

LOCAL ASYMPTOTIC EQUIVALENCE OF PURE STATES ENSEMBLES AND QUANTUM GAUSSIAN WHITE NOISE

BY CRISTINA BUTUCEA, MĂDĂLIN GUȚĂ AND MICHAEL NUSSBAUM¹

Université Paris-Saclay, University of Nottingham and Cornell University

Dedicated to Richard D. Gill on the occasion of his 66th birthday

Quantum technology is increasingly relying on specialised statistical inference methods for analysing quantum measurement data. This motivates the development of “quantum statistics”, a field that is shaping up at the overlap of quantum physics and “classical” statistics. One of the less investigated topics to date is that of statistical inference for infinite dimensional quantum systems, which can be seen as quantum counterpart of nonparametric statistics. In this paper, we analyse the asymptotic theory of quantum statistical models consisting of ensembles of quantum systems which are identically prepared in a pure state. In the limit of large ensembles, we establish the local asymptotic equivalence (LAE) of this i.i.d. model to a quantum Gaussian white noise model. We use the LAE result in order to establish minimax rates for the estimation of pure states belonging to Hermite–Sobolev classes of wave functions. Moreover, for quadratic functional estimation of the same states we note an elbow effect in the rates, whereas for testing a pure state a sharp parametric rate is attained over the nonparametric Hermite–Sobolev class.

1. Introduction. A striking insight of quantum mechanics is that randomness is a fundamental feature of the physical world at the microscopic level. Any observation made on a quantum system such as an atom or a light pulse, results in a nondeterministic, stochastic outcome. The study of the *direct map* from the system’s state or preparation to the probability distribution of the measurement outcomes, has been one of the core topics in traditional quantum theory. In recent decades, the focus of research has shifted from fundamental physics toward applications at the interface with information theory, computer science and metrology, sharing the paradigm that individual quantum systems are carriers of a new type of information [47].

In many quantum protocols, the experimenter has incomplete knowledge and control of the system and its environment, or is interested in estimating an external

Received May 2017; revised December 2017.

¹Supported in part by NSF Grant DMS-1407600.

MSC2010 subject classifications. Primary 62B15; secondary 62G05, 62G10, 81P50.

Key words and phrases. Le Cam distance, local asymptotic equivalence, quantum Gaussian process, quantum Gaussian sequence, quantum states ensemble, nonparametric estimation, quadratic functionals, nonparametric sharp testing rates.

field parameter which affects the system dynamics. In this case, one deals with a *statistical inverse problem* of inferring unknown state parameters from the measurement data obtained by probing a large number of individual quantum systems. The theory and practice arising from tackling such questions is shaping up into the field of *quantum statistics*, which lies at the intersection of quantum theory and statistical inference [1, 6, 31, 32, 34, 51].

One of the central problems in quantum statistics is state estimation: given an ensemble of identically prepared, independent systems with unknown state, the task is to estimate the state by performing appropriate measurements and devising estimators based on the measurement data. A landmark experiment aimed at creating multipartite entangled states [29] highlighted the direct practical relevance of efficient estimation techniques for large dimensional systems, the complexity of estimating large dimensional states and the need for solid statistical methodology in computing reliable “error bars”. This has motivated the development of new methods such as compressed sensing and matrix ℓ_1 -minimisation [19, 23, 24], spectral thresholding for low rank states [11], confidence regions [15, 16, 18, 55, 58].

Another important research direction is toward developing a *quantum decision theory* as the overall mathematical framework for inference involving quantum systems seen as a form of “statistical data”. Typically, the route to finding the building blocks of this theory starts with a decision problem (e.g., testing between two states, or estimating certain parameters of a state) and the problem of finding optimal measurement settings and statistical procedures for treating the (classical, random) measurement data. For instance, in the context of asymptotic binary hypothesis testing, two key results are the quantum Stein lemma [33, 50] and the quantum Chernoff bound [2, 3, 45, 49]. As in the classical case, they describe the exponential decay of appropriate error probabilities for optimal measurements, and they provide operational interpretations for quantum relative entropy, and respectively quantum Chernoff distance. Similarly, an important problem in state estimation is to identify measurements which allow for the smallest possible estimation error. A traditional approach has been to establish a “quantum Cramér–Rao bound” (QCRB) [9, 32, 34] for the covariance of unbiased estimators, where the right side is the inverse of the “quantum Fisher information matrix”, the latter depending only on the structure of the quantum statistical model. However, while the QCRB is achievable asymptotically for one-dimensional parameters, this is not the case for multi-parameter models due to the fact that the measurements which are optimal for different one-dimensional components, are generally incompatible with each other.

