



**HAL**  
open science

# Optimal Fidelity Estimation from Binary Measurements for Discrete and Continuous Variable Systems

Omar Fawzi, Aadil Oufkir, Robert Salzmann

► **To cite this version:**

Omar Fawzi, Aadil Oufkir, Robert Salzmann. Optimal Fidelity Estimation from Binary Measurements for Discrete and Continuous Variable Systems. 2024. hal-04840385

**HAL Id: hal-04840385**

**<https://hal.science/hal-04840385v1>**

Preprint submitted on 16 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal Fidelity Estimation from Binary Measurements for Discrete and Continuous Variable Systems

Omar Fawzi \*

Aadil Oufkir †

Robert Salzmann \*

September 9, 2024

## Abstract

Estimating the fidelity between a desired target quantum state and an actual prepared state is essential for assessing the success of experiments. For pure target states, we use functional representations that can be measured directly and determine the number of copies of the prepared state needed for fidelity estimation. In continuous variable (CV) systems, we utilise the Wigner function, which can be measured via displaced parity measurements. We provide upper and lower bounds on the sample complexity required for fidelity estimation, considering the worst-case scenario across all possible prepared states. For target states of particular interest, such as Fock and Gaussian states, we find that this sample complexity is characterised by the  $L^1$ -norm of the Wigner function, a measure of Wigner negativity widely studied in the literature, in particular in resource theories of quantum computation. For discrete variable systems consisting of  $n$  qubits, we explore fidelity estimation protocols using Pauli string measurements. Similarly to the CV approach, the sample complexity is shown to be characterised by the  $L^1$ -norm of the characteristic function of the target state for both Haar random states and stabiliser states. Furthermore, in a general black box model, we prove that, for any target state, the optimal sample complexity for fidelity estimation is characterised by the smoothed  $L^1$ -norm of the target state. To the best of our knowledge, this is the first time the  $L^1$ -norm of the Wigner function provides a lower bound on the cost of some information processing task.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Main results . . . . .	3
1.2	Related work . . . . .	7
1.3	Outline of the article . . . . .	9
<b>2</b>	<b>Preliminaries</b>	<b>10</b>
2.1	Notation . . . . .	10
2.2	$L^q$ -spaces and smoothed $L^1$ -norm . . . . .	10
2.3	Information theoretic divergences . . . . .	10
2.4	Wigner functions for continuous variable systems . . . . .	11

---

\*Univ Lyon, Inria, ENS de Lyon, LIP, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

†Institute for Quantum Information, RWTH Aachen University, Aachen, Germany

<b>3</b>	<b>Black box overlap estimation</b>	<b>13</b>
3.1	Proof of the upper bound on the sample complexity $N_{\varepsilon,\delta}(f, \mathcal{S})$	15
3.2	Proof of the lower bound on the sample complexity $N_{\varepsilon,\delta}(f, \mathcal{S})$	17
<b>4</b>	<b>Fidelity estimation of continuous variable systems</b>	<b>22</b>
4.1	Matching upper and lower bounds for $N_{\varepsilon,\delta}^W(\rho)$ for examples of states	24
4.1.1	Sample complexity for Fock states	26
4.1.2	Sample complexity for spike states	30
4.1.3	Sample complexity for Gaussian states	33
4.1.4	Proof of Theorem 4.4	34
<b>5</b>	<b>Fidelity estimation for discrete variable systems</b>	<b>34</b>
5.1	Matching upper and lower bounds for $N_{\varepsilon,\delta}^P(\rho)$ for example of states	37
5.1.1	Sample complexity for Haar random states	38
5.1.2	Sample complexity for stabiliser states	40
<b>A</b>	<b>Some facts about the smoothed <math>L^1</math>-norm</b>	<b>40</b>
A.1	Lower bound on smoothed $L^1$ -norm for functions with rapidly decaying tail	42

## 1 Introduction

With the possibility to produce a variety of complex and highly entangled quantum states in experiments, it is essential to develop reliable methods certifying that a desired target state has been prepared successfully. For a pure target state  $\rho = |\psi\rangle\langle\psi|$  and an actually prepared state  $\sigma$ , an operationally meaningful way to achieve such certification is by estimating the fidelity  $F(\rho, \sigma)$  through measurements on the state  $\sigma$ . For example, performing the projective measurement  $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$ , allows fidelity estimation using a number of copies of  $\sigma$  that depends only on the desired accuracy and success probability, and is independent of the target state or system size. However, for most states  $|\psi\rangle\langle\psi|$  of interest, performing this measurement is not possible in practice in an efficient and reliable manner. Therefore, it is sensible to express the fidelity as the overlap of certain functions representing the quantum states, which can be directly measured with current experimental techniques [FL11, dSLCP11]:

For continuous variable (CV) quantum systems with  $m \in \mathbb{N}$  degrees of freedom, e.g. an  $m$ -mode photonic system, we use the phase space representation of the fidelity between  $\rho$  and  $\sigma$  in terms of their *Wigner functions*  $W_\rho$  and  $W_\sigma$  given by

$$F(\rho, \sigma) = \pi^m \int_{\mathbb{C}^m} W_\rho(\alpha) W_\sigma(\alpha) d\alpha. \quad (1.1)$$

Assuming full knowledge of the pure target state  $\rho$ , the fidelity can hence be estimated by directly measuring the Wigner function  $W_\sigma(\alpha)$  at different points in phase space  $\alpha \in \mathbb{C}^m$  through *Wigner tomography*. This can be achieved by utilising that the Wigner function of  $\sigma$  is given by the expectation value of the *displaced parity operator* [Roy77], i.e.

$$W_\sigma(\alpha) = \left(\frac{2}{\pi}\right)^m \text{Tr}\left(D(\alpha)\Pi D^*(\alpha) \sigma\right), \quad (1.2)$$

with  $D(\alpha)$  denoting the *displacement operator* and  $\Pi = (-1)^N$  the *parity operator* with  $N$  being the *m-mode number operator* (c.f. Section 2.4 for definition). The observable  $(2/\pi)^m D(\alpha)\Pi D^*(\alpha)$  has measurement outcomes  $\{(2/\pi)^m, -(2/\pi)^m\}$ , showing that we can estimate  $W_\sigma(\alpha)$  by some binary measurement on the state  $\sigma$ . A practical scheme for performing these displaced parity measurements relying on Ramsey interferometry has been proposed in [LD97] and experimentally first implemented in [NRO<sup>+</sup>00, BAM<sup>+</sup>02, VKL<sup>+</sup>13a]. Directly measuring the Wigner function in this way and performing fidelity estimation using (1.1) has become conventional in practice, see [VKL<sup>+</sup>13b, WGR<sup>+</sup>16, RBG<sup>+</sup>24, MEC<sup>+</sup>24, MDR<sup>+</sup>24] and references therein. In the following, we refer to the above-described paradigm as the *Wigner model of fidelity estimation*.

On the other hand, for discrete variable (DV) quantum systems consisting of  $n$  qubits, we consider the representation of the fidelity given by

$$F(\rho, \sigma) = \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_\rho(P) \chi_\sigma(P), \quad (1.3)$$

with dimension of the overall system  $d = 2^n$ ,  $\mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n}$  denoting the set of Pauli strings of length  $n$  and

$$\chi_\sigma(P) = \text{Tr}(P\sigma) \quad (1.4)$$

the *characteristic function* of the state  $\sigma$ . Again, assuming full knowledge of the pure target state  $\rho$ , we can estimate the fidelity by measuring Pauli strings with binary outcome in  $\{1, -1\}$  on the state  $\sigma$ . We refer to this paradigm as the *Pauli model of fidelity estimation* in the following.

For both of these settings, efficient fidelity estimation algorithms are known for some special families of states, e.g., well-conditioned states including stabiliser states [FL11]. However, as far as we are aware, no lower bound for fidelity estimation is known. This raises the following natural question:

*Are there states for which fidelity estimation is hard?*

More specifically, in the Wigner model of fidelity estimation, are there states for which the sample complexity is arbitrarily large? In the Pauli model, are there states for which the number of samples required is exponential in  $n$ ? Our results give an affirmative answer to both of these questions using Fock states (Eq. (1.10)) and Haar-random  $n$ -qubit states (Eq. (1.17)). Going beyond such hard states, we can ask:

*Which property of  $\rho$  quantifies the sample complexity of fidelity estimation?*

We identify the smoothed  $L^1$ -norm as quantifying the optimal sample complexity for fidelity estimation in a general black box model (Eq. (1.7)).

## 1.1 Main results

We now describe more precisely the setup for our results. We analyse the *sample complexity of fidelity estimation in the Wigner- and Pauli model*, respectively denoted by  $N_{\varepsilon, \delta}^W(\rho)$  and  $N_{\varepsilon, \delta}^P(\rho)$ . They are defined as the minimal number of copies of the state  $\sigma$  needed to estimate the fidelity  $F(\rho, \sigma)$  with accuracy  $\varepsilon > 0$  and probability of failure at most  $\delta > 0$  in the worst case over all states  $\sigma$  using the paradigms described around (1.1) and (1.3) respectively.

For this, we unify the discussion for continuous and discrete variable systems by considering general overlaps of functions  $f$  and  $g$  on some measurable space  $\Omega$  with measure  $\mu$ . Inspired from (1.1) and (1.3), we choose for CV systems  $\Omega_W = \mathbb{C}^m$  with rescaled Lebesgue measure  $d\mu_W(\alpha) = \pi^m d\alpha$  and for DV systems  $\Omega_P = \mathcal{P}_n$ <sup>1</sup> with  $\mu_P = \frac{1}{d}\mu_{\text{count}}$  and  $\mu_{\text{count}}$  denoting the counting measure.

The function  $f$  corresponds to the Wigner- or characteristic function of the target state  $\rho$  in the Wigner- or Pauli model respectively and hence full knowledge of this function is assumed. On the other hand, for the function  $g$ , which plays the role of  $W_\sigma$  or  $\chi_\sigma$  in the above, we assume to have *black box access*: That is for each  $\lambda \in \Omega$  we can obtain samples of a random variable  $Y_\lambda$  with binary outcome space and which satisfies  $\mathbb{E}[Y_\lambda] = g(\lambda)$ .

The objective is then to estimate the overlap

$$\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) = \int_{\Omega} f(\lambda) \mathbb{E}[Y_\lambda]d\mu(\lambda) \quad (1.5)$$

from as few samples of  $Y_\lambda$  at different points  $\lambda$  as possible. We call this task *black box overlap estimation* in the following.

Here,  $f$  is assumed to be fixed and the function  $g$ , for which black box access is provided, is allowed to be any element of some fixed set of functions  $\mathcal{S}$ . The *sample complexity of black box overlap estimation*, denoted by  $N_{\varepsilon,\delta}(f, \mathcal{S})$ , is then given by the minimal number of samples of  $Y_\lambda$  needed to estimate the overlap (1.5) with accuracy  $\varepsilon > 0$  and probability of failure at most  $\delta > 0$  in the worst case over all  $g \in \mathcal{S}$ . In particular, choosing  $\Omega$  and  $\mu$  for CV and DV quantum systems as outlined above and denoting the set of possible Wigner- and characteristic functions by  $\mathcal{S}_W$  and  $\mathcal{S}_P$  respectively, we can make the following identifications:

$$\begin{aligned} \text{CV systems:} & \quad N_{\varepsilon,\delta}^W(\rho) = N_{\varepsilon,\delta}(W_\rho, \mathcal{S}_W), \\ \text{DV systems:} & \quad N_{\varepsilon,\delta}^P(\rho) = N_{\varepsilon,\delta}(\chi_\sigma, \mathcal{S}_P). \end{aligned} \quad (1.6)$$

We provide upper and lower bounds on the sample complexity of black box overlap estimation  $N_{\varepsilon,\delta}(f, \mathcal{S})$  for a large class of sets of functions  $\mathcal{S}$ :

### Black box overlap estimation:

For general measurable spaces  $\Omega$  with measure  $\mu$  we prove upper and lower bounds on the sample complexity  $N_{\varepsilon,\delta}(f, \mathcal{S})$  for  $f \in L^2(\Omega, \mu)$  and all sets of functions  $\mathcal{S} \subset \mathcal{S}_{\text{max}}$  where for some fixed number  $r > 0$  the set  $\mathcal{S}_{\text{max}}$  consists of all functions  $g$  satisfying  $\|g\|_{L^2(\Omega, \mu)} \leq 1$  and  $\|g\|_{L^\infty(\Omega, \mu)} \leq r$ . These conditions on the sets  $\mathcal{S}$  and  $\mathcal{S}_{\text{max}}$  are motivated by the Wigner- and Pauli model as for all quantum states  $\rho$  of a continuous (or discrete) variable system the corresponding Wigner (or characteristic) function satisfies  $\|W_\rho\|_{L^2(\mathbb{C}^m, \mu_W)} \leq 1$  and  $\|W_\rho\|_{L^\infty(\mathbb{C}^m, \mu_W)} \leq (2/\pi)^m$  (or  $\|\chi_\rho\|_{L^2(\mathcal{P}_n, \mu_P)} \leq 1$  and  $\|\chi_\rho\|_{L^\infty(\mathcal{P}_n, \mu_P)} \leq 1$ ). Hence, under the appropriate choice of  $\Omega$  and  $\mu$  as above, the sets of Wigner- and characteristic functions for continuous and discrete variable quantum systems respectively provide examples of such sets  $\mathcal{S}$ , i.e.  $\mathcal{S}_W \subset \mathcal{S}_{\text{max}}$  and  $\mathcal{S}_P \subset \mathcal{S}_{\text{max}}$ .

For all such sets  $\mathcal{S}$ , we provide an explicit algorithm for estimating the overlap (1.5) yielding an upper bound on  $N_{\varepsilon,\delta}(f, \mathcal{S})$  (c.f. Theorem 3.1) and furthermore, through some information theoretic argument, the corresponding lower bound (c.f. Theorem 3.3). For the particular choice of  $\mathcal{S} = \mathcal{S}_{\text{max}}$ ,

---

<sup>1</sup>More precisely, we choose in Section 5 the set  $\Omega_P = \mathcal{P}_n \setminus \{\mathbb{1}\}$  instead of  $\mathcal{P}_n$  as on the identity string the characteristic function of any state  $\sigma$  trivially satisfies  $\chi_\sigma(\mathbb{1}) = 1$  and hence no information is gained when measuring the observable  $\mathbb{1}$  in  $\sigma$ .

these upper and lower bounds can be shown to be essentially matching in their scaling in the relevant variables  $f, \varepsilon, \delta$  since they are given as

$$\left(\frac{\|f\|_1^{(2\varepsilon)}}{2\varepsilon}\right)^2 \log\left(\frac{1}{3\delta}\right) \lesssim N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) \lesssim \inf_{\varepsilon' \in [0, \varepsilon]} \left(\frac{\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'}\right)^2 \log\left(\frac{1}{\delta}\right). \quad (1.7)$$

Here, we have denoted the *smoothed*  $L^1$ -norm of the function  $f$  by

$$\|f\|_1^{(\varepsilon)} = \inf_{\substack{\tilde{f} \in L^1(\Omega) \cap L^2(\Omega), \\ \|f - \tilde{f}\|_2 \leq \varepsilon}} \|\tilde{f}\|_1. \quad (1.8)$$

Note that (1.7) provides an *instance-optimal* characterisation for the task of black box overlap estimation, i.e., it identifies precisely the property of  $f$  that governs the sample complexity as the smoothed  $L^1$ -norm. The study of instance-optimal bounds is well-developed in the learning literature, for example for the problem of identity testing, one of the flagship results is that classical identity testing is characterised by a smoothed version of the  $L^{2/3}$ -quasinorm of the distribution to be tested [VV17]; see also [CLO22] for results about quantum identity testing in this direction.

For functions  $f$  with rapidly decaying tail as defined in Section 3 and discussed in detail in Appendix A.1, which include for example exponentially decaying functions on Euclidean space, we find that the sample complexity is characterised by the usual  $L^1$ -norm as

$$N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) = \Theta\left(\left(\frac{\|f\|_1}{\varepsilon}\right)^2 \log\left(\frac{1}{\delta}\right)\right), \quad (1.9)$$

with precise statement given in Corollary 3.4.

**Fidelity estimation in Wigner model:** In Theorem 4.1 and 4.2 we give upper and lower bounds on the sample complexity of fidelity estimation in the Wigner model,  $N_{\varepsilon,\delta}^W(\rho)$ , for pure states  $\rho$  on  $L^2(\mathbb{R}^m)$  from the ones established for the abstract framework of black box overlap estimation. In particular, the upper bound involves the smoothed  $L^1$ -norm of  $W_\rho$ , giving that for every fixed pure state  $\rho$  and in the worst case over all states  $\sigma$ , the fidelity  $F(\rho, \sigma)$  can be estimated by performing displaced parity measurements on a finite number of copies of  $\sigma$ , i.e.  $N_{\varepsilon,\delta}^W(\rho) < \infty$ .

Showing for general pure states  $\rho$  that these upper and lower bounds on  $N_{\varepsilon,\delta}^W(\rho)$  are essentially matching similarly to (1.7) and (1.9) is, however, hard as the set of Wigner functions,  $\mathcal{S}_W$ , is a proper and complicated subset of  $\mathcal{S}_{\max} = B_1^2(0) \cap B_{(2/\pi)^m}^\infty(0)$ . Hence, we focus on interesting examples of states of continuous variable systems and establish for those that the sample complexity can, up to constants, be expressed in terms of the  $L^1$ -norm of the corresponding Wigner function:

Firstly, we consider the Fock states  $(|n\rangle\langle n|)_{n \in \mathbb{N}}$ , i.e. the eigenstates of the harmonic oscillator or number operator. In this case, we find in Proposition 4.5 that their sample complexity of fidelity estimation in the Wigner model is fully characterised as

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) = \Theta\left(\left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon}\right)^2 \log\left(\frac{1}{\delta}\right)\right). \quad (1.10)$$

Here, the  $L^1$ -norm of the Wigner functions of the Fock states satisfies the scaling behaviour

$$\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n} \quad (1.11)$$

as shown in (4.19) below by straightforwardly employing known asymptotic expressions of the  $L^p$ -norms of the Laguerre polynomials in [Mar82]. Note that shows that there exist states with arbitrarily large sample complexity.

As another example, we construct in Section 4.1.2 the so-called *spike states* for any  $m$ -mode continuous variable quantum system. These are pure states  $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$  on the Hilbert space  $L^2(\mathbb{R}^m)$  whose Wigner functions are uniformly vanishing as  $n \rightarrow \infty$ , i.e.

$$\|W_{\rho_n}\|_{\infty} \xrightarrow[n \rightarrow \infty]{} 0. \quad (1.12)$$

Furthermore, we show that  $W_{\rho_n} \in L^1(\mathbb{C}^m)$  for all  $n \in \mathbb{N}$  with  $L^1$ -norms blowing up and scaling reciprocally to their  $L^\infty$ -norms, i.e.

$$\|W_{\rho_n}\|_1 \sim \|W_{\rho_n}\|_{\infty}^{-1}. \quad (1.13)$$

Using (1.12) and (1.13) together with the established bounds on  $N_{\varepsilon, \delta}^W(\rho)$ , we show in Proposition 4.9 that also for the spike states the corresponding  $L^1$ -norm characterises the sample complexity of fidelity estimation, i.e. precisely

$$N_{\varepsilon, \delta}^W(\rho_n) = \Theta \left( \left( \frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right). \quad (1.14)$$

We believe that the existence of a sequence of pure states satisfying (1.12) and (1.13) can be of independent interest.

For  $\rho$  being a Gaussian state, we know that  $W_{\rho} \geq 0$  and hence by normalisation of the Wigner function that  $\|W_{\rho}\|_1 = 1$ . From this, we see in Proposition 4.10 that the sample complexity of fidelity estimation in the Wigner model satisfies for all pure Gaussian states  $\rho$  satisfies

$$N_{\varepsilon, \delta}^W(\rho) = \Theta \left( \frac{1}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right) \quad (1.15)$$

and hence that for all states  $\sigma$ , the fidelity  $F(\rho, \sigma)$  can efficiently be estimated using displaced parity measurements on a few copies of  $\sigma$ .

In addition, looking at bounds that only depend on the  $L^1$ -norm of the state, we find in Theorem 4.4 a characterisation of the sample complexity in the worst case over all pure states  $\rho$  whose Wigner function has  $L^1$ -norm less or equal than some fixed threshold value. In particular, we see for all  $t \geq 1$

$$\sup_{\substack{\rho \text{ pure state,} \\ \|W_{\rho}\|_1 \leq t}} N_{\varepsilon, \delta}^W(\rho) = \Theta \left( \left( \frac{t}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right), \quad (1.16)$$

showcasing that also in the considered worst case, the sample complexity has the same scaling behaviour as the ones for the Fock, spike and Gaussian states portrayed in (1.10), (1.14) and (1.15) respectively.

