Improving Quantum Neural Networks exploration by Noise-Induced Equalization

Francesco Scala,[1, *] Giacomo Guarnieri,[2, 3] and Aurelien Lucchi[1]

[1]*Department of Mathematics and Computer Science, University of Basel (Switzerland)*
[2]*Dipartimento di Fisica "A. Volta," Università di Pavia, via Bassi 6, 27100 Pavia (Italy)*
[3]*INFN Sezione di Pavia, Via Agostino Bassi 6, I-27100, Pavia, Italy*
(Dated: November 13, 2025)

Quantum noise is known to strongly affect quantum computation, thus potentially limiting the performance of currently available quantum processing units. Even models of quantum neural networks based on variational quantum algorithms, which were designed to cope with the limitations of state-of-the art noisy hardware capabilities, are affected by noise-induced barren plateaus, arising when the noise level becomes too strong. However, the generalization performances of such quantum machine learning algorithms can also be positively influenced by a proper level of noise, despite its generally detrimental effects. Here, we propose a pre-training procedure to determine the quantum noise level leading to desirable optimization landscape properties. We show that an optimal quantum noise level induces an "equalization" of variational parameters: the least important parameters gain relevance in the computation, while the most relevant ones lose it. We analyse this noise-induced equalization through the lens of the Quantum Fisher Information Matrix, thus providing a recipe that allows to estimate the noise level inducing the strongest equalization. Then, we report on extensive numerical simulations providing evidence of the beneficial noise effects in the neighborhood of the best equalization, often leading to improved generalization.

I. INTRODUCTION

In the noisy intermediate-scale quantum era [1], quantum processing units are inherently affected by various sources of noise [2]. Accurately modeling and understanding quantum noise is essential for the advancement of quantum computing, quantum information, and quantum machine learning (QML) [3–6]. A significant research effort is currently focusing on mitigating the detrimental effects of noise to preserve the reliability of quantum algorithms [7–9]. Among the several areas in which Quantum Computing research is currently pursued, QML has attracted considerable interest due to its integration of quantum computation with classical optimization techniques in the so-called variational quantum algorithms [10]. However, noise can practically hinder the potential advantages of QML over classical machine learning models [11, 12], and severely impact the trainability [13, 14] of parameterized quantum circuits [10, 15, 16], including Quantum Neural Networks (QNNs) [17, 18]. In particular, noise-induced barren plateaus (BPs) [19–21] can cause the gradient of the loss function to vanish, thus making classical optimization ineffective. As a consequence, researchers have been investigating the trainability conditions of QNNs in various scenarios [22, 23], and novel techniques have been developed to cope with these challenges [24, 25].

Among the different origins of BPs, noise stands out due to its distinctive nature. A recent study [21] explores the impact of noise on overparametrization [26–28], providing valuable insights into how a considerable amount of realistic noise can induce BPs. In essence, noise induces an exponential suppression of the model's ability to explore the Hilbert space, thus severely limiting its expressivity. This phenomenon can be analyzed through the Quantum Fisher Information Matrix (QFIM) [29–33], whose rank determines the number of meaningful directions in parameter space that are accessible for optimization [27, 28].

Classical machine learning models [34] are known to exploit noise to induce good generalization properties with different techniques, such as noise injection [35–39], stochastic gradient descent [40, 41], data augmentation [42] and dropout [43, 44]. In fact, even if a given model is very well optimized on training data, there is no guarantee that it will perform just as well on unseen data. Overfitting occurs when a model memorizes random noise in the training data, rather than learning the underlying patterns that allow it to make accurate predictions on new, unseen examples. [45–47]. Moreover, a careful analysis of the Fisher Information Matrix (FIM) of learning models leads to the conclusion that well-conditioned FIMs at initialization are often associated with a more favourable optimization landscape (more uniform curvature) [48], thus setting the stage for better generalization [49]. Recently in QML, there has also been an increasing interest in understanding how incoherent processes may help, for example, to avoid BPs [50], or to escape saddle points [51], to provide generative modelling [52] or reinforcement learning [53], or having better generalization capabilities [54–59].

In this work, we provide a novel procedure to identify a beneficial quantum noise level, p^*, before the onset of noise-induced BPs, positively reshaping the optimization landscape and in which vicinity generalization capabilities may be enhanced. While quantum noise on average dampens the effect of parameters on the computation leading to BPs, our results show that modest noise levels increase the relevance of the less influential parameters
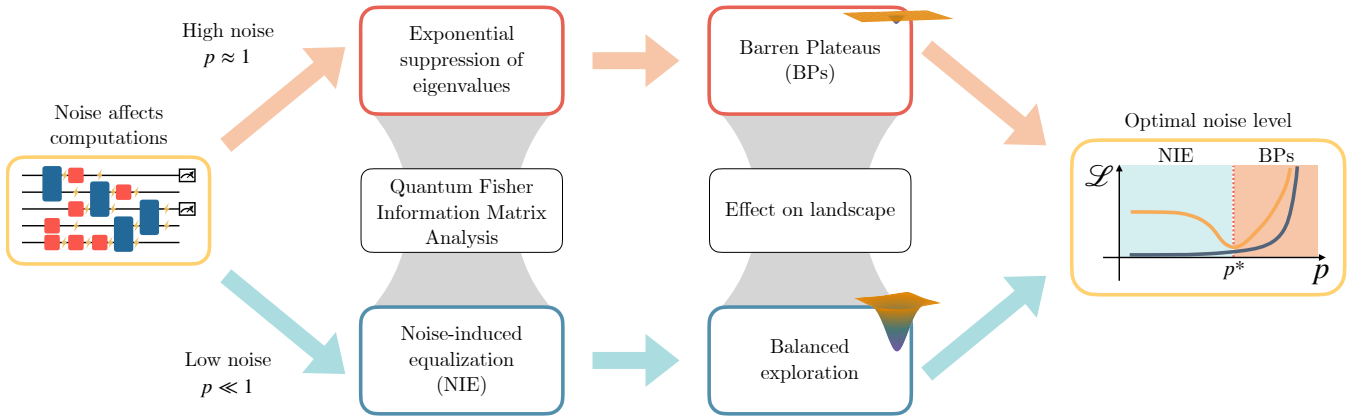
FIG. 1. Schematic illustration of the noise analysis proposed in this work. The action of noise on the QNN algorithmic performances can be analyzed via the Quantum Fisher Information Matrix. If noise is too strong, the eigenvalues of the QFIM are exponentially suppressed, which is known to lead to the BPs issue. However, small levels of noise induce an *equalization* process, in which smaller eigenvalues gain more and more importance, thus allowing a more balanced exploration of the cost landscape. We argue that an optimal level of noise exists, p^*, which gives the best equalization and leads to improved generalization performances of the QNN.

with respect to the noiseless case, thus making the curvature of the landscape more uniform across different directions. We refer to this phenomenon as noise-induced "equalization" (NIE), which we investigate by analysing the eigenspectrum of the QFIM by means of a newly introduced spectrally-resolved measure. Furthermore, we conjecture and numerically verify that in the neighborhood of the noise level yielding the best equalization, superior generalization is promoted, as the reshaping of the optimization landscape favors the parameter space exploration over its exploitation. We notice that the improvement in generalization performances is not a direct property of the QFIM spectrum at initialization; rather, it is a knock-on effect induced by enabling a smoother training dynamics, which in turn tends to land the optimization in flatter — and thus often more generalizable — parameter space regions. The protocol presented in this work is applied before training, and it only depends on the model design, which makes it applicable to various settings and datasets. Finally, we corroborate the quality of the proposed procedure by comparing its estimated optimal noise level to the one obtained by means of a recently proposed generalization bound that also depends on the QFIM spectrum [57]. We show that our procedure allows for a better estimate of useful noisy regimes as compared to the mentioned generalization bound.

The manuscript is organized as follows: in Sec. II we introduce the Quantum Fisher Information Matrix, QNNs as QML models and how to describe quantum noise; in Sec. III we discuss the impact of noise on QNNs, and we define the concept of noise-induced equalization; in Sec. IV we validate our theories by numerical simulations and in Sec. V we discuss our findings and their potential impact; finally, in Sec. VI we describe the details of our implementations.

II. BACKGROUND KNOWLEDGE

In this section, we provide an introduction to the general concept of "information matrix", and in particular to the QFIM, as well as to the quantum algorithms known as QNNs. For completeness, we also provide an outline of the theoretical description of QNN architectures and their overparametrization. Finally, we summarize the noise channel formalism as well as the different kinds of prototypical quantum noise models.

A. Quantum Fisher Information

The optimisation of a parameterized quantum circuits corresponds to adjusting a set of circuit parameters so as to prepare a desired target quantum state.

The natural framework of this problem is the so-called quantum parameter-estimation theory [32], in which the Quantum Fisher Information Matrix (QFIM) quantifies how sensitive the quantum state is upon changes of the parameters, in analogy with the classical Fisher Information Matrix (FIM) [29–33]. This quantity plays a foundational role in quantum metrology (via the Quantum Cramér–Rao bound) and has recently also been applied to analyse overparameterization of parameterized quantum circuits [27, 28]. Thus the QFIM provides a powerful and principled entry point for studying how circuit parameterisation maps to state space and ultimately to algorithmic performance. However, it is important to recognise that, in the multi-parameter quantum regime, the QFIM is not the only tool and additional bounds become relevant. For instance, the Holevo Cramér–Rao bound gives the most general lower bound on the covariance of unbiased estimators when measurement incompatibilities or collective strategies matter [60].

For pure states, the QFIM can be derived from the quantum fidelity, a contrast function quantifying the overlap between two quantum states. For a parameterized family of pure states, $|\psi(\boldsymbol{\theta})\rangle$, where $\boldsymbol{\theta}$ represents the parameters array, one of the possible measures of quantum fidelity between two states is defined as the squared overlap:

$$f(|\psi(\boldsymbol{\theta})\rangle, |\psi(\boldsymbol{\theta}')\rangle) = |\langle\psi(\boldsymbol{\theta}')|\psi(\boldsymbol{\theta})\rangle|^2. \tag{1}$$

One can then define a fidelity-based distance, $d_f = 2(1 - f)$, from which the QFIM is derived as the Hessian of this distance [31]:

$$\mathcal{F}_{ij} = \frac{\partial^2}{\partial\delta\theta_i \partial\delta\theta_j} d_f(|\psi(\boldsymbol{\theta})\rangle, |\psi(\boldsymbol{\theta}+\delta\boldsymbol{\theta})\rangle), \tag{2}$$

where $\delta\boldsymbol{\theta}$ represents a small shift of parameters. Hence, the QFIM elements are explicitly given by:

$$[\mathcal{F}(\boldsymbol{\theta})]_{ij} = 4\text{Re}\{\langle\partial_i\psi(\boldsymbol{\theta})|\partial_j\psi(\boldsymbol{\theta})\rangle + \\ - \langle\partial_i\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|\partial_j\psi(\boldsymbol{\theta})\rangle\}, \tag{3}$$

in which $|\partial_i\psi(\boldsymbol{\theta})\rangle = \partial|\psi(\boldsymbol{\theta})\rangle/\partial\theta_i$.

For mixed states, the QFIM generalizes using the Bures distance and Uhlmann fidelity. The Bures distance provides a measure of the dissimilarity between two density matrices ρ and σ as

$$d_B(\rho, \sigma) = \sqrt{2(1 - \sqrt{F(\rho, \sigma)})}, \tag{4}$$

in which $F(\rho, \sigma)$ is the so called Uhlmann fidelity

$$F(\rho, \sigma) = \left(\text{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2, \tag{5}$$

which quantifies the maximal overlap between their purifications. Hence, the QFIM for mixed states can be derived by considering the spectral decomposition of the density matrix $\rho = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k|$, where λ_k are the eigenvalues and $|\psi_k\rangle$ the corresponding eigenstates. Then, the QFIM incorporates contributions from diagonal and off-diagonal terms, respectively [29, 31]:

$$[\mathcal{F}(\boldsymbol{\theta})]_{ij} = \sum_{\substack{k \\ \lambda_k \neq 0}} \left[\frac{(\partial_i\lambda_k)(\partial_j\lambda_k)}{\lambda_k} + 4\lambda_k\text{Re}\{\langle\partial_i\psi_k|\partial_j\psi_k\rangle\}\right] + \\ - \sum_{\substack{k,l \\ \lambda_k,\lambda_l \neq 0}} \frac{8\lambda_k\lambda_l}{\lambda_k + \lambda_l}\text{Re}\{\langle\partial_i\psi_l|\psi_k\rangle\langle\psi_k|\partial_j\psi_l\rangle\}. \tag{6}$$

We notice that the latter matrix is positive semidefinite, real, and symmetric thus inducing a proper metric onto the parameterized manifold.

Extending the concept of the QFIM to non-pure quantum states allows capturing the sensitivity to parameter variations in real-world scenarios where quantum noise is also present. A crucial property for what follows is the general contractivity of the QFIM under the action of a quantum channel $\Lambda_t[\cdot]$(i.e. a completely positive and trace preserving dynamical map acting on the state space)

$$\mathcal{F}(\Lambda_t[\rho(\boldsymbol{\theta})]) \preccurlyeq \mathcal{F}(\rho(\boldsymbol{\theta})), \tag{7}$$

where $\preccurlyeq$ represents the Löwner ordering of matrices. Eq.(7), known as Data-Processing Inequality, physically implies that the ability to discriminate two quantum states can only be degraded under the action of noise. A special case of the Data-Processing Inequality implies also that the classical Fisher information $\mathcal{I}$ extracted after a quantum measurement process is always upper bounded by the quantum Fisher information [31]:

$$\mathcal{I}(\text{Tr}[\rho(\boldsymbol{\theta})O]) \preccurlyeq \mathcal{F}(\rho(\boldsymbol{\theta})), \tag{8}$$

implying that the information that can be extracted from a state is always smaller than the information contained in the state itself.

Since the QFIM provides at least the same amount of information as the classical FIM, in the following, we will analyze QNNs, defined as parameterized quantum circuits, by using the QFIM instead of the classical FIM. The analysis we propose is based on the QFIM eigenspectrum. As just introduced, the QFIM describes the sensitivity to parameter variations of a certain quantum model. This sensitivity can also be interpreted as the importance of different parameters in the computation. Hence, analyzing the eigenspectrum of the QFIM corresponds to conducting the study in a rotated framework, in which the QFIM is diagonal. Notice that this approach considers only linearly independent directions in the parameters space, which may be obtained by linear combinations of the directions defined by the variational parameters.

B. Quantum Neural Networks

Given a dataset $\{\mathbf{x}_i, y_i\}_{i=1}^M$, where data points $\{\mathbf{x}_i\}_{i=1}^M$ are sampled from a distribution $\mathcal{D}$ with corresponding labels $\{y_i\}_{i=1}^M$, we define a QNN model by selecting an observable (i.e., a Hermitian operator) O and computing its expectation value with respect to a L-layer parameterized quantum state, represented as a density matrix $\rho_L(\mathbf{x}_i, \boldsymbol{\theta})$:

$$f(\mathbf{x}_i, \boldsymbol{\theta}) = \text{Tr}\left[O\rho(\mathbf{x}_i, \boldsymbol{\theta})\right]. \tag{9}$$

The density matrix can be represented as

$$\rho_L(\mathbf{x}_i, \boldsymbol{\theta}) = U_L(\mathbf{x}_i, \boldsymbol{\theta})\rho_0 U_L^\dagger(\mathbf{x}_i, \boldsymbol{\theta}), \tag{10}$$

in which $\rho_0 = (|0\rangle\langle0|)^{\otimes n}$ is the initial quantum register state. The evolution operator $U_L(\mathbf{x}_i, \boldsymbol{\theta})$ is then explicitly given by

$$U_L(\mathbf{x}_i, \boldsymbol{\theta}) = \prod_{l=0}^L U_l(\boldsymbol{\theta})S_l(\mathbf{x}_i), \tag{11}$$

where $U_l(\boldsymbol{\theta})$ represent the trainable parameterized unitaries, and $S_l(\mathbf{x}_i)$ are encoding operations that embed the classical data points, $\mathbf{x}_i$. Here, L denotes the number of layers (i.e., the depth) of the QNN. While we focus on scalar output functions for simplicity, this model can be readily extended to produce a vector-valued output, $\boldsymbol{f}$, by assigning different observables O_j to each component f_j.

In practical QNN implementations, quantum computations are inevitably subject to noise. This can be modeled by the action of a quantum channel $\mathcal{N}_l$ [2], which may affect the system after each encoding operation S_l and after each trainable unitary U_l. Analitically, any quantum channel $\mathcal{N}$ acting on a density matrix ρ can be described using the Kraus decomposition as

$$\mathcal{N}(\rho) = \sum_k K_k \rho K_k^\dagger, \qquad (12)$$

where K_k are the Kraus operators associated with the channel and satisfy the completeness relation $\sum_k K_k^\dagger K_k = I$ to ensure trace preservation. This representation provides a convenient and general framework to model various noise processes, such as depolarization, dephasing, or amplitude damping, by specifying the appropriate set of K_k (see Appendix B). Using this framework, noisy models can then be trained according to the known variational quantum algorithm scheme [10]. In this work, we investigate the impact of these different noise channels on QNN training and generalization.

Noisy models can then be trained according to the known variational quantum algorithm scheme [10].

A QNN is said to be overparameterized when the rank of the QFIM reaches its maximal value for all the elements in the training set [28]. This corresponds to the situation in which adding more parameters to the model does not increase the rank of its associated QFIM. Notice that this definition of overparametrization only depends on features of the QNN model, while it is independent from the particular loss function employed, or the given variational problem to be solved. In particular, the rank of the QFIM can be considered as one of the possible definitions of the effective dimension of the QNN [27, 57, 61]:

$$d_{eff} = \mathrm{rank}\left[\mathcal{F}(\boldsymbol{\theta})\right]. \qquad (13)$$

The maximal achievable dimension for pure states is $d_{eff}^{max} = 2^{n+1} - 2$, which corresponds to the number of independent real parameters in the state vector describing the quantum state. For mixed states this value is enhanced to $d_{eff}^{max} = 2^{2n+1}$, since their full description requires defining the corresponding density matrices. Alternatively, the effective dimension can also be defined as the number of QFIM eigenvalues that are above a given threshold.

III. THE IMPACT OF NOISE ON QNNS

After having introduced the QFIM and the main models employed to approximate the description of the different quantum noise channels, here we start diving into the effects of such sources of noise on the QNN performances. First, we briefly report some preliminary results, mostly known from the literature [21], highlighting the effects of noise on overparameterized QNNs and motivating the analysis conducted in this work. Then, the main contribution of the present work will be introduced, i.e., the NIE procedure and its relationship with the performance of the QNN learning model. As already anticipated, the QFIM will be the main tool employed to analyse the effects of noise.

A. Motivation of the study

Computing the eigenspectrum of the QFIM associated to a specific QNN provides valuable insights into how variations in the parameter space may impact the output quantum state. In the case of overparameterized quantum models, only a subset of all the parameters "actively" contributes to changing the quantum state, while the other parameters act redundantly in the same directions within the parameter space. Mathematically, such a redundancy results in null eigenvalues and a saturated rank of the QFIM. Building upon previous theoretical knowledge [62], Ref. [21] studied the effect of depolarizing noise on overparameterized QNNs by means of the QFIM.