These difficulties can be overcome by developing a fundamental theory of comparison and convergence of quantum statistical models, as an extension of its classical counterpart [43, 56]. While classical “data processing” is described by randomisations, physical transformations of quantum systems are described by *quantum channels* [47]. Following up on this idea, Petz and Jencova [37] have obtained

a general characterisation of *equivalent models*, as families of states that are related by quantum channels in both directions. This naturally leads to the notion of *Le Cam distance* between quantum statistical models as the least trace-norm error incurred when trying to map one model into another via quantum channels [38]. In this framework, the asymptotic theory of state estimation can be investigated by adopting ideas from the classical *local asymptotic normality* (LAN) theory [43]. Quantum LAN theory [26, 27, 38] shows that the sequence of models describing large samples of identically prepared systems can be approximated by a simpler *quantum Gaussian shift model*, in the neighbourhood of an interior point of the parameter space. The original optimal state estimation problem is then solved by combining LAN theory with known procedures for estimation of Gaussian states [21, 25, 28].

In this paper, we extend the scope of the quantum LAN theory to cover *non-parametric* quantum models; more precisely, we will be interested in the set of *pure states* (one-dimensional projections) on *infinite dimensional Hilbert spaces*. Infinite dimensional systems such as light pulses, free particles are commonly encountered in quantum physics, and their estimation is an important topic in quantum optics [44]. The minimax results derived in this paper can serve as a benchmark for the performance of specific methods such as for instance quantum homodyne tomography [1, 10] by comparing their risk with the minimax risk derived here.

The paper is organised as follows. In Section 2, we review the basic notions of quantum mechanics needed for understanding the physical context of our investigation. In particular, we define the concepts of state, measurement and quantum channel which can loosely be seen as quantum analogs of probability distribution and Markov kernels, respectively. We further introduce the formalism of quantum Gaussian states, the Fock spaces and second quantisation, which establish the quantum analogs of Gaussian distributions, Gaussian sequences and Gaussian processes in continuous time. In Section 3.1, we introduce the general notion of a quantum statistical model and the Le Cam distance between two models. In particular, in Section 3.2 we define the i.i.d. and Gaussian quantum models which are analysed in the remainder of the paper. In Appendix A.1 [12], we review results in classical statistics on nonparametric asymptotic equivalence which serve as motivation and comparison to our work.

One of the main results is Theorem 4.1 giving the *local asymptotic equivalence* (LAE) between the nonparametric i.i.d. pure states model and the Gaussian shift model. This extends the existing local asymptotic normality theory from parametric to nonparametric (infinite dimensional) models. Section 5 details three applications of the LAE result in Theorem 4.1. In Section 5.1, we derive the asymptotic minimax rates and provide concrete estimation procedures for state estimation with respect to the trace-norm and Bures distances, which are analogues of the norm-one and Hellinger distances, respectively. The main results are Theorems 5.1 and 5.3 which deal with the upper and respectively lower bound for a

model consisting of an ensemble of n independent identically prepared systems in a pure state belonging to a Hermite–Sobolev class $S^\alpha(L)$ of wave functions. In Theorem 5.1, we describe a specific measurement procedure which provides an estimator whose risk attains the nonparametric rate $n^{-\alpha/(2\alpha+1)}$. The lower bound follows by using the LAE result to approximate the model with a Gaussian one, combined with the lower bound for the corresponding quantum Gaussian model derived in Theorem 5.2. In Section 5.2, we consider the estimation of a state functional corresponding to the expectation of a power $N^{2\beta}$ of the number operator. Theorems 5.4 and 5.5 establish the upper and lower bounds for functional estimation for the Hermite–Sobolev class $S^\alpha(L)$. The minimax rates are $n^{-1/2}$ (parametric) if $\alpha \geq 2\beta$, and $n^{-1+\beta/\alpha}$ if $\beta < \alpha < 2\beta$. In Section 5.3, we investigate nonparametric testing between a single state and a composite hypothesis consisting of all states outside a ball of shrinking radius. Surprisingly, we find that the minimax testing rates are parametric, in contrast to the nonparametric estimation rates. This fact is closely related to the fact that the optimal estimation and testing measurements are incompatible with each other, so that no single measurement strategy can allow for minimax estimation and testing in the same time. Results on the minimax optimal rate for testing and the sharp asymptotics are given in Theorems 5.6 and 5.7, respectively. Further discussion on these topics and proofs of all results are presented in Appendix A and B in [12], respectively.