Negativity of the Wigner function of quantum states is an indication of non-classicality [KŻ04, AGPF18, TCJ20] as it portrays the fact that the Wigner function deviates from being a proper

probability distribution, implies that the quantum state cannot be written as a convex combination of coherent or, more generally, Gaussian states and furthermore is equivalent to contextuality [BCE22]. It has been shown that Wigner negativity is necessary for quantum advantage, as computations involving only positive Wigner functions are efficiently classically simulable [BSBN02, ME12, VWFE13]. The  $L^1$ -norm of the Wigner function can be regarded as a quantification of Wigner negativity [KŻ04] with it being equal to 1 if and only if the Wigner function is non-negative and otherwise being strictly greater than 1. However, in this context, the  $L^1$ -norm of the Wigner function only provides an upper bound on the cost of classical simulation [ME12, VWFE13, PWB15a, HFT24] as there are states with large  $L^1$ -norm that can still be simulated efficiently [GÁCF20, CFF23]. Interestingly, for the fidelity estimation problem we consider here, we can achieve lower bounds on the sample cost in terms of the Wigner  $L^1$ -norm. The results presented here in (1.7), (1.10), (1.14), (1.15) and (1.16) show an interesting connection between the complexity of learning properties of a quantum state and its non-classicality and we hope the techniques we introduce here, such as the lower bound techniques and the smoothed  $L^1$ -norm, find further applications in the context of non-classicality.

### Fidelity estimation in Pauli model:

Applying again the results from the abstract paradigm of black box overlap estimation to the Pauli model of fidelity estimation yields upper and lower bounds on the corresponding sample complexity  $N_{\varepsilon,\delta}^P(\rho)$  for all pure states  $\rho$  on a  $n$ -qubit system, c.f. Theorem 5.1 and Theorem 5.3.

In Proposition 5.5 we use these to find for Haar random pure states  $\rho$  with probability at least  $1 - 1/d$  that the corresponding sample complexity is characterised as

$$N_{\varepsilon,\delta}^P(\rho) = \tilde{\Theta} \left( \left( \frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) = \tilde{\Theta} \left( \frac{d}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right), \quad (1.17)$$

with  $L^1$ -norm defined by  $\|\chi_\rho\|_1 = \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{1\}} |\chi_\rho(P)|$ . Here, we used  $\tilde{\Theta}$  to denote the fact that the upper and lower bounds in (1.17) match up to factors that scale logarithmically in the leading order term. In particular, the result implies that in the worst case over all pure states  $\rho$  the sample complexity blows up as  $d \rightarrow \infty$ , i.e.  $\sup_\rho N_{\varepsilon,\delta}^P(\rho) \xrightarrow{d \rightarrow \infty} \infty$ . The  $L^1$ -norm of  $\chi_\rho$  has appeared in the literature under the name of st-norm in [Cam11] and it was shown in [HFH<sup>+</sup>22] that it provides a lower bound on a magic measure obtained via an encoding into continuous variable systems.

For stabiliser states  $\rho$  on  $n$  qubits, we have that  $\|\chi_\rho\|_1 = \frac{d-1}{d} = \Theta(1)$  (Proposition 5.6). Hence,  $N_{\varepsilon,\delta}^P(\rho)$  is independent of the system size and the fidelity  $F(\rho, \sigma)$  can be estimated efficiently by performing Pauli measurements on a few copies of  $\sigma$ . This was previously shown in [FL11].

## 1.2 Related work

In [FL11, dSLCP11], employing the overlap formula (1.3), upper bounds on the sample complexity of fidelity estimation in the Pauli model have been established. In particular in [FL11] the upper bound in Proposition 5.6 for the sample complexity for stabiliser states has already been found as well as the worst case bound  $\sup_{\rho \text{ pure}} N_{\varepsilon,\delta}^P(\rho) = \mathcal{O} \left( \frac{1}{\varepsilon^2 \delta^2} + \frac{d}{\varepsilon^2 \delta} \log \left( \frac{1}{\delta} \right) \right)$  (using the Markov argument around Eq. (10) in [FL11]) or  $\sup_{\rho \text{ pure}} N_{\varepsilon,\delta}^P(\rho) = \mathcal{O} \left( \frac{1}{\varepsilon^2 \delta} + \frac{d}{\varepsilon^4} \log \left( \frac{1}{\delta} \right) \right)$  (using the “truncating bad events” technique)<sup>2</sup>. In Remark 5.2, we improve these worst case bounds by establishing

<sup>2</sup>In the main text of [FL11] the authors claim to have proven  $\sup_{\rho \text{ pure}} N_{\varepsilon,\delta}^P(\rho) = \mathcal{O} \left( \frac{1}{\varepsilon^2 \delta} + \frac{d}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right)$ . However, considering their proof in the appendix (in particular the “Truncating bad events” section), the error parameter  $\beta$



$\sup_{\rho \text{ pure}} N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}\left(\frac{d}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right)$ . Furthermore, we show that this worst case sample complexity is tight (up to a logarithmic factor in the dimension) for Haar random pure state  $\rho$  with high probability (c.f. Proposition 5.5).

In [GPR23], an upper bound on the sample complexity in terms of the  $L^1$ -norm is given for the problem of estimating the expectation values of observables in the Pauli model. In this work, we use a similar sampling rule to derive (with a simpler analysis) our upper bounds. Moreover, we provide nearly matching lower bounds in the case of observables being either Haar random pure states (c.f. Proposition 5.5) or stabilizer states (c.f. Proposition 5.6) as well as all pure observables in the low accuracy regime (c.f. Remark 5.4).

In [dSLCP11, SEL+22] using the overlap formula in terms of Wigner functions (1.1) for fidelity estimation of CV systems is discussed, though without establishing bounds on the corresponding sample complexity. The work [SEL+22] as well as [PWB15b], in which a method for estimating the probabilities of outcomes of a quantum circuit is presented, rely on a sampling rule similar to ours (c.f. (3.12)). Moreover, in both works it is claimed that this sampling procedure is optimal as it has the smallest variance. We stress that this optimality proof has implicit assumptions: the estimator should be of a particular form (c.f. [PWB15b, Eq. (14)] or [SEL+22, Eq. (5)]), the sampling distribution remains the same during different steps, and the global estimator is given by the empirical one. In fact, we know that for some CV states for which the Wigner function has infinite  $L^1$ -norm [Sim92, Theorem 2], estimators satisfying these assumptions have infinite variance but nonetheless our bound (1.7) gives an algorithm with finite sample complexity. Here, we prove our lower bounds on the sample complexity (c.f. Theorems 3.3, 4.2 and 5.3) without making these assumptions. In particular, we allow the learner to choose adaptively the sampling procedures. This allows us to prove the optimality of the sampling rule (3.12) (applied on the corresponding smoothing function  $\tilde{f}$ ) for  $\mathcal{S} = \mathcal{S}_{\max}$  (c.f. (1.7)), for the Fock, spike and Gaussian states as well as the worst case in the Wigner model (c.f. (1.10), (1.14), (1.15) and (1.16)) and for Haar random and stabiliser states in the Pauli model (c.f. (1.17) and Proposition 5.6).

In [CDG+20] the authors consider fidelity estimation of CV systems using heterodyne measurements<sup>3</sup> instead of performing Wigner tomography through displaced parity measurements, and establish upper bounds on the corresponding sample complexity.

The recent paper [HPS24] considers the task of quantum state certification, i.e. the decision problem of whether the actually prepared state  $\sigma$  is close or far from the target state  $\rho$ , on  $n$  qubit systems from Pauli string measurements. Incorporating the full  $n$ -bit string of measurement outcomes obtained from measuring each Pauli matrix for a given  $P \in \mathcal{P}_n$  individually enables them to prove a  $\mathcal{O}(n^2) = \mathcal{O}(\log^2(d))$  upper bound on the corresponding sample complexity for almost all Haar random pure states. This is in sharp contrast to the lower bound we found in (1.17), which is exponential in the number of qubits due to only using the binary measurement outcome from the full Pauli string  $P \in \mathcal{P}_n$ .

Quantum state tomography for continuous variable systems for different measurement setups has been extensively studied: In [LR09] quantum state tomography using homodyne detection,<sup>4</sup> a

---

introduced there needs to scale as  $\varepsilon$  in order estimate the fidelity with the desired accuracy leading to a  $1/\varepsilon^4$  of their actual upper bound.

<sup>3</sup>Heterodyne measurements are given by performing the POVM  $(\frac{1}{\pi^m}|\alpha\rangle\langle\alpha|)_{\alpha \in \mathbb{C}^m}$  defined through the coherent states.

<sup>4</sup>Homodyne detection is given by measuring the observable  $X_\theta = \cos(\theta)X + \sin(\theta)P$  where,  $X$  and  $P$  are the standard position and momentum operators and  $\theta$  is an angle, which can be freely chosen and can be seen as rotating phase space.

practically implementable measurement scheme, which differs from displaced parity measurements considered here, has been studied. Furthermore, continuous variable tomography under energy constraints on the unknown state has been a topic of interest in the literature: Restricting to classical shadow tomography and using homodyne, heterodyne, photon number resolving<sup>5</sup> or photon parity measurements, [GAG<sup>+</sup>23, BDLR24] establish performance guarantees by providing upper bounds on the corresponding sample complexity provided the unknown state satisfies certain energy constraints with respect to the photon number operator. In the recent paper [MMB<sup>+</sup>24], lower bounds (for general mixed states) and upper bounds (for pure and Gaussian states) on the sample complexity of full quantum state tomography using general measurements are provided. There, the authors prove their results again under the assumption that the unknown state satisfies an energy constraint. Interestingly, their lower bound implies that the scaling of the sample complexity in terms of error parameter  $\varepsilon$  worsen when the number of modes  $m$  increases. Note that in comparison, our analysis here does not rely on an energy constraint on either the target state  $\rho$  or prepared state  $\sigma$ . In fact, our results are for arbitrary states  $\sigma$  and involve only the (smoothed)  $L^1$ -norm of the Wigner function of  $\rho$ , which can remain bounded for arbitrarily high energy of  $\rho$ .<sup>6</sup>

### 1.3 Outline of the article

The rest of the article is outlined as follows:

- In Section 2 we discuss some preliminary facts which are needed for the derivations of the rest of the paper.
- In Section 3 we formally define the task of black box overlap estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 3.1 and 3.3 and its classification for functions with rapidly decaying tail in Corollary 3.4.
- In Section 4 we formally define the task of fidelity estimation in the Wigner model, discuss its connection to the abstract black box estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 4.1 and 4.2. In Section 4.1 we discuss instances for which these upper and lower bounds can be shown to be matching up to constants including the Fock states in Section 4.1.1, spike states in Section 4.1.2, Gaussian states in Section 4.1.3 and the worst case behaviour of the sample complexity in Theorem 4.4.
- In Section 5 we formally define the task of fidelity estimation in the Pauli model, discuss its connection to the abstract black box estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 5.1 and 5.3. We then continue to show in Section 5.1 that these upper and lower bounds are essentially matching for Haar random pure states in Proposition 5.5 and for stabiliser states in Proposition 5.6.
- In the Appendix we study properties of the smoothed  $L^1$ -norm. In Lemma A.1 we show continuity in the sense of  $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$ . We then provide convergence rates for this limit for functions with rapidly decaying tail in Appendix A.1 with explicit example being the Wigner functions of the Fock states.

---

<sup>5</sup>Photon number resolving measurements are given by performing the POVM  $(\frac{1}{\pi^m} D(\alpha)|n\rangle\langle n|D(\alpha)^*)_{\alpha \in \mathbb{C}^m}$  for different photon numbers  $n \in \mathbb{N}$ .

<sup>6</sup>This can for example be seen by considering the displaced state  $\rho_\alpha = D(\alpha)\rho D(\alpha)^*$  which satisfies  $\|W_{\rho_\alpha}\|_1 = \|W_\rho\|_1$  for all  $\alpha \in \mathbb{C}^m$  but  $\text{Tr}(N\rho_\alpha) \xrightarrow{|\alpha| \rightarrow \infty} \infty$ .

## 2 Preliminaries

In this section we discuss some preliminary definitions and facts which are needed for the derivations of the main results of this paper.

### 2.1 Notation

For  $X$  being a set and  $f, g : X \rightarrow \mathbb{R}$  being two functions we write  $f \lesssim g$  if there exists  $C \geq 0$  such that for all  $x \in X$  we have  $f(x) \leq Cg(x)$  and furthermore  $f = \mathcal{O}(g)$  if  $|f| \lesssim |g|$ . Similarly we write  $f \gtrsim g$  if there exists  $c > 0$  such that  $f(x) \geq cg(x)$  for all  $x \in X$  and  $f = \Omega(g)$  if  $|f| \gtrsim |g|$ . We denote by  $f \sim g$  that  $f \lesssim g$  and  $g \gtrsim f$  and by  $f = \Theta(g)$  that  $f = \mathcal{O}(g)$  and  $f = \Omega(g)$ .

### 2.2 $L^q$ -spaces and smoothed $L^1$ -norm

Let  $\Omega$  be some measurable space with some measure  $\mu$ . In the following we denote for  $q \in [1, \infty]$  by  $L^q(\Omega, \mathbb{C})$  and  $L^q(\Omega, \mathbb{R})$  the Banach spaces of complex and real valued  $L^q$ -functions respectively. For brevity we often denote for the case of complex valued functions  $L^q(\Omega) \equiv L^q(\Omega, \mathbb{C})$ . We denote the ball of real valued  $L^q$ -functions with radius  $r \geq 0$  centered around some  $f \in L^q(\Omega, \mathbb{R})$  by

$$B_r^q(f) = \left\{ g \in L^q(\Omega, \mathbb{R}) \mid \|f - g\|_q \leq r \right\} \quad \text{where} \quad \|f\|_q = \left( \int_{\Omega} |f(\lambda)|^q d\mu(\lambda) \right)^{1/q}. \quad (2.1)$$

The *smoothed  $L^1$ -norm*<sup>7</sup> of a function  $f \in L^2(\Omega, \mathbb{R})$  for  $\varepsilon \geq 0$  is defined to be

$$\|f\|_1^{(\varepsilon)} = \inf_{\substack{\tilde{f} \in L^1(\Omega, \mathbb{R}) \\ \|f - \tilde{f}\|_2 \leq \varepsilon}} \|\tilde{f}\|_1. \quad (2.2)$$

As seen in Lemma A.1 in the Appendix, the smoothed  $L^1$ -norm satisfies the natural continuity property  $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$ .

Note that for all  $f \in L^2(\Omega, \mathbb{R})$  and  $\varepsilon > 0$  the respective smoothed  $L^1$ -norm is always finite, i.e.  $\|f\|_1^{(\varepsilon)} < \infty$ , which follows since  $L^1(\Omega, \mathbb{R}) \cap L^2(\Omega, \mathbb{R})$  is dense in  $L^2(\Omega, \mathbb{R})$ .<sup>8</sup>

### 2.3 Information theoretic divergences

The Kullback-Leibler (KL) divergence between two probability distributions  $P$  and  $Q$  on  $\mathcal{X}$  is defined by

$$\text{KL}(P\|Q) = \int_{\mathcal{X}} dP(x) \log \left( \frac{dP}{dQ}(x) \right)$$

whenever  $P$  is absolutely continuous with respect to  $Q$  and  $\text{KL}(P\|Q) = +\infty$  otherwise.

<sup>7</sup>Note that the smoothed  $L^1$ -norm is not a norm itself as it does not satisfy triangle inequality or faithfulness due to the employed infimum.

<sup>8</sup>Density can be seen by defining for  $c > 0$  the function  $\tilde{f}_c = f \mathbf{1}_{\{|f| \geq c\}}$  and noting that by the dominated convergence theorem it approximates  $f$  in  $L^2$ -norm as  $\lim_{c \rightarrow 0} \int_{\Omega} |\tilde{f}_c(\lambda) - f(\lambda)|^2 d\mu(\lambda) = \lim_{c \rightarrow 0} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) = 0$ . Furthermore, note that for all  $c > 0$  we have  $\tilde{f}_c \in L^1(\Omega, \mathbb{R})$  as  $\int_{\{|f| \geq c\}} |f(\lambda)| d\mu(\lambda) \leq \|f\|_2^2 / c$ .

For  $p, q \in [0, 1]$ , we write  $\text{KL}(p\|q) = \text{KL}(\text{Ber}(p)\|\text{Ber}(q))$ , where  $\text{Ber}(p)$  denotes a Bernoulli distribution with parameter  $p$ . Note that if  $\delta \leq 1/2$ , and  $p \geq 1 - \delta$  and  $q \leq \delta$  we have  $\text{KL}(p\|q) \geq \text{KL}(1 - \delta\|\delta) \geq \log(1/(3\delta))$ .

Furthermore, the  $\chi^2$ -divergence is defined by

$$\chi^2(P\|Q) = \int_{\mathcal{X}} \left( \frac{dP}{dQ}(x) - 1 \right)^2 Q(dx) \quad (2.3)$$

if  $P$  is absolutely continuous with respect to  $Q$  and  $+\infty$  otherwise.

These divergences satisfy the well-known bound

$$\text{KL}(P\|Q) \leq \chi^2(P\|Q). \quad (2.4)$$

## 2.4 Wigner functions for continuous variable systems

For  $m \in \mathbb{N}$  we consider a  $m$ -mode quantum harmonic oscillator on the Hilbert space  $\mathcal{H} = L^2(\mathbb{R}^m)$ . For a point in phase space  $\alpha \in \mathbb{C}^m \cong \mathbb{R}^{2m}$  the *displacement operator* is defined as

$$D(\alpha) = e^{\sum_{k=1}^m (\alpha_k a_k^* - \alpha_k^* a_k)} = \bigotimes_{k=1}^m e^{\alpha_k a_k^* - \alpha_k^* a_k}, \quad (2.5)$$

where we denoted by  $a_k$  and  $a_k^*$  the canonical bosonic annihilation and creation operators of the  $k^{\text{th}}$  mode (see e.g. [BR97, Chapter 5.2] or [Tes14, Chapter 8.3] for a definition and discussion). Note that  $D(\alpha)$  is a unitary operator with  $D^*(\alpha) = D(-\alpha)$  and satisfying the relation  $D(\alpha)D(\beta) = e^{(\alpha \cdot \beta^* - \alpha^* \cdot \beta)/2} D(\alpha + \beta)$ , where we denoted the real dot product between two complex vectors  $\alpha, \beta \in \mathbb{C}^m$  by  $\alpha \cdot \beta = \sum_{k=1}^m \alpha_k \beta_k$ .

For a state  $\rho$  we can define its *characteristic function* as

$$\chi_\rho(\alpha) := \text{Tr}(\rho D(\alpha)). \quad (2.6)$$

The characteristic function is complex valued and satisfies  $\chi_\rho^*(\alpha) = \chi_\rho(-\alpha)$  and furthermore  $\chi_\rho \in L^q(\mathbb{C}^m) \cong L^q(\mathbb{R}^{2m})$  for all  $q \in [2, \infty]$  [Wer84]. Therefore, we can define the so-called *Wigner function* of  $\rho$  as the (symplectic) Fourier transform of its characteristic function, i.e. formally<sup>9</sup> through the integral expression

$$W_\rho(\alpha) := \frac{1}{\pi^{2m}} \int_{\mathbb{C}^m} e^{\alpha \cdot \beta^* - \alpha^* \cdot \beta} \chi_\rho(\beta) d\beta = \frac{1}{\pi^{2m}} \int_{\mathbb{C}^m} e^{2i\Im(\alpha \cdot \beta^*)} \chi_\rho(\beta) d\beta, \quad (2.7)$$

with  $d\beta \equiv d\Re(\beta)d\Im(\beta)$  denoting the Lebesgue measure on  $\mathbb{C}^m$ . Note, that the Wigner function (2.7) is real valued and can have positive and negative values. For  $\rho = |\psi\rangle\langle\psi|$  being a pure state, the Wigner function is directly given by

$$W_\rho(x, p) = \left( \frac{2}{\pi} \right)^m \int_{\mathbb{R}^m} \psi^*(x + y) \psi(x - y) e^{i2p \cdot y} dy, \quad (2.8)$$

---

<sup>9</sup>The integral expression (2.7) is well-defined for  $\chi_\rho$  being a Schwartz function or more generally in  $L^1(\mathbb{C}^m) \cap L^2(\mathbb{C}^m)$ . For general functions in  $L^2(\mathbb{C}^m)$  the Fourier transform is then extended by density in the usual way, see e.g. [Tes14, Section 7.1].

where we changed coordinates to position and momentum variables  $(x, p) \in \mathbb{R}^{2m}$  through the relation  $\alpha = \frac{1}{\sqrt{2}}(x + ip)$ . Note the integral in (2.8) is well-defined as the integrand is in  $L^1(\mathbb{R}^m)$  and furthermore that this definition of the Wigner function can be extended to general mixed states by linearity.

