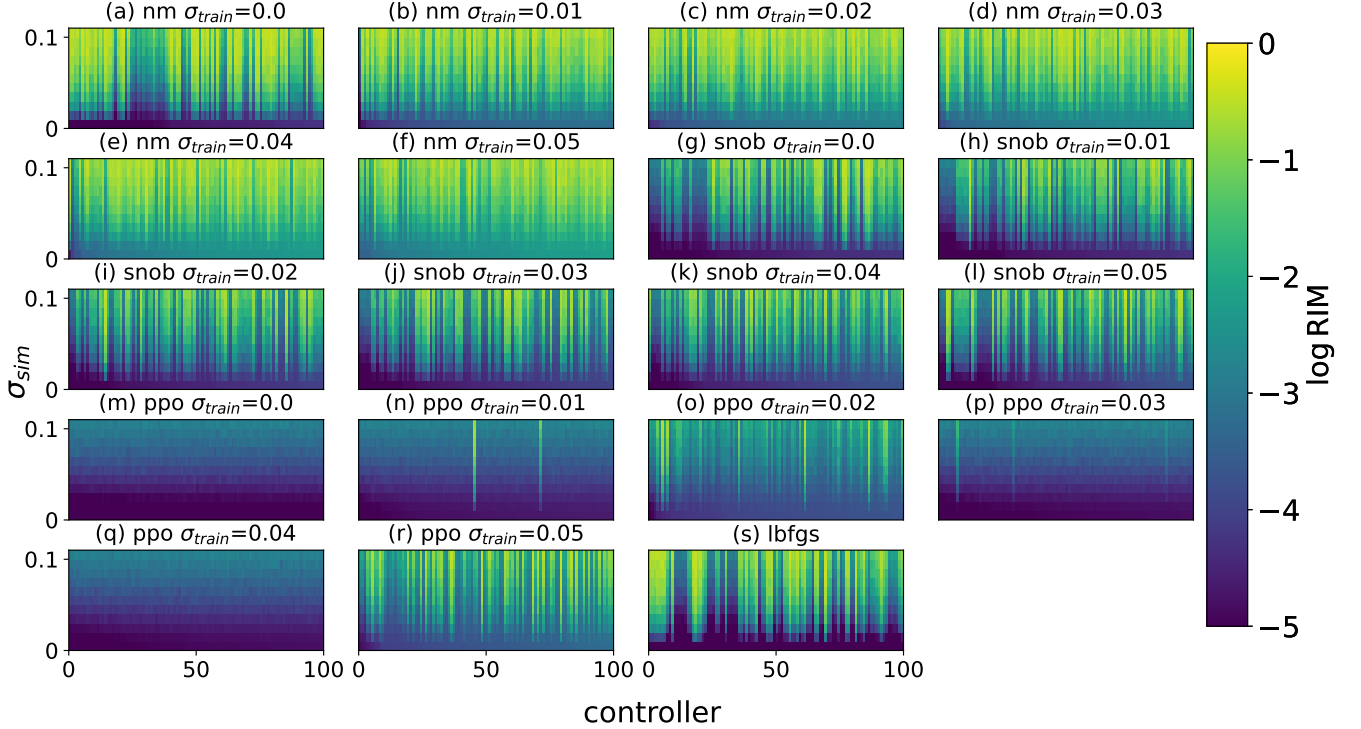
FIG. 6. Individual-controller comparison between (a)-(f) Nelder-Mead, (g)-(k) SNOBFit and (m)-(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity (left) for the case $M = 5$ and the spin transition from $|1\rangle$ to $|3\rangle$. (s) shows the L-BFGS results for the same spin transition problem.

ison as pairwise box-plots in Fig. 7. For both Figs. 6 and 7, we also show L-BFGS results for comparison.

On an individual level, SNOBFit and Nelder-Mead controllers share more characteristics with each other across $\sigma_{\text{train}}$ than the PPO controllers, as observed earlier for the ARIM curves for various control problems (see Fig. 5). On a distributional level, the SNOBFit and Nelder-Mead controllers have high RIM variance within distribution per $\sigma_{\text{train}}$ that is comparable to the L-BFGS controllers as observed from Fig. 7(a). However, the median SNOBFit RIM value per $\sigma_{\text{sim}}$ is higher than L-BFGS so with a bigger left tail. The Nelder-Mead controllers have the most weight on their right tails and are comparatively the worst. However, individually, the controllers found by PPO differ significantly across $\sigma_{\text{train}}$ where notably the RIM values stay uniformly very low for the case $\sigma_{\text{train}} = 0, 0.03, 0.04$ and with minor streaks at $\sigma_{\text{train}} = 0.01, 0.02$. However, the case $\sigma_{\text{train}} = 0.05$ shows pronounced streaks for $\sigma_{\text{sim}} \geq 0.03$. On the distributional level, PPO controllers are generally more robust compared to SNOBFit controllers for the training noises considered here. There is great disparity between the highest performing PPO controllers at $\sigma_{\text{train}} = 0, 0.04, 0.03$ and the SNOBFit and Nelder-Mead controllers at various $\sigma_{\text{train}}$.

Since SNOBFit is a branch and fit method that tries to estimate optimization gradients by constructing local quadratic models around non-local points chosen via

a branch searching algorithm, the overall effect of this strategy, as observed, seems to be in some sense a filtering of the perturbations $S_{\sigma_{\text{train}}}$ in its gradient estimation. The manifestation of this effect is that the controllers at one training noise react similarly, in the sense of RIM, compared to: (a) controllers at various other training noises; and (b) the controllers found by SNOBfit and L-BFGS without any training noise. In contrast, PPO does not filter out the perturbations under $S_{\sigma_{\text{train}}}$ and forms its policy gradient estimates for reinforcement learning from fidelity function evaluations incorporating those perturbations. This is likely the reason differentiating it from SNOBFit. Moreover, the model estimation of the fidelity landscape is done non-linearly using a fixed two-layer linear ($100 \times 100$ dimensional) neural-network. PPO, in some sense, learns the training noise and gets better with it in some cases, but it also performs worse at some training noise levels.

For Nelder-Mead, there are fewer defence mechanisms compared to PPO and SNOBFit for large noise perturbations that might affect the quality of the estimated gradient direction and hence the deterioration of RIM quality with higher noise is unavoidable. This is observed for most of the spin transfer problems that we studied. However, there is a possibility of smoothing in the landscape or a similar positive bias induced by training noise that could have a regularizing effect on the optimization landscape exploration. This can be seen in the case
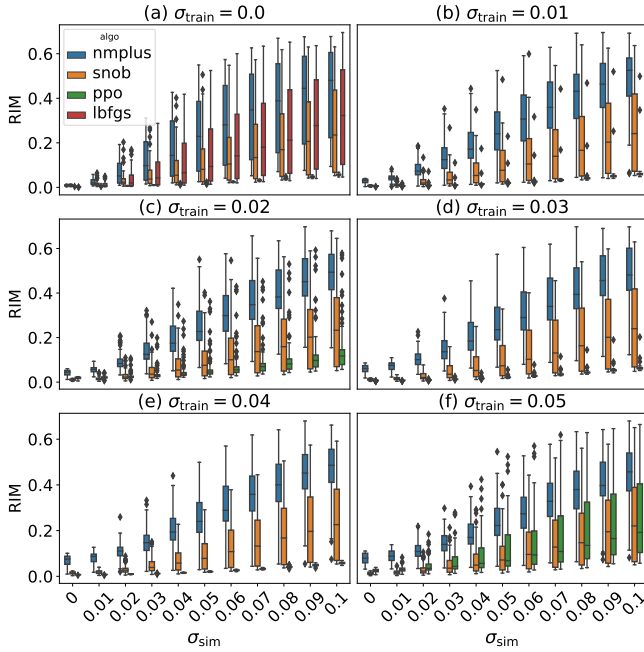
FIG. 7. Box plots of the RIM for the 100 controllers shown in Fig. 6 found by PPO, SNOBFit, and PPO, in this order, for various $\sigma_{\text{train}}$ (a)-(f). For the case $\sigma_{\text{train}} = 0$ in (a), we also show L-BFGS box plots as a final reference. On the distributional level, PPO controllers are generally the more robust of the three w.r.t. the RIM, but there is high variance across $\sigma_{\text{train}}$.