The main outcome of this latter analysis is that the QFIM entries (as well as the eigenvalues) undergo an exponential suppression in the case of depolarizing noise, either global or local, combined with general unital Pauli channels. In particular, for the latter case, the scaling is exponential in both the probability of applying depolarizing noise (p) and the total number of noisy gates in the circuit. This exponential decay of the QFIM implies that the QNN becomes insensitive to parameter variations as the level of depolarizing noise increases, thus resulting in a flat loss landscape. This behaviour explains the rise of noise-induced barren plateaus.

Moreover, it was also noticed that small local depolarizing noise (possibly combined with unital Pauli channels) may increase the QFIM rank in overparameterized QNNs. This corresponds to an increase in the value of previously null eigenvalues, even though the QFIM, and hence the average of all the eigenvalues, is exponentially suppressed overall. We stress that the null eigenvalues increase only for small noise intensities. This effect stops when these eigenvalues become comparable to the non-zero ones. After this threshold in noise level, they start to be exponentially suppressed too. This leads to two different regimes. First, the null eigenvalues become non-trivial, and the QNN can be considered as *quasi-overparameterized*, meaning that noise enables

the exploration of new directions, effectively reducing the level of overparametrization. However, as the noise level gets higher and higher, *all* the eigenvalues are exponentially suppressed, which ultimately results in noise-induced barren plateaus.

In practice, this implies that in the overparameterized regime, the noiseless QFIM becomes ill-conditioned due to its smallest eigenvalue vanishing, whereas the presence of moderate quantum noise improves its conditioning. This phenomenon strongly resembles what was noticed in Ref. [48] for classical NNs: linear networks have ill-conditioned FIM, while nonlinear NNs do not possess null FIM eigenvalues. This change in FIM conditioning was shown to improve the effectiveness of first-order optimization methods. Given that quantum noise can be considered as a quantum nonlinearity, this motivates the study of how this phenomenon actually affects the performance of QNN learning models. In what follows, we are going to argue that this quasi-overparameterized regime is part of a more general phenomenon in which the least important parameters gain relevance with respect to the most important ones. We name this process the "noise-induced equalization", as detailed in the following.

B. Noise-induced equalization

Inspired by these previous results on overparameterized QNN, here we aim at giving an answer to the following question: "*what are the consequences of this change in the relative importance of the parameters?*" by further analyzing the QFIM eigenspectrum.

Before presenting the numerical results in the next Section, we introduce a new, useful QFIM-based framework to better interpret the NIE phenomenon. In particular, the QFIM is computed as the Hessian of the fidelity between quantum states, for this reason, it can be considered as an indicator of the flatness/steepness of the state space. Consequently, its eigenvalues can also be related to the steepness in the "state landscape" with respect to a certain direction (in the diagonalized mapping). Sorting the eigenspectrum by intensity would correspond to assigning a *steepness rank* r to all the different directions. It has to be noticed that such a rank (and in general eigenvalues) would depend on the specific QFIM under analysis, which changes for different input data $\mathbf{x}$, variational parameters $\boldsymbol{\theta}$ and, if present, the quantum noise level p, which implies that r depends on all these variables as well. This means that the QFIM is a *local* object sensitive to changes in the underlying data and model parameters, as well as quantum noise. For this reason, here we are going to define a direction-blind, but spectrally resolved, information measure which will then be evaluated at multiple points in order to draw some useful insights into the *global* landscape. Hence, given a QNN with P parameters and with an associated QFIM, let us denote with $\{\lambda_r\}_{r=1}^{P}$ the ordered eigenspectrum of the QFIM such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_P$. We note

that in cases of degenerate eigenvalues, we manually and arbitrarily assign an relative rank between them, e.g., if $\lambda_a = \lambda_b$, we set $r(a) > r(b)$. After fully sorting the eigenspectrum, the original directions no longer matter and we focus solely on the magnitudes of the ordered eigenvalues. In other words, our goal is to compare the overall distortion of the manifold without regard to which specific directions in the space are more or less distorted. Suppose that each operation in the quantum model is noisy, i.e., that every gate is followed by a quantum channel with a noise level quantified by p. To understand the effect of p on the given model, we can study the "rank-wise" change in the importance of each direction in the parameter space, $I_r(p)$, with respect to the noiseless case $(p = p_0 = 0)$, which is explicitly defined as:

$$I_r(p) = \frac{\lambda_r(p)}{\lambda_r(p_0)} . \tag{14}$$

For the above reasons, Eq. (14) can be interpreted as a change in steepness in the quantum state space, allowing us to identify how distorted the manifold is in different directions, thus offering a finer characterization at the single eigenvalue level. In this sense, $I_r(p)$ provides a novel, spectrally-resolved perspective on the degradation of quantum information, going beyond previously studied aggregated measures [21, 27, 57, 61] and enabling a more detailed understanding of how information geometry deforms with increasing noise.

The choice for this specific measure is guided by multiple factors. First, from the Data-Processing Inequality and the Löwner order relation we know that the trace of the QFIM can only decrease under the action of quantum noise (for depolarizing noise, we know that it is moreover exponentially suppressed [21]). The same argument also excludes the average of eigenvalues (and other global quantities like the determinant [57]) among the candidates for measuring the effect of noise on the spectrum. Hence, we may think of splitting the eigenspectrum into two parts, i.e., large and small eigenvalues, then excluding the largest eigenvalues, and computing the average of the small eigenvalues in order to characterise their behaviour. While this approach may seem a valuable path, how to determine a proper splitting of the eigenspectrum remains unclear for a generic QNN. In fact, while such a splitting may appear quite evident for overparameterized QNNs, generalizing this procedure is non-trivial. By considering the measure defined in Eq. (14), instead, we can compare each noisy eigenvalue to its noiseless version, thus giving a clear quantification of how much the specific element of the QFIM eigenspectrum is affected by noise, in particular, whether it is increased or lowered, respectively. Notice that the possibility of some of the smallest eigenvalues increasing does not contradict the Löwner ordering relation between the noiseless and noisy QFIM. In fact, assuming that the noise will decrease the trace of the QFIM, the noiseless QFIM only *weakly majorize* [63] the noisy QFIM, meaning that the following inequality holds for the partial sums of decreas-

ingly ordered eigenvalues

$$\sum_{r=P}^{k} \lambda_r(p) \le \sum_{r=P}^{k} \lambda_r(p_0) \quad \text{for } k = P, \dots, 1\,. \quad (15)$$

We now define the concept of equalization and specify the conditions under which it is considered optimal, a regime in which improved exploration is expected. The equalization definition is based on empirical deductions derived from extensive numerical simulations (see the next Section). In particular, we numerically observe that certain levels of noise increase the least important eigenvalues, while the most relevant eigenvalues are damped. In fact, this is the process we define "noise-induced equalization" (NIE), which we formalize as follows:

Observation (Noise-induced equalization). *Given a QNN model whose associated QFIM ordered eigenspectrum is $\{\lambda_r\}_{r=1}^{P}$, for all noise levels $p \ge 0$ there exists an integer $R \in [0, P]$ such that*

$$\begin{cases} I_r(p) > 1 & \forall r \le R\,, \\ I_r(p) \le 1 & \forall r > R\,. \end{cases} \quad (16)$$

Here, notice that $R = 0$ is also included as a possibility, to account for the situation in which all the eigenvalues lose importance (i.e., they are exponentially suppressed). The latter is the case of high noise levels. As a consequence of the former Observation we can define the level of noise inducing the best equalization:

Definition (Best NIE). *Given a QNN model, the best noise-induced equalization arises when subject to quantum noise of level p^*:*

$$p^* = \frac{1}{R_{max}} \sum_{r=1}^{R_{max}} \mathbb{E}_{\mathbf{x},\theta} \left[p_r^* \right]\,, \quad (17)$$

with $R_{max} = \max_p R$ and $p_r^* = \arg\max_p I_r(p)$.

Here, R_{max} is taken to consider all and only the eigenvalues that can acquire importance by changing the noise level. Then, p^* is obtained by averaging the noise values leading to the maximal gain ($p_r^* = \arg\max_p I_r(p)$) per each eigenvalue that can acquire importance ($r \le R_{max}$). In Fig. 2, we show how our measure varies for different eigenvalues as a function of the quantum noise. The averaging over inputs and parameters is taken in order to cancel the dependency on inputs and specific points in the landscape. While this is necessary to determine a unique noise level over the landscape, getting rid of data dependence may limit the power of the method in the context of generalization, as it is known that such performances are influenced by the data distribution [64–66].

It is worth emphasizing that Eq. (17) is different from using the $\arg\max$ of $\bar{I}_{R_{max}}(p) = \sum_r^{R_{max}} \mathbb{E}_{\mathbf{x},\theta}(I_r)/R_{max}$, as it would allow to only select noise levels among the tested ones.

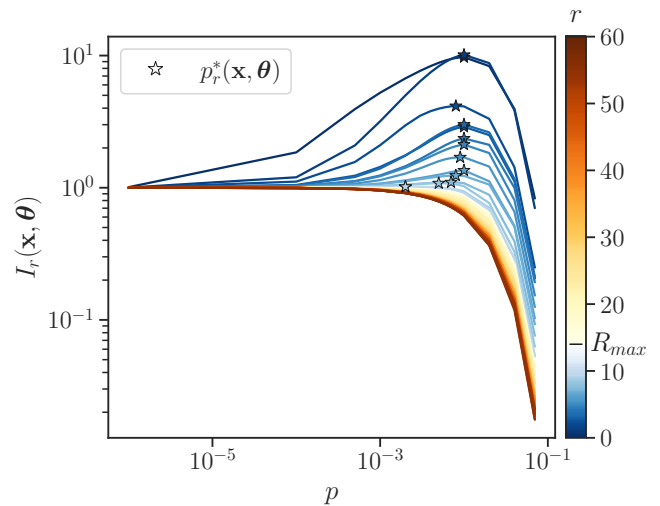FIG. 2. Change in steepness $I_r(\mathbf{x},\boldsymbol{\theta})$ for the for input $\mathbf{x}$ and parameter vector $\boldsymbol{\theta}$, for an underparametrized QNN of $n = 5$ qubits subject to dephasing noise.

Following this definition, we conjecture that around the noise level p^* the QNN should experience improved performances. The intuition behind this conjecture is that noise equalization allows to reduce the hyper-specialization of the most influential directions in parameter space, while reinforcing the weakest ones, i.e., reducing exploitation in favor of exploration.

In Appendix C, we approach the effect of quantum noise from a theoretical perspective. In particular, after introducing the necessary background knowledge on Dynamical Lie Algebras (DLAs) in closed and open quantum systems, we give insights into how noise can induce equalization in two different situations: when the generators of the unitary dynamics either commute with the noise superoperators or when they do not. When noise introduces new generators in the quantum dynamics, new directions are explored if they do not commute with the ones already present in the noiseless setting. This reduces the ratio between the number of variational parameters and the number of directions that can be explored, effectively reducing the exploitation of redundant directions. This is the reason why the null eigenvalues are activated when we apply noise to overparameterized QNNs. Anyway, this does not explain how equalization takes place in underparameterized QNNs, where less important eigenvalues are strengthened without any activation of new eigenvalues. To provide a partial explanation to this phenomenon, inspired by Ref. [67], we analytically show in Appendix D how the QFIM eigenvalues are affected by noise for two toy models with noise operators commuting with the generators of the DLA associated with the quantum circuit, thus not enlarging the DLA. This allows us to prove the existence of an optimal level of noise p^* and demonstrate how information flows in directions that were not active in the noiseless setting.

Hence, there can be situations in which noise enhances expressivity, potentially enabling more complex transformations, even in the presence of decoherence. Noise-induced equalization could then be regarded as the sweet spot between enhanced expressivity (strenghtened low eigenvalues) and detrimental noise effects given by loss of coherence (weakened high eigenvalues). Ultimately, this could allow the improvement of generalization capabilities for QNN models.

From the point of view of space curvature, this can also be intuitively seen as a reshaping of the landscape where extremely steep directions (associated with the most influential parameters) are smoothened, while flatter directions (linked to the least important parameters) gain some steepness. This may result in a loss landscape in parameter space in which minima are wider and flatter, a property which has been associated to better generalization performance [49]. Hence, we conjecture that such an *equalization* in the QFIM eigenvalues related to QNNs (not necessarily overparameterized) might be the reason behind the improved generalization properties that have been numerically observed, e.g., in Ref. [68], and analytically predicted with generalization bounds in Refs. [57–59].

In Sec. IV, we provide numerical evidence to confirm our anticipated conjecture, demonstrating its applicability to both overparameterized and underparameterized QNNs. Our results show that NIE of the eigenspectrum occurs in both cases, thus providing a key pre-training insight into the behavior of QNNs. Through this phenomenon, we derive an estimate for p^* based on our conjecture, and validate that such a noise level actually often leads to desirable generalization performance across various use cases.

C. Comparison with generalization bound

As already mentioned, the connection between noise and generalization of QNN has started to appear in the recent literature. In particular, in the case of Ref. [68] empirical conclusions have been drawn after numerical evidence in selected cases, while in Refs. [57–59] the authors have tried to quantify the regularizing power of noise by means of generalization bounds. In contrast to these approaches, in the present work we study the origin of such a phenomenon, which allows us to give an estimate of a noise level, p^*, before actually training the QNN model, where we argue that improved performances may be observed. In fact, to further strengthen our claim, in Sec. IV we apply our estimation procedure to the very same dataset considered in Ref. [68] to check whether we have compatible findings. Moreover, Ref. [57] provides a generalization bound for noisy quantum circuits depending on the QFIM spectrum among other quantities, but the dependency of these elements on noise and the specific role of QFIM is not discussed. Understanding how the quantum noise explicitly enters the gener-

alization bound could potentially lead to an alternative method to estimate p^*.

In Appendix E, we explicitly report the expression of this bound, grouping together in the quantity $B(p)$ all the noise-dependent terms. The main quantities affected by noise are: (i) the effective dimension (d_{eff}), (ii) the square root of the QFIM determinant (m), and (iii) the Lipschitz constant of the model (L_f). Specifically, we realize how the first two quantities are affected by noise through their dependence on the QFIM. In fact, we know that the QFIM eigenvalues undergo an equalization process, from which they are exponentially damped by noise. This implies that both the effective dimension (derived by the rank of the QFIM) and the determinant will change under the action of noise. The noise progressively smoothens the optimization landscape until it is completely flat (noise-induced BPs). As the Lipschitz function L_f dominates the gradient norm, increasing the noise will reduce the gradient norm and, consequently, L_f. In Appendix J, we are going to monitor the variation of $B(p)$ on increasing levels of noise, explicitly showing after numerical simulations that its minimum occurs for a non-trivial level of noise. As shown in Sec. IV, this noise level could be used as a *broad estimate* of p^*, since it is obtained from a bound. More specifically, this kind of theoretical tool associates good generalization performances with a small generalization gap. However, the latter may also result from underfitting, making it rather vacuous.

To conclude this section, we remark that since the noise levels on current hardware are lower than the ones required for regularization [68], noisy dynamics can be obtained by exploiting ancillary qubits via Stinespring dilation [69]. With the present framework, such noise levels could be obtained by simulations anticipating the training procedure, after which incoherent dynamics can be artificially introduced in the quantum circuits for regularization purposes in QML applications.

IV. NUMERICAL RESULTS

In this section, we provide numerical evidence that specific levels of quantum noise, in the neighbourhood of a certain p^*, can induce regularization in QNNs suffering from overfitting. In particular, via a pre-training analysis of the QFIM eigenspectrum, we are able to estimate the noise level p^*, typical of the particular noise under investigation and model design, which will correspond to the regime where the optimization landscape is smoother. This strategy is advantageous, as it allows one to find a noisy good operating regime without extensive repetitions of the training procedure using hyperparameter grid search.

We start by presenting how noise affects the QFIM eigenspectrum in both underparameterized and overparameterized QNNs. In particular, we show that different noise levels have different effects on the eigenvalues and
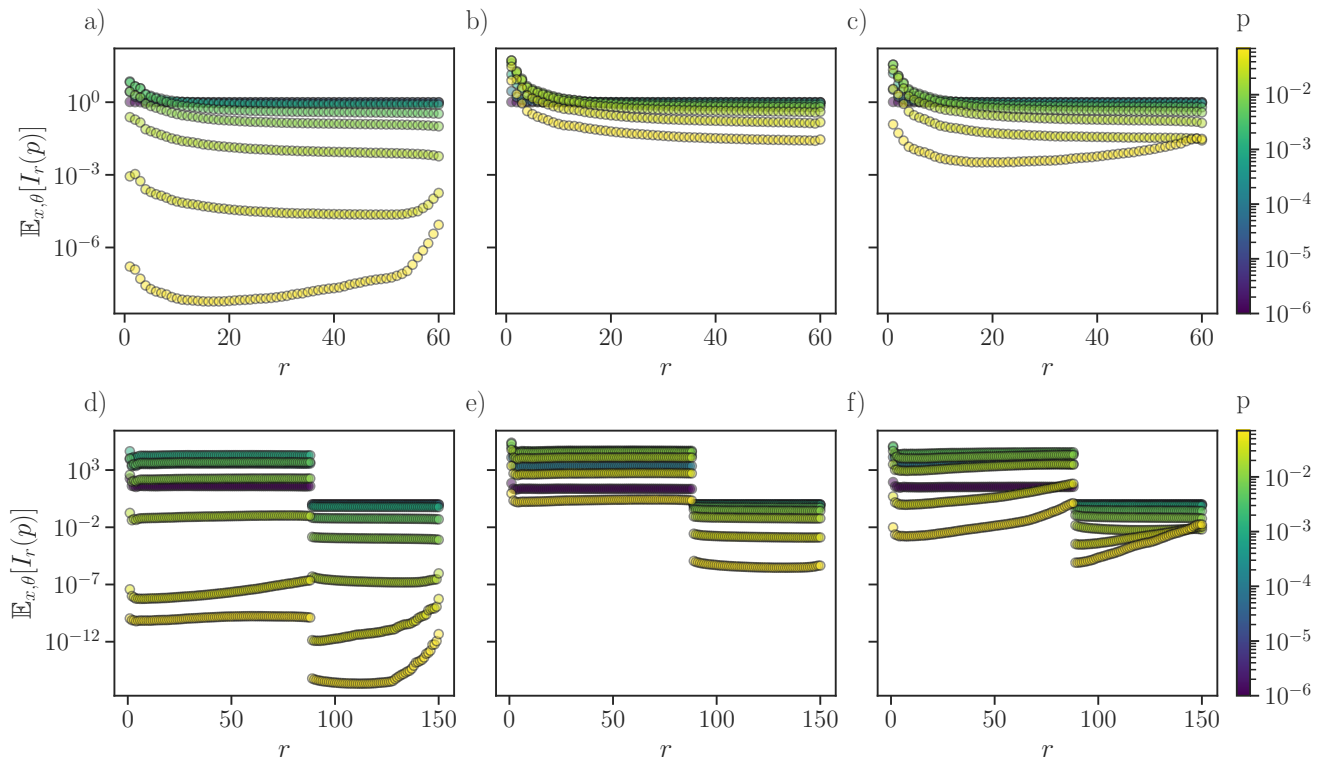
FIG. 3. Relative change of the averaged QFIM eigenvalues λ_m under different levels of noise p with respect to the noiseless case ($p = 0$) for a), d) depolarizing, b), e) dephasing and c), f) amplitude-damping noise. The first row represents the relative change for an underparameterized model (60 parameters), while the second for an overparameterized one (150 parameters). The average is computed over all the inputs of the training set (noisy sinusoidal dataset) and 5 different parameter vectors.

from this changing behavior we can estimate the noise level p^* inducing the strongest equalization of variational parameters. This is studied for depolarizing, dephasing and amplitude-damping noise. Once we have the optimal level p^*, we train the quantum model for different noise values and show that the equalization of parameters can induce a regularizing effect, effectively reducing overfitting. Eventually, we analogously try to find a good noisy regime by means of the generalization bound given in Eq. (E1).