The Wigner function  $W_\rho$  for any state  $\rho$  is a *quasiprobability distribution* as it can have positive and negative values and satisfies the normalisation<sup>10</sup>

$$\int_{\mathbb{C}^m} W_\rho(\alpha) d\alpha = \frac{1}{2^m} \int_{\mathbb{R}^{2m}} W_\rho(x, p) dx dp = 1. \quad (2.9)$$

An important class of states which have non-negative Wigner function are *Gaussian states* with defining property being that their Wigner (or equivalently characteristic) function is a Gaussian function. Notably, by Hudson's theorem [Hud74, SC83], a pure state has non-negative Wigner function if and only if it is Gaussian. On the other hand, the set of general mixed states with non-negative Wigner function has been shown to be strictly bigger than the convex hull of all Gaussian states and is itself hard to characterise [MKC09, FMcv11].

The Hilbert-Schmidt inner product between two states  $\rho$  and  $\sigma$  can by Plancherel's identity be expressed as the inner product of the corresponding characteristic- or Wigner functions as

$$\text{Tr}(\rho\sigma) = \frac{1}{\pi^m} \int_{\mathbb{C}^m} \chi_\rho^*(\alpha) \chi_\sigma(\alpha) d\alpha = \pi^m \int_{\mathbb{C}^m} W_\rho(\alpha) W_\sigma(\alpha) d\alpha. \quad (2.10)$$

Hence, in particular we know for  $\rho$  being a state that

$$\|W_\rho\|_2 = \frac{1}{\pi^{m/2}} \|\rho\|_2 \leq \frac{1}{\pi^{m/2}} \quad (2.11)$$

where we denoted the Hilbert-Schmidt norm of  $\rho$  by  $\|\rho\|_2$  and used the fact that it is dominated by the trace norm. Alternatively to (2.7), we can write the Wigner function of  $\rho$  as the expectation value of the *displaced parity operator* [Roy77], i.e.

$$W_\rho(\alpha) = \left(\frac{2}{\pi}\right)^m \text{Tr}\left(\rho D(\alpha) \Pi D^*(\alpha)\right) \quad (2.12)$$

with  $\Pi$  being the *parity operator* defined through the relation  $(\Pi\psi)(x) := \psi(-x)$  for all  $\psi \in L^2(\mathbb{R}^m)$  or alternatively in Fock basis

$$\Pi = \sum_{n_1, \dots, n_m \in \mathbb{N}_0} (-1)^{n_1 + \dots + n_m} |n_1, \dots, n_m\rangle \langle n_1, \dots, n_m|.$$

From that we see that the Wigner function is bounded with

$$\|W_\rho\|_\infty \leq \left(\frac{2}{\pi}\right)^m. \quad (2.13)$$

From (2.11) and (2.13) and Hölder's inequality we see that  $W_\rho \in L^q(\mathbb{C}^m)$  for all  $q \in [2, \infty]$ . However, for all  $q \in [1, 2)$  there exists explicit (pure) states  $\rho$  such that  $W_\rho \notin L^q(\mathbb{C}^m)$  [Sim92, Theorem 2]. The set of functions  $\psi \in L^2(\mathbb{R}^m)$  whose Wigner function of the corresponding (possibly unnormalised) pure state satisfies  $W_{|\psi\rangle\langle\psi|} \in L^1(\mathbb{C}^m)$  is hence a proper subset of  $L^2(\mathbb{R}^m)$ . It is called the *Feichtinger algebra* and possesses rich mathematical properties which are extensively studied in the literature [DG17, Chapter 7].

---

<sup>10</sup>To be precise, this property holds rigorously for all states  $\rho$  such that  $W_\rho \in L^1(\mathbb{C}^m)$  [DG17, Proposition 18], which is a proper subset of the set of all states as discussed below.

### 3 Black box overlap estimation

We consider some measurable space  $\Omega$  with measure  $\mu$  and employ the notation of Section 2.2. Let  $r > 0$  be some number, which is fixed throughout the whole section. We say one has *black box access to some function*  $g \in B_r^\infty(0)$  if for every  $\lambda \in \Omega$  one can obtain samples from the unique<sup>11</sup> random variable  $Y_\lambda \equiv Y_\lambda(g)$  with outcome space  $\{r, -r\}$  and

$$\mathbb{E}[Y_\lambda] = g(\lambda). \quad (3.1)$$

In the following we consider subsets  $\mathcal{S} \subseteq B_r^\infty(0)$  and assume to have black box access to the functions  $g \in \mathcal{S}$ .

With that terminology the task of *black box overlap estimation* is defined to be the following: Let  $f \in L^2(\Omega, \mathbb{R})$  and  $\mathcal{S} \subseteq L^2(\Omega, \mathbb{R}) \cap B_r^\infty(0)$ . Then, provided full access to the function  $f$  and black box access to a function  $g \in \mathcal{S}$ , the goal is to estimate

$$\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) = \int_{\Omega} f(\lambda) \mathbb{E}[Y_\lambda]d\mu(\lambda) \quad (3.2)$$

using as few samples as possible of the random variable  $Y_\lambda \equiv Y_\lambda(g)$  for different points  $\lambda \in \Omega$ . More precisely, for  $N \in \mathbb{N}$ , a *N-query algorithm*  $\mathcal{A}$  for the task of black box overlap estimation takes points  $\lambda_1, \dots, \lambda_N \in \Omega$  and for each  $\lambda_t$  a sample  $y_{\lambda_t} \in \{r, -r\}$  from  $Y_{\lambda_t}$ . Here, we allow for adaptive algorithms meaning that the points  $\lambda_t$ , or their distributions when chosen at random, can depend on  $f$  and the previous points  $\lambda_1, \dots, \lambda_{t-1}$  and corresponding samples, i.e.

$$\begin{aligned} \lambda_1 &\equiv \lambda_1(f) && \longrightarrow y_{\lambda_1} \text{ sample from } Y_{\lambda_1} \\ \lambda_2 &\equiv \lambda_2(f, \lambda_1, y_{\lambda_1}) && \longrightarrow y_{\lambda_2} \text{ sample from } Y_{\lambda_2} \\ &\vdots && \vdots \\ \lambda_N &\equiv \lambda_N(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_{N-1}, y_{\lambda_{N-1}}) && \longrightarrow y_{\lambda_N} \text{ sample from } Y_{\lambda_N}. \end{aligned} \quad (3.3)$$

The algorithm then outputs a value  $\phi_{\mathcal{A}}(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_N, y_{\lambda_N}) \in \mathbb{R}$ . We demand that  $\mathcal{A}$  works for any function  $g \in \mathcal{S}$ , i.e. for all  $g \in \mathcal{S}$  fixed its output should estimate the corresponding overlap (3.2). We say the algorithm  $\mathcal{A}$  has accuracy  $\varepsilon > 0$  with failure probability bounded by  $\delta > 0$  if for all  $g \in \mathcal{S}$

$$\mathbb{P} \left( \left| \phi_{\mathcal{A}}(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_N, y_{\lambda_N}) - \int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) \right| > \varepsilon \right) \leq \delta.$$

The optimal sample complexity of black box overlap estimation for the function  $f$  and set  $\mathcal{S}$  is hence defined as

$$N_{\varepsilon, \delta}(f, \mathcal{S}) = \inf \left\{ N \in \mathbb{N} \mid \exists N\text{-query algorithm with accuracy } \varepsilon \text{ and failure probability bounded by } \delta \right\}.$$

Theorem 3.1 below, with proof given in Section 3.1, proposes an algorithm achieving this objective and provides an upper bound on the needed number of points  $\lambda$  to take samples from  $Y_\lambda$ . Furthermore, Theorem 3.3 provides corresponding lower bounds and is proved in Section 3.2. We

<sup>11</sup>Note that the condition on  $Y_\lambda$  uniquely determines its distribution to be  $\mathbb{P}(Y_\lambda = r) = \frac{1+g(\lambda)/r}{2}$  and  $\mathbb{P}(Y_\lambda = -r) = \frac{1-g(\lambda)/r}{2}$

see that in the case of  $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$ , the optimal sample complexity is characterised by the smoothed  $L^1$ -norm of  $f$  (see (1.8) for the definition) as we show for all  $\varepsilon, \delta > 0$

$$\left( \frac{r \|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \log \left( \frac{1}{3\delta} \right) \leq N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) \leq 2 \inf_{\varepsilon' \in [0, \varepsilon]} \left( \frac{r \|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right). \quad (3.4)$$

**Theorem 3.1 (Upper bound for black box overlap estimation)** *Let  $f \in L^2(\Omega, \mathbb{R})$ ,  $r > 0$  and  $\mathcal{S} \subseteq B_1^2(0) \cap B_r^\infty(0)$ . Then we have*

$$N_{\varepsilon, \delta}(f, \mathcal{S}) \leq 2 \inf_{\varepsilon' \in [0, \varepsilon]} \left( \frac{r \|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right) \quad (3.5)$$

*The upper bound holds true even when restricting to non-adaptive algorithms, i.e. where the  $\lambda_t$  in (3.3) can depend on the function  $f$  but not on the previous points  $\lambda_1, \dots, \lambda_{t-1}$  and the corresponding samples  $y_{\lambda_1}, \dots, y_{\lambda_{t-1}}$ .*

**Remark 3.2** *The algorithm proposed in the proof of Theorem 3.1 also works more generally if instead of having access to the two-outcome random variable  $Y_\lambda$ , we could sample from a bounded random variable  $\tilde{Y}_\lambda$  satisfying*

$$|\tilde{Y}_\lambda| \leq r \quad \text{and} \quad \mathbb{E}[\tilde{Y}_\lambda] = g(\lambda) \quad (3.6)$$

*for all  $\lambda \in \Omega$ . In fact, for this slightly changed model of black box overlap estimation, we obtain the same upper bound as (3.5) for the corresponding sample complexity.*

**Theorem 3.3 (Lower bound for black box overlap estimation)** *Let  $f \in L^2(\Omega, \mathbb{R})$ ,  $r > 0$  and  $\mathcal{S} \subseteq L^2(\Omega, \mathbb{R}) \cap B_r^\infty(0)$ . Then for  $\varepsilon, \delta > 0$  the sample complexity of black box overlap estimation is lower bounded as*

$$N_{\varepsilon, \delta}(f, \mathcal{S}) \geq \sup_{\substack{g_1, g_2 \in \mathcal{S}, \\ |f(g_1 - g_2)| > 2\varepsilon}} \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_\infty^{-1} \log \left( \frac{1}{3\delta} \right). \quad (3.7)$$

*Furthermore, for  $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$  and  $f \in \mathcal{S}_{\max}$  with  $\|f\|_2 = 1$  this gives*

$$N_{\varepsilon, \delta}(f, \mathcal{S}) \geq \left( \frac{r \|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \log \left( \frac{1}{3\delta} \right). \quad (3.8)$$

For a special class of functions  $f$ , satisfying a certain decay of tail condition, we can show in Corollary 3.4 below that the upper and lower bounds in (3.4) match up to constants. Here, it is important to have good control over the behaviour of the smoothed  $L^1$ -norm  $\|f\|_1^{(\varepsilon)}$  for different values of  $\varepsilon$ . As we see below, the sample complexity of black box overlap estimation is in this case then characterised by the usual  $L^1$ -norm of the function  $f$ .

For  $\gamma, \kappa > 0$  and  $\Omega_0 \subseteq \Omega$  a finite measure set, we say a measurable function  $f$  has a *rapidly decaying tail of order  $(\gamma, \kappa)$  outside  $\Omega_0$*  if for all  $\delta > 0$  we have

$$\int_{\{|f| \leq \delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \kappa \delta^\gamma. \quad (3.9)$$

Note that a function  $f \in L^2(\Omega, \mathbb{R})$  satisfying this condition can be also shown to be in  $L^1(\Omega, \mathbb{R})$ . Functions satisfying this condition are discussed in detail in Appendix A.1. In particular, in Lemma A.2 it is shown that their smoothed  $L^1$ -norm satisfies the lower bound

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \kappa' \varepsilon^\beta - \varepsilon \sqrt{\mu(\Omega_0)} \quad (3.10)$$

with  $\beta := \frac{\gamma}{\gamma+1} \in (0, 1]$  and  $\kappa' := 2\kappa^{\frac{1}{1+\gamma}}$ .

**Corollary 3.4** *Let  $f \in \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$  with  $\|f\|_2 = 1$  and rapidly decaying tail of order  $(\gamma, \kappa)$  outside  $\Omega_0$  for some  $\gamma, \kappa > 0$  and finite measure set  $\Omega_0 \subseteq \Omega$  such that  $\kappa, \sqrt{\mu(\Omega_0)} \leq c\|f\|_1$  for some universal constant  $c \geq 0$ . Then we have*

$$N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) = \Theta \left( \left( \frac{r\|f\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) \quad (3.11)$$

for  $\varepsilon, \delta > 0$  small enough, i.e. for all  $\varepsilon, \delta \in (0, c']$  where  $c'$  depends on  $c$  but not on  $f$ .

**Proof.** We get from Theorem 3.1 and the definition of the smoothed  $L^1$ -norm in (2.2) the upper bound

$$N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) \leq 2 \left( \frac{r\|f\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right).$$

On the other hand from Lemma A.2 and the fact that  $\kappa, \sqrt{\mu(\Omega_0)} \leq \mathcal{O}(\|f\|_1)$ , we have

$$\|f\|_1^{(2\varepsilon)} \geq \|f\|_1 (1 - C\varepsilon^\beta)$$

for some  $C \geq 0$  independent of  $f$  and  $\varepsilon$ . Hence, by the lower bound (3.8) in Theorem 3.3 we find for all  $\varepsilon, \delta > 0$  small enough

$$N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) \geq \left( \frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \log \left( \frac{1}{3\delta} \right) \gtrsim \left( \frac{r\|f\|_1}{2\varepsilon} \right)^2 \log \left( \frac{1}{3\delta} \right).$$

■

### 3.1 Proof of the upper bound on the sample complexity $N_{\varepsilon, \delta}(f, \mathcal{S})$

In this section we prove the upper bound on the sample complexity of black box overlap estimation given in Theorem 3.1. The proof works by providing a specific algorithm for the estimation task: Here, the tester chooses random points  $\lambda \in \Omega$  by sampling from the distribution

$$p(\lambda) = \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1}, \quad (3.12)$$

where  $\tilde{f}$  is essentially the minimiser in the optimisation of the smoothed  $L^1$ -norm in (2.2). For every chosen  $\lambda$ , a sample of the bounded random variable  $Y_\lambda \|\tilde{f}\|_1 \text{sgn}(f(\lambda))$  is obtained. Using



these samples, an empirical estimator is calculated which, provided the number of samples is high enough, is close to the overlap

$$\int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda)$$

by Hoeffding's inequality [Hoe63]. This overlap approximates the desired one in (3.2) since  $\tilde{f}$  and the target function  $f$  are close in  $L^2$ -norm.

In [FL11, dSLCP11] a similar algorithm has been considered for the concrete case of fidelity estimation in the Pauli model. Formulating their approach in terms in the abstract black box overlap estimation framework, the authors consider the distribution

$$p_2(\lambda) = \frac{f^2(\lambda)}{\|f\|_2^2}, \quad (3.13)$$

instead the distribution  $p(\lambda)$  above, for sampling  $\lambda \in \Omega$ . For every chosen  $\lambda$  they then take samples from the random variable  $\|f\|_2^2 Y_\lambda / f(\lambda)$  to construct an empirical estimator for the overlap (3.2). This random variable is no longer bounded, leading to a worse upper bound on the corresponding sample complexity for many functions  $f$ . Recently, [GPR23] use the same sampling distribution as ours for the problem of estimation the expectation values of observables in the Pauli model. However, their analysis is different than ours.

**Proof of Theorem 3.1.** Let  $\varepsilon > 0$ ,  $\varepsilon' \in [0, \varepsilon)$  and  $\tilde{f} \in L^2(\Omega, \mathbb{R}) \cap L^1(\Omega, \mathbb{R})$  such that  $\|f - \tilde{f}\|_2 \leq \varepsilon'$ . Note that such a function always exists since  $L^2(\Omega, \mathbb{R}) \cap L^1(\Omega, \mathbb{R})$  is dense in  $L^2(\Omega, \mathbb{R})$  as remarked in Section 2.2. Consider the random variable  $\Lambda$  with outcome space  $\Omega$  and distributed according to the probability density

$$p(\lambda) = \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1}.$$

Let  $g \in \mathcal{S} \subseteq B_1^2(0) \cap B_r^\infty(0)$  for which we assume to have black box access and  $Y_\lambda \equiv Y_\lambda(g)$  the corresponding random variable with outcome space  $\{r, -r\}$  satisfying (3.1). Consider the random variable

$$X = Y_\Lambda \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\Lambda)).$$

Note that by definition we have

$$\mathbb{E}[X] = \mathbb{E}[Y_\Lambda \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\Lambda))] = \int_{\Omega} \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1} g(\lambda) \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\lambda)) d\mu(\lambda) = \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda).$$

For  $N \in \mathbb{N}$  to be determined later we take samples  $\lambda_1, \dots, \lambda_N$  of  $\Lambda$  and then for each  $\lambda_i$  a sample  $y_{\lambda_i}$  of  $Y_{\lambda_i}$ . Furthermore, we define the empirical average

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N y_{\lambda_i} \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\lambda_i)).$$

Noting that by definition and assumption (3.1) the random variable  $X$  is bounded as

$$|X| \leq r \|\tilde{f}\|_1,$$

we can employ Hoeffding's inequality [Hoe63] which gives

$$\mathbb{P} \left( \left| \bar{X} - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| \geq \varepsilon - \varepsilon' \right) \leq \exp \left( -\frac{N(\varepsilon - \varepsilon')^2}{2r^2\|\tilde{f}\|_1^2} \right). \quad (3.14)$$

Furthermore note that by Cauchy Schwarz inequality and the assumption that  $\|g\|_2 \leq 1$  we have

$$\left| \int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| \leq \|f - \tilde{f}\|_2 \leq \varepsilon'.$$

Hence, we see that for  $\delta > 0$  a total number of

$$N = 2 \left( \frac{r\|\tilde{f}\|_1}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right)$$

samples suffices to have with probability greater than  $1 - \delta$ :

$$\left| \bar{X} - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| \leq \varepsilon - \varepsilon' \quad \text{and} \quad \left| \int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| \leq \varepsilon'$$

and thus, by the triangle inequality, give a good approximation of the desired overlap as

$$\begin{aligned} \left| \int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) - \bar{X} \right| &\leq \left| \int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| + \left| \bar{X} - \int_{\Omega} \tilde{f}(\lambda)g(\lambda)d\mu(\lambda) \right| \\ &\leq \varepsilon' + \varepsilon - \varepsilon' = \varepsilon. \end{aligned}$$

Since  $\varepsilon' \in [0, \varepsilon)$  and  $\tilde{f}$  under the constraint above were arbitrary, this shows (3.5).  
■

### 3.2 Proof of the lower bound on the sample complexity $N_{\varepsilon, \delta}(f, \mathcal{S})$

Here, we prove the lower bound on the sample complexity of black box overlap estimation given in Theorem 3.3. The first part given in (3.7) follows by some information theoretic argument in which we show hardness of the estimation task by embedding the problem of distinguishing two hypotheses. For the second part in (3.8) showing the lower bound for the choice  $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$ , we use the established (3.7) together with a specific test function  $g \in \mathcal{S}_{\max}$ , whose existence is provided in Lemma 3.5 below. We first state and prove this lemma and then continue to give the proof of Theorem 3.3.

**Lemma 3.5** *Let  $f \in L^2(\Omega, \mathbb{R})$  and  $\varepsilon \in (0, \|f\|_2)$ . Then there exists  $g \in L^2(\Omega, \mathbb{R}) \cap L^\infty(\Omega, \mathbb{R})$  such that the following three points hold:*

1.  $\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) \geq \varepsilon$ ,
2.  $\|g\|_2 \leq 1$ ,
3.  $\|g\|_\infty \leq \frac{\varepsilon}{\|f\|_1^{(\varepsilon)}}$ .

**Remark 3.6** *In the context of Sections 4 and 5 below we consider the specific choices of  $\mathcal{S}$  being either the set of Wigner functions or characteristic functions for valid quantum states on  $L^2(\mathbb{R}^m)$  or  $(\mathbb{C})^{\otimes n}$  respectively. There, the considered  $f$  is a Wigner or characteristic function of some pure state  $\rho$ . However, the function  $g$  constructed in Lemma 3.5 does in general not correspond the Wigner or characteristic function of some state  $\sigma$ . Hence, in the context of fidelity estimation for continuous or discrete variable systems of Sections 4 and 5, we cannot use the lemma above to establish the corresponding the lower bound (3.8) when  $\sigma$  is assumed to be a valid state.*

**Proof of Lemma 3.5.** In the following we explicitly construct a function  $\tilde{f} \in L^1(\Omega, \mathbb{R}) \cap L^2(\Omega, \mathbb{R})$  with

$$\|f - \tilde{f}\|_2 \leq \varepsilon \quad (3.15)$$

and furthermore  $g \in L^2(\Omega, \mathbb{R}) \cap L^\infty(\Omega, \mathbb{R})$  satisfying points 1 and 2 and furthermore

$$\|g\|_\infty \leq \frac{\varepsilon}{\|\tilde{f}\|_1} \leq \frac{\varepsilon}{\|f\|_1^{(\varepsilon)}},$$

where the last inequality followed by the definition of the smoothed  $L^1$ -norm in (2.2).