$M = 7$ from Sec. IV B and, here, to a lesser degree, for Nelder-Mead and SNOBFit, but more for PPO.

### C. Comparison of Control Algorithms with Unconstrained Resources

We now look at the behavior of our control algorithms when their target functional calls are unconstrained to address the motivating question (d). That is, we seek to understand an algorithm's ability to optimize the ARIM without constraints on the number of function calls. We contrast two settings: (i) stochastic target function evaluation where for each function evaluation a new Hamiltonian is drawn according to the noise model; (ii) non-stochastic target function evaluations, where the evaluation is over $k$ perturbed, but fixed Hamiltonians, pre-drawn from the noise model. Here, (i) corresponds to producing a high number of quantum devices with slightly different Hamiltonians and choosing one randomly each time we measure the fidelity of a controller under optimization, while (ii) optimizes one quantum device with a slightly uncertain Hamiltonian. Scenario (ii) is the more realistic one in the current quantum device landscape, but the stochastic setting will become more relevant as quantum devices are mass-produced. We find (cf. Fig. 8) that non-stochastic target functionals allow

all control algorithms to optimize the ARIM. When considering stochastic target functionals, they fail with the exception of PPO. Furthermore, PPO takes fewer function calls to asymptotically optimize the ARIM in both settings.

We can further ascertain what effects the training noise level $\sigma_{\text{train}}$ or the number of function calls have on whether a control algorithm optimizes the ARIM successfully. This was inconclusive for some problem instances or noise levels in the preceding section. Moreover, note that in order for L-BFGS [28] to be applicable to the noisy target functionals in both settings, we use a version of L-BFGS that approximates the Hessian using forward differences for both settings.

We fix the control problem to be the end-to-middle $M = 5$ transition, as in preceding sections. We specifically look at how the ARIM for the top 100 controllers changes as a function of the number of function calls for $\sigma_{\text{train}} = 0, 0.05, 0.1$ in steps of $10^6$ function calls up to $40 \times 10^6$. The ARIM is computed in exactly the same way as before for each step. For the stochastic setting (i), we maintain a ranking of top 100 controllers via the stochastic fidelity function evaluation. For (ii) we maintain the ranking through the deterministic RIM value obtained for $k = 100$ Hamiltonians with single structured perturbations of strength $\sigma_{\text{train}}$. This choice was obtained by comparing the RIM for $k = 10, 100, 10000$ to the RIM for $k = 10000$ for a validation set of Hamiltonians computed during optimization where the variance in the RIM decreases by $O(1/k)$.

The results for the average ARIM for all $\sigma_{\text{sim}}$ vs. function calls are displayed in Fig. 8. The first thing to notice is that, for all $\sigma_{\text{train}}$, all algorithms improve the ARIM with increasing function calls in the non-stochastic setting. PPO has the fastest convergence for all $\sigma_{\text{train}}$ after $10^7$ function calls followed by L-BFGS, then Nelder-Mead and finally SNOBFit. SNOBFit is quicker for less than $10^7$ function calls but its ARIM improvements diminish faster than the rest. This means that agnostic of the choice of the control algorithm, we can optimize the ARIM in setting (ii) for certain noise levels (environment or otherwise), but at the expense of an increased number of total function calls.

In the stochastic setting (i), with the exception of PPO, increasing the training noise reduces the asymptotic ability of the control algorithm to optimize the ARIM for all the $\sigma_{\text{train}}$ considered herein. L-BFGS, followed by Nelder-Mead and SNOBFit, is the most prone to performance deterioration at higher $\sigma_{\text{train}}$, which makes sense due to the differences in their reliance on (estimated) gradient information. Moreover, the objective function in these cases is no longer the RIM but just a noisy infidelity value, so even for PPO we see a selection bias for low infidelity but not low RIM as the number of function calls increase after the local optimum in average ARIM is achieved. This causes an increase in the average ARIM for function calls beyond the local optimum point for all the $\sigma_{\text{train}}$. Thus, the average ARIM is no longer reliably
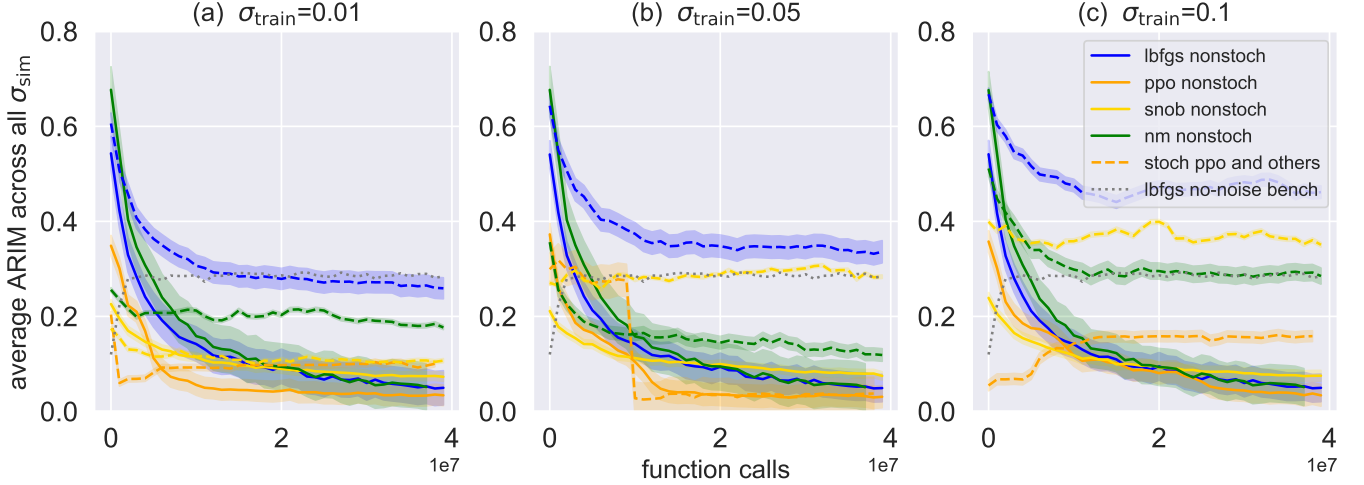
FIG. 8. Asymptotic control algorithm ARIM performance when the number of fidelity function calls is unconstrained. Algorithms where the optimization objective is stochastic and evaluated singularly are plotted with dotted lines and those corresponding to the setting where 100 non-stochastic de-randomized fidelity function evaluations are used to compute a target RIM are plotted as solid lines. The ARIM is computed and averaged over a $\sigma_{\text{sim}}$ scale-set used previously for the $M = 5$ end-to-middle spin transfer problem. Plots (a)-(c) correspond to the training noise $\sigma_{\text{train}} = 0, 0.05, \ldots, 0.1$, where the curves are plotted using 100 controllers ranked by the objective function evaluation and are updated every $10^6$ function calls. For the non-stochastic case, all control algorithms optimize the average ARIM asymptotically but this is not cost competitive with the stochastic setting where PPO performance converges to the local optimum for all noise levels with fewer function calls. The training noise level can help the landscape exploration process, as we see that it positively affects PPO in (a), (c) and Nelder-Mead in (b).