As a use case, we select a regression task with a noisy sinusoidal dataset. It is built by drawing with uniform probability 50 data samples x, which is then divided into 30% training samples and 70% test samples. The analytical expression describing the labels that we assign to these points is the following:

$$y = \sin(\pi x) + \epsilon \,, \qquad (18)$$

where $x \in [-1, 1]$ and ϵ is an additive white Gaussian noise with amplitude equal to 0.4, zero mean and a standard deviation of 0.5. In addition, we also verify the applicability of our method for the diabetes dataset analysed in Ref. [68] and another two-dimensional regression dataset. In the Appendix we provide additional details about the datasets (Appendix F), quantum models (Ap-

pendix G) and additional experiments and analysis (Appendices H, I and J).

A. QFIM suppression

As introduced in Sec. II A the QFIM is a powerful tool to describe how sensitive a parametric quantum model is to small variations of its parameters. It is then reasonable to use this instrument to gain a deeper understanding of how noise affects the action of a QNN, as already done for overparameterized QNNs in Ref. [21]. In particular, we focus our attention on how much the eigenspectrum of the QFIM changes under the action of different kinds of quantum noise with respect to the noiseless case ($p = 0$) for both underparameterized and overparameterized QNNs. This analysis is carried out using our novel spectrally-resolved measure $I_r(p)$, defined in Eq. (14). More in detail, we are interested in seeing the effects on average on the landscapes. For this purpose, we would need the expectation of $I_r(p)$ with respect to both the data and parameters distribution. To approximate this, $\mathbb{E}_{x,\theta}$ is calculated as the average over training samples and 5 parameters initializations. One could also retain some data samples and perform this analysis based on a
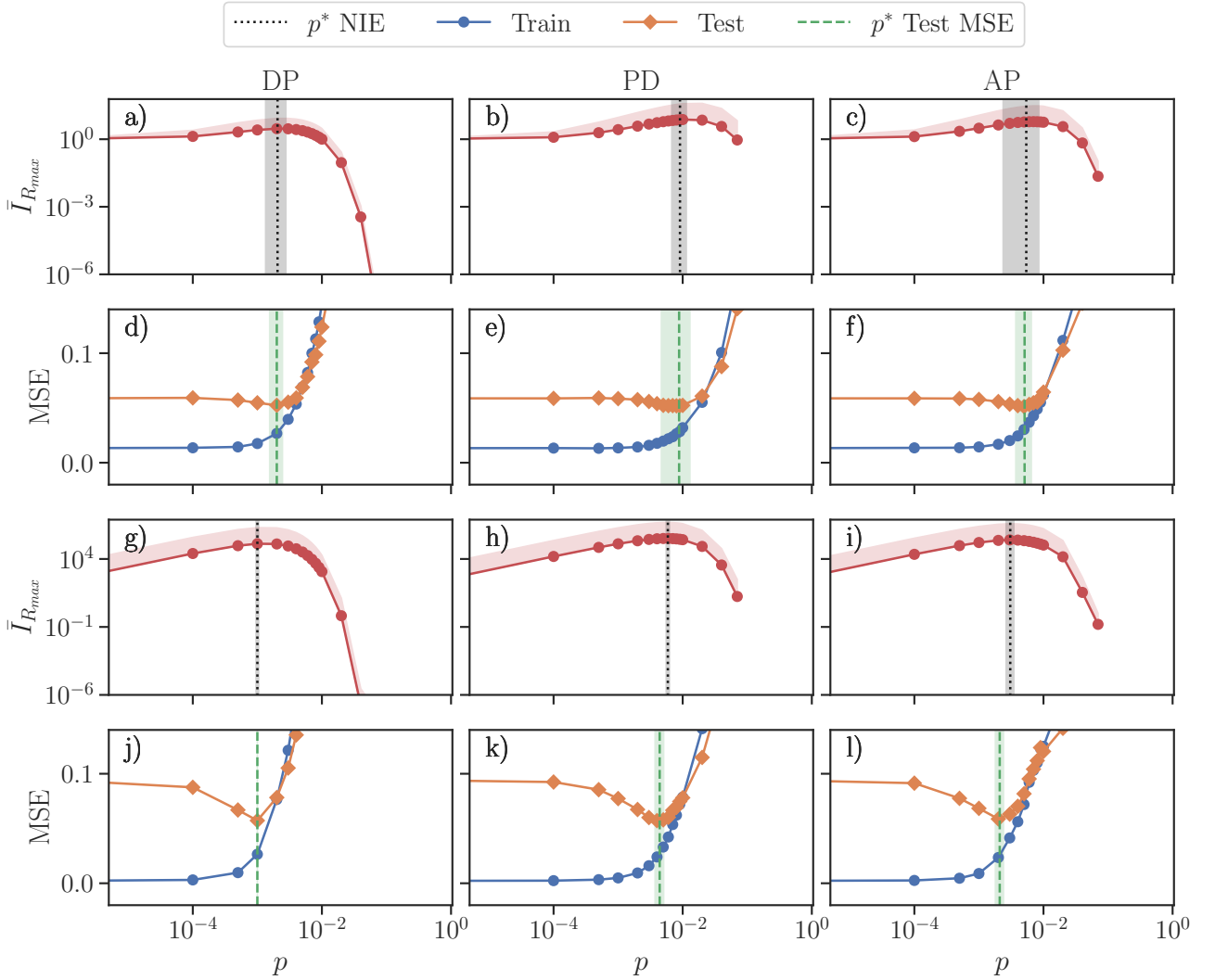
FIG. 4. **Sinusoidal** Comparison of optimal noise level p^* obtained from NIE and test MSE on noisy sinusoidal dataset. Columns correspond to different types of noise: depolarizing (DP), phase damping (PD) and amplitude damping (AD). First and second rows refer to the underparameterized model, while third and fourth to the overparameterized one. We can see that p^* estimated with the two methods (vertical lines) are in good agreement.

validation set.

In Fig. 3, we show the relative change of the QFIM eigenvalues λ_m under different levels of noise p for depolarizing, dephasing and amplitude-damping noise for both underparameterized and overparameterized models of $n = 5$ qubits with maximal expressivity (i.e. over-parametrization threshold at $2^{5+1} - 2 = 62$ [27]). The dataset is a noisy sinuoidal and the circuit is a Hardware Efficient Ansatz (HEA) [70] (see Appendix F and Appendix G for more details). The eigenvalues are indexed in increasing order. In all these plots, we can notice that there is a first phase where some of the least relevant eigenvalues are increased by the presence of noise, while eigenvalues with higher index m either decrease or stay stable. Since the QFIM eigenvalues are associated with a linearly independent direction in the state space, this means that parameters (in the diagonalized framework)

with a growing eigenvalue are gaining importance in the computation with respect to the noiseless case. After a certain threshold value p^* then the whole eigenspectrum is suppressed exponentially, as already shown in Ref. [21]. Here we want to stress that the increasing importance of the least relevant variational parameters takes place not only in overparameterized QNNs, but is a more general phenomenon occurring also in underparameterized models. This opens the question of whether such a phenomenon could also happen in other settings/systems outside QML and what the consequences could be.

B. Noise regularization

We now proceed with the verification of our conjecture, i.e. the critical point p^*, where the least important pa-

	DP	PD	AD
NIE	$(2.1 \pm 0.7)10^{-3}$	$(8.9 \pm 2.5)10^{-3}$	$(5.5 \pm 3.0)10^{-3}$
Test MSE	$(2.0 \pm 0.4)10^{-3}$	$(8.8 \pm 4.1)10^{-3}$	$(5.1 \pm 1.4)10^{-3}$
$B(p)$	$(7.7 \pm 0.5)10^{-3}$	$(2.3 \pm 0.7)10^{-2}$	$(1.2 \pm 0.4)10^{-2}$
Gen. gap	$(5.5 \pm 1.0)10^{-3}$	$(2.2 \pm 0.6)10^{-2}$	$(1.24 \pm 0.5)10^{-2}$

TABLE I. **Sinusoidal** Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) for sinusoidal dataset estimated with different methods for underparameterized QNN.

	DP	PD	AD
NIE	$(1.0 \pm 0.0)10^{-3}$	$(5.9 \pm 0.3)10^{-3}$	$(3.1 \pm 0.4)10^{-3}$
Test MSE	$(1.0 \pm 0.0)10^{-3}$	$(4.4 \pm 0.7)10^{-3}$	$(2.1 \pm 0.3)10^{-3}$
$B(p)$	$(1.8 \pm 0.4)10^{-3}$	$(1.0 \pm 0.01)10^{-2}$	$(6.0 \pm 0.6)10^{-3}$
Gen. gap	$(2.0 \pm 0.0)10^{-3}$	$(9.2 \pm 0.6)10^{-3}$	$(8.7 \pm 1.6)10^{-3}$

TABLE II. **Sinusoidal** Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) for sinusoidal dataset estimated with different methods for overparameterized QNN.

rameters have the maximal relative increase compared to the noiseless case, is also the noise level determining desirable generalization performances achievable by applying quantum noise only. The noise level p^* is determined by averaging the noise values leading to the maximal gain per each eigenvalue that can acquire importance as defined in Eq. (17). At this point, we need to specify what we mean by *desirable generalization properties*, as there is no simple unique measure to evaluate it. In fact, one may consider desirable either the situation where the test error is minimal or the one where the generalization gap (the difference between training and test error) closes. While looking at these situations, one still has to take into account the value of the error on the training set, if this becomes too large, a low test error (compared to the training one) or a closing gap might be related to underfitting the data, which is not beneficial in the end. For this reason, we compare p^* obtained by the NIE analysis with both the noise level leading to the minimal test error, measured in terms of Mean Squared Error (MSE) (see Sec. VI), and the one associated with a closing generalization gap. The comparison of these estimates for different quantum noise channels and QNN depth is presented both in Fig. 4 for a visual understanding and in Tabs. I-II for a more compact and precise assessment. We would like to remark that we do not make use of any technique that might induce regularization (as stochastic gradient descent, L2 regularization or shot noise) other than quantum noise in order to identify its influence.

More in detail, Figs. 4a-f display results for an underparameterized QNN, while Figs. 4g-l gather the same information for the same model in the overparameterized regime. As anticipated, in the first and third row of Fig. 4 we show how $\bar{I}_{R_{max}}(p)$ changes upon variation of the noise level. In particular, here we plot the average trend over different QFIM matrices (five random parameter vectors per training data), and the shade represents one standard deviation. To have a better estimate, the optimal level of p^* is computed individually on different QFIMs, averaging over different inputs and variational parameters as described in Eq. (17). The vertical dotted line highlights our estimate of the optimal level of noise p^*, while the shaded band represent one standard deviation.

Similarly, in the second and fourth row, we show the value of the final MSE on training and test data, averaged over 10 different initializations, together with the estimate of the best noise level according to such values. Specifically, p^* is determined for the single runs and then averaged. Different columns in Fig. 4 report results for different kinds of quantum noise, depolarizing, phase and amplitude damping, respectively.

We can notice a good agreement between the two different estimates in all the configurations. This confirms that the NIE-based procedure allows an estimation of the optimal noise level p^* with a neighbourhood corresponding to a dip in the test MSE. In particular, our approximation of p^* in most cases exactly coincides with what is found to be the level of noise inducing the minimal test MSE (see Tab. I and Tab. II). When the two levels do not exactly match, this could be due to multiple factors, such as the finite number of test samples and initializations with which we evaluate the test MSE, the finite number of training samples and parameters configurations used to compute the average eigenspectrum and the discrete set of noise levels studied. All these factors contribute to the size of the error bar, allowing for compatibility between the approximations. We note that for the overparameterized QNN, the error in the NIE-based determination of p^* is much smaller with respect to the underparameterized case. This is strictly due to the noise enabling new directions to be explored, which happens irrespective of the different training samples or parameters configurations.

We also investigate the feasibility of determining p^* using the generalization bound given in Appendix E. In particular, we focus on a single term in the bound $B(p)$, which is the only noise-dependent term. This approach proves viable since $B(p)$ (and hence the bound) exhibits a minimum (see numerical exoeriments in Appendix J) for nontrivial values of p whose value is reported in Tabs. I-II. Nonetheless, several limitations arise due to the quantities required for computing the bound. First, the determinant becomes problematic in models with a large number of parameters, as many parameters are associated with eigenvalues smaller than 1. This situation yields a determinant that is nearly zero, causing the bound to diverge to infinity and violating one of the theorem's assumptions ($\sqrt{\det(\mathcal{F}(\theta))} \geq m > 0$). In contrast, when the determinant remains significantly different from zero, as observed in QNNs with fewer parameters, increasing noise induces barren plateaus that suppress the gradient

of the model. This suppression results in the Lipschitz constant L_f approaching zero, thereby biasing the minimum toward $p = 1$.

Furthermore, generalization bounds assess the gap between training and optimal performance, where a closed gap is ideally indicative of optimal generalization. However, the gap may also narrow as a result of deteriorating training loss, rather than an improvement in generalization performance. By looking at Tabs. I-II it is possible to see a fair agreement between the values of p^* estimated from the closing generalization gap and the one from generalization bound. Unfortunately, most of the times this happens for values of training MSE that are way worse than the initial ones. These inherent limitations become particularly evident when analyzing the diabetes dataset in Appendix H.

Since the proposed approach relies solely on a subset of the eigenspectrum, it circumvents these issues while requiring fewer computational resources, rendering it suitable for a wide range of QNN architectures.

V. DISCUSSION

In this work, we have investigated the impact of quantum noise on the generalization properties of QNNs. In particular, we have shown a correlation between the improvement of generalization performance of a QNN and the noise-induced equalization (NIE) effect in the eigenspectrum of the Quantum Fisher Information Matrix (QFIM), whereby the least important parameters gain relevance while the most relevant ones lose it. This result can be intuitively explained by combining previous results on quantum noise and overparametrization [21], conditioning of neural networks [48] and the relationship between wide minima and generalization [49]. Since the noise level inducing the best equalization can be deduced based on the previously mentioned QFIM analysis, we propose this as an effective protocol allowing to determine the noise level that leads to enhanced exploration and consequently desirable generalization performance for the given QNN model.

More in detail, we have numerically showcased the NIE effect by introducing the spectrally-resolved measure $I_r(p)$, which quantifies the relative change of importance in the parameters as a function of noise. Then, we identified an optimal noise level, p^*, obtained by averaging the noise values leading to the maximal increase of importance ($\arg\max_p I_r(p)$) per each eigenvalue that can acquire importance ($r \leq R_{max}$), via Eq. (17). Remarkably, p^* is, in most of the analyzed cases, compatible with the noise level that yields the most beneficial generalization performance, reinforcing the idea that noise can play a constructive role in improving generalization. This method has a significant advantage over other common regularization techniques, as it allows for the optimal noise level (i.e., p^*) to be determined without extensive hyperparameter grid search requiring repetitions

of the training procedure. Furthermore, our results are in agreement with the noise levels experimentally determined in Ref. [68] for the diabetes dataset, suggesting that the NIE effect is not a merely theoretical curiosity, but it could actually be observed in practical implementations of QML models.

A comparison with existing generalization bounds highlights the limitations of previous theoretical results. Specifically, Refs. [58, 59] provide bounds that suggest an improvement in generalization with increasing noise. However, these bounds are only applicable in the case of Stochastic Gradient Descent (SGD) optimizer, and they do not provide a method to determine an optimal noise level (p^*). In fact, their predictions become vacuous when noise-induced barren plateaus impede model training. In contrast, the bound introduced in Ref. [57] provides an interesting dependence on the determinant of the QFIM. However, such a dependence is left implicit, while the explicit dependence on the noise level is not discussed. By numerically computing this bound as a function of noise, we observed that it exhibits a minimum, thus suggesting the existence of an optimal noise level. Nevertheless, this approach is hindered by intrinsic limitations related to the determinant of the QFIM, as well as to the effect of noise-induced barren plateaus, which prevent an accurate estimation of p^*. We stress that our procedure only depends on the QNN design, making it extremely versatile and applicable to disparate datasets and optimizers.

Looking forward, several directions remain open for future research. Given the known suppression of the eigenspectrum under noise, an interesting avenue would be to analytically describe the initial growth of the least important eigenvalues. As a first step, in this work, we analytically showed that the enhancement of low eigenvalues is indeed possible for small toy problems. A combination of this growth with the known exponential decay mechanisms [21] could potentially lead to an analytical determination of p^*.

Computational improvements for this pre-training analysis could be achieved in the QFIM calculation by leveraging techniques such as SPSA [71], Stein's identity [72] or classical shadows [73] similarly to what has been done in Ref. [74]. Moreover, it would be interesting to apply a similar investigation leveraging the weighted approximate metric tensor [74] instead of the QFIM. This would take into account additional information coming from the observable possibly leading to more accurate estimate of the optimal noise level p^*, especially for shallower circuits where the locality of the hamiltonian could imply a limited light-cone contribution.

An intrinsic limitation of the introduced technique lies in the fact that we seek better QFIM conditioning on average over the optimization landscape, but then learning is usually initialized at random points that might have different conditioning with respect to the previously analyzed points. A potential enhancement could stem from the combination with meta-learning techniques, similar to the ones employed in Ref. [75].

Finally, a broader perspective concerns the implications of noise-induced equalization beyond the scope of quantum machine learning. In fact, it was recently shown in Ref. [67] that incoherent dynamics can lead to metrological advantages in quantum sensing. This is the same principle at the heart of the NIE. It remains an open question whether similar effects could manifest in other quantum paradigms related to quantum sensing like, for example, quantum thermodynamics. Investigating these aspects in different fields could further enable a deeper understanding of the interplay between noise, optimization, and generalization in quantum models.

VI. METHODS

Here we provide some insights into practical details related to this work, leaving the most technical part in the appendices. Numerical simulations of quantum circuits are performed in Python with Pennylane [76] in combination with JAX [77]. For noisy simulations, we execute circuits in the density matrix formalism, applying a noise channel after each gate, while for noiseless we rely on statevector simulations. The QFIM matrix is derived from quantum circuits by leveraging JAX as well.