For that, note first of all that by the dominated convergence theorem we have  $\lim_{c \rightarrow 0} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) = 0$  and hence we can define

$$c_\star = \sup \left\{ c \in [0, \infty] \mid \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) \leq \varepsilon^2 \right\} > 0.$$

Note that since  $\varepsilon < \|f\|_2$  we have  $c_\star < \infty$ . Furthermore, since for all  $\lambda \in \Omega$  we have  $\lim_{c \uparrow c_\star} |f(\lambda)|^2 \mathbf{1}_{\{|f| < c\}}(\lambda) = |f(\lambda)|^2 \mathbf{1}_{\{|f| < c_\star\}}(\lambda)$ , we have again by the dominated convergence theorem

$$\int_{\{|f| < c_\star\}} |f(\lambda)|^2 d\mu(\lambda) = \lim_{c \uparrow c_\star} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) \leq \varepsilon^2. \quad (3.16)$$

Therefore, we can define

$$\tilde{f} = f \mathbf{1}_{\{|f| \geq c_\star\}} \neq 0 \quad (3.17)$$

which by (3.16) approximates  $f$  in  $L^2$ -norm, i.e. satisfies (3.15). Note that by definition we have that  $\tilde{f} \in L^1(\Omega, \mathbb{R})$  as

$$\|\tilde{f}\|_1 = \int_{\{|f| \geq c_\star\}} |f(\lambda)| d\mu(\lambda) \leq \frac{1}{c_\star} \int_{\{|f| \geq c_\star\}} |f(\lambda)|^2 d\mu(\lambda) \leq \frac{\|f\|_2^2}{c_\star}.$$

We define the function  $g$  differently depending on whether  $c_\star < \frac{\varepsilon^2}{\|f\|_1}$  and  $c_\star \geq \frac{\varepsilon^2}{\|f\|_1}$ . In the first case  $c_\star < \frac{\varepsilon^2}{\|f\|_1}$ , we define

$$h = f \mathbf{1}_{\left\{ |f| < \frac{\varepsilon^2}{\|f\|_1} \right\}}$$

and the normalised function  $g = h/\|h\|_2$ . By definition of  $c_\star$  and using  $\frac{\varepsilon^2}{\|\tilde{f}\|_1} > c_\star$  we have

$$\|h\|_2 = \sqrt{\int_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} |f(\lambda)|^2 d\mu(\lambda)} > \varepsilon \quad (3.18)$$

and hence

$$\|g\|_\infty = \frac{1}{\|h\|_2} \left\| |f| \mathbf{1}_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} \right\|_\infty \leq \frac{\varepsilon}{\|\tilde{f}\|_1}.$$

Furthermore, using (3.18) again we see

$$\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) = \sqrt{\int_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} |f(\lambda)|^2 d\mu(\lambda)} \geq \varepsilon.$$

Therefore, the function  $g$  satisfies the points 1 - 3.

On the other hand, in the case  $c_\star \geq \frac{\varepsilon^2}{\|\tilde{f}\|_1}$  define

$$g = \frac{\varepsilon}{\|\tilde{f}\|_1} \operatorname{sgn}(f) \mathbf{1}_{\{|f| \geq c_\star\}}.$$

Clearly, we have  $\|g\|_\infty \leq \frac{\varepsilon}{\|\tilde{f}\|_1}$  and furthermore by definition of  $\tilde{f}$  in (3.17)

$$\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda) = \frac{\varepsilon}{\|\tilde{f}\|_1} \int_{\{|f| \geq c_\star\}} |f(\lambda)|d\mu(\lambda) = \varepsilon.$$

Lastly, we have using  $c_\star \geq \frac{\varepsilon^2}{\|\tilde{f}\|_1}$  that also

$$\|g\|_2^2 = \frac{\varepsilon^2}{\|\tilde{f}\|_1^2} \mu(\{|f| \geq c_\star\}) \leq \frac{\varepsilon^2}{\|\tilde{f}\|_1^2 c_\star} \int_{\{|f| \geq c_\star\}} |f(\lambda)|d\mu(\lambda) = \frac{\varepsilon^2}{\|\tilde{f}\|_1 c_\star} \leq 1$$

and hence, also in this case, the defined function  $g$  satisfies the points 1 - 3. ■

**Proof of Theorem 3.3.** Let  $g_1, g_2 \in \mathcal{S}$  such that

$$\left| \int_{\Omega} f(\lambda) (g_1(\lambda) - g_2(\lambda)) d\mu(\lambda) \right| > 2\varepsilon. \quad (3.19)$$

If such functions do not exist, (3.7) is trivially satisfied. Consider for  $i = 1, 2$  and  $\lambda \in \Omega$  the random variable  $Y_\lambda^{(i)}$  with outcome space  $\{r, -r\}$ , with  $r > 0$  being the number which was fixed at the beginning of Section 3, and distributed according to

$$\mathbb{P}(Y_\lambda^{(i)} = r) = \frac{1 + g_i(\lambda)/r}{2}, \quad \mathbb{P}(Y_\lambda^{(i)} = -r) = \frac{1 - g_i(\lambda)/r}{2}. \quad (3.20)$$

Note, both pairs  $(g_1, (Y_\lambda^{(1)})_{\lambda \in \Omega})$  and  $(g_2, (Y_\lambda^{(2)})_{\lambda \in \Omega})$  satisfy the assumption (3.1). We consider two situations:

$\mathcal{H}_1$  : The tester has black box access to  $g_1$  through the random variable  $Y_\lambda^{(1)}$ ,

$\mathcal{H}_2$  : The tester has black box access to  $g_2$  through the random variable  $Y_\lambda^{(2)}$ .

Hence, if

$$\left(g, (Y_\lambda)_{\lambda \in \Omega}\right) \in \left\{ \left(g_1, \left(Y_\lambda^{(1)}\right)_{\lambda \in \Omega}\right), \left(g_2, \left(Y_\lambda^{(2)}\right)_{\lambda \in \Omega}\right) \right\},$$

and using (3.19) a black box overlap estimation algorithm with success probability  $1 - \delta$  can find which pair is present only by estimating the overlap  $\int_\Omega f(\lambda)g(\lambda)d\mu(\lambda)$  up to additive error  $\varepsilon$ . Let  $N$  be a sufficient number of steps for this test and let  $\lambda_1, \dots, \lambda_N \in \Omega$  be the points the tester chooses to sample the random variable  $Y_\lambda$ . Let  $y_{\lambda_1}, \dots, y_{\lambda_N} \in \{r, -r\}$  be the corresponding obtained samples. As explained around (3.3), we allow for the algorithm to be adaptive, i.e. at each step  $t$  the variable  $\lambda_t$  can depend on the previous points  $\lambda_1, \dots, \lambda_{t-1}$ , outcomes  $y_{\lambda_1}, \dots, y_{\lambda_{t-1}}$  as well as on the function  $f$ , i.e.

$$\lambda_t \equiv \lambda_t(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_{t-1}, y_{\lambda_{t-1}}).$$

We denote  $Z_t = (\lambda_t, Y_{\lambda_t})$  and let  $\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N}$  be the distribution of  $Z_1, \dots, Z_N$  under  $\mathcal{H}_1$  and  $\mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}$  the law of  $Z_1, \dots, Z_N$  under  $\mathcal{H}_2$ . Let  $\mathcal{E}$  be the event that the tester returns the first hypothesis. We have by the data processing inequality of the KL divergence:

$$\begin{aligned} \text{KL} \left( \mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \middle\| \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N} \right) &\geq \text{KL} \left( \mathbb{P}_{\mathcal{H}_1}(\mathcal{E}) \middle\| \mathbb{P}_{\mathcal{H}_2}(\mathcal{E}) \right) \\ &\geq \text{KL} (1 - \delta \middle\| \delta) \\ &\geq \log \left( \frac{1}{3\delta} \right). \end{aligned} \tag{3.21}$$

On the other hand, we have using the notation  $\mathbf{Z}_{<t} = (Z_1, \dots, Z_{t-1})$  for  $t \in [N]$  that

$$\begin{aligned} \text{KL} \left( \mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \middle\| \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N} \right) &\stackrel{(a)}{=} \sum_{t=1}^N \mathbb{E}_{\mathbf{Z}_{<t} \sim \mathbb{P}_{\mathcal{H}_1}} \left[ \text{KL} \left( \mathbb{P}_{\mathcal{H}_1}^{Z_t | \mathbf{Z}_{<t}} \middle\| \mathbb{P}_{\mathcal{H}_2}^{Z_t | \mathbf{Z}_{<t}} \right) \right] \\ &\stackrel{(b)}{=} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[ \text{KL} \left( \mathbb{P}_{\mathcal{H}_1}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \middle\| \mathbb{P}_{\mathcal{H}_2}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \right) \right] \\ &\stackrel{(c)}{\leq} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[ \chi^2 \left( \mathbb{P}_{\mathcal{H}_1}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \middle\| \mathbb{P}_{\mathcal{H}_2}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \right) \right] \\ &\stackrel{(d)}{=} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[ \frac{1}{2} \left( \frac{(1 + g_1(\lambda_t)/r)^2}{1 + g_2(\lambda_t)/r} + \frac{(1 - g_1(\lambda_t)/r)^2}{1 - g_2(\lambda_t)/r} \right) - 1 \right] \\ &= \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[ \frac{(g_1(\lambda_t) - g_2(\lambda_t))^2}{r^2 - g_2^2(\lambda_t)} \right] \\ &\leq N \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_\infty, \end{aligned} \tag{3.22}$$

where in (a), we have used the chain rule of the KL divergence; in (b), we have used that the distribution of  $\lambda_t$  conditioned on  $\mathbf{Z}_{<t}$  does not depend on  $\mathcal{H}_1$  or  $\mathcal{H}_2$  combined again with the chain rule; in (c), we have used the well known inequality  $\text{KL} \leq \chi^2$  (c.f. (2.4)); in (d) we have used the definition of  $Y_\lambda$  (c.f. (3.20)) and the definition of the  $\chi^2$ -divergence (c.f. (2.3)).

Therefore, combining both inequalities (3.21) and (3.22), we obtain

$$\log\left(\frac{1}{3\delta}\right) \leq \text{KL}\left(\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \parallel \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}\right) \leq N \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_\infty$$

hence

$$N \geq \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_\infty^{-1} \log\left(\frac{1}{3\delta}\right).$$

Since  $g_1, g_2 \in \mathcal{S}$  under the constraint (3.19) were arbitrary this shows (3.7).

To prove (3.8) in the case of  $\mathcal{S} = \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$  and  $f \in \mathcal{S}_{\max}$  with  $\|f\|_2 = 1$ , we first use that since  $g_2 = 0 \in \mathcal{S}_{\max}$  we can conclude from (3.7) that

$$N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) \geq \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |ffg| > 2\varepsilon}} \frac{r^2}{\|g\|_\infty^2} \log\left(\frac{1}{3\delta}\right).$$

Furthermore, we can, without loss of generality, restrict to  $2\varepsilon < \|f\|_2 = 1$  as otherwise  $\|f\|_1^{(2\varepsilon)} = 0$  and hence (3.8) is trivially satisfied.

Under this constraint, we want to show

$$\sup_{\substack{g \in \mathcal{S}_{\max}, \\ |ffg| > 2\varepsilon}} \frac{1}{\|g\|_\infty^2} = \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |ffg| \geq 2\varepsilon}} \frac{1}{\|g\|_\infty^2}. \quad (3.23)$$

Note that the left hand side of (3.23) is clearly less or equal than the right hand side, so we only need to show the opposite inequality. Let  $(\varepsilon_n)_{n \in \mathbb{N}} \subseteq (0, \infty)$  be a sequence such that  $\varepsilon_n > \varepsilon$  and  $\lim_{n \rightarrow \infty} \varepsilon_n = \varepsilon$ . Define furthermore

$$\kappa_n = \frac{2(\varepsilon_n - \varepsilon)}{1 - 2\varepsilon} \quad (3.24)$$

which for  $n \in \mathbb{N}$  large enough satisfies  $\kappa_n \in (0, 1]$  and to which we shall restrict in the following. Let  $g \in \mathcal{S}_{\max}$  such that  $|\int_\Omega f(\lambda)g(\lambda)d\mu(\lambda)| \geq 2\varepsilon$  and define

$$g_n = (1 - \kappa_n)g + \kappa_n \text{sgn}\left(\int_\Omega f(\lambda)g(\lambda)d\mu(\lambda)\right) f \quad (3.25)$$

Note first of all that since  $f, g \in \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$  we also have by triangle inequality that  $g_n \in \mathcal{S}_{\max}$ . Moreover, we have

$$\left| \int_\Omega f(\lambda)g_n(\lambda)d\mu(\lambda) \right| = (1 - \kappa_n) \left| \int_\Omega f(\lambda)g(\lambda)d\mu(\lambda) \right| + \kappa_n \geq 2\varepsilon_n > 2\varepsilon.$$

Since, we also have

$$\|g_n\|_\infty \leq (1 - \kappa_n)\|g\|_\infty + \kappa_n\|f\|_\infty \xrightarrow{n \rightarrow \infty} \|g\|_\infty,$$

this shows (3.23).

To conclude (3.8) note that we can restrict to  $\delta \leq 1/3$  as otherwise (3.8) is trivially satisfied. For the case  $\frac{2\varepsilon}{\|f\|_1^{(2\varepsilon)}} > r$  the statement follows by simply lower bounding the supremum in (3.23) by choosing the feasible function  $g = f$ . On the other hand, for the case  $\frac{2\varepsilon}{\|f\|_1^{(2\varepsilon)}} \leq r$ , we can employ Lemma 3.5 and use the function  $g$  (with parameter  $2\varepsilon$ ) from there as a feasible function which gives

$$N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) \geq \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |ffg| \geq 2\varepsilon}} \frac{r^2}{\|g\|_\infty^2} \log\left(\frac{1}{3\delta}\right) \geq \left(\frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon}\right)^2 \log\left(\frac{1}{3\delta}\right)$$

and finishes the proof. ■

## 4 Fidelity estimation of continuous variable systems

In this section we consider for  $m \in \mathbb{N}$  an  $m$ -mode harmonic oscillator on the Hilbert space  $\mathcal{H} = L^2(\mathbb{R}^m)$ . We study the task of estimating the fidelity between a desired pure state  $\rho$  on  $L^2(\mathbb{R}^m)$ , of which we have a theoretical description, and a general state  $\sigma$ , which is prepared experimentally, by performing measurements on  $\sigma$ . Since  $\rho$  is a pure state, we can write the fidelity by (2.10) as

$$F(\rho, \sigma) = \pi^m \int_{\mathbb{C}^m} W_\rho(\alpha) W_\sigma(\alpha) d\alpha \quad (4.1)$$

Assume that we are able to perform measurements of the displaced parity operator  $(2/\pi)^m D(\alpha)\Pi D(\alpha)$  with outcomes in  $\{(2/\pi)^m, -(2/\pi)^m\}$  on the state  $\sigma$  for chosen values of  $\alpha \in \mathbb{C}^m$ . Here, the values of  $\alpha$  can be chosen adaptively and hence depend on the previous values and corresponding measurement outcomes. In the following we refer to this model as the *Wigner model of fidelity estimation*.

For this model,  $\varepsilon > 0$  and  $\delta > 0$  being the allowed additive error and probability of failure respectively and pure state  $\rho$  fixed, which we call the *instanced based setup*, we denote the *sample complexity of fidelity estimation in the Wigner model* by  $N_{\varepsilon,\delta}^W(\rho)$ : That is  $N_{\varepsilon,\delta}^W(\rho)$  is the optimal number of measurements on different copies of the unknown state  $\sigma$  needed to estimate  $F(\rho, \sigma)$  in the worst case over all states  $\sigma$ .

Due to the relation (2.12), the assumption that we are able to perform measurements of the displaced parity operator on  $\sigma$  exactly means that we have black box access to the Wigner function  $W_\sigma$ . Hence, estimating the fidelity (4.1) in this setup becomes a special instance of black box overlap estimation as discussed in Section 3: In particular we can make the identifications

$$\begin{aligned} \Omega &\mapsto \mathbb{C}^m \cong \mathbb{R}^{2m} && \text{and} && d\mu(\lambda) &\mapsto \pi^m d\alpha \\ r &\mapsto (2/\pi)^m \\ f(\lambda) &\mapsto W_\rho(\alpha) && \text{to which we have full access} \\ g(\lambda) &\mapsto W_\sigma(\alpha) && \text{to which we have black box access} \end{aligned}$$

$$Y_\lambda \mapsto Y_\alpha \quad \text{measurement outcome of observable } (2/\pi)^m D(\alpha)\Pi D(\alpha) \text{ on state } \sigma. \quad (4.2)$$

Since  $\mu \equiv \mu_W$  is given by the Lebesgue measure rescaled by  $\pi^m$ , we introduce the corresponding  $L^q$ -norms for  $q \in [1, \infty)$  for notational clarity and to distinguish them from the usual  $L^q$ -norms by

$$\|W_\rho\|_{L^q(\mathbb{C}^m, \mu_W)} = \left( \int_{\mathbb{C}^m} |W_\rho(\alpha)|^q d\mu_W(\alpha) \right)^{1/q} = \pi^{m/q} \|W_\rho\|_q. \quad (4.3)$$

Note in particular, with that convention, we have from (2.11) for all states  $\rho$

$$\|W_\rho\|_{L^2(\mathbb{C}^m, \mu_W)} \leq 1$$

with equality if the state is pure.

For the subset  $\mathcal{S}$  employed in Section 3, we are interested here in the particular choice  $\mathcal{S}_W$  consisting of all Wigner functions  $W_\sigma$  for valid states  $\sigma$ . Using (2.11) and (2.13), we see that  $\mathcal{S}_W \subseteq B_1^2(0) \cap B_{(2/\pi)^m}^\infty(0)$ , again with balls with respect to the rescaled Lebesgue measure  $d\mu(\alpha) = \pi^m d\alpha$ .

With the notation above, the sample complexity of fidelity estimation can be written as

$$N_{\varepsilon, \delta}^W(\rho) \equiv N_{\varepsilon, \delta}(W_\rho, \mathcal{S}_W) \quad (4.4)$$

with  $N_{\varepsilon, \delta}(W_\rho, \mathcal{S}_W)$  being defined in Section 3.