asymptotically optimized. Moreover, sharp transitions in the average ARIM are also typically reported in classical reinforcement learning contexts and is linked to sharp improvements in the reward by the algorithm [81]. There is a clear effect for the presence of noise improving convergence for PPO at $\sigma_{\text{train}} = 0.1$ compared to $\sigma_{\text{train}} = 0.05$, but it also clear that given enough resources, PPO converges to some local optimum regardless of the noise level, which is theoretically reasonable as it optimizes a discounted RIM by design (see App. B 1).

As a baseline, we also plot the asymptotic no-noise L-BFGS ARIM average curve. Contrasting the two settings for a single control algorithm, the point at which there is an advantage for non-stochastic optimization is around $10^7$ function calls for the algorithms excluding L-BFGS with noise. This occurs at about $3 \times 10^6$ function calls for setting (ii) compared with setting (i). For the regime below $10^7$ function calls, setting (i) has a clear advantage over (ii) for PPO and SNOBFit.

The takeaway message is that noise, as a hyperparameter, can positively affect the landscape and aid an optimization algorithm like PPO (and Nelder-Mead for $M = 7$ in Fig. 5(d),(h)) to converge quicker. The precise value varied across the control problems that we studied when the optimization resources were constrained. However, asymptotically, when considering computational resources, the noise level does not matter and PPO is able to optimize a proxy ARIM (a collection of discounted RIMs) in the stochastic setting even for the highest lev-

els of $\sigma_{\text{train}}$ that we considered. Nelder-Mead is similarly capable for lower levels of $\sigma_{\text{train}} < 0.05$ and to a lesser extent. There is also the case that an algorithm's meta-strategy (concerning how to process landscape information, e.g., compute second-order gradient directions for L-BFGS) could be more suitable for particular spin transfer problem instances.

## V. CONCLUSION

We have presented a statistical generalization of the fidelity of a quantum controller in the form of the robustness-infidelity measure $(\text{RIM}_p)$ under arbitrary noise based on the $p$-order Wasserstein distance from the ideal distribution which would be impervious to the structured noise. This measure can be evaluated using physical experiments or Monte Carlo simulations. The $\text{RIM}_p$ is the $p$-th root of the $p$-th raw moment of the infidelity distribution and for $p = 1$ the infidelity measure reduces to the average infidelity. Using a metrization argument, one can justify why $\text{RIM}_1$ (average infidelity) is a good, practical robustness measure for quantum control problems with a nice interpretation as the area under the curve of the cumulative distribution of the infidelity. The RIM framework is illustrated in detail on general time-independent Hamiltonians and unstructured perturbations. However, it can be computed on any object that generates a fidelity distribution over $[0, 1]$, includ-

ing time-dependent and structured perturbations. This could be further explored by looking at methods of aggregating outcomes from piecewise time-constant perturbed Hamiltonians and other extensions to open quantum systems. The RIM concept has been generalized to define an algorithmic RIM (ARIM) to compare the performance of algorithms in terms of finding robust controllers.

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

Under constrained resources, for a stochastic fidelity functional, PPO performed better than SNOBFit and Nelder-Mead when trained with added structured noise during the optimization at certain levels of this added training noise, resulting in lower ARIM curves for a set of control problems on the XX model. SNOBFit exhibited low variance in performance across training noises, suggesting that it is filtering out the added noise. Nelder-Mead exhibits similar ARIM clustering by training noise with less than optimal performance for all but one problem. With unconstrained resources, the variance of PPO ARIM performance disappears across the noise levels, and it is able to optimize the ARIM asymptotically compared to the rest since its reward accumulation strategy implicitly optimizes a discounted RIM. This can also be observed from its consistency statistic matrix where the initial infidelities obtained by PPO are correlated the most with the RIM at low noise levels compared to the other algorithms. In contrast, for optimization over a non-stochastic target RIM, i.e., an ensemble average infidelity over a fixed ensemble, all algorithms are capable of optimizing this objective asymptotically. However, this approach is expensive and less advantageous compared to PPO for all training noises and Nelder-Mead and SNOB-Fit for low training noises. Our results also show that for the more general stochastic setting, e.g. shot noise, PPO is a promising approach to obtain robust controllers.

A limitation of this work is that we require computing multiple controllers per control problem. In simulation, this further involves a large number of 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, though only applied on the XX model here using uncorrelated noise, 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.

Our code and data are available at this url. [82]

[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. Royal Soc. London. Series A: Mathematical, Physical and Engineering Sciences **452**, 2551 (1996).
[5] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation (2009), 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, Quantum gates on individually-addressed atomic qubits subject to noisy transverse motion (2020), 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. Rezakhani, in *IEEE Conf. Decision and Control* (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. Applied **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 **597**, 209 (2021).

[22] R. Blumel, N. Grzesiak, N. Pisenti, K. Wright, and Y. Nam, Power-optimal, stabilized entangling gate between trapped-ion qubits (2021), arXiv:1905.09292.

[23] Q. Ansel, S. J. Glaser, and D. Sugny, J. Phys. A: Mathematical and Theoretical **54**, 085204 (2021).

[24] E. Jonckheere, S. Schirmer, and F. Langbein, Int. J. Robust and Nonlinear Control **28**, 2383 (2018).

[25] C. Villani, *Optimal transport: old and new*, Vol. 338 (Springer, 2009).

[26] S. G. Schirmer, E. A. Jonckheere, and F. C. Langbein, IEEE Trans. Automatic Control **63**, 2523 (2018).

[27] C. Brif, R. Chakrabarti, and H. Rabitz, New J. Phys. **12**, 075008 (2010).

[28] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, ACM Trans. Math. Softw. **23**, 550–560 (1997).

[29] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms (2017), arXiv:1707.06347.

[30] J. Nelder, The Computer Journal **7**, 13 (1965).

[31] W. Huyer and A. Neumaier, ACM Trans. Math. Softw. **35**, 10.1145/1377612.1377613 (2008).

[32] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, Nature methods **17**, 261 (2020).

[33] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. De Jong, in *IEEE Int. Conf. Quantum Computing and Engineering* (2020) pp. 267–277.

[34] M. G. Safonov, A. J. Laub, and G. L. Hartmann, IEEE Trans. Automatic Control **AC-26**, 47 (1981).