Notation. Following physics convention, the vectors of a Hilbert space \mathcal{H} will be denoted by the “ket” $|v\rangle$, so that the inner product of two vectors is the “bra-ket” $\langle u|v\rangle \in \mathbb{C}$ which is linear with respect to the right entry and anti-linear with respect to the left entry. Similarly, $M := |u\rangle\langle v|$ is the rank one operator acting as $M : |w\rangle \mapsto M|w\rangle = \langle v|w\rangle|u\rangle$. We denote by $\mathcal{L}(\mathcal{H})$ the space of bounded linear operators on \mathcal{H} which is a C^* -algebra with respect to the operator norm $\|A\| := \sup_{\psi \neq 0} \|A\psi\|/\|\psi\|$. Additionally, $\mathcal{T}_1(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ is the space of Hilbert–Schmidt (or trace-class) operators equipped with the norm-one $\|\tau\|_1 := \text{Tr}(|\tau|)$, where the operator $|\tau| := (\tau^*\tau)^{1/2}$ is the absolute value of τ , and τ^* is the adjoint of τ . Finally, we denote by $\mathcal{T}_2(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ the space of Hilbert–Schmidt operators equipped with the norm-two $\|\tau\|_2^2 := \text{Tr}(|\tau|^2)$, which is a Hilbert space with respect to the inner product $(\tau, \sigma) := \text{Tr}(\tau^*\sigma)$.

2. Quantum mechanics background. In this section, we review some basic notions of quantum mechanics (QM), in as much as it is required for understanding the subsequent results of the paper. Since QM is a probabilistic theory of quantum phenomena, it is helpful to approach the formalism from the perspective of analogies and differences with “classical” probability. We refer to [47] for more details on the quantum formalism.

2.1. *States, measurements, channels.* The QM formalism assigns to each quantum mechanical system (e.g., an atom, light pulse, quantum spin) a complex Hilbert space \mathcal{H} , called the space of states. For instance, the finite dimensional space \mathbb{C}^d is the Hilbert space of a system with d “energy levels”, while $L^2(\mathbb{R})$ is the space of “wave functions” of a particle moving in one dimension, or of a monochromatic light pulse. The *state* of a quantum system is represented mathematically by a density matrix.

DEFINITION 1. Let \mathcal{H} be the Hilbert space of a quantum system. A density matrix (or state) on \mathcal{H} is a linear operator $\rho : \mathcal{H} \rightarrow \mathcal{H}$ which is positive (i.e., it is self-adjoint and has nonnegative eigenvalues), and has trace one.

We denote by $\mathcal{S}(\mathcal{H})$ the convex space of states on \mathcal{H} . Its linear span is the space of trace class operators $\mathcal{T}_1(\mathcal{H})$, which is the noncommutative analogue of the space of absolutely integrable functions on a probability space $L_1(\Omega, \Sigma, \mathbb{P})$. For any states ρ_1 or ρ_2 , the convex combination $\lambda\rho_1 + (1 - \lambda)\rho_2$ is also a state which corresponds to randomly preparing the system in either the state ρ_1 or ρ_2 with probabilities λ , and respectively $1 - \lambda$. The extremal elements of the convex set $\mathcal{S}(\mathcal{H})$ are the one-dimensional projections $P_\psi = |\psi\rangle\langle\psi|$ where $|\psi\rangle$ is a normalised vector, that is, $\|\psi\| = 1$. Such states are called *pure* (as opposed to mixed states which are convex combinations of pure ones), and are uniquely determined by the vector $|\psi\rangle$. Conversely, the vector $|\psi\rangle$ is fixed by the state up to a complex phase factor, that is, $|\psi\rangle$ and $|\psi'\rangle := e^{i\phi}|\psi\rangle$ represent the same state.