Formulating the achievability result of black box estimation, Theorem 3.1, for this special case leads to following result:

**Theorem 4.1 (Upper bound on fidelity estimation for CV systems)** *Let  $\rho$  be a pure state on  $L^2(\mathbb{R}^m)$  to which we have a theoretical description. Then, for any state  $\sigma$ , we can estimate the fidelity  $F(\rho, \sigma)$  with precision  $\varepsilon > 0$  and failure probability at most  $\delta > 0$  by measuring the displaced parity operator on a number of*

$$N_{\varepsilon, \delta}^W(\rho) \leq 2^{2m+1} \inf_{\varepsilon' \in [0, \varepsilon)} \left( \frac{\|W_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right) \quad (4.5)$$

*copies of the state  $\sigma$ . Note that by the argument around (2.2), this in particular shows  $N_{\varepsilon, \delta}^W(\rho) < \infty$ . Furthermore, the upper bound holds true even when restricting to non-adaptive algorithms.*

**Proof.** Here, with the conventions from (4.2), the smoothed  $L^1$ -norm in Theorem 3.1 is given by  $\|W_\rho\|_{L^1(\mathbb{C}^m, \mu_W)}^{(\varepsilon')} = \pi^m \|W_\rho\|_1^{(\varepsilon')}$ . Hence, we can immediately read off from (3.5) that

$$N_{\varepsilon, \delta}^W(\rho) \leq 2 \left( \frac{2}{\pi} \right)^{2m} \inf_{\varepsilon' \in [0, \varepsilon)} \left( \frac{\|W_\rho\|_{L^1(\mathbb{C}^m, \mu)}^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right) = 2^{2m+1} \inf_{\varepsilon' \in [0, \varepsilon)} \left( \frac{\|W_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right).$$

■

Furthermore, from the lower bound on the sample complexity of black box estimation in Theorem 3.3, we find the corresponding lower bound on  $N_{\varepsilon, \delta}^W(\rho)$ :



**Theorem 4.2 (Lower bound on fidelity estimation for CV systems)** *Let  $\rho$  be a pure state on  $L^2(\mathbb{R}^m)$  to which we have a theoretical description. Then for  $0 < \varepsilon < 1/2$  and  $0 < \delta \leq 1/3$  we have that the sample complexity of fidelity estimation in the Wigner model is lower bounded as*

$$\begin{aligned} N_{\varepsilon, \delta}^W(\rho) &\geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(W_{\sigma_1} - W_{\sigma_2})^2}{r^2 - W_{\sigma_2}^2} \right\|_{\infty}^{-1} \log \left( \frac{1}{3\delta} \right) \\ &\geq \left( \frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) \geq 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \log \left( \frac{1}{3\delta} \right) \end{aligned} \quad (4.6)$$

**Remark 4.3** *Note that the restriction  $0 < \varepsilon < 1/2$  in Theorem 4.2 is sensible as for allowed additive error  $\varepsilon \geq 1/2$  one can always output  $1/2$  to estimate  $F(\rho, \sigma)$  with demanded precision without performing any measurements on  $\sigma$ . In other words, for  $\varepsilon \geq 1/2$  we have the trivial result*

$$N_{\varepsilon, \delta}^W(\rho) = 0.$$

### Proof of Theorem 4.2.

Using (4.1) and the identifications (4.2), we immediately get from (3.7) in Theorem 3.3 that

$$N_{\varepsilon, \delta}^W(\rho) \geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(W_{\sigma_1} - W_{\sigma_2})^2}{r^2 - W_{\sigma_2}^2} \right\|_{\infty}^{-1} \log \left( \frac{1}{3\delta} \right). \quad (4.7)$$

We now focus on proving the second inequality in (4.6). Consider for that  $\sigma_2$  in (4.7) to be the sequence  $(\rho_n)_{n \in \mathbb{N}}$  from Lemma 4.7 below. Since  $\rho$  is pure, (4.36) yields  $\lim_{n \rightarrow \infty} F(\rho, \rho_n) = 0$ . Combining this with the fact that  $\lim_{n \rightarrow \infty} \|W_{\rho_n}\|_{\infty} = 0$ , c.f. (4.35), this gives for every state  $\sigma$  that

$$\left\| \frac{(W_{\sigma} - W_{\rho_n})^2}{r^2 - W_{\rho_n}^2} \right\|_{\infty} \leq \frac{\|W_{\sigma} - W_{\rho_n}\|_{\infty}^2}{r^2 - \|W_{\rho_n}\|_{\infty}^2} \xrightarrow{n \rightarrow \infty} \left( \frac{\|W_{\sigma}\|_{\infty}}{r} \right)^2$$

and hence

$$N_{\varepsilon, \delta}^W(\rho) \geq \left( \frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) > 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \log \left( \frac{1}{3\delta} \right) = \left( \frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) \geq 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \log \left( \frac{1}{3\delta} \right), \quad (4.8)$$

Here, the last equality follows due to  $\varepsilon < 1/2$  and the same argument as for (3.23). Note that the corresponding function  $g_n$  defined in (3.25) in the context here is the convex combination of two Wigner functions and hence a valid Wigner function itself. ■

## 4.1 Matching upper and lower bounds for $N_{\varepsilon, \delta}^W(\rho)$ for examples of states

In this section we discuss two examples of families of states for which the upper and lower bounds on  $N_{\varepsilon, \delta}^W(\rho)$  provided in Theorem 4.1 and Theorem 4.2 match. In particular we give a full characterisation of the sample complexity of fidelity estimation in the Wigner model for

1. the Fock states  $(|n\rangle\langle n|)_{n \in \mathbb{N}}$  corresponding to the eigenstates of the (1-mode) harmonic oscillator,

2. the *spike states*, which we define in Section 4.1.2,
3. Gaussian states,

in terms of the  $L^1$ -norms of the corresponding Wigner functions.

In the case of the Fock states the key insight is that the corresponding Wigner functions can explicitly be expressed in terms of Laguerre polynomials, c.f. (4.16) below. Combining this with the results on the asymptotic behaviour of the Laguerre polynomials in [AW65, Sze75, Mar82, Lan00], we show in Proposition 4.5 that the upper bound in Theorem 4.1 and the first lower bound in Theorem 4.2 match, giving that

$$N_{\varepsilon, \delta}^W(|n\rangle\langle n|) = \Theta \left( \left( \frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) = \Theta \left( \frac{n}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right).$$

The spike states on the other hand are a sequence of pure states  $(\rho_n)_{n \in \mathbb{N}}$  on  $L^2(\mathbb{R}^m)$  constructed in such a way, that their Wigner functions uniformly vanish as  $n$  increases, i.e.

$$\|W_{\rho_n}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (4.9)$$

Furthermore, we show in Lemma 4.7 that the  $L^1$ -norms of their Wigner functions blow up and exactly scale reciprocally to the  $L^{\infty}$ -norms, i.e.

$$\|W_{\rho_n}\|_1 \sim \|W_{\rho_n}\|_{\infty}^{-1}. \quad (4.10)$$

Hence, we can use the upper bound in Theorem 4.1 and the second lower bound in Theorem 4.2 showing that

$$N_{\varepsilon, \delta}^W(\rho_n) = \Theta \left( \left( \frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right). \quad (4.11)$$

We believe the existence of such a sequence of pure states satisfying (4.9) and (4.10) to be potentially of independent interest.

Lastly, for Gaussian states  $\rho$ , the  $L^1$ -norm of the corresponding Wigner function is by the normalisation (2.9) equal to 1, i.e.  $\|W_{\rho}\|_1 = 1$ . Combining this with the lower bound provided in Theorem 4.2 we see in Proposition 4.10

$$N_{\varepsilon, \delta}^W(\rho) = \Theta \left( \frac{1}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right) \quad (4.12)$$

for all pure Gaussian states  $\rho$ .

As a consequence of these results we find the following worst case characterisation of the sample complexity of fidelity estimation in the Wigner model:

**Theorem 4.4 (Sample complexity of fidelity estimation in Wigner model in worst case)**  
*Let  $0 < \varepsilon < 1/2$  and  $0 < \delta < 1/4$ . Then we have for all  $t \geq 1$*

$$\sup_{\substack{\rho \text{ pure state,} \\ \|W_{\rho}\|_1 \leq t}} N_{\varepsilon, \delta}^W(\rho) = \Theta \left( \left( \frac{t}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right), \quad (4.13)$$

where the constants in the  $\Theta(\cdot)$  can be  $m$ -dependent. In particular, this shows that in worst case over all pure states  $\rho$  the sample complexity of fidelity estimation in the Wigner model is infinite, i.e.

$$\sup_{\rho \text{ pure state}} N_{\varepsilon, \delta}^W(\rho) = \infty. \quad (4.14)$$

In the following sections we formally state and prove the above mentioned results for the Fock, spike states and Gaussian states and then provide the proof of Theorem 4.4.

#### 4.1.1 Sample complexity for Fock states

**Proposition 4.5 (Sample complexity for Fock states)** *Let  $0 < \varepsilon, \delta < 1/4$  and  $(|n\rangle\langle n|)_{n \in \mathbb{N}}$  be the 1-mode Fock states. Then we have the following characterisation of the corresponding sample complexity of fidelity estimation in the Wigner model:*

$$N_{\varepsilon, \delta}^W(|n\rangle\langle n|) = \Theta \left( \left( \frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) = \Theta \left( \frac{n}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right). \quad (4.15)$$

**Remark 4.6** *As seen in Proposition A.4 in the Appendix, the smoothed  $L^1$ -norm of the Wigner functions of the Fock states satisfy*

$$\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \geq \|W_{|n\rangle\langle n|}\|_1 (1 - C\varepsilon^\beta)$$

for all  $\varepsilon > 0$ ,  $\beta \in (0, 1/2)$  and some  $C > 0$  possibly dependent on  $\beta$  but independent of  $n$  and  $\varepsilon$ . As discussed in the beginning of Appendix A.1, this gives

$$\inf_{\varepsilon' \in [0, \varepsilon]} \frac{\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} = \Theta \left( \frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right).$$

Hence, for the Fock states, the upper bound on the sample complexity in terms of the smoothed  $L^1$ -function in Theorem 4.1 does essentially not become weaker when expressing it directly through the non-smoothed  $L^1$ -norm. This further justifies that we find in Proposition 4.5 a characterisation of the sample complexity of fidelity estimation for the Fock states in terms of the  $L^1$ -norm of the corresponding Wigner functions.

For the proof of Proposition 4.5, we use the fact that the corresponding Wigner functions can be expressed in terms of the Laguerre polynomials  $L_n$  as [LP95, Equation 36][Sch, Equation 7.3]

$$W_{|n\rangle\langle n|}(\alpha) = \left( \frac{2}{\pi} \right) (-1)^n e^{-2|\alpha|^2} L_n(4|\alpha|^2). \quad (4.16)$$

Using the results of [Mar82] on the  $L^p$ -norms of the Laguerre polynomials, we can deduce the following scaling of the  $L^1$ -norms of the Wigner function of the Fock states:

$$\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n}. \quad (4.17)$$

This can be combined with Theorem 4.1 yielding the desired upper bound on  $N_{\varepsilon,\delta}^W(|n\rangle\langle n|)$ . On the other hand, using the multiple asymptotic expressions of the Laguerre polynomials from [Sze75, AW65, Lan00] for different regions in phase space, we show explicitly below that

$$\left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{(2/\pi)^2 - W_{|n\rangle\langle n|}^2} \right\|_{\infty} \lesssim \frac{1}{n}.$$

Applying this to the first lower bound in Theorem 4.2 for specific choices of  $\sigma_1$  and  $\sigma_2$  provides the corresponding lower bound of  $N_{\varepsilon,\delta}^W(|n\rangle\langle n|)$  in Proposition 4.5.

**Proof of Proposition 4.5.** We first start by proving the corresponding upper bound in (4.15). We use the upper bound of Theorem 4.1 which gives

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) \lesssim \left( \frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right).$$

The Wigner function of the Fock states  $|n\rangle\langle n|$  can be written explicitly as [LP95, Equation 36][Sch, Equation 7.3]

$$W_{|n\rangle\langle n|}(\alpha/2) = r(-1)^n e^{-|\alpha|^2/2} L_n(|\alpha|^2) \quad (4.18)$$

with  $L_n$  denoting the *Laguerre polynomial* of  $n^{\text{th}}$  order and  $r = 2/\pi$ . From that we see that the scaling of the corresponding  $L^1$ -norms is given as

$$\|W_{|n\rangle\langle n|}\|_1 \sim \int_0^{\infty} e^{-s^2/2} |L_n(s^2)| s ds = \int_0^{\infty} e^{-s^2/2} |L_n(s)| ds \sim \sqrt{n}, \quad (4.19)$$

where we used [Mar82, Lemma 1] for the last step. Hence, we have shown

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) \lesssim \frac{n}{\varepsilon^2} \log \left( \frac{1}{\delta} \right). \quad (4.20)$$

Let us finish the proof by providing the corresponding lower bound in (4.15): Since  $\varepsilon < 1/4$  we can pick states  $\sigma_1^{(n)} = (1 - 3\varepsilon)|n\rangle\langle n| + 3\varepsilon|n+2\rangle\langle n+2|$  and  $\sigma_2^{(n)} = |n\rangle\langle n|$ , which satisfy

$$\left| F(|n\rangle\langle n|, \sigma_1^{(n)}) - F(|n\rangle\langle n|, \sigma_2^{(n)}) \right| = 3\varepsilon > 2\varepsilon$$

Hence, we see from the first lower bound in Theorem 4.2 that

$$\begin{aligned} N_{\varepsilon,\delta}^W(|n\rangle\langle n|) &\gtrsim \left\| \frac{(W_{\sigma_1^{(n)}} - W_{\sigma_2^{(n)}})^2}{r^2 - W_{\sigma_2^{(n)}}^2} \right\|_{\infty}^{-1} \log \left( \frac{1}{3\delta} \right) \\ &= \frac{1}{9\varepsilon^2} \left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{r^2 - W_{|n\rangle\langle n|}^2} \right\|_{\infty}^{-1} \log \left( \frac{1}{3\delta} \right). \end{aligned} \quad (4.21)$$

In the following, we use estimate the quotient appearing in infinity norm in the second line of (4.21) to find the desired lower bound on the sample complexity. We do this by using the explicit form of the Wigner functions in (4.18) and provide bounds in three different regions of phase space, i.e. for phase space points  $\alpha$  satisfying

1.  $|\alpha|^2 \leq 1/\tilde{n}$ ,
2.  $|\alpha|^2 \in [\tilde{n}^{-1}, 1]$  or
3.  $|\alpha|^2 > 1$ .

Here, we introduced the short hand notation  $\tilde{n} = n + 1/2$ . For the first two regions we use the *Hilb's type formula* [Sze75, Theorem 8.22.4] which gives for  $|\alpha|^2 \in [0, 1]$

$$e^{-|\alpha|^2/2} L_n(|\alpha|^2) = J_0(2\sqrt{\tilde{n}}|\alpha|) + \begin{cases} \mathcal{O}(|\alpha|^4 \log(|\alpha|^{-2}n^{-1})), & \text{for } |\alpha|^2 \in [0, \tilde{n}^{-1}] \\ \mathcal{O}(n^{-3/4}), & \text{for } |\alpha|^2 \in [\tilde{n}^{-1}, 1] \end{cases} \quad (4.22)$$

Here, the *Bessel function* of order 0 was denoted by

$$J_0(y) = \sum_{k=0}^{\infty} \frac{(-1)^k (y/2)^{2k}}{(k!)^2} = 1 - \left(\frac{y}{2}\right)^2 + R(y), \quad (4.23)$$

with the last equation holding by Taylor's theorem including up to third order for small  $y$ , say  $y \in [0, 3]$ , with  $R(y)$  being the corresponding remainder. The remainder satisfies

$$|R(y)| \leq \frac{y^4}{4!} \quad (4.24)$$

which follows by the fact<sup>12</sup> that  $|J_0^{(4)}(y)| \leq 1$  together with the standard estimate for the remainder in Taylor's theorem [Apo91, Theorem 7.7]. Let us focus on the first phase space region close to the origin, i.e.  $|\alpha|^2 \leq 1/\tilde{n}$ , in which case we get from the series expression of the Bessel function

$$\begin{aligned} \left| J_0(2\sqrt{\tilde{n}+2}|\alpha|) - J_0(2\sqrt{\tilde{n}}|\alpha|) \right| &= \left| \sum_{k=0}^{\infty} \frac{(-1)^k ((\tilde{n}+2)^k - \tilde{n}^k) |\alpha|^{2k}}{(k!)^2} \right| \lesssim \sum_{k=1}^{\infty} \frac{\tilde{n}^{k-1} |\alpha|^{2k}}{k!} \\ &\leq |\alpha|^2 e^{\tilde{n}|\alpha|^2} \lesssim |\alpha|^2, \end{aligned}$$

where we used for the first inequality that  $|(\tilde{n}+2)^k - \tilde{n}^k| = |\sum_{l=0}^{k-1} \tilde{n}^l 2^{k-l} \binom{k}{l}| \leq \tilde{n}^{k-1} 3^k$  together with the fact that  $3^k/k! \leq C$  for some  $C > 0$ . From this and (4.22) we see

$$\begin{aligned} r^{-2} (W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2 &= \left( e^{-|\alpha|^2/2} L_{n+2}(|\alpha|^2) - e^{-|\alpha|^2/2} L_n(|\alpha|^2) \right)^2 \\ &\lesssim (|\alpha|^2 + \mathcal{O}(|\alpha|^4 \log(|\alpha|^{-2}n^{-1})))^2 \lesssim |\alpha|^4, \end{aligned}$$

where we used in the last inequality that  $|\alpha|^2 \log(|\alpha|^{-2}) \leq C$  and  $|\alpha|^2 \log(n) \leq C$  for some  $C > 0$ . Furthermore, again for  $|\alpha|^2 \leq 1/\tilde{n}$ , we have

$$\begin{aligned} 1 - r^{-2} W_{|n\rangle\langle n|}^2(\alpha/2) &= 1 - e^{-|\alpha|^2} L_n^2(|\alpha|^2) \\ &= 1 - (J_0(2\sqrt{\tilde{n}}|\alpha|) + \mathcal{O}(|\alpha|^4 \log(|\alpha|^{-2}n^{-1})))^2 \end{aligned}$$

---

<sup>12</sup>This can be seen by denoting the Bessel function of order  $n \in \mathbb{N}$  by  $J_n$  and using that  $J_0' = -J_1$  and the recurrence relation  $J_n' = \frac{1}{2}(J_{n-1} - J_{n+1})$  [AS64, Equation 9.1.27] combined with the fact that  $|J_n(y)| \leq 1$  for all  $y \in \mathbb{R}$  [AS64, Equation 9.1.60].

$$\begin{aligned}
&= 1 - (1 - n|\alpha|^2 + R(\sqrt{\tilde{n}}|\alpha|) + \mathcal{O}(|\alpha|^4 \log(|\alpha|^{-2}n^{-1})))^2 \\
&\geq 1 - (1 - cn|\alpha|^2)^2 \geq cn|\alpha|^2
\end{aligned}$$

where the first inequality holds for  $n$  large enough and some  $0 < c < 1$  and we used (4.24). Dividing both we see

$$\sup_{0 \leq |\alpha|^2 \leq 1/\tilde{n}} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \lesssim \sup_{0 \leq |\alpha|^2 \leq \frac{1}{\tilde{n}}} \frac{|\alpha|^2}{n} \lesssim \frac{1}{n^2}. \quad (4.25)$$

For the other two regions in phase space we use [AW65], in particular the table on page 699, which gives

$$\sup_{\alpha \in \mathbb{C}} \left( e^{-|\alpha|^2/2} (L_{n+2}(|\alpha|^2) - L_n(|\alpha|^2)) \right) = \mathcal{O}(n^{-1/2}). \quad (4.26)$$

Let us focus first on second region in phase space, i.e.  $|\alpha|^2 \in [\tilde{n}^{-1}, 1]$ : For those phase space points we use the fact that  $|J_0(y)| \leq \frac{0.8}{y^{1/3}}$  [Lan00] for all  $y \in \mathbb{R}$  which gives by (4.18) and (4.22) that

$$1 - r^{-2}W_{|n\rangle\langle n|}^2(\alpha/2) = 1 - \left( J_0(2\sqrt{\tilde{n}}|\alpha|) + \mathcal{O}(n^{-3/4}) \right)^2 \geq c \quad (4.27)$$

for  $n$  large enough and some  $c > 0$ . Dividing now (4.26) by (4.27) this shows

$$\begin{aligned}
&\sup_{|\alpha|^2 \in [\tilde{n}^{-1}, 1]} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \\
&\leq \sup_{|\alpha|^2 \in [\tilde{n}^{-1}, 1]} \frac{\left( e^{-|\alpha|^2/2} (L_{n+2}(|\alpha|^2) - L_n(|\alpha|^2)) \right)^2}{c} \lesssim \frac{1}{n}.
\end{aligned} \quad (4.28)$$

For the last phase space region, we use again [AW65] which gives for  $|\alpha| > 1$ ,

$$e^{-|\alpha|^2/2} L_n(|\alpha|^2) = \mathcal{O}(n^{-1/4})$$

and hence using again (4.18)

$$1 - r^{-2}W_{|n\rangle\langle n|}^2(\alpha/2) \geq c$$

for  $n$  large enough and some  $c > 0$ . Hence, combining this again with (4.26) we see

$$\sup_{|\alpha| > 1} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \lesssim \frac{1}{n} \quad (4.29)$$

Therefore, combining (4.25), (4.28) and (4.29) we see

$$\left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{r^2 - W_{|n\rangle\langle n|}^2} \right\|_{\infty} \lesssim \frac{1}{n}$$

and therefore, from (4.21)

$$N_{\varepsilon, \delta}^W(|n\rangle\langle n|) \gtrsim \frac{n}{\varepsilon^2} \log \left( \frac{1}{3\delta} \right).$$

which together with the established (4.20) finishes the proof. ■

### 4.1.2 Sample complexity for spike states

In this section we define the *spike states* and analyse the scaling behaviour of the  $L^1$ - and  $L^\infty$ -norms of their Wigner functions. In particular for every number of modes  $m \in \mathbb{N}$  the spike states are a sequence of pure states  $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$  on  $L^2(\mathbb{R}^m)$  whose Wigner functions satisfy

$$\|W_{\rho_n}\|_\infty \xrightarrow{n \rightarrow \infty} 0 \quad (4.30)$$

Here, the idea of the construction is to pick wave functions  $\psi_n$  with increasing uncertainty in both position and momentum which then leads to a further and further spreaded out Wigner function. We pick such wave functions by super positioning ‘spikes’ in position space, which each have a high momentum uncertainty. In the particular construction of Lemma 4.7 these ‘spikes’ are simply given by Gaussian functions. Super-positioning now sufficiently many at increasing distance also leads to high position uncertainty and to a point wise uniformly vanishing Wigner function.