[35] M. Arjovsky, S. Chintala, and L. Bottou, in *Int. Conf. Machine Learning* (2017) pp. 214–223.

[36] A. Pratelli, Annales de l'Institut Henri Poincaré, (B) Probability and Statistics **43**, 1 (2007).

[37] A. Ramdas, N. G. Trillos, and M. Cuturi, Entropy **19**, 10.3390/e19020047 (2017).

[38] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).

[39] J. Doyle, in *IEE Proc. D Control Theory and Applications*, Vol. 129 (6) (1982) pp. 242–250.

[40] S. C. Hou, M. A. Khan, X. X. Yi, D. Dong, and I. R. Petersen, Phys. Rev. A **86**, 022321 (2012).

[41] Z. Shen, M. Hsieh, and H. Rabitz, J. Chem. Phys. **124**, 204106 (2006).

[42] E. Barnes, X. Wang, and S. D. Sarma, Scientific Reports **5**, 1 (2015).

[43] M. Demiralp and H. Rabitz, Phys. Rev. A **57**, 2420 (1998).

[44] K. Zhou and J. C. Doyle, *Essentials of robust control*, Vol. 104 (Prentice Hall, Upper Saddle River, NJ, 1998).

[45] S. G. Schirmer, F. C. Langbein, C. A. Weidner, and E. Jonckheere, IEEE Trans. Automatic Control (2022), in press, arXiv:2008.13691.

[46] E. Jonckheere, S. Schirmer, and F. Langbein, IEEE Trans. Automatic Control **62**, 6568 (2017), arXiv:1706.03247.

[47] S. G. Schirmer, F. C. Langbein, C. A. Weidner, and E. A. Jonckheere, in *IEEE Conf. Decision and Control* (2021) pp. 4158–4163.

[48] C. Spearman, American J. Psychology **15**, 72 (1904).

[49] F. C. Langbein, S. Schirmer, and E. Jonckheere, IEEE Conf. Decision and Control , 6454 (2015).

[50] E. Lieb, T. Schultz, and D. Mattis, Annals of Physics **16**, 407 (1961).

[51] S. Bose, Contemporary Physics **48**, 13 (2007).

[52] M. Christandl, N. Datta, A. Ekert, and A. J. Landahl, Phys. Rev. Lett. **92**, 187902 (2004).

[53] J. Zhang, G. L. Long, W. Zhang, Z. Deng, W. Liu, and Z. Lu, Phys. Rev. A **72**, 012331 (2005).

[54] M. Bellec, G. M. Nikolopoulos, and S. Tzortzakis, Opt. Lett. **37**, 4504 (2012).

[55] 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).

[56] T. Graß and M. Lewenstein, EPJ Quantum Technology **1**, 1 (2014).

[57] R. E. Barfknecht, T. Mendes-Santos, and L. Fallani, Phys. Rev. Research **3**, 013112 (2021).

[58] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, Annals of Mathematical Statistics **27**, 642 (1956).

[59] M. Brownnutt, M. Kumph, P. Rabl, and R. Blatt, Rev. Mod. Phys. **87**, 1419 (2015).

[60] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Res. **172**, 296 (2005).

[61] A. Asl and M. L. Overton, Behavior of limited memory bfgs when applied to nonsmooth functions and their nesterov smoothings (2020), arXiv:2006.11336.

[62] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Phys. Rev. A **84**, 022305 (2011).

[63] H.-J. M. Shi, Y. Xie, R. Byrd, and J. Nocedal, A noise-tolerant quasi-newton algorithm for unconstrained optimization (2021), arXiv:2010.04352.

[64] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980.

[65] Z. T. Wang, Y. Ashida, and M. Ueda, Phys. Rev. Lett. **125**, 100401 (2020).

[66] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, npj Quantum Information **6**, 1 (2020).

[67] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, npj Quantum Information **5**, 1 (2019).

[68] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* (MIT press, 2018).

[69] I. Khalid, C. A. Weidner, E. A. Jonckheere, S. G. Schirmer, and F. C. Langbein, Reinforcement learning vs. gradient-based optimisation for robust energy landscape control of spin-1/2 quantum networks (2021), arXiv:2109.07226.

[70] C. A. Weidner and D. Z. Anderson, Phys. Rev. Lett. **120**, 263201 (2018).

[71] E. Zahedinejad, S. Schirmer, and B. C. Sanders, Phys. Rev. A **90**, 032310 (2014).

[72] 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).

[73] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011).

[74] E. Jonckheere, F. C. Langbein, and S. Schirmer, Quantum Information Processing **13**, 1607 (2014).

[75] W.-L. Loh, Annals of Statistics **24**, 2058 (1996).

[76] M. Stein, Technometrics **29**, 143 (1987).

[77] A. B. Owen, Monte Carlo book: the quasi-Monte Carlo parts (2019).

[78] M. G. Kendall, *Rank Correlation Methods: 3d Ed* (C. Griffin, 1962).

[79] A. Agresti, *Analysis of ordinal categorical data*, Vol. 656 (John Wiley & Sons, 2010).

[80] B. Efron, J. American Statistical Association **82**, 171 (1987).

[81] E. J. Rafols, M. B. Ring, R. S. Sutton, and B. Tanner, in *IJCAI* (2005) pp. 835–840.

[82] Https://qyber.black/spinnet/robchar/-/tree/public.

[83] P. Thomas, in *Int. Conf. Machine Learning* (2014) pp. 441–448.

## Appendix A: RIM$_p$ Calculations

### 1.   $p$-th Order RIM

Following Thm. 1, we can write the RIM$_p$ of order $p$ as

$$\text{RIM}_p = \left( \int_0^1 |Q_{\mathbf{P}(\mathcal{F})}(z) - 1|^p dz \right)^{\frac{1}{p}} \quad \text{(A1)}$$

by noting that $Q_{\delta_1}(z) = 1$ for $z \in (0,1]$ by definition, and both terms in the integrand are trivially 0 at $z = 0$. A change of variable $z = C_{\mathbf{P}(\mathcal{F})}(f)$, where $C_{\mathbf{P}(\mathcal{F})}(f) = \mathbf{P}(\mathcal{F} \leq f)$ gives $dz = \frac{dC_{\mathbf{P}(\mathcal{F})}(f)}{df} df = \mathbf{P}(\mathcal{F} = f) df$ and that $Q_{\mathbf{P}(\mathcal{F})}(C_{\mathbf{P}(\mathcal{F})}(f)) = f$. The domain of integration remains invariant and for fidelity measures with support in $[0,1]$ can be extended to $[-\infty, \infty]$. We obtain,