For what concerns the discrete set of noise levels studied, different settings have been between sinusoidal and diabetes datasets. In particular, for the sinusoidal dataset the noise levels sampled are:

$$p_{list}^{sin} = \{10^{-6}, 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3},$$
$$3 \cdot 10^{-3}, 4 \cdot 10^{-3}, 5 \cdot 10^{-3}, 6 \cdot 10^{-3},$$
$$7 \cdot 10^{-3}, 8 \cdot 10^{-3}, 9 \cdot 10^{-3}, 10^{-2},$$
$$2 \cdot 10^{-2}, 4 \cdot 10^{-2}, 7 \cdot 10^{-2}\}.$$

While for diabetes dataset we have added additional noise levels that we studied in Ref. [68]:

$$p_{list}^{diab} = p_{list}^{sin} \cup \{10^{-2.75}, 10^{-2.5}, 10^{-2.25}, 10^{-1.75}, 3 \cdot 10^{-2},$$
$$10^{-1.5}, 5 \cdot 10^{-2}, 10^{-1.25}, 8 \cdot 10^{-2}, 9 \cdot 10^{-2},$$
$$10^{-1}, 10^{-0.75}, 10^{-0.5}, 10^{-0.25}, 1\}$$

The analysis of NIE is conducted on multiple QFIM. In particular, per each training data, we compute the QFIM at 5 random points of the parameter space. In Sec. IV B, we propose to compute p^* as the mean of $\arg\max_p I_r(p) \quad \forall r \leq R_{max}$ on single runs. We point out that R_{max} is not unique across different QFIMs and depends on the specific input $\mathbf{x}$ and parameter vector θ. The selection of R_{max} as the minimum over the different runs is done for convenience, as when averaging and computing the standard deviation, the number of eigenvalues taken into account would be the same. One could have also set R_{max} as the average of the different $R_{max}^{(j)}$, or the closest integer to the mean of the various $R_{max}^{(j)}$, anyway, this might lead to include some eigenvalues that

are not increased in all the analyzed QFIMs. In addition, the denominator of $I_r(p)$ is $\max\{10^{-10}, \lambda_r(p_0)\}$ to avoid numerical issues. We invite the interested reader to check out the available code [78] for further technical details.

We perform numerical simulations for training quantum machine learning models using the Adam optimizer [79] with hyperparameters $\eta = 0.01$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. An important difference from Ref. [68] is that we do not employ batch gradient as this is known to have a positive influence on generalization [40], while our goal is to show the genuine regularizing effect of quantum noise. The cost function employed is the Mean Squared Error (MSE), which can be written as:

$$\mathcal{L}(\boldsymbol{\theta}) = \text{MSE}(\mathbf{x}, y) = \frac{1}{M} \sum_{i=1}^{M} (f(\mathbf{x}_i, \boldsymbol{\theta}) - y_i)^2 \qquad (19)$$

where $f(\mathbf{x}_i, \boldsymbol{\theta})$ are the predicted outputs, y_i are the true outputs (labels), and M is the number of samples. We trained the models on 10 different initializations of the parameters, which were chosen to be unrelated to the 5 initializations used in our pre-training analysis of the QFIM. This approach allows for a more general assessment of the optimal level of noise p^* leading to the best regime in terms of generalization. To determine this optimal level, we estimated the mean of the argmin of the final test MSE (FTMSE) determined on single runs:

$$p_j^* = \arg\min_p \text{FTMSE}_j, \qquad (20)$$

while the error on the estimation is computed as the standard deviation.

Ultimately, to estimate the generalization bound given in Eq. (E1), we take the effective dimension d_{eff} as average the rank of the QFIM (as prescribed by Eq. (13)) over training data each evaluated at 5 random points in the parameter space. We approximate the Lipschitz required by the bound constant as the maximum gradient over these same configurations.

CODE AVAILABILITY STATEMENT

Code to reproduce the results and to create all figures and tables presented in this manuscript is available at Github repository [78].

AKNOWLEDGMENTS

AI TOOLS DISCLAIMER

ChatGPT was used to improve the readability of parts of the paper. No new content was created by the AI tool. The authors have checked all texts and take full responsibility for the result.

[1] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).

[2] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).

[3] P. Wittek, *Quantum Machine Learning* (Academic Press, Boston, 2014).

[4] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[5] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nature Computational Science **2**, 567 (2022).

[6] J. Preskill, Beyond nisq: The megaquop machine (2025), arXiv:2502.17368 [quant-ph].

[7] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[8] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[9] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).

[10] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nature Reviews Physics **3**, 625 (2021).

[11] D. S. França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17**, 1221 (2021).

[12] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, Limitations of variational quantum algorithms: A quantum optimal transport approach, PRX Quantum **4**, 010309 (2023).

[13] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat Commun **9**, 10.1038/s41467-018-07090-4 (2018).

[14] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, Quantum Science and Technology **7**, 045015 (2022).

[15] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5** (2014).

[16] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **5**, 019601 (2020).

[17] S. Mangini, F. Tacchino, D. Gerace, D. Bajoni, and C. Macchiavello, Quantum computing models for artificial neural networks, Europhysics Letters **134**, 10002 (2021).

[18] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nat Comput Sci **1**, 403 (2021).

[19] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat Commun **12**, 10.1038/s41467-021-27045-6 (2021).

[20] M. Schumann, F. K. Wilhelm, and A. Ciani, Emergence of noise-induced barren plateaus in arbitrary layered noise models (2023), arXiv:2310.08405 [quant-ph].

[21] D. García-Martín, M. Larocca, and M. Cerezo, Effects of noise on the overparametrization of quantum neural networks, Phys. Rev. Res. **6**, 013295 (2024).

[22] M. Oliv, A. Matic, T. Messerer, and J. M. Lorenz, Evaluating the impact of noise on the performance of the variational quantum eigensolver (2022), arXiv:2209.12803 [quant-ph].

[23] K. Borras, S. Y. Chang, L. Funcke, M. Grossi, T. Hartung, K. Jansen, D. Kruecker, S. Kühn, F. Rehm, C. Tüysüz, and S. Vallecorsa, Impact of quantum noise on the training of quantum generative adversarial networks, Journal of Physics: Conference Series **2438**, 012093 (2023).

[24] A. Gu, A. Lowe, P. A. Dub, P. J. Coles, and A. Arrasmith, Adaptive shot allocation for fast convergence in variational quantum algorithms (2021), arXiv:2108.10434 [quant-ph].

[25] K. Ito, Latency-aware adaptive shot allocation for runtime efficient variational quantum algorithms (2023), arXiv:2302.04422 [quant-ph].

[26] B. T. Kiani, S. Lloyd, and R. Maity, Learning unitaries by gradient descent (2020), arXiv:2001.11897 [quant-ph].

[27] T. Haug, K. Bharti, and M. Kim, Capacity and quantum geometry of parametrized quantum circuits, PRX Quantum **2**, 10.1103/prxquantum.2.040309 (2021).

[28] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, Nat Comput Sci **3**, 542 (2023).

[29] J. Liu, H.-N. Xiong, F. Song, and X. Wang, Fidelity susceptibility and quantum fisher information for density operators with arbitrary ranks, Physica A: Statistical Mechanics and its Applications **410**, 167 (2014).

[30] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, Quantum fisher information matrix and multiparameter estimation, Journal of Physics A: Mathematical and Theoretical **53**, 023001 (2019).

[31] J. J. Meyer, Fisher Information in Noisy Intermediate-Scale Quantum Applications, Quantum **5**, 539 (2021).

[32] L. Pezzè and A. Smerzi, Advances in multiparameter quantum sensing and metrology (2025), arXiv:2502.17396 [quant-ph].

[33] M. Scandi, P. Abiuso, J. Surace, and D. De Santis, Quantum fisher information and its dynamical nature, Reports on Progress in Physics (2023).

[34] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning* (MIT press, 2018).

[35] A. Camuto, M. Willetts, U. Simsekli, S. J. Roberts, and C. C. Holmes, Explicit regularisation in gaussian noise injections, Advances in Neural Information Processing Systems **33**, 16603 (2020).

[36] Y. Li and F. Liu, Adaptive gaussian noise injection regularization for neural networks, in *International Symposium on Neural Networks* (Springer, 2020) pp. 176–189.

[37] N. Levi, I. M. Bloch, M. Freytsis, and T. Volansky, Noise injection node regularization for robust learning, arXiv preprint arXiv:2210.15764 (2022).

[38] A. Orvieto, H. Kersting, F. Proske, F. Bach, and A. Lucchi, Anticorrelated noise injection for improved generalization, in *International Conference on Machine Learning* (PMLR, 2022) pp. 17094–17116.

[39] A. Orvieto, A. Raj, H. Kersting, and F. Bach, Explicit regularization in overparametrized models via noise injection, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2023) pp. 7265–7287.

[40] S. Smith, E. Elsen, and S. De, On the generalization benefit of noise in stochastic gradient descent, in *International Conference on Machine Learning* (PMLR, 2020) pp. 9058–9067.

[41] A. Sclocchi, M. Geiger, and M. Wyart, Dissecting the effects of sgd noise in distinct regimes of deep learning, in *International Conference on Machine Learning* (PMLR, 2023) pp. 30381–30405.

[42] C. M. Bishop, Training with noise is equivalent to tikhonov regularization, Neural computation **7**, 108 (1995).

[43] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, Regularization of neural networks using dropconnect, in *International conference on machine learning* (PMLR, 2013) pp. 1058–1066.

[44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research **15**, 1929 (2014).

[45] D. M. Hawkins, The problem of overfitting, Journal of chemical information and computer sciences **44**, 1 (2004).

[46] X. Ying, An overview of overfitting and its solutions, in *Journal of physics: Conference series*, Vol. 1168 (IOP Publishing, 2019) p. 022022.

[47] L. Banchi, J. Pereira, and S. Pirandola, Generalization in quantum machine learning: A quantum information standpoint, PRX Quantum **2**, 040321 (2021).

[48] J. Pennington and P. Worah, The spectrum of the fisher information matrix of a single-hidden-layer neural network, in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018).

[49] C. Baldassi, C. Lauditi, E. Malatesta, G. Perugini, and R. Zecchina, Unveiling the structure of wide flat minima in neural networks, Physical review letters **127 27**, 278301 (2021).

[50] A. Sannia, F. Tacchino, I. Tavernelli, G. L. Giorgi, and R. Zambrini, Engineered dissipation to mitigate barren plateaus, npj Quantum Information **10**, 81 (2024).

[51] J. Liu, F. Wilde, A. A. Mele, L. Jiang, and J. Eisert, Stochastic noise can be helpful for variational quantum algorithms (2023), arXiv:2210.06723 [quant-ph].

[52] M. Parigi, S. Martina, and F. Caruso, Quantum-noise-driven generative diffusion models, Advanced Quantum Technologies , 2300401 (2024).

[53] M. L. Olivera-Atencio, L. Lamata, and J. Casado-Pascual, Impact of amplitude and phase damping noise on quantum reinforcement learning: Challenges and opportunities (2025), arXiv:2503.24069 [quant-ph].

[54] F. Scala, A. Ceschini, M. Panella, and D. Gerace, A general approach to dropout in quantum neural networks, Advanced Quantum Technologies , 2300220 (2023).

[55] V. Heyraud, Z. Li, Z. Denis, A. Le Boité, and C. Ciuti, Noisy quantum kernel machines, Phys. Rev. A **106**, 052421 (2022).

[56] F. Hu, G. Angelatos, S. A. Khan, M. Vives, E. Türeci, L. Bello, G. E. Rowlands, G. J. Ribeill, and H. E. Türeci, Tackling sampling noise in physical systems for machine learning applications: Fundamental limits and eigentasks, Phys. Rev. X **13**, 041020 (2023).

[57] B. Khanal and P. Rivas, Data-dependent generalization bounds for parameterized quantum models under noise, The Journal of Supercomputing **81**, 611 (2025).

[58] C. Zhu, H. Yao, Y. Liu, and X. Wang, Optimizer-dependent generalization bound for quantum neural networks, arXiv preprint arXiv:2501.16228 (2025).

[59] J. Yang, W. Xie, and X. Xu, Stability and generalization of quantum neural networks, arXiv preprint arXiv:2501.12737 (2025).

[60] F. Albarelli, J. F. Friel, and A. Datta, Evaluating the holevo cramér-rao bound for multiparameter quantum metrology, Physical review letters **123**, 200503 (2019).

[61] A. Abbas, D. Sutter, A. Figalli, and S. Woerner, Effective dimension of machine learning models, arXiv preprint arXiv:2112.04807 (2021).

[62] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, Journal of Mathematical Physics **57** (2016).

[63] R. A. Horn and C. R. Johnson, *Topics in matrix analysis* (Cambridge university press, 1994).

[64] J. Langford and J. Shawe-Taylor, Pac-bayes & margins, Advances in neural information processing systems **15** (2002).

[65] G. K. Dziugaite and D. M. Roy, Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data, arXiv preprint arXiv:1703.11008 (2017).

[66] O. Rivasplata, I. Kuzborskij, C. Szepesvári, and J. Shawe-Taylor, Pac-bayes analysis beyond the usual bounds, Advances in Neural Information Processing Systems **33**, 16833 (2020).

[67] L. P. García-Pintos, Noise-enhanced quantum clocks and global field sensors, arXiv preprint arXiv:2507.02071 (2025).

[68] W. Somogyi, E. Pankovets, V. Kuzmin, and A. Melnikov, Method for noise-induced regularization in quantum neural networks, arXiv preprint arXiv:2410.19921 (2024).

[69] W. F. Stinespring, Positive functions on c-algebras, Proceedings of the American Mathematical Society **6**, 211 (1955).

[70] L. Leone, S. F. Oliviero, L. Cincio, and M. Cerezo, On the practical usefulness of the hardware efficient ansatz, Quantum **8**, 1395 (2024).

[71] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information, Quantum **5**, 567 (2021).

[72] M. Halla, Estimation of quantum fisher information via stein's identity in variational quantum algorithms (2025), arXiv:2502.17231 [quant-ph].

[73] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050–1057 (2020).

[74] C. Shi, V. Dunjko, and H. Wang, Weighted approximate quantum natural gradient for variational quantum eigensolver (2025), arXiv:2504.04932 [quant-ph].

[75] M. Ait Haddou and M. Bennai, Sculpting quantum landscapes: Fubini-study metric conditioning for geometry-aware learning in parameterized quantum circuits, Research Square preprint 10.21203/rs.3.rs-7101280/v1 (2025).

[76] V. Bergholm, J. Izaac, M. Schuld, *et al.*, Pennylane: Automatic differentiation of hybrid quantum-classical computations (2018).

[77] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python+NumPy programs (2018).

[78] F. Scala, Improving Quantum Neural Networks exploration by Noise-Induced Equalization, https://github.com/fran-scala/public-noise-induced (2025).

[79] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[80] P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak, and I. Sutskever, Deep double descent: Where bigger models and more data hurt, Journal of Statistical Mechanics: Theory and Experiment **2021**, 124003 (2021).

[81] Y. Du, Y. Yang, D. Tao, and M.-H. Hsieh, Problem-dependent power of quantum neural networks on multiclass classification, Physical Review Letters **131**, 140601 (2023).

[82] M. Kempkes, A. Ijaz, E. Gil-Fuster, C. Bravo-Prieto, J. Spiegelberg, E. van Nieuwenburg, and V. Dunjko, Double descent in quantum machine learning (2025), arXiv:2501.10077 [quant-ph].

[83] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, Understanding deep learning (still) requires rethinking generalization, Communications of the ACM **64**, 107 (2021).

[84] G. Dirr, U. Helmke, I. Kurniawan, and T. Schulte-Herbrüggen, Lie-semigroup structures for reachability and control of open quantum systems: Kossakowski-lindblad generators form lie wedge to markovian channels, Reports on Mathematical Physics **64**, 93 (2009).

[85] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, Journal of Mathematical Physics **17**, 821 (1976).

[86] G. Lindblad, On the generators of quantum dynamical semigroups, Communications in mathematical physics **48**, 119 (1976).

[87] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research **12**, 2825 (2011).

Appendix A: Generalization and overfitting

In this appendix, we provide a concise introduction to generalization, the overfitting problem, and theoretical approaches to addressing these issues. This section is not meant to be exhaustive, but rather intended as a gentle outline for readers new to the subject.

In supervised machine learning, the goal is to learn a function that maps input data $\mathbf{x}_i \in \mathbb{R}^m$ to corresponding outputs y_i, by minimizing a loss function over a training dataset. However, evaluating the model solely on the training data is not sufficient: what ultimately matters is the model's performance on unseen data drawn from the same underlying distribution $\mathcal{D}$. This ability is referred in the field as *generalization*, a crucial property for machine learning because in real-world applications, models will be exposed to new data not seen before [34]. This motivates the division of data into a *training set* and a *test set*, where the training set is used to optimize the model parameters, and the test set serves as a proxy to estimate the so-called *true risk*, defined as the expected loss over the entire distribution:

$$R(f) = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}} \left[\mathcal{L}(f(\mathbf{x}), y) \right]. \tag{A1}$$

Since $\mathcal{D}$ is unknown, we estimate $R(f)$ using the *empirical risk* on a finite sample:

$$R_S(f) = \frac{1}{M} \sum_{i=1}^{M} \mathcal{L}(f(\mathbf{x}_i), y_i), \tag{A2}$$

with $(\mathbf{x}_i, y_i)$ belonging either to the training set or the test set, depending on the context.

In the case of Mean Squared Error (MSE) loss, we can better understand generalization via a theoretical framework called *bias-variance decomposition*. As a matter of fact, a central issue in machine learning is to reach a tradeoff between *bias*, accounting for limitation of the learning algorithm, and *variance*, accounting for sensitivity to fluctuations in the training data. For regression tasks, the expected squared error at a test point $\mathbf{x}$ can be decomposed as:

$$\mathbb{E}\left[(f(\mathbf{x}) - y)^2\right] = \underbrace{(\mathbb{E}[f(\mathbf{x})] - y)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[(f(\mathbf{x}) - \mathbb{E}[f(\mathbf{x})])^2\right]}_{\text{Variance}} + \underbrace{\sigma^2}_{\text{Irreducible noise}}, \tag{A3}$$

where the expectations are taken over the randomness in the training set. A highly complex model typically exhibits low bias but high variance, meaning it can closely fit the training data but may perform poorly on unseen data—a phenomenon known as *overfitting* [45, 46].

Overfitting occurs when a model fits the training data too closely, including its noise or spurious patterns, rather than capturing the underlying structure of the data distribution. Expressive models are particularly susceptible to this. Model expressiveness increases with the number of parameters and the nonlinearity of the function class. For instance, in quantum machine learning, parameterized quantum circuits with many layers, entangling gates and data re-uploading may become increasingly expressive and prone to overfitting.