Although the quantum state encodes all information about the preparation of the system, it is not a directly observable property. Instead, any measurement produces a *random outcome* whose distribution depends on the state, and thus reveals in a probabilistic way a certain aspect of the system’s preparation. The simplest type of measurement is determined by an orthonormal basis (ONB) $\{|i\rangle\}_{i=1}^{\dim \mathcal{H}}$ and a set of possible outcomes $\{\lambda_i\}_{i=1}^{\dim \mathcal{H}}$ in the following way: the outcome is a random variable X taking the value λ_i with probability given by the diagonal elements of ρ in this particular basis

$$\mathbb{P}_\rho([X = \lambda_i]) = \rho_{ii} = \langle i|\rho|i\rangle.$$

More generally, a measurement M with outcomes in a measurable space (Ω, Σ) is determined by a positive operator valued measure.

DEFINITION 2. A positive operator valued measure (POVM) is a map $M : \Sigma \rightarrow \mathcal{L}(\mathcal{H})$ having the following properties:

- (1) positivity: $M(E) \geq 0$ for all events $E \in \Sigma$
- (2) σ -additivity: $M(\bigcup_i E_i) = \sum_i M(E_i)$ for any countable set of mutually disjoint events E_i
- (3) normalization: $M(\Omega) = \mathbf{1}$.

The outcome of the corresponding measurement associated to M has probability distribution

$$\mathbb{P}_\rho(E) = \text{Tr}(\rho M(E)), \quad E \in \Sigma.$$

The most important example of a POVM is that associated to the measurement of an observable, the latter being represented mathematically by a self-adjoint operator $A : \mathcal{H} \rightarrow \mathcal{H}$. The spectral theorem shows that such operators can be “diagonalised”, that is, they have a spectral decomposition

$$A = \int_{\sigma(A)} x P(dx),$$

where $\sigma(A)$ is the spectrum of A , and $\{P(E) : E \in \Sigma\}$ is the collection of spectral projections of A . The corresponding measurement has outcome $a \in \sigma(A)$ with probability distribution $\mathbb{P}_\rho[a \in E] = \text{Tr}(\rho P(E))$.

Unlike “classical” systems which can be observed without disturbing their state, quantum systems are typically perturbed by the measurement, so the system needs to be re-prepared in order to obtain more information about the state. In this sense, the system can be seen as a “quantum sample” which it can be converted into a “classical” sample only by performing a measurement. Thus, a measurement can be seen as a “quantum-to-classical randomisation”, that is, a linear map \mathcal{M} which sends a state ρ to the probability density $\mathcal{M}(\rho) \equiv p_\rho := \frac{d\mathbb{P}_\rho}{d\mathbb{P}}$ with respect to a reference measure \mathbb{P} . The latter can be taken to be \mathbb{P}_{ρ_0} for a strictly positive density matrix ρ_0 . The following lemma summarises this perspective on measurements.

LEMMA 2.1. *Let \mathcal{H} be a Hilbert space, and let (Ω, Σ) be a measurable space. For any fixed state $\rho_0 > 0$ on \mathcal{H} , there is a one-to-one correspondence between POVMs M over (Ω, Σ) and quantum-to-classical randomisations, that is, linear maps*

$$\mathcal{M} : \mathcal{T}_1(\mathcal{H}) \rightarrow L_1(\Omega, \Sigma, \mathbb{P})$$

which are positive and normalised (maps states into probability densities). The correspondence is given by

$$\mathbb{P}_\rho(E) = \text{Tr}(M(E)\rho) = \int_A p_\rho(\omega) \mathbb{P}_{\rho_0}(d\omega), \quad \mathcal{M}(\rho) \equiv p_\rho := \frac{d\mathbb{P}_\rho}{d\mathbb{P}}.$$

For comparison, recall that a linear map $R : L_1(\Omega', \Sigma', \mathbb{P}') \rightarrow L_1(\Omega, \Sigma, \mathbb{P})$ is a stochastic operator if it maps probability densities into probability densities [54]. Typically, such maps arise from Markov kernels and describe randomizations of dominated statistical experiments (models).

While a measurement is a quantum-to-classical randomization, a “quantum-to-quantum randomization” describes how the system’s state changes as a result of time evolution or interaction with other systems. The maps describing such transformations are called quantum channels.

DEFINITION 3. A quantum channel between systems with Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 is a trace preserving, completely positive linear map $T : \mathcal{T}_1(\mathcal{H}_1) \rightarrow \mathcal{T}_1(\mathcal{H}_2)$.