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

$$= \left( \int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F}=f)|f-1|^p df \right)^{\frac{1}{p}}$$

$$= \left( \int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F}=f)(1-f)^p df \right)^{\frac{1}{p}}$$

as $f \leq 1$ switch order and drop $|\cdot|$

$$= \left( \sum_{k=0}^p \binom{p}{k} (-1)^k \int_{-\infty}^{\infty} \mathbf{P}(\mathcal{F}=f) f^k df \right)^{\frac{1}{p}}$$

using the binomial theorem

$$= \left( \sum_{k=0}^p \binom{p}{k} (-1)^k \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})} [f^k] \right)^{\frac{1}{p}} \quad \text{(A2)}$$

as a sum of expectations of various powers of the fidelity. For example, using Eq. (A2), we obtain for $p = 1$, $\text{RIM}_1 = 1 - \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})}[f]$ and for $p = 2$,

$$\text{RIM}_2 = \sqrt{1 - 2\mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})}[f] + Var(f) + \mathbb{E}^2_{f \sim \mathbf{P}(\mathcal{F})}[f]}$$

using Eq. A2, $Var(X) = \mathbb{E}_{X \sim \mathbf{P}}[X^2] - \mathbb{E}^2_{X \sim \mathbf{P}}[x]$

$$= \sqrt{\text{RIM}_1 + Var(f) - \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})}[f] \, \text{RIM}_1}$$

$$= \sqrt{Var(f) + \text{RIM}_1^2} \quad \text{(A3)}$$

expanding $\text{RIM}_1$ and simplifying

and likewise for $p = 3$,

$$\text{RIM}_3 = \big( \text{RIM}_1^3 + 3\, Var(f)$$
$$+ \mathbb{E}^3_{f \sim \mathbf{P}(\mathcal{F})}[f] - \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})}[f^3] \big)^{\frac{1}{3}} \quad \text{(A4)}$$

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$_p$ Estimator

Here we propose a probably approximately correct (PAC) error bound for an estimation $\widehat{\text{RIM}}_p$ of RIM$_p$ in Eq. (A2) based on an empirical estimate $\widehat{\mathbf{P}}(\mathcal{F})$ of its generating probability distribution $\mathbf{P}(\mathcal{F})$. With probability at least $1 - \delta/2$,

$$|\widehat{\text{RIM}}_p - \text{RIM}_p| \quad \text{(A5)}$$

$$= \left| \mathbb{E}_{f \sim \widehat{\mathbf{P}}(\mathcal{F})} [(1-f)^p]^{\frac{1}{p}} - \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})} [(1-f)^p]^{\frac{1}{p}} \right|$$

$$\leq |\mathbb{E}_{f \sim \widehat{\mathbf{P}}(\mathcal{F})} [(1-f)^p] - \mathbb{E}_{f \sim \mathbf{P}(\mathcal{F})} [(1-f)^p]|^{\frac{1}{p}}$$

$$= \left| \int_0^1 \widehat{\mathbf{P}}(\mathcal{F}=f)(1-f)^p df - \int_0^1 \mathbf{P}(\mathcal{F}=f)(1-f)^p df \right|^{\frac{1}{p}}$$

$$\leq \left( \int_0^1 |\widehat{\mathbf{P}}(\mathcal{F}=f) - \mathbf{P}(\mathcal{F}=f)|(1-f)^p df \right)^{\frac{1}{p}}$$

$$= \left( \int_0^1 \left| \widehat{\mathbf{P}}(\mathcal{F}=f) - \mathbb{E}_{\widehat{\mathbf{P}} \sim \mathbf{D}} \left[ \widehat{\mathbf{P}}(\mathcal{F}=f) \right] \right| (1-f)^p df \right)^{\frac{1}{p}}$$

$$\leq \frac{C^{\frac{1}{p}}}{p+1} = \frac{1}{p+1} \left( \frac{\log \frac{4}{\delta}}{2n} \right)^{\frac{1}{2p}} \quad \text{(A6)}$$

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

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

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

### 3. Relative Order of RIM$_p$

Using Lyuapunov's inequality that states that $\mathbb{E}\left[|X|^q\right]^{1/q} - \mathbb{E}\left[|X|^p\right]^{1/p} \geq 0$ for $q \geq p > 0$ for some $\mathbb{E}\left[|X|^t\right] < \infty$ we can easily show that

$$\text{RIM}_q - \text{RIM}_p \tag{A7}$$

$$= \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^q\right]^{\frac{1}{q}} - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^p\right]^{\frac{1}{p}}$$

$$= \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[|(1-f)|^q\right]^{\frac{1}{q}} - \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[|(1-f)|^p\right]^{\frac{1}{p}}$$

$$\geq 0. \tag{A8}$$

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

We can also derive a lower bound on $\text{RIM}_p$. For some $p' \geq p$, we have

$$\text{RIM}_{p'} \leq \text{RIM}_p^{\frac{p}{p'}} = \mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^p\right]^{\frac{1}{p'}}$$

$$= \frac{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^p\right]^{\frac{1}{p}}}{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^p\right]^{\frac{1}{p}-\frac{1}{p'}}} \tag{A9}$$

$$\leq \frac{\text{RIM}_p}{\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)\right]^{1-\frac{p}{p'}}}$$

$$\leq \frac{\text{RIM}_p}{(\min_f 1-f)^{1-\frac{p}{p'}}} \tag{A10}$$

where the relation in the second 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 between a higher order RIM and a lower order RIM. This discrepancy is still concavely dependent on $p$ and $p'$.

We can arrive at equivalence relations for RIMs of different order by noting that $\mathbb{E}_{f\sim\mathbf{P}(\mathcal{F})}\left[(1-f)^p\right] \geq m \sup_f (1-f) = m$ 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. Eq. (A9) yields

$$\text{RIM}_{p'} \leq m^{\left(\frac{1}{p'}-\frac{1}{p}\right)}\text{RIM}_p. \tag{A11}$$

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

$$\widehat{\text{RIM}}_{p'} \leq n^{\left(\frac{1}{p}-\frac{1}{p'}\right)}\widehat{\text{RIM}}_p \tag{A12}$$

This implies that RIMs of different orders are equivalent. We also note that the higher order RIMs increase the measure's sensitivity to outliers greatly, even though growth in the RIM is logarithmic in $p$. For most practical purposes, the first order RIM measure should be sufficient for performance measurements, especially in the quantum technologies setting.

### Appendix B: Optimization Algorithms

### 1. PPO Optimizes a Discounted RIM$_1$

We follow the standard finite-horizon Markov Decision Process (MDP) 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 proximal policy optimization (PPO) algorithm uses a clip objective to update the policy $\pi_\theta$ parameters $\theta$ with first-order constraints that minimize policy distributional divergence. The policy objective is

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

where $\pi(\cdot)$ is the policy probability distribution. $A_{\pi_{\theta_k}}$ are the advantage estimates

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

in terms of the value function $V_\phi(s_t) = \mathbb{E}_\pi\left[\sum_{i=0}^{T-1} \gamma^i r_{t+i+1}|s = s_t\right]$ where $\phi$ 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

$$\phi_{k+1} \propto \arg\max_\phi \sum_{\tau\in\mathbf{D}} \sum_{t=0}^T \left(V_\phi(s_t) - \sum_{i=t}^T \gamma^i r_i(s_t^\tau)\right)^2. \tag{B2}$$

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

$$A_{\pi_{\theta_k}}(s_t, a_t) =$$
$$V_{\phi_k}(s_t) - \left(V_{\phi_k}(s_T) + \sum_{i=t}^{T-1} r_t\right) = V_{\phi_k}(s_t) - \widehat{V_{\phi_k}}(s_t). \tag{B3}$$

The value function can be written in terms of an expectation under the policy, as an average reward: $V_\phi(s_t) = T\mathbb{E}_\pi\left[\frac{1}{T}\sum_{i=t}^T r_i | s = s_t\right]$. The optimal value function defined by $V_*(s_t) = \max_\pi V_\phi(s_t)$, which is maximized if the policy is optimal $\pi_\theta = \pi_{\theta^*}$ at $\theta = \theta^*$. Near optimality, the advantages are approximately 0 as there should be no advantages conferred to the optimal policy $\pi_{\theta^*}$ 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 $\mathrm{RIM}_1$ 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 [83].