Classical machine learning model's capacity was studied in terms of the *interpolation threshold*. It refers to the regime in which the model has enough parameters to perfectly fit (i.e., interpolate) the training data, i.e. $P = M$ where P is the number of parameters in the model and M is the training set size. Classical results suggest that generalization should degrade when $P > M$, but recent empirical and theoretical developments have shown that models can generalize well even beyond this threshold, a phenomenon called *double-descent* [80]. Nonetheless, for classical machine learning, the interpolation threshold is often associated with the onset of overfitting, particularly when the dataset is small or noisy. For what concerns quantum models, quantum neural networks were shown to be unable to reach double-descent [81] while recently Kempkes et al. [82] demonstrated that this is achievable in quantum kernel methods. For this reason, overfitting remains a longstanding challenge in QML, with ongoing research focusing on developing novel methods to mitigate or prevent it.

To theoretically approach the model's performances on unseen data, one can make use of *generalization bounds*, which provide probabilistic guarantees on how close $R_S(f)$ is to $R(f)$ for a given model (hypotesis) class $\mathcal{F}$ formalizing the relationship between empirical and true risk. A typical form of such bounds is:

$$R(f) \leq R_S(f) + \mathcal{C}(\mathcal{F}, M, \delta), \tag{A4}$$

which holds with high probability $1 - \delta$, where $\mathcal{C}$ is a complexity term that depends on the richness of $\mathcal{F}$, the number of training examples M, and the confidence level δ. It is worth highlighting that the function $\mathcal{C}$ approaches 0 as the number of samples M tends to infinity.

One way to quantify the complexity of a hypothesis class is through the *Rademacher complexity*, which measures how well functions in $\mathcal{F}$ can fit random noise [34]. Given a sample $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M}$, the empirical Rademacher complexity is defined as:

$$\hat{\mathfrak{R}}_S(\mathcal{F}) = \mathbb{E}_{\boldsymbol{\sigma}} \left[\sup_{f \in \mathcal{F}} \frac{1}{M} \sum_{i=1}^{M} \sigma_i f(\mathbf{x}_i) \right], \tag{A5}$$

where $\sigma_i \in \{-1, +1\}$ are independent Rademacher variables. Intuitively, a high Rademacher complexity indicates a model class capable of fitting arbitrary labels, suggesting a high risk of overfitting. We point out that the bound in Eq. (E1) belongs to the Rademacher complexity type.

In both classical and quantum settings, bounding the generalization gap via Rademacher complexity or related tools (such as VC dimension or covering numbers [34]) is common practice for designing architectures that generalize well, even if this bounds have been shown to be somewhat vacuous [83]. In this work, we study the effect of quantum noise on generalization performances and refer to the generalization bound given in Ref. [57], showing that, in contrast with our apporoach, this does not allow to accurately estimate a good noisy operating regime.

Appendix B: Noise channels

A useful tool allowing us to describe the evolution of quantum mechanical systems is the quantum operation (channel) formalism [2], which is particularly effective for characterizing quantum noise sources. Here, we summarize the essential features of depolarizing, phase damping, and amplitude damping noise, which are the key types of quantum noise channels employed in the following of this work.

Depolarizing noise arises from random unitary rotations on the quantum state. This type of noise tends to isotropically reduce the coherence of the quantum state, effectively spreading the information uniformly across all possible outcomes, and reducing the original state to the completely mixed state, i.e., $\rho_{\mathrm{mix}} = \mathbb{1}/2^d$. For a single qubit, this type of quantum noise can be mathematically described by the depolarizing channel:

$$\mathcal{N}_{depol}(\rho, p) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$$
$$= p\frac{\mathbb{1}}{2} + (1-p)\rho, \tag{B1}$$

which is quantitatively characterized by the probability that a depolarizing event occurs, p. On the Bloch sphere, depolarizing noise can be visualized as the isotropic shrinking of the sphere towards the origin.

Phase damping (or dephasing) noise, on the other hand, is associated with the random introduction of phase errors in the quantum state. Unlike depolarizing noise, phase damping preserves the amplitude information but disrupts the relative phase relationships between different components of the quantum state. The dephasing channel, describing the action of the homonym noise, can be represented with a Kraus notation for a single qubit as:

$$\mathcal{N}_{deph}(\rho, p) = \sum_i E_{i,deph} \rho E_{i,deph}^\dagger \quad , \quad \text{where} \tag{B2}$$

$$E_{0,deph} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \quad , \quad E_{1,deph} = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{bmatrix} ,$$

with the degree of dephasing determined by the probability parameter, p. In Eq. (B2), E_0 leaves the state $|0\rangle$ unchanged, but it reduces the amplitude of state $|1\rangle$, while E_1 destroys $|0\rangle$ and reduces the amplitude of state $|1\rangle$. Dephasing noise can be visualized as shrinking the Bloch sphere into an ellipsoid, in which the z-axis is left unchanged while the other two axes are contracted.

Amplitude noise, also defined amplitude damping channel, represents a different facet of quantum noise. This noise source involves the loss of amplitude information, leading to the decay of the quantum state. The amplitude-damping channel is particularly relevant in describing processes in which the quantum system interacts with its environment, causing the loss of energy and coherence. For a single qubit, it can be described by the following Kraus decomposition:

$$\mathcal{N}_{amp}(\rho, p) = \sum_i E_{i,amp} \rho E_{i,amp}^\dagger \quad , \quad \text{where} \tag{B3}$$

$$E_{0,amp} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \quad , \quad E_{1,amp} = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix} .$$

From Eq. (B3) it is possible to understand why amplitude noise is associated to energy loss: E_1 in Eq. (B3) turns the state $|0\rangle$ into $|1\rangle$, which corresponds to the process of losing energy due to the interaction with an environment; E_0 leaves the state $|0\rangle$ unchanged, but it reduces the amplitude of state $|1\rangle$, as for the dephasing noise.

In this work, we will consider noisy quantum learning models where noise acts after each quantum gate, be it a single qubit gate or an entangling gate (in this latter case, two single qubit noise channels act on the considered qubits). As all these noise channels depend on the parameter p, we will refer to it as the noise level, thus assuming all the noise channels to be characterized by the same parameter p, with obvious generalization.

Appendix C: Growth of the Dynamical Lie Algebra in Noisy Quantum Circuits

In this appendix, we analyze how quantum noise can affect the dynamics of a quantum system through the Dynamical Lie Algebra (DLA) associated with the quantum circuit of interest and potentially lead to the noise-induced equalization (NIE). Different effects may occur depending on whether there is commutation or not between Hamiltonian dynamics and noise-induced dissipative generators and they can be rigorously understood by studying the evolution under the Lindblad master equation and applying the Baker-Campbell-Hausdorff (BCH) expansion [84]. We also discuss how the dimension of the DLA relates to the accessible directions in the system's evolution.

The Dynamical Lie Algebra in the Noiseless Case

In the absence of noise, a closed quantum system evolves under the Schrödinger equation. When the evolution is driven by a finite set of time-independent Hamiltonians $\{H_j\}$, the time evolution operators (quantum gates) take the form

$$U_j = e^{-iH_j t_j}, \quad j = 1, 2, \ldots \tag{C1}$$

The Dynamical Lie Algebra $\mathfrak{g}$ is defined as the smallest Lie algebra closed under commutators and containing the skew-Hermitian generators iH_j. It determines the set of effective Hamiltonians that can be synthesized through combinations of the available gates. If the generators iH_j do not commute, then their products generate additional directions in $\mathfrak{g}$ through the BCH formula:

$$e^A e^B = e^{A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] - \frac{1}{12}[B,[A,B]] + \cdots} \quad . \tag{C2}$$

The nested commutators imply that the reachable set of unitaries expands beyond the span of the original Hamiltonians, depending on the algebraic structure of their commutators. The *dimension of the DLA* corresponds to the number of linearly independent skew-Hermitian operators generated from $\{iH_j\}$ and their nested commutators. Each independent direction in this algebra represents a possible trajectory in the system's unitary evolution space. For a n-qubit system, the maximum DLA is $\mathfrak{su}(2^n)$, which has dimension $4^n - 1$; a lower-dimensional DLA means limited controllability and expressiveness.

In the unitary case, the evolution of a quantum state is restricted to the unitary orbit of the initial pure state:

$$\rho(t) = U(t)\rho(0)U^\dagger(t) \quad , \tag{C3}$$

where $U = \prod_j U_j$. This evolution preserves the eigenvalues of ρ, so the reachable set is confined to a lower-dimensional manifold. The number of independent real parameters in a pure state (modulo global phase) is:

$$\dim_\mathbb{R}(\mathbb{CP}^{2^n-1}) = 2^{n+1} - 2. \tag{C4}$$

This is far smaller than the DLA dimension $4^n - 1$, highlighting that the DLA describes the possible dynamics, not the static configuration space.

Open-System Evolution: The Lindblad Master Equation

When the system interacts with an environment, the dynamics are no longer unitary. Instead, the time evolution of the density matrix ρ is governed by the Lindblad master equation [85, 86]:

$$\frac{d\rho}{dt} = -i[H, \rho] + \sum_k \left(L_k \rho L_k^\dagger - \frac{1}{2}\left\{ L_k^\dagger L_k, \rho \right\} \right) \quad . \tag{C5}$$

Here, H is the system Hamiltonian, and the Lindblad operators L_k describe the dissipative interaction with the environment (e.g., depolarizing, dephasing, amplitude damping). The solution of this equation for time-independent generators is given by:

$$\rho(t) = e^{\mathcal{L}t}[\rho(0)] \tag{C6}$$

where $\mathcal{L}$ is the Liouvillian superoperator, which acts linearly on the space of density matrices:

$$\mathcal{L}[\rho] = \mathcal{L}_H + \mathcal{L}_\mathcal{D} = -i[H, \rho] + \sum_k \left(L_k \rho L_k^\dagger - \frac{1}{2}\left\{ L_k^\dagger L_k, \rho \right\} \right) \tag{C7}$$

The Liouvillian defines a semigroup of completely positive, trace-preserving maps. While the dynamics are no longer represented by Lie groups of unitaries, the structure of $\mathcal{L}$ still allows for algebraic analysis via Lindbladian algebras, which extend the concept of DLAs to open systems. Unlike unitary evolution, the Lindbladian can change the eigenvalues of ρ, enabling transitions from pure to mixed states and expanding the reachable set of states beyond unitary orbits [84]. As a matter of fact, the space $\mathfrak{D}_{2^n}$ of all density matrices (trace-one, positive semidefinite $2^n \times 2^n$ matrices) has real dimension

$$\dim_\mathbb{R}(\mathfrak{D}_{2^n}) = 4^n - 1 \quad , \tag{C8}$$

matching that of $\mathfrak{su}(2^n)$, including both pure and mixed states.

Dynamical Lie Algebra Growth Induced by Noise

In the noisy setting, new generators emerge from the dissipative Lindblad terms. When the noise-induced generators $\mathcal{L}_\mathcal{D}$ do not commute with the original Hamiltonian generators $\mathcal{L}_H$, the algebra of effective dynamical generators expands through their nested commutators. This is analogous to the noiseless case, where consecutive quantum gates correspond to a product of exponentials of non-commuting generators and are combined via the Baker-Campbell-Hausdorff (BCH) formula. Specifically, if we consider two noisy gates modeled by superoperators $\mathcal{L}_1$ and $\mathcal{L}_2$, their consecutive application corresponds to:

$$e^{\mathcal{L}_1 t_1} e^{\mathcal{L}_2 t_2} = \exp\left(t_1 \mathcal{L}_1 + t_2 \mathcal{L}_2 + \frac{t_1 t_2}{2}[\mathcal{L}_1, \mathcal{L}_2] + \cdots \right), \tag{C9}$$

where we used the BCH expansion. The DLA is then generated by both the Hamiltonian part and the dissipative part of the Liouvillian:

$$\mathfrak{g}_{\text{noisy}} = \text{Lie}\left(\{iH_j\} \cup \{\mathcal{D}_k\}\right) \tag{C10}$$

where $\mathcal{D}_k$ denotes the superoperators associated with the dissipators:

$$\mathcal{D}_k[\rho] := L_k \rho L_k^\dagger - \frac{1}{2}\left\{L_k^\dagger L_k, \rho\right\} \tag{C11}$$

At this stage, two cases may arise: the generators of the unitary dynamics either commute with the noise superoperators or they do not. Commutation occurs, for instance, when the jump operators are eigenoperators of the Hamiltonian's adjoint action, that is, $[H, L_k] = \omega_k L_k$, $\omega_k \in \mathbb{R}$, which includes the special case $[H, L_k] = 0$ for all k. In the next section, we will show analytically tractable toy examples where, in the case of $[H, L] = 0$, NIE takes place. On the other hand, the non-commutativity between Hamiltonian and noise superoperators allows for additional directions to emerge through commutators such as $[\mathcal{L}_H, \mathcal{D}_k]$, $[\mathcal{D}_k, \mathcal{D}_l]$, etc. This leads to a growth of the DLA, enriching the space of reachable operations. The generators now include non-Hermitian and non-unitary elements, acting on the space of operators. The effective DLA becomes a subset of the space of superoperators on Hermitian matrices, which has real dimension:

$$\dim_{\mathbb{R}}(\text{End}(\mathfrak{su}(2^n))) = (4^n - 1)^2 \quad, \tag{C12}$$

where the notation $\text{End}(\mathfrak{su}(2^n))$ stands for the space of endomorphisms, i.e. superoperators that map traceless skew-Hermitian operators to other such operators.

As in the unitary case, the dimension of the noisy DLA reflects the number of independent directions in the Liouvillian evolution space. A higher-dimensional DLA implies that the system can explore a larger portion of the operator or state space, potentially enabling more complex transformations, even in the presence of decoherence. In some cases, noise can paradoxically *enhance controllability*, allowing the system to reach dynamical regimes that would not be accessible with Hamiltonian evolution alone. We argue that the noise-induced equalization stems from the balancing of the dissipative dynamics and the increased controllability provided by quantum noise.

Appendix D: Analytical toy demonstrations of NIE

In this appendix, leveraging the insights from Ref. [67], we provide two illustrative examples that demonstrate the onset of noise-induced equalization (NIE). The derivation relies on the assumption that the generators of the unitary dynamics commute with the noise superoperators. The first example investigates a system with a single tunable parameter, revealing the existence of an optimal level of noise for the NIE. Furthermore, we compute the eigenvalues of the QFIM for systems with multiple variational parameters analytically uncovering the equalization process.

1. State Evolution under Noise: $[H, L] = 0$

We start by analyzing the time evolution of a quantum state subject to decoherence in the case where the system Hamiltonian H and the Lindblad operator L commute, i.e. $[H, L] = 0$. This commutation relation ensures that H and L share a common eigenbasis, which we denote by $\{|E_0\rangle, |E_1\rangle\}$, with corresponding eigenvalues E_j and ℓ_j

$$H|E_j\rangle = E_j|E_j\rangle, \quad L|E_j\rangle = \ell_j|E_j\rangle. \tag{D1}$$

We focus on an initial state given by the nontrivial superposition of $|E_0\rangle$, $|E_1\rangle$, a two-level state, whose density operator is

$$\rho(0) = |\psi(0)\rangle\langle\psi(0)| = \tfrac{1}{2}\big(|E_0\rangle\langle E_0| + |E_1\rangle\langle E_1| + |E_0\rangle\langle E_1| + |E_1\rangle\langle E_0|\big). \tag{D2}$$

Under unitary evolution $U(t) = e^{-iHt}$ and pure dephasing via L at rate γ, the state at time t becomes

$$\rho(t) = e^{\mathcal{L}t}\rho(0) = e^{(-i[H,\cdot] - \frac{1}{2}\gamma\{L^\dagger L, \cdot\} + \gamma L \cdot L^\dagger)t}\rho(0) = \tag{D3}$$

$$= \tfrac{1}{2}\Big(|E_0\rangle\langle E_0| + |E_1\rangle\langle E_1| + e^{-i\Delta E\, t}e^{-\Gamma(t)}|E_0\rangle\langle E_1| + e^{i\Delta E\, t}e^{-\Gamma(t)}|E_1\rangle\langle E_0|\Big), \tag{D4}$$

where $\Delta E = E_1 - E_0$, and $\Gamma(t) = \frac{\gamma t}{2}(\ell_1 - \ell_0)^2$.

We now analyze the spectral properties of the time-evolved density matrix $\rho(t)$. Its normalized eigenstates can be written as

$$|\phi_\pm(t)\rangle = \frac{1}{\sqrt{2}}\left(e^{-i\frac{\Delta E}{2}t}|E_0\rangle \pm e^{i\frac{\Delta E}{2}t}|E_1\rangle\right), \tag{D5}$$

which represent, respectively, the in-phase $(+)$ and out-of-phase $(-)$ superpositions of the energy eigenstates, each evolving with opposite phase factors due to the energy splitting ΔE. With this basis, one finds that $\rho(t)$ has two nonzero eigenvalues given by

$$\lambda_\pm(t) = \tfrac{1}{2}\big(1 \pm |r(t)|\big) = \tfrac{1}{2}\big(1 \pm e^{-\Gamma(t)}\big), \tag{D6}$$

where, for convenience, we introduced the complex decoherence factor $r(t) = e^{-i\Delta E t}e^{-\Gamma(t)}$, encoding both the unitary phase evolution due to the energy difference $\Delta E = E_1 - E_0$ and the exponential damping induced by the dephasing rate $\Gamma(t)$.

To investigate the parameter dependence of these quantities, we compute their derivatives, as these will be needed for deriving the Quantum Fisher Information (QFI). Differentiating the eigenvalues with respect to time yields

$$\partial_t \lambda_\pm = \pm\tfrac{1}{2}(-\Gamma')e^{-\Gamma} = \mp\tfrac{1}{2}\left(\tfrac{\gamma}{2}(\ell_1 - \ell_0)^2\right)e^{-\Gamma(t)},$$

where $\Gamma' = d\Gamma/dt = \frac{\gamma}{2}(\ell_1 - \ell_0)^2$. Similarly, differentiating the eigenstates gives

$$\partial_t |\phi_\pm(t)\rangle = \frac{\pm i\Delta E}{2}|\phi_\mp(t)\rangle,$$

showing that the instantaneous rate of change of each eigenstate is proportional to the energy splitting and points along the orthogonal superposition.

The QFI for single parameter t is of a generic density matrix $\rho = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k|$ is given by the following expression

$$\mathcal{F}(t) = \sum_{\substack{k \\ \lambda_k \neq 0}} \left[\frac{\partial_t \lambda_k}{\lambda_k} + 4\lambda_k \langle \partial_t \psi_k | \partial_t \psi_k \rangle\right] - \sum_{\substack{k,l \\ \lambda_k, \lambda_l \neq 0}} \frac{8\lambda_k \lambda_l}{\lambda_k + \lambda_l}|\langle\psi_k|\partial_t\psi_l\rangle|^2. \tag{D7}$$

Plugging in the evolved state $\rho(t)$ yields

$$\mathcal{F}(t) = \frac{(\Gamma' e^{-\Gamma})^2}{1 - e^{-2\Gamma}} + (\Delta E)^2 e^{-2\Gamma} = \left(\frac{\Gamma'^2}{1 - e^{-2\Gamma}} + (\Delta E)^2\right)e^{-2\Gamma}. \tag{D8}$$

In the limit $\gamma \to 0$, $\Gamma \to 0$, one recovers the noiseless value $\mathcal{F}(0) = (\Delta E)^2$. Here, we can notice that the presence of noise may enhance parameter sensitivity. This enhancement originates from the interplay between the coherent phase evolution and the decay of off-diagonal terms: dephasing redistributes information between populations and coherences, and the derivative of the damping factor $\Gamma'(t)$ contributes a positive term to the QFI. Physically, this means that *moderate noise* can increase the rate at which the state changes with respect to t. However, for *strong noise*, the exponential suppression $e^{-2\Gamma}$ dominates, leading to the expected decay of $\mathcal{F}(t)$.