More precisely, the spike states are defined as follows: Let  $\phi \in L^2(\mathbb{R}^m)$  be the normalised (within  $L^2(\mathbb{R}^m)$ , i.e.  $\|\phi\|_2 = 1$ ) Gaussian function centered at the origin, i.e. for  $x \in \mathbb{R}^m$

$$\phi(x) = \left(\frac{2}{\pi}\right)^{m/4} e^{-x^2}. \quad (4.31)$$

Consider for  $n \in \mathbb{N}$  the wave function defined for  $x \in \mathbb{R}^m$  by

$$\psi_n(x) = \frac{1}{\sqrt{c_n}} \sum_{k=1}^n \phi(x - \mu_k), \quad (4.32)$$

with

$$\mu_k = n 3^k e_1, \quad (4.33)$$

and  $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^m$  and where we drop the  $n$  dependence for brevity. Furthermore,  $c_n$  being the normalisation constant such that  $\|\psi_n\|_2 = 1$  which in particular satisfies

$$c_n = \sum_{k,l=1}^n \int_{\mathbb{R}^m} \phi(x - \mu_k) \phi(x - \mu_l) dx \geq \sum_{k=1}^n \int_{\mathbb{R}^m} |\phi(x - \mu_k)|^2 dx = n. \quad (4.34)$$

The  $n^{\text{th}}$  *spike state* is defined to be the corresponding pure state  $\rho_n = |\psi_n\rangle\langle\psi_n|$ .

The particular spreading of the spikes parameterised by the  $\mu_k$  in (4.33) is chosen in such a way that for every fixed point in phase space essentially at most only two spikes contribute to the Wigner function in the formula (2.8) (see the proof of Lemma 4.7 for more details).

Notably, we provide control of the speed of convergence for the limit (4.30) and also on the blow up of the corresponding  $L^1$ -norms. In particular, we see that the the sequence of pure states satisfies

$$\|W_{\rho_n}\|_\infty^{-1} \sim \|W_{\rho_n}\|_1$$

(c.f. Remark 4.8). This is then used as key insight for proving the worst case characterisation (4.13) of the sample complexity in Theorem 4.2.

In the following lemma we provide the precise scaling behaviour of the  $L^1$ - and  $L^\infty$ -norms of the spike states.

**Lemma 4.7 (Asymptotic behaviour of spike states)** *The pure states  $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$  on  $L^2(\mathbb{R}^m)$  defined by (4.32) have Wigner functions satisfying  $W_{\rho_n} \in L^1(\mathbb{C}^m)$  and*

$$\|W_{\rho_n}\|_{\infty} \leq 4 \left(\frac{2}{\pi}\right)^m \frac{1}{n} \quad \text{and} \quad \|W_{\rho_n}\|_1 \leq n. \quad (4.35)$$

Moreover, the corresponding sequence  $(\psi_n)_{n \in \mathbb{N}}$  converges weakly to 0 in  $L^2(\mathbb{R}^m)$ , i.e. for all  $\varphi \in L^2(\mathbb{R}^m)$  we have

$$\lim_{n \rightarrow \infty} \langle \varphi, \psi_n \rangle = 0. \quad (4.36)$$

**Remark 4.8** *Note that the  $n$ -dependence in of the bounds (4.35) is tight. This can be seen by using that since  $\rho_n$  is pure we have  $\|W_{\rho_n}\|_2^2 = \pi^{-m}$  by (2.11) and hence*

$$\|W_{\rho_n}\|_{\infty}^{-1} \leq \pi^m \|W_{\rho_n}\|_1.$$

Using this and the established bounds in (4.35) shows

$$\|W_{\rho_n}\|_{\infty} \sim \frac{1}{n} \quad \text{and} \quad \|W_{\rho_n}\|_1 \sim n \quad (4.37)$$

Before giving the proof of Lemma 4.7, we show how we can use it to obtain a characterisation of the sample complexity  $N_{\varepsilon, \delta}^W(\rho_n)$  from the upper bound and lower bounds in Theorems 4.1 and 4.2:

**Proposition 4.9 (Sample complexity for spike states)** *Let  $0 < \varepsilon < 1/2$ ,  $0 < \delta < 1/4$  and  $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$  be the spike states defined in (4.32). Then we have the following characterisation of the corresponding sample complexity of fidelity estimation in the Wigner model:*

$$N_{\varepsilon, \delta}^W(\rho_n) = \Theta \left( \left( \frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right),$$

where the constants in the  $\Theta(\cdot)$  can be  $m$ -dependent.

**Proof.** By the upper bound provided in Theorem 4.1 we immediately see

$$N_{\varepsilon, \delta}^W(\rho_n) \leq 2^{2m+1} \left( \frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \leq 2^{2m+1} \left( \frac{n}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right),$$

where for the second inequality we have used Lemma 4.7.

To prove the corresponding lower bound we take  $n' = \lceil n/\varepsilon \rceil$  and state  $\sigma_n = 2\varepsilon\rho_n + (1 - 2\varepsilon)\rho_{n'}$ , which satisfies  $F(\rho_n, \sigma_n) \geq 2\varepsilon$ . Furthermore, we have by Lemma 4.7 that

$$\|W_{\sigma_n}\|_{\infty} \leq 2\varepsilon\|W_{\rho_n}\|_{\infty} + (1 - 2\varepsilon)\|W_{\rho_{n'}}\|_{\infty} \leq 16 \left(\frac{2}{\pi}\right)^m \frac{\varepsilon}{n}.$$

Therefore, using the second lower bound in Theorem 4.2 we see that

$$N_{\varepsilon, \delta}^W(\rho_n) \geq \left(\frac{2}{\pi}\right)^{2m} \frac{1}{\|W_{\sigma_n}\|_{\infty}^2} \log \left( \frac{1}{3\delta} \right) \geq \frac{1}{16} \left(\frac{n}{\varepsilon}\right)^2 \log \left( \frac{1}{3\delta} \right)$$



which finishes the proof. ■

**Proof of Lemma 4.7.** We proof the result with phase space being parametrised with position and momentum variables  $(x, p) \in \mathbb{R}^{2m}$  which corresponds to the change of coordinates given by  $\alpha = \frac{1}{\sqrt{2}}(x + ip)$ .

Using (2.8) and the explicit form of the wave functions in (4.32), we can calculate the Wigner function of  $\rho_n$  at phase space point  $(x, p) \in \mathbb{R}^{2m}$  by

$$\begin{aligned} W_{\rho_n}(x, p) &= \frac{1}{c_n} \left(\frac{2}{\pi}\right)^{3m/2} \sum_{k, l=1}^n \int_{\mathbb{R}^m} e^{-(x-\mu_k+y)^2-(x-\mu_l-y)^2} e^{i2p \cdot y} dy \\ &= \frac{1}{c_n} \left(\frac{2}{\pi}\right)^{3m/2} \sum_{k, l=1}^n e^{-(x-\mu_k)^2-(x-\mu_l)^2+\frac{(\mu_k-\mu_l)^2}{2}} \int_{\mathbb{R}^m} e^{-2(y-(\mu_k-\mu_l)/2)^2} e^{i2p \cdot y} dy \\ &= \frac{1}{c_n} \left(\frac{2}{\pi}\right)^m \sum_{k, l=1}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} e^{ip \cdot (\mu_k-\mu_l)-p^2/2}, \end{aligned} \quad (4.38)$$

where for the last line the standard relation for Fourier transforming a Gaussian function (see e.g. [Tes14, Lemma 7.3]).

Note that because of the specific choice of  $\mu_k$  in (4.33), we have for  $x \in \mathbb{R}^m$  fixed that there exists at most one  $(k_x, l_x) \in \mathbb{N}^2$  with  $k_x \geq l_x$  such that for the euclidean norm we have

$$\left\| x - \frac{\mu_{k_x} + \mu_{l_x}}{2} \right\| \leq \frac{n}{2}, \quad (4.39)$$

which can be seen by the following argument: Assume for contradiction that there exists  $(k, l) \neq (i, j)$  with  $k \geq l$  and  $i \geq j$  and such that  $\|x - (\mu_k + \mu_l)/2\| \leq n/2$  and  $\|x - (\mu_i + \mu_j)/2\| \leq n/2$ . This hence gives

$$\|\mu_k + \mu_l - \mu_i - \mu_j\| = n \left| 3^k + 3^l - 3^i - 3^j \right| \leq 2n \quad (4.40)$$

Without loss of generality we can assume that  $i > k$  as for  $k = i$  we immediately also have  $l = j$  from (4.40). Using  $k \geq l$  this gives

$$3^i + 3^j \geq 3^i \geq 3^{k+1} \geq 3^k + 3^l + 3,$$

which contradicts (4.40) and hence proves the desired claim.

Using hence for  $(x, p) \in \mathbb{R}^{2m}$  the uniqueness of  $(k_x, l_x) \in \mathbb{N}^2$  with  $k_x \geq l_x$  and satisfying (4.39) and plugging this into (4.38) gives together with (4.34)

$$\begin{aligned} |W_{\rho_n}(x, p)| &\leq \frac{1}{n} \left(\frac{2}{\pi}\right)^m \sum_{k, l=1}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} \\ &\leq \frac{2}{n} \left(\frac{2}{\pi}\right)^m \left( e^{-2(x-(\mu_{k_x}+\mu_{l_x})/2)^2} + \sum_{\substack{k, l=1 \\ k > l, (k, l) \neq (k_x, l_x)}}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} \right) \\ &\leq \frac{2}{n} \left(\frac{2}{\pi}\right)^m \left( 1 + n^2 e^{-n^2/2} \right) \leq \frac{4}{n} \left(\frac{2}{\pi}\right)^m, \end{aligned}$$

where we used  $t^2 e^{-t^2/2} \leq 1$  for all  $t \in \mathbb{R}$  and which hence shows the  $L^\infty$ -bound in (4.35).

On the other hand, using again the explicit form of the Wigner function of  $\rho_n$  in (4.38), we see

$$\begin{aligned} \|W_{\rho_n}\|_1 &= \int_{\mathbb{C}^m} |W_\rho(\alpha)| d\alpha = 2^{-m} \int_{\mathbb{R}^{2m}} |W_\rho(x, p)| dx dp \\ &\leq \frac{1}{n\pi^m} \sum_{k,l=1}^n \int_{\mathbb{R}^{2m}} e^{-2(x-(\mu_k+\mu_l)/2)^2} e^{-p^2/2} dx dp \\ &= \frac{n}{\pi^m} \int_{\mathbb{R}^{2m}} e^{-x^2} e^{-p^2} dx dp = n. \end{aligned}$$

Lastly we want to show the weak convergence of the sequence  $(\psi_n)_{n \in \mathbb{N}}$  to 0, i.e. (4.36). For that let  $\varphi \in L^2(\mathbb{R})$  and  $\varepsilon > 0$  and pick  $R \geq 0$  such that

$$\int_{\|x\| > R} |\varphi(x)|^2 dx < \varepsilon,$$

which is possible due to the dominated convergence theorem. Furthermore, note that we have for the Gaussian function  $\phi$  defined in (4.31),  $k \in [n]$  and  $n3^k \geq R$

$$\int_{\|x\| \leq R} |\phi(x - \mu_k)|^2 dx \leq \sqrt{\frac{2}{\pi}} \int_{-R}^R e^{-2(x_1 - n3^k)^2} dx_1 \leq \sqrt{\frac{2}{\pi}} \int_{-R+n3^k}^{\infty} e^{-2x_1^2} dx_1 \leq e^{-2(n3^k - R)^2}$$

where the first inequality follows by definition of  $\mu_k$  in (4.33) and second inequality follows by standard Gaussian tail bounds. Hence, using the Cauchy Schwarz inequality and the fact that  $\psi_n$  is normalised in  $L^2$ , we see for  $n$  large enough

$$\begin{aligned} |\langle \varphi, \psi_n \rangle| &\leq \sqrt{\int_{\|x\| > R} |\varphi(x)|^2 dx} + \frac{\|\varphi\|_2}{\sqrt{c_n}} \sum_{k=1}^n \sqrt{\int_{\|x\| \leq R} |\phi(x - \mu_k)|^2 dx} \\ &\leq \sqrt{\varepsilon} + \frac{\|\varphi\|_2}{\sqrt{c_n}} \sum_{k=1}^n e^{-(n3^k - R)^2} \xrightarrow{n \rightarrow \infty} \sqrt{\varepsilon}. \end{aligned}$$

Since,  $\varepsilon > 0$  was arbitrary, this shows (4.36). ■

### 4.1.3 Sample complexity for Gaussian states

**Proposition 4.10** *Let  $\rho$  be a pure Gaussian state on  $L^2(\mathbb{R}^m)$  and  $0 < \varepsilon, \delta < 1/4$ . Then*

$$N_{\varepsilon, \delta}^W(\rho) = \Theta \left( \frac{1}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right), \quad (4.41)$$

where the constants in the  $\Theta(\cdot)$  can be  $m$ -dependent.

**Proof.** For  $\rho$  being a Gaussian state, we have that  $W_\rho \geq 0$  and hence by normalisation (2.9)

$$\|W_\rho\|_1 = 1. \quad (4.42)$$

Therefore, we see from Theorem 4.1 the upper bound

$$N_{\varepsilon,\delta}^W(\rho) = \mathcal{O}\left(\frac{1}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right).$$

For the corresponding lower bound we use the second inequality in (4.6) of Theorem 4.2 for the sequence of feasible states given by  $\sigma_n = 2\varepsilon\rho + (1 - 2\varepsilon)\rho_n$  where  $(\rho_n)_{n \in \mathbb{N}}$  is the sequence of spike states defined in Section 4.1.2 which satisfies  $\lim_{n \rightarrow \infty} \|W_{\rho_n}\|_\infty = 0$  and hence  $\lim_{n \rightarrow \infty} \|W_{\sigma_n}\|_\infty = 2\varepsilon\|W_\rho\|_\infty \leq 2\varepsilon$ . ■

#### 4.1.4 Proof of Theorem 4.4

In this section we give the proof of Theorem 4.4, which essentially follows the same argument as the proof of Proposition 4.9.

**Proof of Theorem 4.4.** Let  $t \geq 1$  and use Theorem 4.1 for the corresponding upper bound as

$$\begin{aligned} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon,\delta}(W_\rho, \mathcal{S}_W) &\leq 2^{2m+1} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} \left(\frac{\|W_\rho\|_1}{\varepsilon}\right)^2 \log\left(\frac{1}{\delta}\right) \\ &\leq 2^{2m+1} \left(\frac{t}{\varepsilon}\right)^2 \log\left(\frac{1}{\delta}\right). \end{aligned}$$

For the corresponding lower bound in (4.13) we take the second lower bound in Theorem 4.2 for the specific choice  $\rho \equiv \rho_n$  where  $n = \lfloor t \rfloor$  and  $(\rho_n)_{n \in \mathbb{N}}$  being the spike states defined and discusses in Section 4.1.2. Note that  $\rho_n$  is feasible in (4.13) as by (4.35) we have  $\|W_{\rho_n}\|_1 \leq n \leq t$ . Furthermore, let  $\sigma_n = 2\varepsilon\rho_n + (1 - 2\varepsilon)\rho_{n'}$  where  $n' = \lceil \frac{t}{\varepsilon} \rceil$ . Note that for that choice we have  $F(\rho_n, \sigma_n) \geq 2\varepsilon$  and furthermore by (4.35) and the fact that  $1/n \leq 2/t$  since  $t \geq 1$  that

$$\|W_{\sigma_n}\|_\infty \leq 2\varepsilon\|W_{\rho_n}\|_\infty + (1 - 2\varepsilon)\|W_{\rho_{n'}}\|_\infty \leq C \left(\frac{2}{\pi}\right)^m \frac{\varepsilon}{t}$$

for some  $C \geq 0$  independent of  $t, \varepsilon$  and  $m$ . Hence,  $\sigma_n$  is a feasible state in (4.6) and we obtain

$$\begin{aligned} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon,\delta}(W_\rho, \mathcal{S}_W) &\geq N_{\varepsilon,\delta}(W_{\rho_n}, \mathcal{S}_W) \geq \left(\frac{2}{\pi}\right)^{2m} \frac{1}{\|W_{\sigma_n}\|_\infty^2} \log\left(\frac{1}{3\delta}\right) \\ &\geq C' \left(\frac{t}{\varepsilon}\right)^2 \log\left(\frac{1}{3\delta}\right) \end{aligned}$$

for some  $C' \geq 0$ . ■

## 5 Fidelity estimation for discrete variable systems

We consider for  $n \in \mathbb{N}$  an  $n$ -qubit quantum system on the Hilbert space  $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$  with dimension  $d = 2^n$ . The corresponding set of Pauli strings is denoted by  $\mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n}$ . Note that  $\mathcal{P}_n$  forms an orthogonal basis on the space of operators on  $(\mathbb{C}^2)^{\otimes n}$  and we have for all  $P, Q \in \mathcal{P}_n$

$$\frac{1}{d} \text{Tr}(PQ) = \delta_{PQ}. \quad (5.1)$$

For  $\rho$  a state on  $(\mathbb{C}^2)^{\otimes n}$  we can define its *characteristic function* as

$$\begin{aligned}\chi_\rho : \mathcal{P}_n &\rightarrow [-1, 1] \\ P &\mapsto \text{Tr}(P\rho).\end{aligned}$$

In the following we consider the task of estimating the fidelity between a desired  $n$ -qubit pure state  $\rho$ , of which we have a theoretical description, and a general state  $\sigma$ , which we prepared experimentally, by performing measurements on  $\sigma$ . Since  $\rho$  is a pure state, we can write the fidelity by using (5.1) as

$$F(\rho, \sigma) = \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_\rho(P) \chi_\sigma(P) = \frac{1}{d} + \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \chi_\rho(P) \chi_\sigma(P) \quad (5.2)$$

Assume that we are able to measure any observable  $P \in \mathcal{P}_n$  with outcome in  $\{1, -1\}$  on the state  $\sigma$ . Here, the observables  $P$  can be chosen adaptively and hence depend on the previous ones and their corresponding measurement outcomes. Since  $\chi_\sigma(\mathbb{1}) = 1$  we can restrict to Pauli measurements with  $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$  as for  $P = \mathbb{1}$  a tester does not gain any information. In the following we refer to this model as the *Pauli-parity model of fidelity estimation*.

For this model,  $\varepsilon > 0$  and  $\delta > 0$  being the allowed additive error and probability of failure respectively and pure state  $\rho$  fixed, which we call the *instanced based setup*, we denote the *sample complexity of fidelity estimation in the Pauli-parity model* by  $N_{\varepsilon, \delta}^P(\rho)$ : That is  $N_{\varepsilon, \delta}^P(\rho)$  is the optimal number of measurements on different copies of the unknown state  $\sigma$  needed to estimate  $F(\rho, \sigma)$  in the worst case over all states  $\sigma$ .

Due to the definition of the characteristic function  $\chi_\sigma$ , the assumption that we are able to perform measurements of the observables  $P \in \mathcal{P}_n$  on  $\sigma$  exactly means that we have black box access to the  $\chi_\sigma$ . Hence, estimating the fidelity (5.2) in this setup becomes special instance of black box overlap estimation as discussed in Section 3: In particular we can make the identifications

$$\begin{aligned}\Omega &\mapsto \mathcal{P}_n \setminus \{\mathbb{1}\} & \text{and} & & d\mu(\lambda) &\mapsto \frac{1}{d} d\mu_{\text{count}}(P) & \text{with } \mu_{\text{count}} & \text{being the counting measure on } \mathcal{P}_n \setminus \{\mathbb{1}\} \\ r &\mapsto 1 \\ f(\lambda) &\mapsto \chi_\rho(P) & & & & \text{to which we have full access} \\ g(\lambda) &\mapsto \chi_\sigma(P) & & & & \text{to which we have black box access} \\ Y_\lambda &\mapsto Y_P & & & & \text{measurement outcome of observable } P \in \mathcal{P}_n \setminus \{\mathbb{1}\} \text{ on state } \sigma.\end{aligned} \quad (5.3)$$

In particular this leads to the following convention of  $L^q$ -norms for  $q \in [1, \infty)$

$$\|\chi_\sigma\|_q := \left( \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\rho(P)|^q \right)^{1/q}. \quad (5.4)$$

and

$$\|\chi_\sigma\|_\infty := \max_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\rho(P)|. \quad (5.5)$$

With these definition we have  $\|\chi_\sigma\|_2 = \sqrt{\text{Tr}(\sigma^2) - 1/d} \leq 1$  and hence by the Cauchy-Schwarz inequality

$$\|\chi_\rho\|_1 = \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{1\}} |\chi_\rho(P)| \leq \sqrt{d} \|\chi_\rho\|_2 \leq \sqrt{d}. \quad (5.6)$$

For the subset  $\mathcal{S}$  employed in Section 3, we are interested here in the particular choice  $\mathcal{S}_P$  consisting of all characteristic functions  $\chi_\sigma$  for valid states  $\sigma$ . We see that  $\mathcal{S}_P \subseteq B_1^2(0) \cap B_1^\infty(0)$ .