### 2. More Consistency Statistic Plots

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 Fig. 9(a)-(f) ((a) is Fig. 4) and $|1\rangle$ to $|4\rangle$ in Fig. 10(a)-(f) for multiple training noise levels. Note that for each subplot the L-BFGS curve is always the same at $\sigma_{\mathrm{train}} = 0$. The controllers found by PPO at $\sigma_{\mathrm{train}} = 0.05$ for are less consistent for some noise levels than others e.g. $\sigma_{\mathrm{sim}} \geq 0.04$ compared with the controllers found by PPO at $\sigma_{\mathrm{train}} = 0.04$. This is also true for SNOBFit and Nelder-Mead. Moreover, the decline in the correlation values is smoothest for PPO compared with the rest for over mostly all twelve instances shown in both figures. With more training noise, Nelder-Mead

is more closer in consistency of the controllers found to L-BFGS sometimes e.g. Fig. 9(a,b) but slowly 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. For most PPO instances, the consistency statistic is highest for $\sigma_{\mathrm{sim}} \leq 0.04$ and thus the infidelity rank order is a good predictor of RIM rank order for higher $\sigma_{\mathrm{sim}}$ which was not observed for any of the other algorithms. Finally, note that this analysis does not reveal anything about how high the RIM values are for the controllers (a drawback of the non-parametric test) and should be processed as companion plots to the figures where these explicit values are shown.

### 3. 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. Fig. 11 shows the case ($M = 5$ and transition $|1\rangle$ to $|4\rangle$) where all the controllers found are effectively not very robust comparatively. This is likely either due to incomplete sampling the space of possible controllers or their non-existence. Note that SNOBFit and PPO are similar in their RIM degradation as observed from Fig. 11(e). We also provide some more cases ($M = 5$ and transition $|1\rangle$ to $|5\rangle$) and ($M = 6$ and transitions: $|1\rangle$ to $|4\rangle$, $|1\rangle$ to $|6\rangle$) for algorithm comparison of controllers under noisy training analysis done in Sec. IV B to highlight some of the variation of controller quality in terms of RIM for different regimes of noise and spin chain transitions observed in the main ARIM comparison presented in Fig. 5.
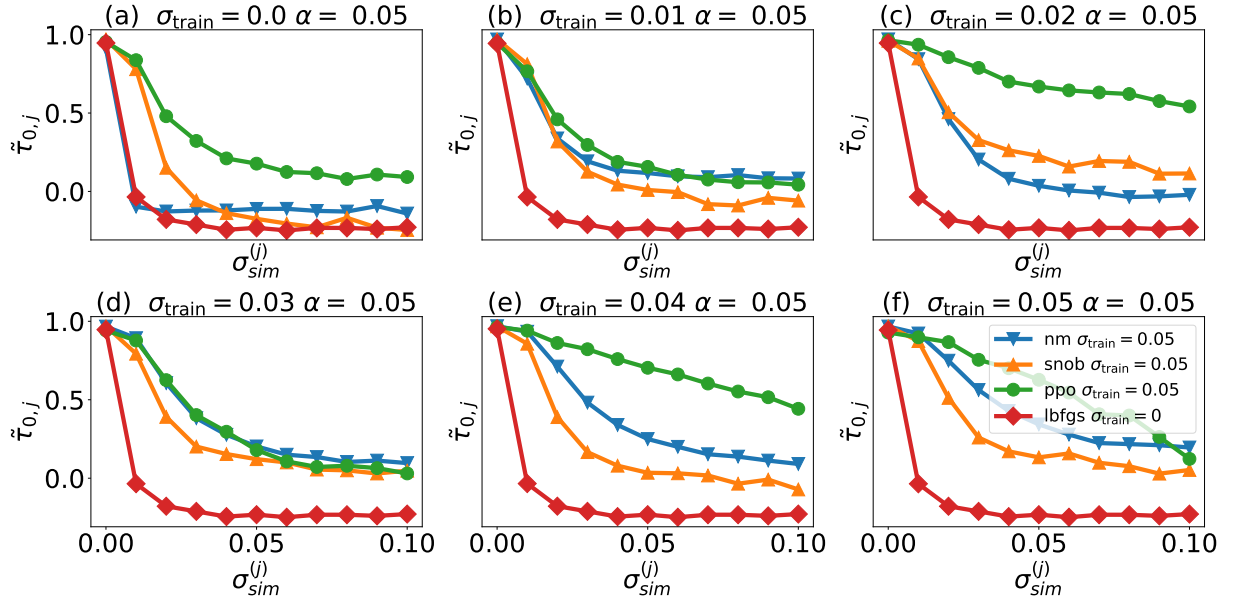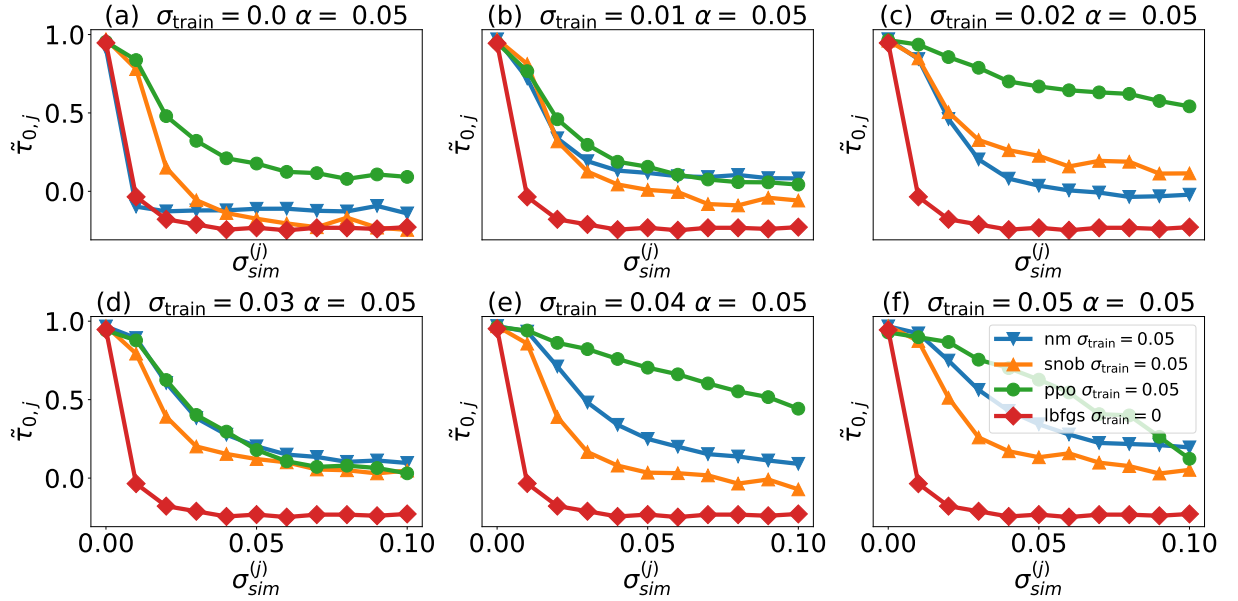
FIG. 9. Consistency statistic $\tilde{\tau}_{0,j}$ for all algorithms at $\sigma_{\text{train}} = 0.0, \ldots, 0.05$ for discrepancy parameter $\alpha = 0.05$ for $M = 5$ and the transition from $|1\rangle$ to $|3\rangle$. Case (a) was presented in the main text. For (f), $\sigma_{\text{train}} = 0.04$, PPO is actually more robust in terms of ARIM growth compared with (e) as seen from their positions in Fig. 5(b). Characteristically, most ARIM-robust PPO controllers show high rank consistency in the region $0 \leq \sigma_{\text{sim}} \leq 0.04$. Nelder-Mead is consistently similar to L-BFGS for all plots except (e,f) where it shows slightly more consistency than PPO and SNOBFit controllers