At this point, we can try to find the optimal value of γ maximazing the QFI in the limit of $\gamma \ll 1$. In order to do that we need to rewrite $\mathcal{F}$ as:

$$\mathcal{F}(t) = \left(\frac{\Gamma'^2}{1 - e^{-2\Gamma}} + (\Delta E)^2\right)e^{-2\Gamma} = \tag{D9}$$

$$= \left(\frac{\gamma^2 A^2}{1 - e^{-2\gamma At}} + (\Delta E)^2\right)e^{-2\gamma At} \tag{D10}$$

where $\Delta\ell = (\ell_1 - \ell_0)$ and $A = \frac{\Delta\ell^2}{2}$. Expanding to first order in the noise rate γ, the QFI takes the approximate form:

$$\mathcal{F}(t) \approx \left(\frac{\gamma^2 A^2}{2\gamma At} + (\Delta E)^2\right)(1 - 2\gamma At) = \tag{D11}$$

$$= -A^2\gamma^2 + \left(\frac{A}{2t} - 2(\Delta E)^2 At\right)\gamma + (\Delta E)^2 \tag{D12}$$

which is a downward parabola in γ. Now, we can find the coordinate of the maximum γ^* by imposing $d\mathcal{F}/d\gamma = 0$

$$\gamma^* = -\frac{\frac{A}{2t} - 2(\Delta E)^2 At}{-2A^2} = \frac{1 - 4(\Delta E)^2 t^2}{4At} = \frac{1 - 4(\Delta E)^2 t^2}{2\Delta \ell^2 t}. \tag{D13}$$

Here we can see that γ^* depends on the spectrum of the Hamiltonian generator (ΔE), the spectrum of the noise generator ($\Delta \ell$) and the parameter (t). This could mean that in the context of QML where we have many parameters (even if some generators are shared), if the parameters θ are different, the optimal noise level would be, in general, different. We must come up with a collective measure that takes into account everything.

The quadratic dependence on ΔE of the negative term might explain why equalization kills high eigenvalues first and help low ones: γ^* for high eigenvalues would be too small, or even negative (not physically achievable).

2. Multi-parameter Hamiltonian and State

We now switch to a context more similar to the one of QML, where we have many different parameters associated with multiple generators. We again consider a two-level system with orthonormal eigenstates $|E_0\rangle$ and $|E_1\rangle$ of a family of commuting Hamiltonians $\{H_j\}_{j=1}^M$, and a Lindblad operator L that also commutes with each H_j:

$$H_j |E_k\rangle = E_k^{(j)} |E_k\rangle, \quad L |E_k\rangle = \ell_k |E_k\rangle, \quad [H_j, L] = 0, \quad \forall j, k. \tag{D14}$$

Given the multi-parameter vector $\mathbf{t} = (t_1 \ldots t_M)$ We then define the multi-parameter Hamiltonian $H(\mathbf{t}) = \sum_{j=1}^M t_j H_j$, and the initial state

$$\rho(0) = |\psi(0)\rangle\langle\psi(0)| = \frac{1}{2}\big(|E_0\rangle\langle E_0| + |E_1\rangle\langle E_1| + |E_0\rangle\langle E_1| + |E_1\rangle\langle E_0|\big). \tag{D15}$$

Proceeding in similarly to the what done for the single parameter case, the evolution of the initial state is given by

$$\rho(\mathbf{t}) = e^{\sum_j (-it_j[H_j, \cdot] + \gamma t_j \mathcal{D})} \rho(0). \tag{D16}$$

with $\mathcal{D}[\rho] = L\rho L - \frac{1}{2}\{L^2, \rho\}$, leading to

$$\rho(\mathbf{t}) = \frac{1}{2}\Big(|E_0\rangle\langle E_0| + |E_1\rangle\langle E_1| + e^{-i\Delta(\mathbf{t})} e^{-\Gamma(t)} |E_0\rangle\langle E_1| + \text{h.c.}\Big), \tag{D17}$$

where

$$\Delta(\mathbf{t}) = \sum_{j=1}^M t_j \, \delta E^{(j)}, \quad \delta E^{(j)} = E_1^{(j)} - E_0^{(j)}, \quad \Gamma(t) = \frac{\gamma t}{2}(\ell_1 - \ell_0)^2, \quad t = \sum_j t_j. \tag{D18}$$

After generalizing the normalized eigenbasis defined in Eq. (D5) substituting ΔE with $\Delta(\mathbf{t})$, we need to compute derivatives with respect to each parameter t_j. First, we define the following quantities for convenience

$$\partial_{t_j}\Delta = \delta E^{(j)}, \qquad \partial_{t_j}\Gamma = \frac{\gamma}{2}(\ell_1 - \ell_0)^2 = \frac{\gamma}{2}\Delta\ell^2 = \Gamma', \qquad \partial_{t_j} r = \Big(-i\delta E^{(j)} + \Gamma'\Big) r \tag{D19}$$

where $\Gamma' = \partial_t \Gamma$. Then for the eigenvalues we obtain

$$\partial_{t_j}\lambda_\pm = \pm\frac{1}{2}\,\text{Re}\big(r^* \partial_t r\big) = \mp\frac{1}{2}\Gamma' e^{-\Gamma(t)}, \tag{D20}$$

and for the eigenstates:

$$\partial_{t_j}|\phi_\pm\rangle = \frac{\pm i\delta E^{(j)}}{2}|\phi_\mp\rangle. \tag{D21}$$

The multi-parameter QFI matrix (QFIM) is given by

$$\mathcal{F}_{ij}(\mathbf{t}) = \sum_{\substack{k \\ \lambda_k \neq 0}} \left[\frac{(\partial_i \lambda_k)(\partial_j \lambda_k)}{\lambda_k} + 4\lambda_k \text{Re}\left\{\langle\partial_i\psi_k|\partial_j\psi_k\rangle\right\}\right] - \sum_{\substack{k,l \\ \lambda_k, \lambda_l \neq 0}} \frac{8\lambda_k \lambda_l}{\lambda_k + \lambda_l} \text{Re}\left\{\langle\partial_i\psi_l|\psi_k\rangle\langle\psi_k|\partial_j\psi_l\rangle\right\}. \tag{D22}$$

This leads to a QFIM with the following elements:

$$\mathcal{F}_{ij} = \left(\delta E^{(i)} \delta E^{(j)} + \frac{\Gamma'^2}{1 - e^{-2\Gamma}} \right) e^{-2\Gamma} \tag{D23}$$

where on the diagonal we retrieve the single parameter case. We now consider the approximate QFIM for $\gamma \ll 1$

$$\mathcal{F} \approx (1 - 2\Gamma)\, \delta\, \delta^T + \alpha\, u\, u^T, \tag{D24}$$

with $\delta_i = \delta E^{(i)}$, $u = (1 \dots 1)$, $\Gamma = \frac{\gamma t}{2}(\Delta\ell)^2 \ll 1$, $\alpha = \frac{\Gamma'^2}{1-e^{-2\Gamma}} \approx \frac{\gamma(\Delta\ell)^2}{4t}$. Here $\delta\,\delta^T$ and $u\,u^T$ are rank-1 matrices, so F has at most two nonzero eigenvalues while the remaining $M - 2$ are zero. Recall that since H_j commute wiwth each other in the noiseless setting, the QFIM would have rank 1, already implying that noise is transforming one zero eigenvalue to a non-zero one. The nonzero eigenvalues λ_1, λ_2 satisfy $\lambda_1 + \lambda_2 = \operatorname{Tr}\mathcal{F} = T$, and $\lambda_1^2 + \lambda_2^2 = \operatorname{Tr}\mathcal{F}^2 = S$. Then we can write

$$\mathcal{F}^2 = (1 - 2\Gamma)\delta\delta^T\delta\delta^T + 2(1 - 2\Gamma)\alpha\delta\delta^T uu^T + \alpha^2 uu^T uu^T \tag{D25}$$

by defining the scalars

$$a = \delta^T \delta = \sum_i (\delta E^{(i)})^2, \quad b = \delta^T u = \sum_i \delta E^{(i)}, \quad c = u^T u = M.$$

Computing T and S allows to obtain the product $\lambda_1 \lambda_2 = D$ with which we can write the simple quadratic equation $\lambda^2 - T\lambda + D = 0$:

$$T = (1 - 2\Gamma) \operatorname{Tr}(\delta\delta^T) + \alpha \operatorname{Tr}(u\,u^T) = (1 - 2\Gamma)\,a + \alpha\,c,$$

$$S = (1 - 2\Gamma)^2 \operatorname{Tr}(\delta\delta^T \delta\delta^T) + 2(1 - 2\Gamma)\alpha \operatorname{Tr}(\delta\delta^T u\,u^T) + \alpha^2 \operatorname{Tr}(u\,u^T u\,u^T)$$
$$= (1 - 2\Gamma)^2 a^2 + 2(1 - 2\Gamma)\alpha\, b^2 + \alpha^2 c^2.$$

Then $D = \frac{1}{2}\left[T^2 - S \right] = (1 - 2\Gamma)\,\alpha\,(a\,c - b^2)$ and

$$\lambda_\pm = \frac{T \pm \sqrt{T^2 - 4D}}{2} = \frac{(1 - 2\Gamma)\,a + \alpha\,c \pm \sqrt{(1 - 2\Gamma)^2 a^2 - 2(1 - 2\Gamma)\alpha(a\,c - 2b^2) + \alpha^2 c^2}}{2}.$$

We now expand to first order in the small quantities Γ and α (i.e. $\Gamma, \alpha \sim \mathcal{O}(\gamma) \ll 1$) neglecting all second order terms (like Γ^2, α^2, $\alpha\Gamma$). Let $\Delta_Q = T^2 - 4D$

$$T^2 = (a - 2\Gamma a + \alpha c)^2 \approx a^2 - 4\Gamma a^2 + 2\alpha c a, \quad 4D \approx 4\alpha(a\,c - b^2),$$

hence

$$\Delta_Q = T^2 - 4D \approx a^2 - 4\Gamma a^2 - 2\alpha a c + 4\alpha b^2.$$

Thus since $\sqrt{1 - x} \approx 1 - \frac{x}{2}$ for $x \ll 1$ we obtain $\sqrt{\Delta_Q} \approx a\sqrt{1 - 4\Gamma - 2\alpha\frac{c}{a} + 4\alpha\frac{b^2}{a^2}} \approx a\left(1 - 2\Gamma - \alpha\frac{c}{a} + 2\alpha\frac{b^2}{a^2} \right)$, leading to the following eigenvalues expressions

$$\lambda_+ \approx \frac{1}{2}\left[(a - 2\Gamma a + \alpha c) + a(1 - 2\Gamma - \alpha\frac{c}{a} + 2\alpha\frac{b^2}{a^2}) \right]$$
$$= a - 2\Gamma a + \alpha\frac{b^2}{a},$$
$$\lambda_- \approx \frac{1}{2}\left[(a - 2\Gamma a + \alpha c) - a(1 - 2\Gamma - \alpha\frac{c}{a} + 2\alpha\frac{b^2}{a^2}) \right]$$
$$= \alpha\left(c - \frac{b^2}{a} \right).$$

So in conclusion we have

$$\lambda_+ \approx (1 - \gamma t\Delta\ell^2) \sum_i (\delta E^{(i)})^2 + \frac{\gamma\Delta\ell^2}{4t} \frac{\left(\sum_i \delta E^{(i)} \right)^2}{\sum_i (\delta E^{(i)})^2}, \tag{D26}$$

$$\lambda_- \approx \frac{\gamma\Delta\ell^2}{4t} \left(M - \frac{\left(\sum_i \delta E^{(i)} \right)^2}{\sum_i (\delta E^{(i)})^2} \right), \tag{D27}$$

and all other eigenvalues remain zero. Here is interesting to notice that the principal QFI mode (λ_+) starts at $\sum_i (\delta E^{(i)})^2$ when $\gamma = 0$, then is suppressed by $\gamma t (\Delta \ell)^2$, but partly rescued by a "noise-induced" boost. The second non-zero eigenvalue (λ_-) is *purely noise-induced*, it vanishes for $\gamma = 0$ and grows linearly in γ. This eigenvalue can also vanish if all the $\delta E^{(i)}$ are the same, i.e. all the generators H_j are the same, since the ratio $\frac{(\sum_i \delta E^{(i)})^2}{\sum_i (\delta E^{(i)})^2}$ can be related to a signal to noise ratio:

$$\frac{\left(\sum_i \delta E^{(i)}\right)^2}{\sum_i (\delta E^{(i)})^2} = M \left(\frac{1}{\frac{\text{Var}[\delta E]}{\mathbb{E}[\delta E]^2} + 1} \right), \tag{D28}$$

this implies that if there is no variance in the $\delta E^{(i)}$ this ratio will just reduce to M yielding $\lambda_- = 0$. Instead, if some variability is allowed, the ratio will always be smaller than M. The applied approximations also come with some drawback: the analytical expressions for $\lambda_\pm$ are linear in γ, hiding the possibility to have an optimal noise level for the equalization process and not showing the decay of λ_-.

In this simple setting we were able to derive the noise-induced equalization: noise enable a zero eigenvalue to become non-zero whereas the highest eigenvalue is damped. Avoiding neglecting the higher order terms, one could arrive at a cumbersome equation with also γ^2 and find the best noise level.

Appendix E: Explicit generalization bound

In this appendix, we state an adapted version of the theorem given in Ref. [57] and then briefly present the derivation of our rewriting of the generalization bound.

Theorem (Adapted from Ref. [57]). *Let $P, M \in \mathbb{N}$, $\delta \in [0, 1)$ and $D = \{x_i, y_i\}_{i=1}^M$ an i.i.d. collection of data samples and target labels coming from the distribution $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$. Consider a P-dimensional parameter space $\Theta \subset \mathbb{R}^P$ and a class of quantum machine learning model $\mathcal{M}_\Theta = \{f_{\theta,p}(x) : \theta \in \Theta\}$ subject to quantum noise of intensity $p \in [0, 1)$. Assuming that:*

- *The single samples loss $l : \mathcal{Y} \times \mathbb{R} \to [0, 1]$ is Lipschitz continuous in its second argument with constant $0 < L < 1$.*

- *The gradient of the model $\nabla_\theta f_{\theta,p}(x)$ is bounded by the Lipschitz constant L_f: $\|\nabla_\theta f_{\theta,p}(x)\| \leq L_f$, i.e. the model is Lipschitz continuous w.r.t the parameters θ.*

- *Let $\mathcal{F}(\theta)$ denote the quantum Fisher Information Matrix associated with the model. Suppose there exists $m > 0$ such that: $\sqrt{\det(\mathcal{F}(\theta))} \geq m > 0 \;\; \forall \theta \in \Theta$.*

Then, for any $\delta > 0$, with probability at least $1 - \delta$ over the random draw of the i.i.d training set D, the following generalization bound holds uniformly for all $\theta \in \Theta$:

$$R(\theta) - R_S(\theta) \leq \frac{24\pi}{\sqrt{M}} B(p) + 3\sqrt{\frac{\log(2/\delta)}{2M}}, \tag{E1}$$

where $R(\theta) = \mathbb{E}_{(x,y) \sim \mathcal{D}}[l(y, f_{\theta,p}(x)]$ is the expected risk, $R_S(\theta) = \frac{1}{M} \sum_{i=1}^M l(y, f_{\theta,p}(x))$ is the empirical risk and

$$B(p) = \sqrt{d_{eff}} \left[\Gamma\left(\frac{d_{eff}}{2} + 1 \right) \frac{1}{m} \right]^{1/d_{eff}} L_f, \tag{E2}$$

is a term taking into account the effects of quantum noise with $\Gamma(\cdot)$ being the gamma function. In particular, the noise dependence is given by d_{eff}, m and L_f.

The original generalization bound is of the following form:

$$R(\theta) - R_S(\theta) \leq \frac{12\sqrt{\pi d_{eff}} \exp\left\{ \left(\frac{C'}{d_{eff}} \right) \right\}}{\sqrt{M}} + 3\sqrt{\frac{\log(2/\delta)}{2M}}, \tag{E3}$$

with

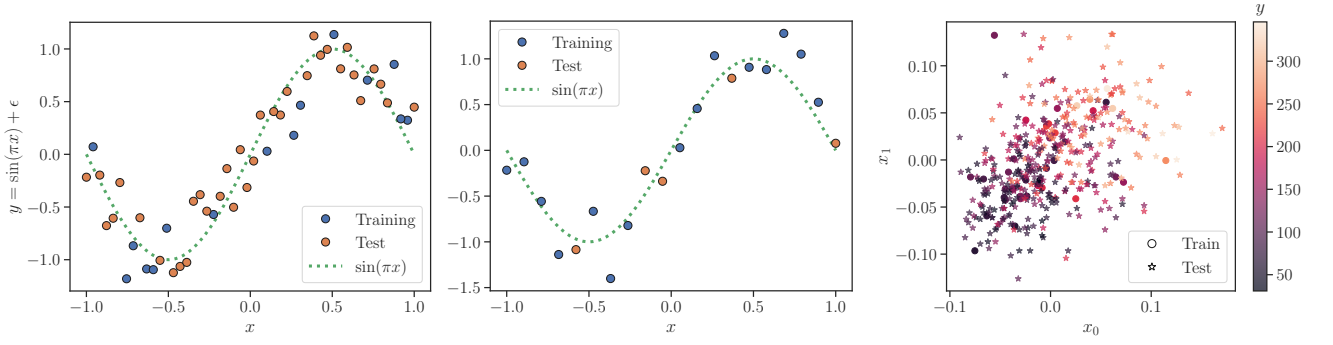$$C' = \log\left(\frac{V_\Theta L_f^{d_{eff}}}{V_{d_{eff}} m} \right) \tag{E4}$$

FIG. 5. Datasets: sinusoidal, sinusoidal 2, diabetes. These are raw data before classical preprocessing.

where V_Θ is the volume of the parameter space Θ and $V_{d_{eff}}$ is the volume of a unit ball in $\mathbb{R}^{d_eff}$. Also in this version of the generalization bound we stress the dependence on d_{eff} instead of the total number of parameters P. This is motivated by two facts:

- even in a noiseless setting, the effective dimension of a quantum model is different from the number of parameters (see overparametrization);

- noise can change the effective role of parameters.