With the notation above, the sample complexity of fidelity estimation can be written as

$$N_{\varepsilon,\delta}^P(\rho) \equiv N_{\varepsilon,\delta}(\chi_\rho, \mathcal{S}_P) \quad (5.7)$$

with  $N_{\varepsilon,\delta}(\chi_\rho, \mathcal{S}_P)$  being defined in Section 3.

Formulating the achievability result Theorem 3.1 for this special case leads to the following theorem:

**Theorem 5.1 (Upper bound on fidelity estimation with Pauli measurements)** *Let  $\rho$  be a pure state on  $(\mathbb{C}^2)^{\otimes n}$  to which we have a theoretical description. Then, for any state  $\sigma$ , we can estimate the fidelity  $F(\rho, \sigma)$  with precision  $\varepsilon > 0$  and failure probability at most  $\delta > 0$  by performing Pauli measurements on a number of*

$$N_{\varepsilon,\delta}^P(\rho) \leq 2 \inf_{\varepsilon' \in [0, \varepsilon)} \left( \frac{\|\chi_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \log \left( \frac{1}{\delta} \right) \quad (5.8)$$

*copies of the state  $\sigma$ . Furthermore, the upper bound holds true even when restricting to non-adaptive algorithms.*

**Remark 5.2 (Worst case bound)** *Using (5.6), we see that the sample complexity in the Pauli model satisfies for all pure states  $\rho$*

$$N_{\varepsilon,\delta}^P(\rho) = \mathcal{O} \left( \frac{d}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right). \quad (5.9)$$

*As we see below, this worst case bound is essentially tight as Theorems 5.3 and 5.5 show a lower bound on the sample complexity for random Haar states which matches with high probability (5.9) up to factors logarithmic in the dimension  $d$ . Furthermore, we believe that this tightness result can be strengthened by not relying on a randomised argument but, possibly inspired by the construction of the spike states in Section 4.1.2, finding explicit examples of states for which the lower bound in Theorem 5.3 below can be shown to match (5.9) up to constants which are independent of the system size. We, however, leave this as an open problem.*

Furthermore, from the lower bound on the sample complexity of black box estimation in Theorem 3.3, we find the corresponding lower bound on  $N_{\varepsilon,\delta}^P(\rho)$ :

**Theorem 5.3 (Lower bound on fidelity estimation with Pauli measurements)** *Let  $\rho$  be a pure state on  $(\mathbb{C}^2)^{\otimes n}$  to which we have a theoretical description. Then for  $0 < \varepsilon < 1/2$  and*

$0 < \delta \leq 1/3$  we have that the sample complexity of fidelity estimation in the Wigner model is lower bounded as

$$\begin{aligned}
N_{\varepsilon, \delta}^P(\rho) &\geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(\chi_{\sigma_1} - \chi_{\sigma_2})^2}{1 - \chi_{\sigma_2}^2} \right\|_{\infty}^{-1} \log \left( \frac{1}{3\delta} \right) \\
&\geq \sup_{\substack{\sigma \text{ state,} \\ |F(\rho, \sigma) - 1/d| > 2\varepsilon}} \frac{1}{\|\chi_{\sigma}\|_{\infty}^2} \log \left( \frac{1}{3\delta} \right). \tag{5.10}
\end{aligned}$$

**Proof.** The first inequality immediately follows from Theorem 3.3. For the second we use the first with the choice  $\sigma_2 = \mathbb{1}/d$  which satisfies  $\chi_{\sigma_2} = 0$  on  $\mathcal{P}_n \setminus \{\mathbb{1}\}$  and  $F(\rho, \sigma_2) = 1/d$ . ■

**Remark 5.4 (Matching upper and lower bound for small  $\varepsilon$ )** *Inspired from the construction in Lemma 3.5 consider for some pure state  $\rho$  and  $\varepsilon > 0$  the operator*

$$\omega = \frac{1}{d} \left( \mathbb{1} + \frac{3\varepsilon}{\|\chi_{\rho}\|_1} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \text{sgn}(\chi_{\rho}(P)) P \right)$$

which has characteristic function  $\chi_{\omega}(P) = \text{Tr}(P\omega) = \frac{3\varepsilon}{\|\chi_{\rho}\|_1} \text{sgn}(\chi_{\rho}(P))$  for all  $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$ . This operator  $\omega$  has normalised trace but is in general not positive semi-definite and hence not a state (c.f. Remark 3.6) unless  $\varepsilon > 0$  is small enough. In particular, a naive operator norm bound shows that  $\omega$  is in fact positive semi-definite if  $0 < \varepsilon \leq \frac{1}{3d(d+1)}$ .

Whenever  $\omega$  is a state we can combine (5.10) with the argument in Lemma 3.5 to show the instanced based lower bound on the sample complexity of fidelity estimation in the Pauli model

$$N_{\varepsilon, \delta}^P(\rho) \geq \left( \frac{\|\chi_{\rho}\|_1}{3\varepsilon} \right)^2 \log \left( \frac{1}{3\delta} \right). \tag{5.11}$$

In fact note that

$$F(\rho, \omega) = \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_{\rho}(P) \chi_{\omega}(P) = \frac{1}{d} \left( 1 + \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \frac{3\varepsilon |\chi_{\rho}(P)|}{\|\chi_{\rho}\|_1} \right) = \frac{1}{d} + 3\varepsilon$$

from which we see that  $\omega$  is a feasible state in the second line of (5.10). Furthermore, since by definition  $\|\chi_{\omega}\|_{\infty} \leq \frac{3\varepsilon}{\|\chi_{\rho}\|_1}$  we can conclude (5.11).

## 5.1 Matching upper and lower bounds for $N_{\varepsilon, \delta}^P(\rho)$ for example of states

In this section we consider two examples of families of states  $\rho$  for which the upper and lower bounds on  $N_{\varepsilon, \delta}^P(\rho)$  provided in Theorems 5.1 and 5.3 can be shown to be essentially matching. In particular we give a full characterisation of the sample complexity of fidelity estimation in the Pauli model for

1. Haar random states,
2. stabiliser states,

in terms of the  $L^1$ -norms of the corresponding characteristic function.

For a Haar random pure state  $\rho = |\Psi\rangle\langle\Psi|$  we find in Proposition 5.5 with probability at least  $1 - 1/d$  that

$$N_{\varepsilon,\delta}^P(\rho) = \tilde{\Theta} \left( \left( \frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) = \tilde{\Theta} \left( \frac{d}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right). \quad (5.12)$$

Here, we used  $\tilde{\Theta}$  to denote the fact that the upper and lower bounds in (5.12) match up to factors that scale logarithmically in the leading order term.

On the other hand, for stabiliser states  $\rho$  on  $n$ -qubits, we have that the  $L^1$ -norm satisfies  $\|\chi_\rho\|_1 \leq 1$  and is hence independent of the system size. Using this we can characterise in Proposition 5.6 the corresponding sample complexity of fidelity estimation in the Pauli model as

$$N_{\varepsilon,\delta}^P(\rho) = \Theta \left( \frac{1}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right). \quad (5.13)$$

The upper bound in (5.13) can already be found in [FL11].

### 5.1.1 Sample complexity for Haar random states

The following proposition shows that the upper and lower bounds of Theorems 5.1 and 5.3 on the sample complexity  $N_{\varepsilon,\delta}^P(\rho)$  match (up to factors logarithmic in the leading order) for Haar random pure states with high probability.

**Proposition 5.5** *Let  $\rho = |\Psi\rangle\langle\Psi|$  be a Haar random pure state on  $(\mathbb{C}^2)^{\otimes n}$ ,  $0 < \varepsilon, \delta \leq 1/5$ . Then with probability at least  $1 - 1/d$ , with  $d = 2^n$ , we have*

$$N_{\varepsilon,\delta}^P(\rho) = \tilde{\Theta} \left( \left( \frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right) \right) = \tilde{\Theta} \left( \frac{d}{\varepsilon^2} \log \left( \frac{1}{\delta} \right) \right). \quad (5.14)$$

Here, we used  $\tilde{\Theta}$  to denote the fact that the upper and lower bounds in (5.14) match up to factors that scale logarithmically in the leading order term.

**Proof.** For the upper bound we use Theorem 5.1 which gives for all pure states  $\rho$

$$N_{\varepsilon,\delta}^P(\rho) \leq 2 \left( \frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \log \left( \frac{1}{\delta} \right).$$

To prove the lower bound, we show in the following that with high probability a Haar random pure state  $\rho = |\Psi\rangle\langle\Psi|$  satisfies

$$\|\chi_\rho\|_\infty \lesssim \sqrt{\frac{\log d}{d}}. \quad (5.15)$$

Let  $U \in \mathbb{U}(d)$  be a Haar unitary such that  $|\Psi\rangle = U^\dagger|0\rangle$ . For a fixed  $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$  we consider the random variable

$$f(U) = |\langle\Psi|P|\Psi\rangle| = |\langle 0|UPU^\dagger|0\rangle|.$$

The function  $f : \mathbb{U}(d) \rightarrow \mathbb{R}$  is 2-Lipschitz with respect to the Frobenius norm denoted  $\|\cdot\|_F$ . Let  $U, V \in \mathbb{U}(d)$  be two unitary matrices, we have

$$\begin{aligned}
|f(U) - f(V)| &= \left| |\langle 0|UPU^\dagger|0\rangle| - |\langle 0|VPV^\dagger|0\rangle| \right| \\
&\stackrel{(a)}{\leq} |\langle 0|(U - V)PU^\dagger|0\rangle| + |\langle 0|VP(U - V)^\dagger|0\rangle| \\
&\stackrel{(b)}{\leq} \|P(U - V)^\dagger|0\rangle\|_2 \|U^\dagger|0\rangle\|_2 + \|V^\dagger|0\rangle\|_2 \|P(U - V)^\dagger|0\rangle\|_2 \\
&\stackrel{(c)}{\leq} 2\|P(U - V)^\dagger\|_F = 2\|U - V\|_F
\end{aligned}$$

where (a) follows from the triangle inequality, (b) follows from the Cauchy Schwarz inequality and (c) uses the fact  $\|A|x\rangle\|_2 \leq \|A\|_F$  for a pure  $|x\rangle$  as well as the Frobenius norm is unitarily invariant. So by the concentration inequality of Lipschitz functions of Haar unitaries [MM13], for all  $s > 0$ :

$$\mathbb{P}(|f(U) - \mathbb{E}(f)| \geq s) \leq \exp\left(-\frac{ds^2}{48}\right).$$

Moreover we have by the Cauchy Schwarz inequality

$$\mathbb{E}(f) \leq \sqrt{\mathbb{E}(f^2)} = \sqrt{\mathbb{E}(\langle \Psi|P|\Psi\rangle^2)} = \sqrt{\frac{\text{Tr}(PP^\dagger) + |\text{Tr}(P)|^2}{d(d+1)}} = \sqrt{\frac{1}{d+1}}.$$

So by the union bound

$$\begin{aligned}
\mathbb{P}\left(\exists P \in \mathcal{P}_n \setminus \{I\} : |\langle \Psi|P|\Psi\rangle| > 2\sqrt{\frac{48 \log d^3}{d}}\right) &\leq \sum_{P \in \mathcal{P}_n \setminus \{I\}} \mathbb{P}\left(|\langle \Psi|P|\Psi\rangle| > 2\sqrt{\frac{48 \log d^3}{d}}\right) \\
&\leq d^2 \mathbb{P}\left(f(U) - \mathbb{E}(f) > 2\sqrt{\frac{48 \log d^3}{d}} - \sqrt{\frac{1}{d+1}}\right) \\
&\leq d^2 \mathbb{P}\left(f(U) - \mathbb{E}(f) > \sqrt{\frac{48 \log d^3}{d}}\right) \\
&\leq d^2 \exp\left(-\frac{d \cdot 48 \log d^3}{48}\right) = \frac{1}{d}.
\end{aligned}$$

Hence, when sampling from the Haar measure, with probability at least  $1 - 1/d$  we find a pure state  $|\Psi\rangle$  such that for all  $P \in \mathcal{P}_n \setminus \{1\}$

$$|\langle \Psi|P|\Psi\rangle| \leq 2\sqrt{\frac{48 \log d^3}{d}}$$

and therefore the corresponding pure state  $\rho = |\Psi\rangle\langle\Psi|$  satisfies (5.15). Consider the state  $\sigma = 5\varepsilon\rho + (1 - 5\varepsilon)\mathbb{1}/d$  which satisfies  $F(\rho, \sigma) - 1/d = 5\varepsilon(1 - 1/d) > 2\varepsilon$  and has characteristic function  $\chi_\sigma = 5\varepsilon\chi_\rho$  on  $\mathcal{P}_n \setminus \{1\}$ . Hence, this state is feasible in the optimisation in the second lower bound provided in Theorem 5.3 which gives using (5.6)

$$N_{\varepsilon, \delta}^P(\rho) \gtrsim \frac{1}{\varepsilon^2 \|\chi_\rho\|_\infty^2} \log\left(\frac{1}{3\delta}\right) \gtrsim \frac{d}{\varepsilon^2 \log d} \log\left(\frac{1}{3\delta}\right) \geq (\log d)^{-1} \left(\frac{\|\chi_\rho\|_1}{\varepsilon}\right)^2 \log\left(\frac{1}{3\delta}\right). \quad (5.16)$$

■



### 5.1.2 Sample complexity for stabiliser states

We say a pure state  $\rho = |\psi\rangle\langle\psi|$  on  $(\mathbb{C}^2)^{\otimes n}$  is a stabiliser state if there are  $2^n$  Pauli strings  $P \in \mathcal{P}_n$  satisfying  $P|\psi\rangle = |\psi\rangle$ .

**Proposition 5.6** *Let  $\rho = |\psi\rangle\langle\psi|$  be a stabiliser state on  $(\mathbb{C}^2)^{\otimes n}$ . Then we have for all  $0 < \varepsilon, \delta < 1/5$  that*

$$N_{\varepsilon, \delta}^P(\rho) = \Theta\left(\frac{1}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right). \quad (5.17)$$

**Proof.** It is well-known that all the  $2^n$  Pauli strings  $P \in \mathcal{P}_n$  which satisfy  $P|\psi\rangle = \psi$  commute and furthermore, for all remaining  $Q \in \mathcal{P}_n$  we have  $\chi_\rho(Q) = \langle\psi, Q\psi\rangle = 0$ . Hence, we have that the  $L^1$ -norm of the characteristic function of  $\rho$  satisfies

$$\|\chi_\rho\|_1 = \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{1\}} |\chi_\rho(P)| = \frac{d-1}{d} \leq 1. \quad (5.18)$$

From Theorems 5.1 we hence see

$$N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}\left(\frac{1}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right). \quad (5.19)$$

For the remaining lower bound we use the second line in (5.10) in Theorem 5.3 for the state  $\sigma = 5\varepsilon\rho + (1 - 5\varepsilon)\mathbb{1}/d$  which is feasible since  $F(\rho, \sigma) - 1/d = 5\varepsilon(1 - 1/d) > 2\varepsilon$  and satisfies  $\|\chi_\sigma\|_\infty = \max_{P \in \mathcal{P}_n \setminus \{1\}} |\chi_\sigma(P)| \leq 5\varepsilon$ . ■

## A Some facts about the smoothed $L^1$ -norm

For  $\Omega$  a measure space with measure  $\mu$ , we have defined in the Section 2.2 the smoothed  $L^1$ -norm of a function  $f \in L^2(\Omega)$  as

$$\|f\|_1^{(\varepsilon)} = \inf_{\substack{\tilde{f} \in L^1(\Omega) \\ \|f - \tilde{f}\|_2 \leq \varepsilon}} \|\tilde{f}\|_1. \quad (A.1)$$

Different from the scope of the rest of the paper, in which we restricted to real valued functions, we allow in this section for real or complex valued functions. Hence, for simplicity, we write as  $L^q(\Omega)$  as the corresponding  $L^q$ -spaces without denoting the field over which the considered functions are taking values.

The following lemma establishes that in the limit  $\varepsilon \rightarrow 0$  the smoothed  $L^1$ -norm converges to the usual (non-smoothed)  $L^1$ -norm.

**Lemma A.1** *For all  $f \in L^2(\Omega)$  we have*

$$\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1, \quad (A.2)$$

where the right hand side is understood as infinity if  $f \notin L^1(\Omega)$ .

**Proof.** We use the relation for a measurable function  $f$

$$\int_{\Omega} |f(\lambda)| d\mu(\lambda) = \sup_{\substack{\Omega_0 \subseteq \Omega \\ \mu(\Omega_0) < \infty}} \int_{\Omega_0} |f(\lambda)| d\mu(\lambda), \quad (\text{A.3})$$

which follows directly from the definition of the Lebesgue integral.

First we consider the case  $f \in L^1(\Omega)$ : Let  $\delta > 0$ . Using (A.3) we can pick a set of finite measure  $\Omega_0 \subseteq \Omega$  such that

$$\int_{\Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \frac{\delta}{2}.$$

Let  $\varepsilon \in (0, \frac{\delta}{2\sqrt{\mu(\Omega_0)}})$  and  $\tilde{f} \in L^1(\Omega) \cap L^2(\Omega)$  be such that  $\|f - \tilde{f}\|_2 \leq \varepsilon$ . Then

$$\begin{aligned} \|f\|_1 &\leq \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \int_{\Omega_0} |\tilde{f}(\lambda)| d\mu(\lambda) + \int_{\Omega_0^c} |f(\lambda)| d\mu(\lambda) \\ &\leq \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \|\tilde{f}\|_1 + \frac{\delta}{2}. \end{aligned}$$

For the first term we use the Cauchy-Schwarz inequality which gives

$$\int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \leq \|f - \tilde{f}\|_2 \sqrt{\mu(\Omega_0)} \leq \frac{\delta}{2}.$$

Plugging this into the above and using that  $\tilde{f}$  was arbitrary under the constraints above gives

$$\|f\|_1^\varepsilon \geq \|f\|_1 - \delta.$$

Noting that the opposite inequality  $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$  is trivially true and using that  $\delta > 0$  was arbitrary, shows  $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$ .

To finish the proof we consider the case  $f \notin L^1(\Omega)$ . We have to show  $\liminf_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \infty$ . Assume for contradiction that there exists  $C > 0$  such that for all  $\varepsilon > 0$  we have the uniform bound

$$\|f\|_1^{(\varepsilon)} \leq C.$$

Let  $\Omega_0 \subseteq \Omega$  be a set of finite measure. Note by similar arguments as above, it holds true that  $\|f \mathbf{1}_{\Omega_0}\|_1^{(\varepsilon)} \leq \|f\|_1^{(\varepsilon)} \leq C$  for all  $\varepsilon > 0$ . Hence, using the established (A.2) for the function  $f \mathbf{1}_{\Omega_0} \in L^1(\Omega)$ , we see

$$\int_{\Omega_0} |f| d\mu(\lambda) = \lim_{\varepsilon \downarrow 0} \|f \mathbf{1}_{\Omega_0}\|_1^{(\varepsilon)} \leq C.$$

But using that  $\Omega_0$  was an arbitrary set of finite measure together with (A.3) shows that  $f \in L^1(\Omega)$ , which gives a contradiction and finishes the proof. ■

## A.1 Lower bound on smoothed $L^1$ -norm for functions with rapidly decaying tail

The smoothed  $L^1$ -norm of a function  $f$ , defined in (2.2), satisfies the trivial upper bound  $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$  for all  $\varepsilon > 0$ . The gap here between smoothed and non-smoothed  $L^1$ -norm can in general be arbitrarily large even for  $L^1$ -functions for which the right hand side is finite. In particular, although by Lemma A.1 we know that  $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$ , in general we have no control on the speed of convergence when taking this limit.

In this section, we, however, restrict to functions with *rapidly decaying tail* (as made precise in (A.6) below) and for which we can provide good control in the relevant lower bound (c.f. Lemma A.2). Here, for instances of interest, we obtain the lower bound

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1(1 - C\varepsilon^\beta) \quad (\text{A.4})$$

for some  $C \geq 0$  and  $\beta \in (0, 1]$  independent of  $\varepsilon$  and the specific function  $f$ . In particular (A.4) holds true for  $f$  being the Wigner functions  $W_{|n\rangle\langle n|}$  of the 1-mode Fock states as shown in Proposition A.4 below.

For functions satisfying (A.4), we find that<sup>13</sup>

$$\inf_{\varepsilon' \in [0, \varepsilon]} \frac{\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} = \Theta\left(\frac{\|f\|_1}{\varepsilon}\right) \quad (\text{A.5})$$

for  $\varepsilon > 0$  small enough. As seen in the previous sections, the sample complexity of black box overlap estimation and fidelity estimation in the Wigner- and Pauli-model can be upper bounded in terms of the smoothed  $L^1$ -norm with quantity of interest being the left hand side of (A.5) (c.f. Theorems 3.1, 4.1 and 5.1). Therefore, by (A.5) we see that for functions satisfying (A.4), we do not weaken this upper bound when choosing  $\varepsilon' = 0$  and providing it in terms of the (non-smoothed)  $L^1$ -norm. This, hence, further justifies that the characterisation of the sample complexity of the Fock states found in Proposition 4.5 is in terms of the usual  $L^1$ -norms of the Wigner functions  $W_{|n\rangle\langle n|}$ .