FIG. 10. 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, effectively, the PPO curves are the most consistent.
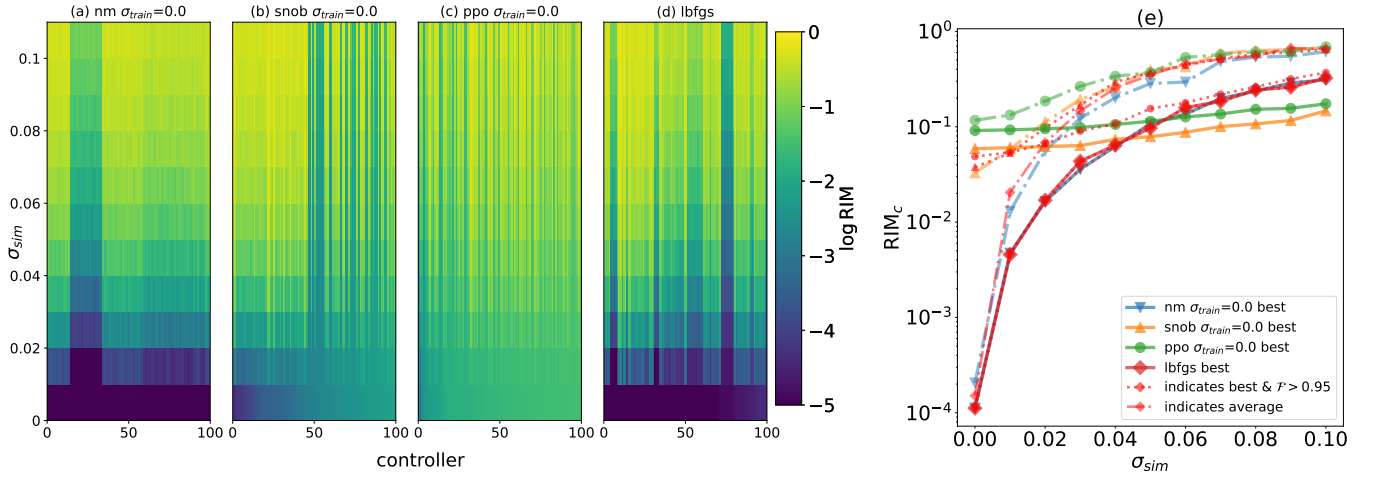
FIG. 11. (a)-(d) 100 controllers found for the XX model, Eq. (14), using Nelder-Mead, SNOBFit, PPO ($\sigma_{\text{train}} = 0$), and L-BFGS for $M = 5$ and the spin transition from $|1\rangle$ to $|5\rangle$. All algorithms find controllers that are not very robust in the sense of RIM. PPO has notably 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 the current noise level and Fig. 12(r) indicates the existence of 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 the Nelder-Mead and L-BFGS both optimize the infidelity to $< 10^{-4}$, however, these best controllers decay very quickly as well.
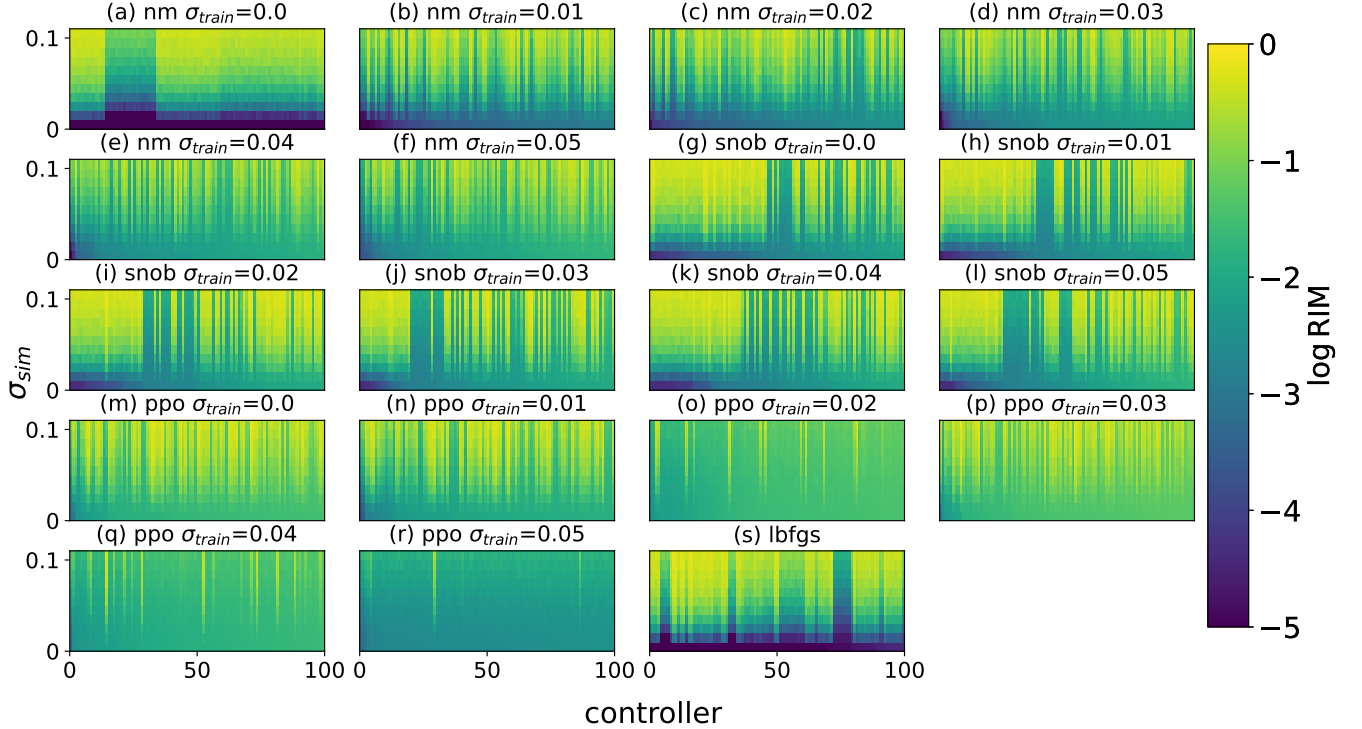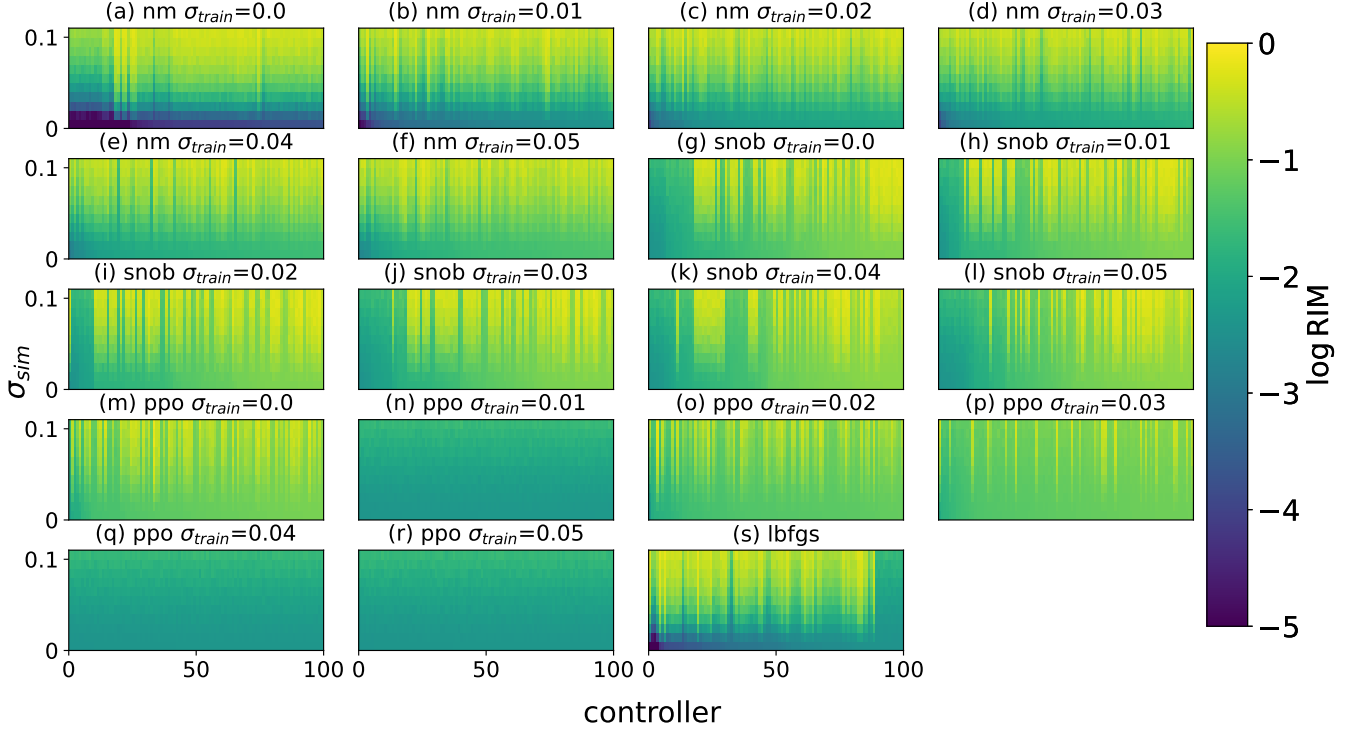