The definition of

$$B(p) = \sqrt{d_{eff}} \left[\Gamma \left(\frac{d_{eff}}{2} + 1 \right) \frac{1}{m} \right]^{1/d_{eff}} L_f \,, \tag{E5}$$

follows from giving the explicit form of V_Θ, $V_{d_{eff}}$ in

$$2\sqrt{\pi} B(p) = \sqrt{d_{eff}} \exp\left\{ \left(\frac{C'}{d_{eff}} \right) \right\}. \tag{E6}$$

In what follows we will use the notation $d = d_{eff}$ for brevity:

$$V_\Theta = (2\pi)^d \quad , \quad V_d = \frac{\pi^{d/2}}{\Gamma \left(\frac{d}{2} + 1 \right)} \tag{E7}$$

$$\frac{V_\Theta}{V_d} = 2^d \pi^{d/2} \Gamma \left(\frac{d}{2} + 1 \right) \tag{E8}$$

then

$$2\sqrt{\pi} B(p) = \sqrt{d} \exp\left\{ \left(\frac{\log \left(2^d \pi^{d/2} \Gamma \left(\frac{d}{2} + 1 \right) \frac{L_f^d}{m} \right)}{d_{eff}} \right) \right\} =$$

$$= \sqrt{d} \left(2^d \pi^{d/2} \Gamma \left(\frac{d}{2} + 1 \right) \frac{L_f^d}{m} \right)^{1/d} =$$

$$= 2\sqrt{\pi} \sqrt{d} \left(\Gamma \left(\frac{d}{2} + 1 \right) \frac{1}{m} \right)^{1/d} L_f \tag{E9}$$

Appendix F: Datasets

In this section, we briefly describe the datasets under study which we also schematically report in Fig. 5. The first dataset analysed is a synthetic sinusoidal dataset. In particular, we generate two different datasets: the first one

is composed of 50 points drawn with uniform probability in the interval $[-1, 1]$ and then divided into 30% training and 70% test samples (sinusoidal), while the second one has 20 samples divided into 75% training and 25% test (sinusoidal2). The analytical expression describing the label that we assign to these points is the following:

$$y = \sin(\pi x) + \epsilon \,, \tag{F1}$$

where $x \in [-1, 1]$ and ϵ is an additive white Gaussian noise with amplitude equal to 0.4, zero mean and a standard deviation of 0.5. In order to properly fit the function, the y variable is rescaled with a `MinMaxScaler` fitted on training data only to span the range $[-1, 1]$.

The second dataset we tackle is a well-known benchmark dataset provided by Scikit-learn [87], with real medical data related to diabetes. It consists of physiological variables measured in patients, which are used to predict a quantitative measure of diabetes progression one year after baseline. It contains ten features, including age, sex, body mass index (BMI), blood pressure, total serum cholesterol, low-density lipoproteins, high-density lipoproteins, total cholesterol to HDL ratio, log of serum triglycerides level (LTG), and glucose level. The target variable represents a numerical value indicating the progression of diabetes. In this case, only BMI and LTG are used as input features. Then, the dataset is divided into 40 train and 400 test samples. Input features are rescaled to fit the range of angles of rotation gates, i.e. $[-\pi, \pi]$, with a `MinMaxScaler` fitted on training data only. Analogously, the target variable is rescaled within $[-1, 1]$.

Appendix G: Quantum Neural Network models

In this Appendix, we provide a detailed description of the quantum neural network (QNN) architectures employed in our study. The QNN employed to analyse the sinusoidal dataset is the same as in Ref. [54], while for the diabetes dataset, we employ the same model as Ref. [68] to show that our procedure is capable of predicting the best noise level in agreement with previous results.

The first QNN model consists of five qubits, all initialized in the computational $|0\rangle$ state. The classical features are encoded through two layers of single-qubit rotations R_Y and R_Z. Since the dataset consists of single-feature data, all qubits encode the same value. The trainable part of the circuit is composed of three sublayers of single-qubit rotations, R_X, R_Z, and R_X, each followed by a sequence of CNOT gates that linearly entangle all qubits. With this elementary layer, we build an underparameterized QNN with $L = 4$ layers and an overparameterized model with $L = 10$ layers, resulting in $P = 60$ and $P = 150$ trainable parameters, respectively. The output of both models is the expectation value of the Z Pauli operator on the first qubit.

The second QNN is used to study the diabetes dataset. The model consists of four qubits, also initialized in the computational $|0\rangle$ state. The encoding process applies two RX gates to the first and third qubits, embedding two classical features into the quantum state. The subsequent variational structure consists of a layer of single-qubit R_Y gates and a ring of symmetric RXX Ising gates that establish entanglement. We alternate this structure $L = 3$ and $L = 5$ times to obtain an underparameterized and an overparameterized QNN, respectively. The output of both models is the expectation value of the $Z^{\otimes 4}$ operator.

In Appendix I, we cross-validate the architectures on the other dataset.

Appendix H: Optimal noise level for diabetes dataset

In this section, we show that we are capable of approximating the optimal level of noise found in Ref. [68] and we extend the analysis also to underparameterized QNNs with the same architecture. We study the diabetes dataset with the second QNN model described in Appendix G. The analysis of the NIE is presented in Fig. 6, while results concerning the estimation of the optimal noise level p^* are reported in Fig. 7 and Tabs. III-IV. The NIE can be appreciated via the analysis of $I_r(p)$ also for this second architecture and dataset. A non-trivial increase for the least important eigenvalues is observed at non-zero noise levels for both underparameterized and overparameterized models under the action of different kinds of quantum noise. Notice that Fig. 7a-b-c-i are missing the noise level $p = 10^{-6}$ due to numerical issues arising when computing the QFIM. Moving to the optimal noise level estimation, we can notice a good agreement between the two values found with the NIE-based procedure and the MSE. The only case in which we find a slight mismatch is for the overparameterized QNN when phase damping noise is present. This might be only a fluctuation due to the particular training-test splitting as the value of p^* given by the NIE-based estimation is close to the noise level found in Ref. [68] while the MSE estimation is quite far (see values in Tab. III).
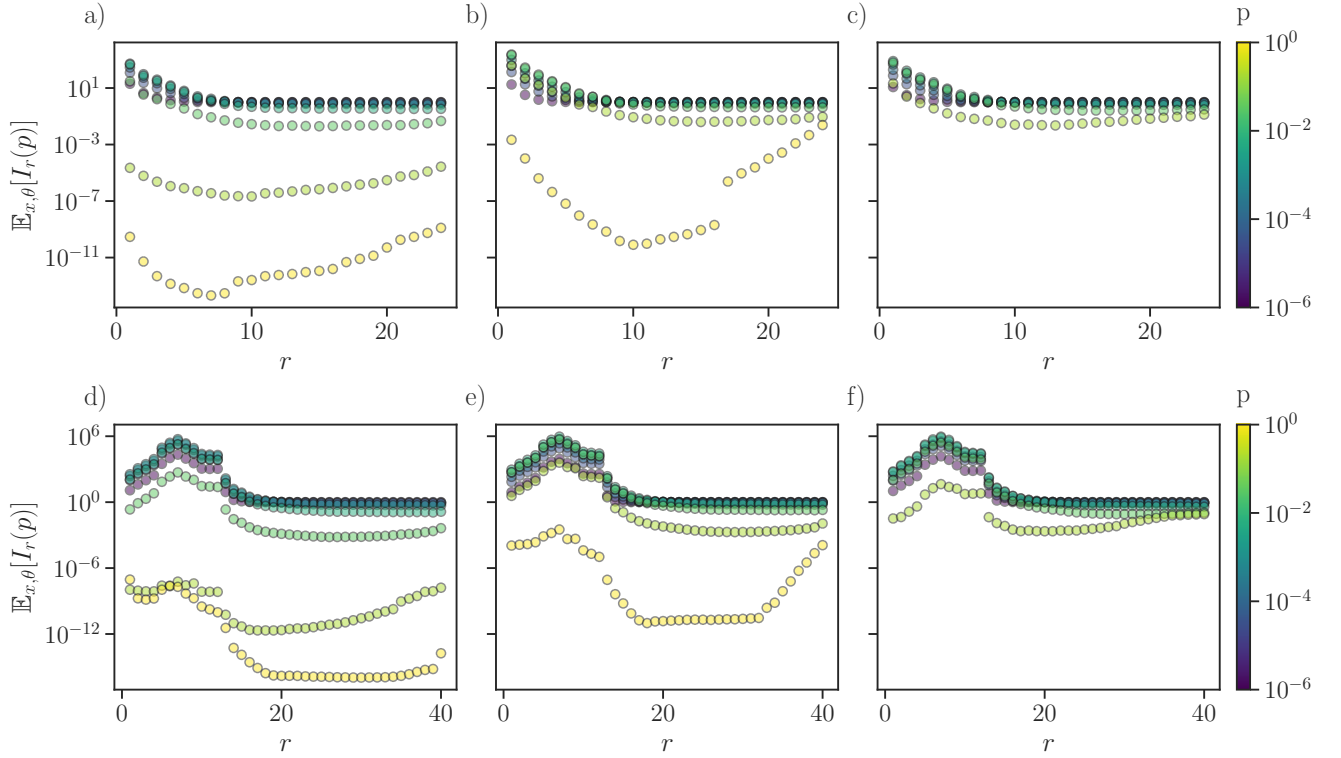
FIG. 6. Relative change of the averaged QFIM eigenvalues λ_m under different levels of noise p with respect to the noiseless case ($p=0$) for a), d) depolarizing, b), e) dephasing and c), f) amplitude-damping noise. The first row represents the relative change for an underprametrized model (24 parameters), while the second for an overparameterized one (40 parameters). The average is computed over all the inputs of the training set (diabetes dataset) and 5 different parameter vectors.

	DP	PD	AD
NIE	$(8.8 \pm 1.3)10^{-3}$	$(3.76 \pm 0.74)10^{-2}$	$(2.45 \pm 0.84)10^{-2}$
Test MSE	$(1.24 \pm 0.76)10^{-2}$	$(9.38 \pm 3.20)10^{-2}$	$(3.01 \pm 1.12)10^{-2}$
Test MSE [68]	0.010	0.056	0.018
$B(p)$	$(1.78 \pm 0.00)10^{-1}$	$(1.90 \pm 0.39)10^{-1}$	$(6.97 \pm 0.19)10^{-2}$
Gen. gap	$(5.38 \pm 0.74)10^{-1}$	$(3.66 \pm 0.98)10^{-1}$	$(5.62 \pm 0.00)10^{-1}$

TABLE III. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) for diabetes dataset estimated with different methods for overparameterized QNN.

	DP	PD	AD
NIE	$(1.41 \pm 0.51)10^{-2}$	$(6.62 \pm 2.20)10^{-2}$	$(4.72 \pm 2.68)10^{-2}$
Test MSE	$(2.10 \pm 0.91)10^{-2}$	$(9.60 \pm 4.73)10^{-2}$	$(3.10 \pm 1.14)10^{-2}$
$B(p)$	$(3.71 \pm 1.86)10^{-1}$	$(4.04 \pm 1.31)10^{-1}$	$(1.00 \pm 0.00)10^{-1}$
Gen. gap	$(7.81 \pm 2.19)10^{-1}$	$(6.25 \pm 3.25)10^{-1}$	$(5.62 \pm 0.00)10^{-1}$

TABLE IV. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) for diabetes dataset estimated with different methods for underparameterized QNN.

Appendix I: Additional numerical experiments

In Fig. 8, we summarize the results of the cross-validation of architectures and datasets when estimating p^* for different types of quantum noise and different estimation methods. In particular, we compare first (first row) and second (second row) QNNs in both under- (left column) and over-parameterized (right column) regimes on sinusoidal
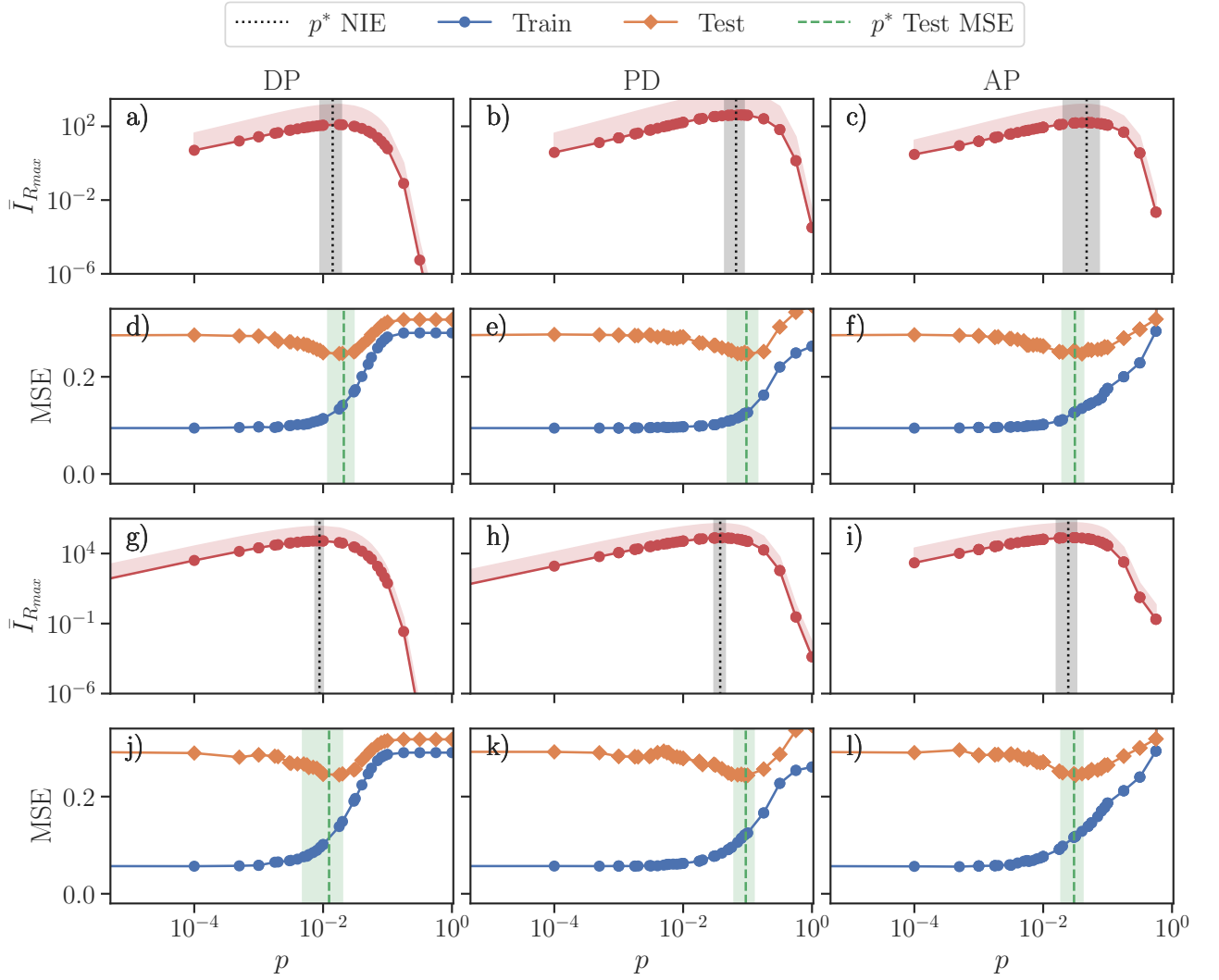
FIG. 7. **Diabetes** Comparison of optimal noise level p^* obtained from NIE and test MSE on diabetes dataset. Columns correspond to different types of noise: depolarizing (DP), phase damping (PD) and amplitude damping (AD). First and second rows refer to the underparameterized model, while third and fourth to the overparameterized one.

and diabetes datasets. The NID estimation appears almost always consistent with the MSE estimation. A substantial discrepancy is observed only for the overparameterized version of the first QNN on the diabetes dataset, highlighting that our pre-training analysis based on averaged quantities on the optimization landscape may fail to find the best regularizing regime. A local approach could maybe lead to better performance in such cases. For what concerns the estimation via generalization gap, as already seen, this is in general not a good estimation method.

Appendix J: Numerical analysis of the generalization bound

We now report a numerical analysis for the generalization bound. In order to estimate p^* from the generalization bound reported in Eq. (E1), we focus on the noise-dependent term $B(p)$ (see Eq. (E2)). In particular, we also study the noise dependence of the single components affected by quantum noise, namely the Lipschitz constant of the quantum model L_f, the (square root of the) determinant of the QFIM and the effective dimension (d_{eff}) of the the quantum model. In particular, we remind the definition of $B(p)$

$$B(p) = \sqrt{d_{eff}} \left[\Gamma \left(\frac{d_{eff}}{2} + 1 \right) \frac{1}{m} \right]^{1/d_{eff}} L_f \,.$$
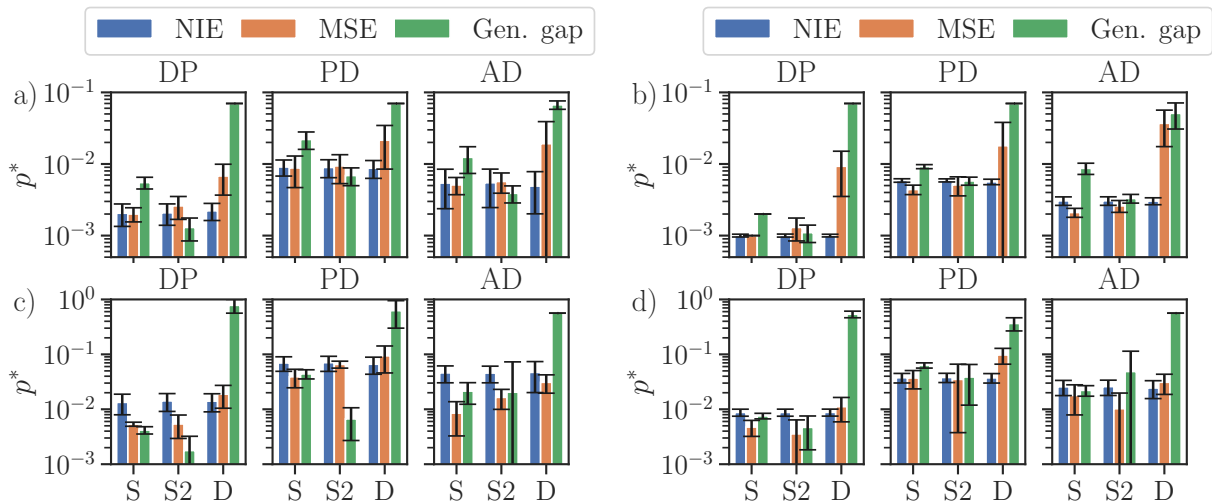
FIG. 8. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) for sinusoidal (S, S2) and diabetes (D) datasets estimated with different methods and QNN architectures: a) first QNN underparameterized, b) first QNN overparameterized, c) second QNN underparameterized, d) second QNN overparameterized.

This quantity is evaluated for 5 random parameter vectors per each one of the M training samples (for a total of $5M$) as a function of noise p. Specifically, in Figs. 9- 12, we plot $B(p)$ in panels a, e, i, the Lipschitz constant in panels b, f, j, the square root of the determinant of the QFIM in panels c, g, k and the effective dimensions in panel d, h, l.

The Lipschitz constant L_f is estimated as the maximal gradient over different samplings (5 random parameter vectors per each one of the M training samples) given that

$$\|\nabla_\theta f_{\theta,p}(x)\| \le L_f \implies \forall p \quad L_f \approx \max_{x,\theta} \|\nabla_\theta f_{\theta,p}(x)\|. \tag{J1}$$

This estimation of the Lipschitz function is quite loose, a more accurate estimate would require exponentially many samples. This highlights how our method is more applicable and less demanding in terms of computational resources.