In the remainder of this section we formally define the notion of functions with rapidly decaying tail and prove the mentioned lower bound on their smoothed  $L^1$ -norm in Lemma A.2. After that we discuss as an example the relevant class of *eventually exponentially decaying* functions on  $\mathbb{R}^m$ , which includes the Wigner functions of Fock states.

Let  $\Omega$  be some measurable space with some measure  $\mu$ . For  $\gamma, \kappa > 0$  and  $\Omega_0 \subseteq \Omega$  a finite measure set, we say a measurable function  $f$  has a *rapidly decaying tail of order  $(\gamma, \kappa)$  outside  $\Omega_0$*  if for all  $\delta > 0$  we have

$$\int_{\{|f| \leq \delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \kappa \delta^\gamma. \quad (\text{A.6})$$

For such functions we have good control on the respective smoothed  $L^1$ -norm as proven in the following lemma. In particular the proof follows similar lines as the one of Lemma A.1, however, providing additionally a convergence rate as  $\varepsilon \rightarrow 0$  by utilising the rapidly decaying tail assumption.

---

<sup>13</sup>This can be seen by  $\inf_{\varepsilon' \in [0, \varepsilon]} \frac{\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \geq \frac{\|f\|_1}{\varepsilon} - C\varepsilon^{-(1-\beta)} = \Omega\left(\frac{\|f\|_1}{\varepsilon}\right)$  together with  $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$ .

**Lemma A.2 (Lower bound on smoothed  $L^1$ -norm)** *Let  $\varepsilon, \gamma, \kappa > 0$  and  $\Omega_0 \subseteq \Omega$  be a set of finite measure. Furthermore, let  $f \in L^2(\Omega)$  with  $\|f\|_2 \leq 1$  have a rapidly decaying tail of order  $(\gamma, \kappa)$  outside  $\Omega_0$ . Then*

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \kappa' \varepsilon^\beta - \varepsilon \sqrt{\mu(\Omega_0)} \quad (\text{A.7})$$

with  $\beta := \frac{\gamma}{\gamma+1} \in (0, 1]$  and  $\kappa' := 2\kappa^{\frac{1}{1+\gamma}}$ .

**Remark A.3** *Note that since  $f$  in Lemma A.2 is supposed to be in  $L^2(\Omega)$  and have a rapidly decaying tail outside a finite measure set, it is immediately also in  $L^1(\Omega)$ . Hence, the right hand side of (A.7) is finite.*

**Proof.** Let  $\tilde{f} \in L^1(\Omega) \cap L^2(\Omega)$  with  $\|f - \tilde{f}\|_2 \leq \varepsilon$  and  $\delta > 0$  to be determined later. We have

$$\begin{aligned} \|f\|_1 &= \int_{\{|f|>\delta\}} |f(\lambda)| d\mu(\lambda) + \int_{\{|f|\leq\delta\} \cap \Omega_0} |f(\lambda)| d\mu(\lambda) + \int_{\{|f|\leq\delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \\ &\leq \int_{\{|f|>\delta\}} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \|\tilde{f}\|_1 + \kappa \delta^\gamma. \end{aligned} \quad (\text{A.8})$$

For the first term we use with the Cauchy-Schwarz inequality together with the fact that  $\|f\|_2 \leq 1$  which gives

$$\int_{\{|f|>\delta\}} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \leq \|f - \tilde{f}\|_2 \sqrt{\mu(\{|f|>\delta\})} \leq \frac{\varepsilon}{\delta}.$$

Similarly, we see for the second term in (A.8) that

$$\int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \leq \varepsilon \sqrt{\mu(\Omega_0)}.$$

Plugging this into (A.8) and using the fact that  $\tilde{f}$  is an arbitrary feasible function in the optimisation corresponding to the smoothed  $L^1$ -norm in (2.2), we see

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \varepsilon \sqrt{\mu(\Omega_0)} - \frac{\varepsilon}{\delta} - \kappa \delta^\gamma.$$

Optimising over  $\delta > 0$  and choosing  $\delta = \left(\frac{\varepsilon}{\kappa}\right)^{1/(\gamma+1)}$  finishes the proof. ■

For  $m \in \mathbb{N}$  and the special case  $\Omega = \mathbb{R}^m$  equipped with the Lebesgue measure, a class of particular interest which satisfy the condition (A.6) are *eventually exponentially decaying functions*, i.e.  $f \in L^1(\mathbb{R}^m)$  which satisfy for some  $t, C > 0$  and  $\Omega_0 \subset \mathbb{R}^m$  of finite measure

$$|f(\alpha)| \leq C e^{-t|\alpha|} \quad (\text{A.9})$$

for all  $\alpha \in \mathbb{R}^m \setminus \Omega_0$ . In fact such functions have rapidly decaying tail of order  $(\gamma, \kappa_{C,\gamma,t,m})$  outside  $\Omega_0$  for every  $\gamma \in (0, 1)$  fixed and  $\kappa_{C,\gamma,t,m}$  being linear in  $C$  and  $\kappa_{C,\gamma,t,m}/C$  depending on  $\gamma, t, m$  but not  $f$ . This essentially follows due to the standard inequality<sup>14</sup>

$$\int_{\{e^{-t|\alpha|} \leq \delta\}} e^{-t|\alpha|} d\alpha \lesssim \int_{\log((1/\delta)^{1/t})}^{\infty} e^{-tr} r^{m-1} dr \lesssim \delta (\log(1/\delta) + 1), \quad (\text{A.10})$$

<sup>14</sup>More precisely, to show that an exponentially decaying function  $f$  satisfies (A.6) for all  $a \in (0, 1)$ , we note that  $\int_{\{|f|\leq\delta < e^{-t|\alpha|}\}} |f(\alpha)| d\alpha \leq \int_0^{\log((1/\delta)^{1/t})} \delta d\alpha = \frac{\delta}{t} \log(1/\delta)$ .

where constants independent of  $\delta$  are hidden in the  $\lesssim$ -notation.

Note that the Wigner functions of the (1-mode) Fock states  $W_{|n\rangle\langle n|}$ , which are discussed in Section 4.1.1 and given in terms of Laguerre polynomials (4.16), are eventually exponentially decaying in the sense of (A.9). This can be seen directly from the asymptotic expressions of the Laguerre polynomials provided in [AW65] (consider there the table on page 699 for convenience). From that we can conclude the mentioned lower bound on their respective smoothed  $L^1$ -norm in the following proposition:

**Proposition A.4** *The Wigner functions of the the 1-mode Fock states  $(|n\rangle\langle n|)_{n \in \mathbb{N}}$  satisfy*

$$\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \geq \|W_{|n\rangle\langle n|}\|_1 (1 - C\varepsilon^\beta) \sim \sqrt{n}(1 - C\varepsilon^\beta) \quad (\text{A.11})$$

for all  $\varepsilon > 0$ ,  $\beta \in (0, 1/2)$  and some  $C > 0$  possibly dependent on  $\beta$  but independent of  $n$  and  $\varepsilon$ .

**Proof.** We use that the Wigner function of the Fock states  $|n\rangle\langle n|$  can be written explicitly as [LP95, Equation 36][Sch, Equation 7.3] in terms of the Laguerre polynomials as

$$W_{|n\rangle\langle n|}(\alpha) = (-1)^n \left(\frac{2}{\pi}\right) e^{-2|\alpha|^2} L_n(4|\alpha|^2).$$

Using the asymptotic expressions for the Laguerre polynomials in [AW65], we see that

$$W_{|n\rangle\langle n|}(\alpha) \leq c e^{-t|\alpha|^2}$$

for all  $|\alpha| \geq s\sqrt{n}$  for some  $s, c, t > 0$  independent of  $n$ . Therefore, by the discussion around (A.9) we see that  $W_{|n\rangle\langle n|}$  has a rapidly decaying tail of order  $(\gamma, \kappa_{c,\gamma,t})$  outside of  $\Omega_0 = B_{s\sqrt{n}}(0) \subset \mathbb{C}$  for all  $\gamma \in (0, 1)$ . Here,  $\kappa_{c,\gamma,s} > 0$  depends linearly on  $c$  and is independent of  $n$ . Noting that the volume of  $\Omega_0$  is of order  $n$ , we see from Lemma A.2 that

$$\|W_{|n\rangle\langle n|}\|_1 - \|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \lesssim \varepsilon^\beta + \sqrt{n}\varepsilon \lesssim \sqrt{n}\varepsilon^\beta,$$

where  $\beta = \frac{\gamma}{\gamma+1} \in (0, 1/2)$  and we are hiding all constants which are independent of  $n$  and  $\varepsilon$  in the  $\lesssim$ -notation. Using the fact that  $\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n}$  as noted in (4.19) as a consequence of [Mar82, Lemma 1], finishes the proof. ■

**Acknowledgments.** The authors would like to thank Ulysse Chabaud, Giulia Ferrini, Niklas Galke, Naixu Guo, Benjamin Huard, Hector Hutin, Ludovico Lami, Lauritz van Luijk and Corentin Moumard for helpful discussions. OF and RS acknowledge funding from the European Research Council (ERC Grant AlgoQIP, Agreement No. 851716), from the QuantERA II Programme of the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101017733 (VERIQ-TAS) as well as the government grant managed by the Agence Nationale de la Recherche under the Plan France 2030 with the reference ANR-22-PETQ-0007. AO acknowledges funding by the European Research Council (ERC Grant Agreement No. 948139).

## References

- [AGPF18] Francesco Albarelli, Marco G Genoni, Matteo GA Paris, and Alessandro Ferraro. Resource theory of quantum non-gaussianity and wigner negativity. *Physical Review A*, 98(5):052350, 2018.

- [Apo91] T.M. Apostol. *Calculus, Volume 1*. Wiley, 1991.
- [AS64] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.
- [AW65] Richard Askey and Stephen Wainger. Mean convergence of expansions in laguerre and hermite series. *American Journal of Mathematics*, 87(3):695–708, 1965.
- [BAM<sup>+</sup>02] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche. Direct measurement of the wigner function of a one-photon fock state in a cavity. *Phys. Rev. Lett.*, 89:200402, Oct 2002.
- [BCE22] Robert I. Booth, Ulysse Chabaud, and Pierre-Emmanuel Emeriau. Contextuality and wigner negativity are equivalent for continuous-variable quantum measurements. *Phys. Rev. Lett.*, 129:230401, Nov 2022.
- [BDLR24] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouze. Classical shadow tomography for continuous variables quantum systems. *IEEE Transactions on Information Theory*, 70(5):3427–3452, 2024.
- [BR97] Ola Bratteli and Derek W. Robinson. *Operator Algebras and Quantum Statistical Mechanics II*. Springer Berlin Heidelberg, 1997.
- [BSBN02] Stephen D. Bartlett, Barry C. Sanders, Samuel L. Braunstein, and Kae Nemoto. Efficient classical simulation of continuous variable quantum information processes. *Phys. Rev. Lett.*, 88:097904, Feb 2002.
- [Cam11] Earl T Campbell. Catalysis and activation of magic states in fault-tolerant architectures. *Physical Review A—Atomic, Molecular, and Optical Physics*, 83(3):032317, 2011.
- [CDG<sup>+</sup>20] Ulysse Chabaud, Tom Douce, Frédéric Grosshans, Elham Kashefi, and Damian Markham. Building Trust for Continuous Variable Quantum States. In Steven T. Flammia, editor, *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, volume 158 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:15, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [CFF23] Cameron Calcluth, Alessandro Ferraro, and Giulia Ferrini. Vacuum provides quantum advantage to otherwise simulatable architectures. *Physical Review A*, 107(6):062414, 2023.
- [CLO22] Sitan Chen, Jerry Li, and Ryan O’Donnell. Toward instance-optimal state certification with incoherent measurements. In *Conference on Learning Theory*, pages 2541–2596. PMLR, 2022.
- [DG17] M.A. De Gosson. *The Wigner Transform*. Advanced Textbooks In Mathematics. World Scientific Publishing Company, 2017.

- [dSLCP11] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Phys. Rev. Lett.*, 107:210404, Nov 2011.
- [FL11] Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.*, 106:230501, Jun 2011.
- [FMcv11] Radim Filip and Ladislav Mišta. Detecting quantum states with a positive wigner function beyond mixtures of gaussian states. *Phys. Rev. Lett.*, 106:200401, May 2011.
- [GÁCFF20] Laura García-Álvarez, Cameron Calcluth, Alessandro Ferraro, and Giulia Ferrini. Efficient simulatability of continuous-variable circuits with large wigner negativity. *Physical Review Research*, 2(4):043322, 2020.
- [GAG<sup>+</sup>23] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. Precision bounds on continuous-variable state tomography using classical shadows, 2023.
- [GPR23] Naixu Guo, Feng Pan, and Patrick Rebentrost. Estimating properties of a quantum state by importance-sampled operator shadows, 2023.
- [HFH<sup>+</sup>22] Oliver Hahn, Alessandro Ferraro, Lina Hultquist, Giulia Ferrini, and Laura García-Álvarez. Quantifying qubit magic resource with Gottesman-Kitaev-Preskill encoding. *Physical Review Letters*, 128(21):210502, 2022.
- [HFT24] Oliver Hahn, Giulia Ferrini, and Ryuji Takagi. Bridging magic and non-gaussian resources via Gottesman-Kitaev-Preskill encoding. *arXiv preprint arXiv:2406.06418*, 2024.
- [Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.
- [HPS24] Hsin-Yuan Huang, John Preskill, and Mehdi Soleimanifar. Certifying almost all quantum states with few single-qubit measurements, 2024.
- [Hud74] R.L. Hudson. When is the wigner quasi-probability density non-negative? *Reports on Mathematical Physics*, 6(2):249–252, 1974.
- [KŻ04] Anatole Kenfack and Karol Życzkowski. Negativity of the wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396, 2004.
- [Lan00] L. J. Landau. Bessel functions: Monotonicity and bounds. *Journal of the London Mathematical Society*, 61(1):197–215, 2000. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1112/S0024610799008352](https://onlinelibrary.wiley.com/doi/pdf/10.1112/S0024610799008352).
- [LD97] L. G. Lutterbach and L. Davidovich. Method for direct measurement of the wigner function in cavity qed and ion traps. *Phys. Rev. Lett.*, 78:2547–2550, Mar 1997.
- [LP95] U. Leonhardt and H. Paul. Measuring the quantum state of light. *Progress in Quantum Electronics*, 19(2):89–130, 1995.

- [LR09] A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81:299–332, Mar 2009.
- [Mar82] Clemens Markt. Mean cesàro summability of laguerre expansions and norm estimates with shifted parameter. *Analysis Mathematica*, 8:19–37, 1982.
- [MDR<sup>+</sup>24] Antoine Marquet, Simon Dupouy, Ulysse Réglade, Antoine Essig, Joachim Cohen, Emanuele Abertinale, Audrey Bienfait, Théau Peronnin, Sébastien Jezouin, Raphaël Lescanne, and Benjamin Huard. Harnessing two-photon dissipation for enhanced quantum measurement and control, 2024.
- [ME12] A. Mari and J. Eisert. Positive wigner functions render classical simulation of quantum computation efficient. *Phys. Rev. Lett.*, 109:230503, Dec 2012.
- [MEC<sup>+</sup>24] A. Marquet, A. Essig, J. Cohen, N. Cottet, A. Murani, E. Albertinale, S. Dupouy, A. Bienfait, T. Peronnin, S. Jezouin, R. Lescanne, and B. Huard. Autoparametric resonance extending the bit-flip time of a cat qubit up to 0.3 s. *Phys. Rev. X*, 14:021019, Apr 2024.
- [MKC09] A. Mandilara, E. Karpov, and N. J. Cerf. Extending hudson’s theorem to mixed quantum states. *Phys. Rev. A*, 79:062302, Jun 2009.
- [MM13] Elizabeth Meckes and Mark Meckes. Spectral measures of powers of random matrices. *Electronic communications in probability*, 18, 2013.
- [MMB<sup>+</sup>24] Francesco Anna Mele, Antonio Anna Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore F. E. Oliviero. Learning quantum states of continuous variable systems, 2024.
- [NRO<sup>+</sup>00] G. Nogues, A. Rauschenbeutel, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, S. Haroche, L. G. Lutterbach, and L. Davidovich. Measurement of a negative value for the wigner function of radiation. *Phys. Rev. A*, 62:054101, Oct 2000.
- [PWB15a] Hakop Pashayan, Joel J Wallman, and Stephen D Bartlett. Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Physical review letters*, 115(7):070501, 2015.
- [PWB15b] Hakop Pashayan, Joel J. Wallman, and Stephen D. Bartlett. Estimating Outcome Probabilities of Quantum Circuits Using Quasiprobabilities. 115(7):070501, 2015.
- [RBG<sup>+</sup>24] U. Réglade, A. Bocquet, R. Gautier, J. Cohen, A. Marquet, E. Albertinale, N. Pankratova, M. Hallén, F. Rautschke, L.-A. Sellem, P. Rouchon, A. Sarlette, M. Mirrahimi, P. Campagne-Ibarcq, R. Lescanne, S. Jezouin, and Z. Leghtas. Quantum control of a cat qubit with bit-flip times exceeding ten seconds. *Nature*, 629(8013):778–783, 2024. Publisher: Nature Publishing Group.
- [Roy77] Antoine Royer. Wigner function as the expectation value of a parity operator. *Phys. Rev. A*, 15:449–450, Feb 1977.



- [SC83] Francisco Soto and Pierre Claverie. When is the wigner function of multidimensional systems nonnegative? *Journal of Mathematical Physics*, 24(1):97–100, 1983. [\\_eprint: https://pubs.aip.org/aip/jmp/article-pdf/24/1/97/19128812/97\\_1\\_online.pdf](https://pubs.aip.org/aip/jmp/article-pdf/24/1/97/19128812/97_1_online.pdf).
- [Sch] Stefan Scheel. Quantum optics. Available at [https://www.lkv.uni-rostock.de/storages/uni-rostock/Alle\\_MNF/Physik\\_Qms/Lehre\\_S](https://www.lkv.uni-rostock.de/storages/uni-rostock/Alle_MNF/Physik_Qms/Lehre_S)
- [SEL<sup>+</sup>22] V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. H. Devoret. Model-free quantum control with reinforcement learning. *Phys. Rev. X*, 12:011059, Mar 2022.
- [Sim92] Barry Simon. The weyl transform and lp functions on phase space. *Proceedings of the American Mathematical Society*, 116(4):1045–1047, 1992.
- [Sze75] G. Szegő. *Orthogonal Polynomials*. American Math. Soc: Colloquium publ. American Mathematical Society, 1975.
- [TCJ20] Kok Chuan Tan, Seongjeon Choi, and Hyunseok Jeong. Negativity of quasiprobability distributions as a measure of nonclassicality. *Phys. Rev. Lett.*, 124:110404, Mar 2020.
- [Tes14] Gerald Teschl. *Mathematical methods in quantum mechanics: with applications to Schrödinger operators*. Number volume 157 in Graduate studies in mathematics. American Mathematical Society, second edition edition, 2014.
- [VKL<sup>+</sup>13a] Brian Vlastakis, Gerhard Kirchmair, Zaki Leghtas, Simon E. Nigg, Luigi Frunzio, S. M. Girvin, Mazyar Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. Deterministically encoding quantum information using 100-photon schrödinger cat states. *Science*, 342(6158):607–610, 2013.
- [VKL<sup>+</sup>13b] Brian Vlastakis, Gerhard Kirchmair, Zaki Leghtas, Simon E. Nigg, Luigi Frunzio, S. M. Girvin, Mazyar Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. Deterministically encoding quantum information using 100-photon schrödinger cat states. *Science*, 342(6158):607–610, 2013.
- [VV17] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.
- [VWFE13] Victor Veitch, Nathan Wiebe, Christopher Ferrie, and Joseph Emerson. Efficient simulation scheme for a class of quantum optics experiments with non-negative wigner representation. *New Journal of Physics*, 15(1):013037, 2013.
- [Wer84] R. Werner. Quantum harmonic analysis on phase space. *Journal of Mathematical Physics*, 25(5):1404–1411, 1984. [\\_eprint: https://pubs.aip.org/aip/jmp/article-pdf/25/5/1404/11262255/1404\\_1\\_online.pdf](https://pubs.aip.org/aip/jmp/article-pdf/25/5/1404/11262255/1404_1_online.pdf).
- [WGR<sup>+</sup>16] Chen Wang, Yvonne Y. Gao, Philip Reinhold, R. W. Heeres, Nissim Ofek, Kevin Chou, Christopher Axline, Matthew Reagor, Jacob Blumoff, K. M. Sliwa, L. Frunzio, S. M. Girvin, Liang Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. A schrödinger cat living in two boxes. *Science*, 352(6289):1087–1091, 2016.