FIG. 12. Individual-controller comparison between (a)-(f) Nelder-Mead, (g)-(k) SNOBFit, (m)-(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity (left) for the case $M = 5$ and the spin transition from $|1\rangle$ to $|5\rangle$. (s) shows the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

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 (left) for the case $M = 6$ and the spin transition from $|1\rangle$ to $|6\rangle$. (s) shows the L-BFGS result for $\sigma_{\text{train}} = 0$.
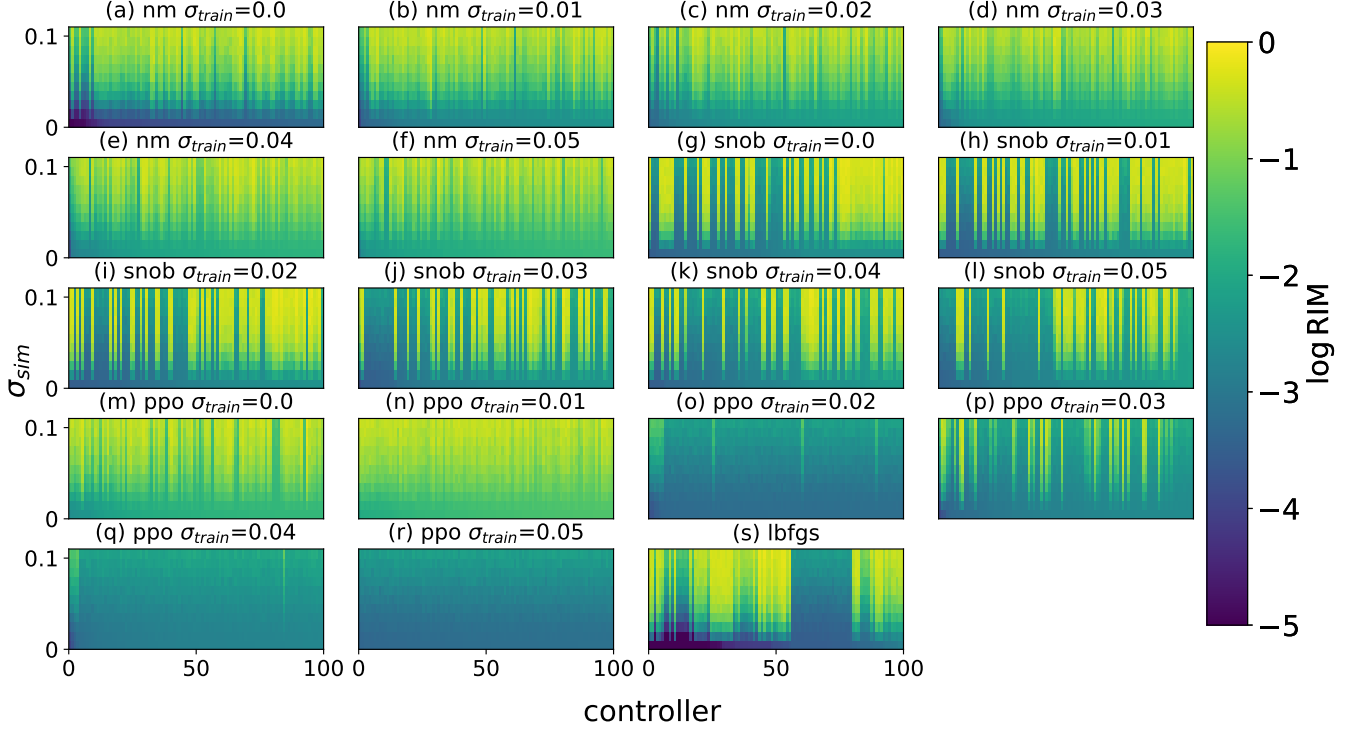


FIG. 14. Individual-controller comparison between (a)-(f) Nelder-Mead, (g)-(k) SNOBFit, (m)-(r) PPO with $\sigma_{\text{train}} = 0, 0.01, \ldots, 0.05$, using 100 controllers ranked by lowest infidelity (left) for the case $M = 6$ and the spin transition from $|1\rangle$ to $|4\rangle$. (s) shows the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