For what concerns the determinant of the QFIM $\mathcal{F}$, we compute the full eigenspectrum, and then we take the product of all the eigenvalues:

$$\det\{\mathcal{F}(x,\theta,p)\} = \prod_{r=1}^{P} \lambda_r. \tag{J2}$$

Since the QFIM depends on the specific input, parameter vector and noise level ($\mathcal{F} = \mathcal{F}(x,\theta,p)$), to analyze its behaviour as a function of the noise only, we compute its average and standard deviation with respect to our finite sampling (5 random parameter vectors per each one of the M training samples). It is worth pointing out that we numerically see many of the eigenvalues being smaller than 1. Consequently, for models with many parameters, this leads to determinants extremely close to 0 or considered equivalent to 0 numerically. In such situations, it will be impossible to estimate the generalization bound of Eq. (E1), as this has an inverse dependence on the square root of the determinant, implying a diverging bound. This is represented by red areas in the Figs. 9- 12.

For what concerns the effective dimension d_{eff}, it is determined as the number of non-trivial direction in the parameter space. This is measured in terms of the number of non-zero eigenvalues of the QFIM, i.e. its rank. In particular, the numerical precision in this case is the one set by the numerical precision of the machine ($\epsilon = 2.22 \cdot 10^{-16}$) times the total number of parameters in the model P.

Appendix K: Dependence on input dataset

Here we test the dependence of the method with respect to the input dataset. We estimate the optimal level of noise via synthetic datasets that are not related to the learning tasks described above. In particular, we create two datasets for each learning model with 15 samples drawn from a uniform distribution in the interval $[-\pi, \pi]$ and from a Gaussian distribution with zero mean and standard deviation equal to 1. For the first QNN architecture, the datasets
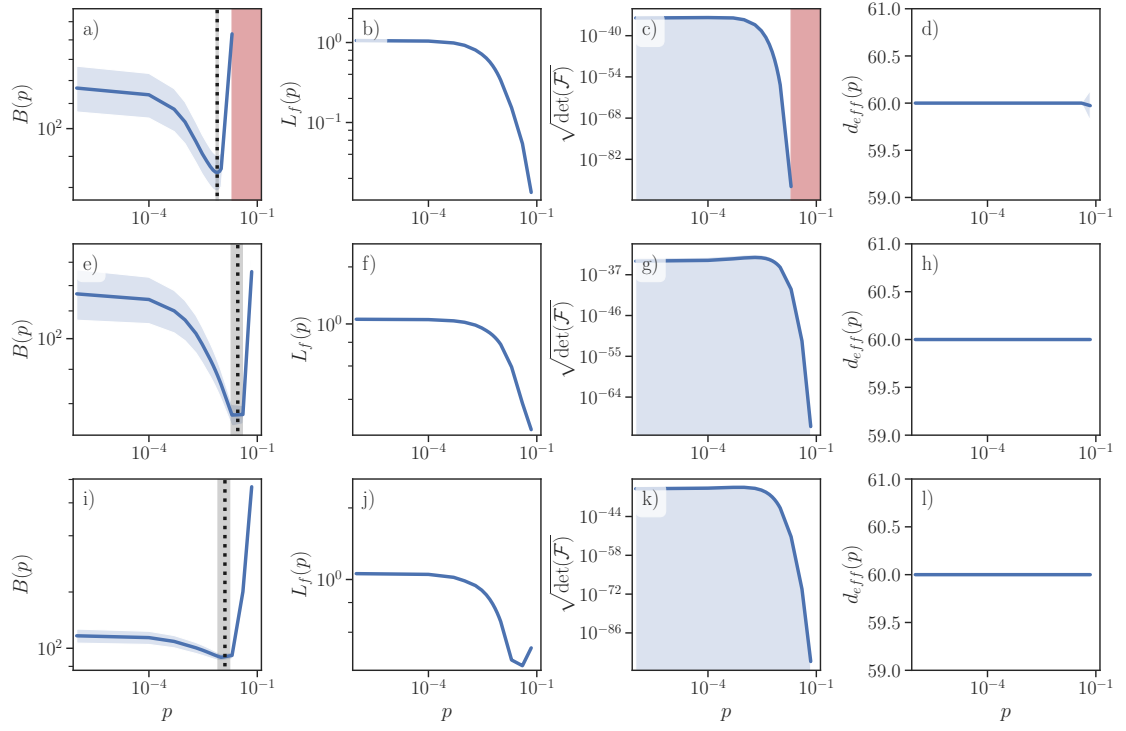
FIG. 9. Generalization bound for the first QNN, in underparameterized regime with sinusoidal dataset, as a function of the noise level p is reported in panel a). The other panels analogously report specific contributions to the generalization bound: b) Lipschitz constant L_f, c) $\sqrt{\det(\mathcal{F})}$, d) effective dimension d_{eff}. The vertical dotted line represents the level of noise minimizing the bound. The shaded areas are one standard deviation. Red areas highlight where the bound is not computable.
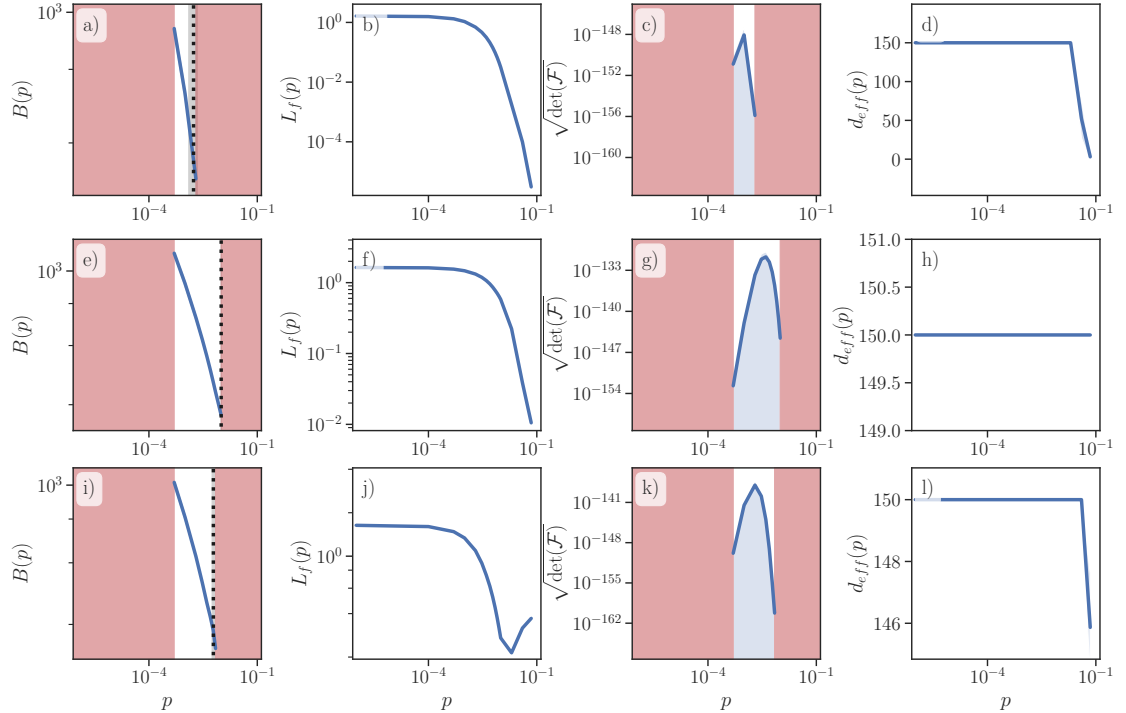

FIG. 10. Generalization bound for the first QNN, in overparameterized regime with sinusoidal dataset, as a function of the noise level p is reported in panel a). The other panels analogously report specific contributions to the generalization bound: b) Lipschitz constant L_f, c) $\sqrt{\det(\mathcal{F})}$, d) effective dimension d_{eff}. The vertical dotted line represents the level of noise minimizing the bound. The shaded areas are one standard deviation. Red areas highlight where the bound is not computable.
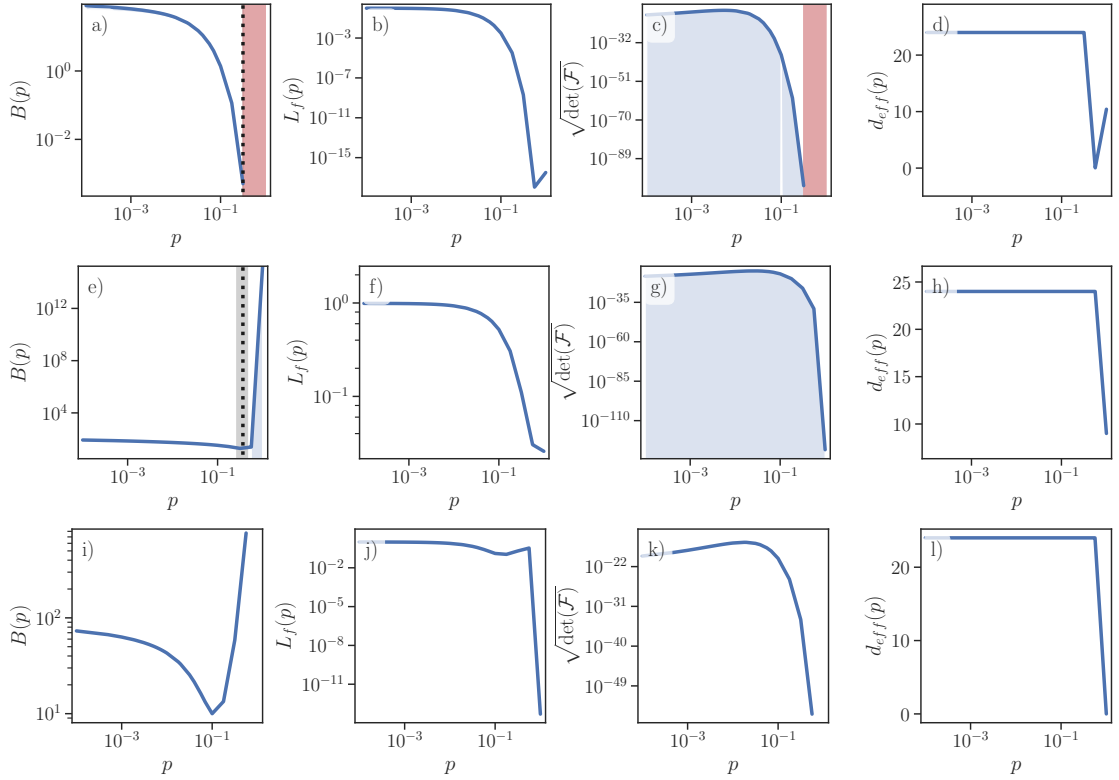
FIG. 11. Generalization bound for the first QNN, in underparameterized regime with diabetes dataset, as a function of the noise level p is reported in panel a). The other panels analogously report specific contributions to the generalization bound: b) Lipschitz constant L_f, c) $\sqrt{\det(\mathcal{F})}$, d) effective dimension d_{eff}. The vertical dotted line represents the level of noise minimizing the bound. The shaded areas are one standard deviation. Red areas highlight where the bound is not computable.

	DP	PD	AD
Sinusoidal	$(2.1 \pm 0.7)10^{-3}$	$(9.1 \pm 2.3)10^{-3}$	$(5.4 \pm 3.0)10^{-3}$
Sinusoidal2	$(2.1 \pm 0.7)10^{-3}$	$(8.9 \pm 2.5)10^{-3}$	$(5.5 \pm 3.0)10^{-3}$
Diabetes	$(2.2 \pm 0.6)10^{-3}$	$(8.7 \pm 2.4)10^{-3}$	$(4.9 \pm 2.9)10^{-3}$
Uniform $[-\pi, \pi]$	$(2.1 \pm 0.7)10^{-3}$	$(8.8 \pm 2.1)10^{-3}$	$(5.0 \pm 2.9)10^{-3}$
Gaussian $(0, 1)$	$(2.1 \pm 0.7)10^{-3}$	$(8.7 \pm 2.4)10^{-3}$	$(5.2 \pm 2.8)10^{-3}$

TABLE V. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) with respect to various input datasets for the first QNN in underparameterized regime.

are single-feature, to reflect the structure of the sinusoidal dataset, while for the second one, the datasets have 2 input features. The estimates of p^* with the NIE procedure for different datasets and models are gathered in Tabs. V-VIII. It is possible to see that the p^* is almost the same when varying the input dataset. This is most likely due to the fact that the average eigenspectrum is the same when changing the dataset.
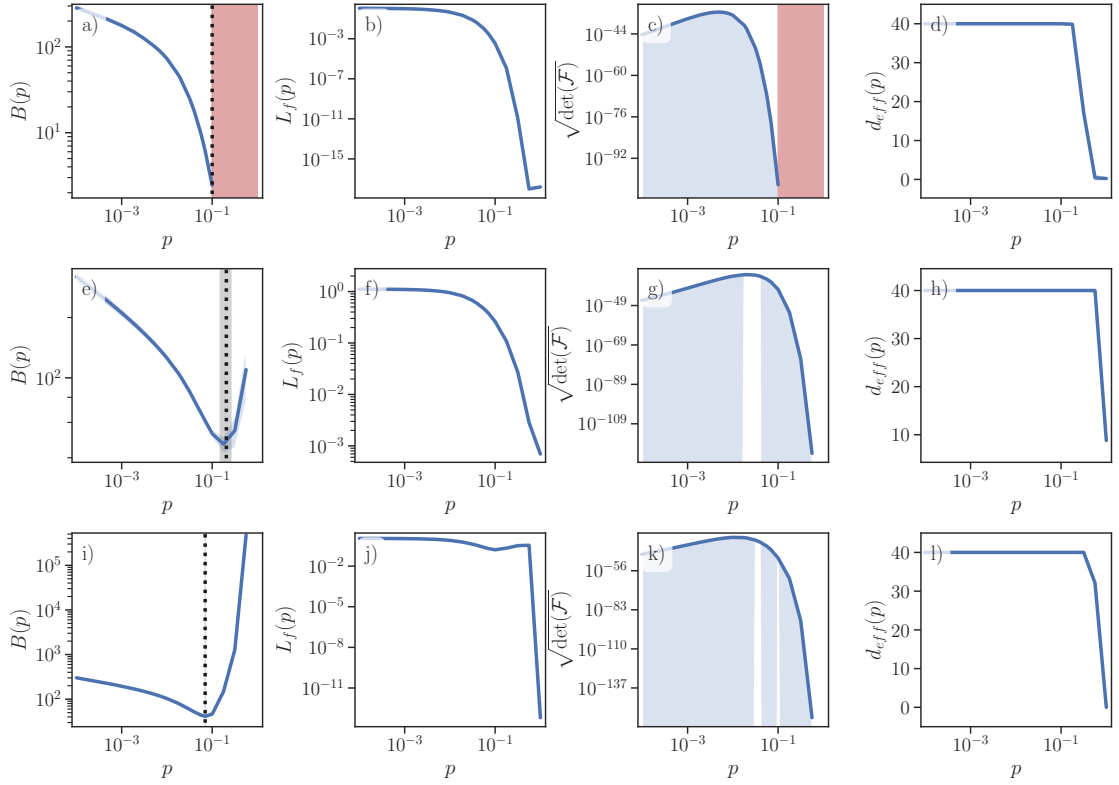
FIG. 12. Generalization bound for the first QNN, in overparameterized regime with diabetes dataset, as a function of the noise level p is reported in panel a). The other panels analogously report specific contributions to the generalization bound: b) Lipschitz constant L_f, c) $\sqrt{\det(\mathcal{F})}$, d) effective dimension d_{eff}. The vertical dotted line represents the level of noise minimizing the bound. The shaded areas are one standard deviation. Red areas highlight where the bound is not computable.

	DP	PD	AD
Sinusoidal	$(1.0 \pm 0.0)10^{-3}$	$(5.9 \pm 0.4)10^{-3}$	$(3.1 \pm 0.4)10^{-3}$
Sinusoidal2	$(1.0 \pm 0.0)10^{-3}$	$(5.9 \pm 0.3)10^{-3}$	$(3.1 \pm 0.4)10^{-3}$
Diabetes	$(1.0 \pm 0.0)10^{-3}$	$(5.6 \pm 0.5)10^{-3}$	$(3.0 \pm 0.3)10^{-3}$
Uniform $[-\pi, \pi]$	$(1.0 \pm 0.0)10^{-3}$	$(5.6 \pm 0.5)10^{-3}$	$(3.0 \pm 0.3)10^{-3}$
Gaussian $(0, 1)$	$(1.0 \pm 0.0)10^{-3}$	$(5.8 \pm 0.4)10^{-3}$	$(3.0 \pm 0.3)10^{-3}$

TABLE VI. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) with respect to various input datasets for the first QNN in overparameterized regime.

	DP	PD	AD
Sinusoidal	$(1.41 \pm 0.53)10^{-2}$	$(6.97 \pm 2.06)10^{-2}$	$(4.60 \pm 1.56)10^{-2}$
Sinusoidal2	$(1.42 \pm 0.51)10^{-2}$	$(7.06 \pm 2.16)10^{-2}$	$(4.56 \pm 1.53)10^{-2}$
Diabetes	$(1.41 \pm 0.51)10^{-2}$	$(6.62 \pm 2.20)10^{-2}$	$(4.72 \pm 2.68)10^{-2}$
Uniform $[-\pi, \pi]$	$(1.46 \pm 0.47)10^{-2}$	$(6.15 \pm 2.54)10^{-2}$	$(4.64 \pm 2.67)10^{-2}$
Gaussian $(0, 1)$	$(1.46 \pm 0.46)10^{-2}$	$(6.59 \pm 2.03)10^{-2}$	$(4.68 \pm 2.02)10^{-2}$

TABLE VII. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) with respect to various input datasets for the second QNN in underparameterized regime.

	DP	PD	AD
Sinusoidal	$(8.8 \pm 1.3)10^{-3}$	$(3.75 \pm 0.74)10^{-2}$	$(2.56 \pm 0.79)10^{-2}$
Sinusoidal2	$(8.8 \pm 1.3)10^{-3}$	$(3.79 \pm 0.72)10^{-2}$	$(2.59 \pm 0.80)10^{-2}$
Diabetes	$(8.8 \pm 1.3)10^{-3}$	$(3.76 \pm 0.74)10^{-2}$	$(2.45 \pm 0.84)10^{-2}$
Uniform $[-\pi, \pi]$	$(8.8 \pm 1.3)10^{-3}$	$(3.71 \pm 0.70)10^{-2}$	$(2.47 \pm 1.02)10^{-2}$
Gaussian $(0, 1)$	$(8.7 \pm 1.3)10^{-3}$	$(3.73 \pm 0.67)10^{-2}$	$(2.51 \pm 0.73)10^{-2}$

TABLE VIII. Comparison of different values of p^* for depolarizing (DP), phase damping (PD) and amplitude damping (AD) with respect to various input datasets for the second QNN in overparameterized regime.