The two properties mentioned above are similar to those of a classical randomization, so in particular T maps states into states. However, unlike the classical case, T is required to satisfy a stronger positivity property: T is *completely positive* if $\text{Id}_m \otimes T$ is positive for all $m \geq 1$, where Id_m is the identity map on the space of m dimensional matrices. This ensures that when the system is correlated with an ancillary system \mathbb{C}^m , and the latter undergoes the identity transformation, the final joint state is still positive, as expected on physical grounds.

The simplest example of a quantum channel is a unitary transformation $\rho \mapsto U\rho U^*$, where U is a unitary operator on \mathcal{H} . More generally, if $|\varphi\rangle \in \mathcal{K}$ is a pure state of an ancillary system, and V is a unitary on $\mathcal{H} \otimes \mathcal{K}$, then

$$\rho \mapsto T(\rho) := \text{Tr}_{\mathcal{K}}(V(\rho \otimes |\varphi\rangle\langle\varphi|)V^*)$$

is a quantum channel describing the system state after interacting with the ancilla. By computing the partial trace $\text{Tr}_{\mathcal{K}}$ over \mathcal{K} with respect to an orthonormal basis $\{|f_i\rangle\}_{i=1}^{\dim \mathcal{K}}$, we obtain the following expression:

$$(1) \quad T(\rho) = \sum_i K_i \rho K_i^*,$$

where K_i are operators on \mathcal{H} defined by $\langle\psi|K_i|\psi'\rangle := \langle\psi \otimes f_i|U|\psi' \otimes \varphi\rangle$. Note that by definition, these operators satisfy the normalisation condition $\sum_i K_i^* K_i = \mathbf{1}$. Conversely, the Kraus theorem shows that any quantum channel is of the form (1) with operators K_i respecting the normalisation condition.

2.2. *Continuous variables, Fock spaces and Gaussian states.* In this section, we look at the class of “continuous variables” (cv) systems, which model a variety of physical systems such as light pulses, or free particles. Such systems play an important role in this work as “carriers” of quantum Gaussian states, and in particular in the local asymptotic equivalence result. We refer to [44] for further reading.

2.2.1. *One mode systems.* We start with the simplest case of a “one-mode” cv system, after which we show how this construction can be extended to more general “multi-mode” cv systems. The Hilbert space of a one-mode system is $L_2(\mathbb{R})$, that is, the space of square integrable wave functions on the real line. On this we define the self-adjoint operators acting on appropriately defined domains as

$$(Q\psi)(q) = q\psi(q), \quad (P\psi)(q) = -i \frac{d\psi(q)}{dq}$$

which satisfy the “canonical commutation relations” $QP - PQ = i\mathbf{1}$. To better understand the meaning of the observable Q , let us consider its measurement for

a pure state $\rho = |\psi\rangle\langle\psi|$ with wave function $|\psi\rangle$. The outcome takes values in \mathbb{R} , and its probability distribution has density with respect to the Lebesgue measure $p_\rho^Q(x) = |\psi(x)|^2$. Similarly, the probability density of the observable P is given by $p_\rho^P(x) = |\tilde{\psi}(x)|^2$, where $\tilde{\psi} \in L^2(\mathbb{R})$ is the Fourier transform of the function $\psi(\cdot)$. When the system under consideration is the free particle, Q and P are usually associated to the position and momentum observables, while for a monochromatic light mode they correspond to the electric and magnetic fields. Note that the distributions of P and Q are not sufficient to identify the state, even in the case of a pure state. However, it turns out that the state is uniquely determined by the collection of probability distributions of all *quadrature* observables $X_\phi := \cos(\phi) \cdot Q + \sin(\phi) \cdot P$ for angles $\phi \in [0, 2\pi]$. To understand this, it is helpful to think of the state of the one-mode cv system as a quantum analogue of a joint distribution of two real valued variables, that is, a 2D distribution. Indeed, in the latter case, the distribution is determined by collection of marginals along all directions in the plane (its Radon transform); this fact is exploited in PET tomography which aims at estimating the 2D distribution from samples of its Radon transform. In the quantum case, since Q and P do not commute with each other, they cannot be measured simultaneously and cannot be assigned a joint distribution in a meaningful way. However, the “quasi-distribution” defined below has some of the desired properties, and is very helpful in visualising the quantum state.

DEFINITION 4. For any state $\rho \in \mathcal{T}_1(L_2(\mathbb{R}))$, we define the *quantum characteristic function* of ρ

$$\tilde{W}_\rho(u, v) := \text{Tr}(\exp(-iuQ - ivP)\rho).$$

The inverse Fourier transform of \tilde{W}_ρ with respect to both variables is called Wigner function W_ρ , or quasi-distribution associated to ρ :

$$W_\rho(q, p) = \frac{1}{(2\pi)^2} \iint \exp(iuq + ivp)\tilde{W}_\rho(u, v) du dv.$$

A consequence of this definition is that the marginal of $W_\rho(q, p)$ along an arbitrary direction with angle ϕ is the probability density of the quadrature X_ϕ introduced above. This is the basis of a quantum state estimation scheme called “quantum homodyne tomography” [1, 44], where the Wigner function plays the role of the 2D distribution from “classical” PET tomography. One of the important differences, however, is that the Wigner functions need not be positive in general, and satisfy other constraints which are specific to the quantum setting and can be exploited in the estimation procedure.

The Wigner function representation offers an intuitive route to defining the notion of Gaussian state.

DEFINITION 5. A state ρ of a one-mode cv system is called Gaussian if its Wigner function W_ρ is a Gaussian probability density, or equivalently if it has the quantum characteristic function

$$\widetilde{W}_\rho(u, v) = \exp\left(- (u, v) \frac{V}{2} (u, v)^T\right) \cdot \exp(iuq_0 + ivp_0),$$

where $(q_0, p_0) \in \mathbb{R}^2$ and V (a real positive 2×2 matrix) are the mean and variance of W_ρ , respectively.

In particular, all the quadratures X_ϕ of a Gaussian state have Gaussian distribution. As consequence of the commutation relation $QP - PQ = i\mathbf{1}$, the observables Q and P cannot have arbitrarily small variance simultaneously; in particular, the covariance matrix V must satisfy the ‘‘uncertainty principle’’ $\text{Det}(V) \geq 1/4$, where the equality is achieved if and only if the state is a pure Gaussian state.

We will be particularly interested in *coherent* states $|G(z)\rangle$ which are *pure* Gaussian states whose Wigner functions have covariance matrix $V = I_2/2$, where I_2 is the 2×2 identity matrix. To give a concrete Hilbert space representation, it is convenient to introduce a special orthonormal basis of $L_2(\mathbb{R})$, consisting of the eigenvectors $\{|0\rangle, |1\rangle, \dots\}$ of the *number operator* $N = a^*a$, with $N|k\rangle = k|k\rangle$. Here, the operators $a^* = (Q - iP)/\sqrt{2}$ and $a = (Q + iP)/\sqrt{2}$ are called *creation and annihilation operators* and act as ‘‘ladder operators’’ on the number basis vectors (or Fock states):

$$\begin{aligned} a|k\rangle &= \sqrt{k}|k - 1\rangle, \\ a^*|k\rangle &= \sqrt{k + 1}|k + 1\rangle. \end{aligned}$$

The coherent states denoted by $|G(z)\rangle$ are obtained by applying the unitary Weyl (displacement) operators to the vacuum state $|0\rangle$

$$(2) \quad |G(z)\rangle = \exp(za^* - \bar{z}a)|0\rangle = \exp(-|z|^2/2) \sum_{k=0}^\infty \frac{z^k}{\sqrt{k!}}|k\rangle,$$

where $z \in \mathbb{C}$ is the eigenvalue of the annihilation operator $a|G(z)\rangle = z|G(z)\rangle$; in particular, the quadrature means are $\langle G(z)|Q|G(z)\rangle = \sqrt{2}\text{Re}(z)$ and $\langle G(z)|P|G(z)\rangle = \sqrt{2}\text{Im}(z)$, and the Wigner function is given by

$$(3) \quad W_{|z\rangle}(q, p) = \frac{1}{\pi} \exp\left(- (q - \sqrt{2}\text{Re}(z))^2 - (p - \sqrt{2}\text{Im}(z))^2\right), \quad q, p \in \mathbb{R}.$$

Equation (2) implies that the number operator N has a Poisson distribution with mean $|z|^2$. Additionally, it can be seen from the Fourier expansion in the second equality that the unitary $\Gamma(\phi) = \exp(i\phi N)$ acts by rotating the coherent states by an angle ϕ in the complex plane, that is, $\Gamma(\phi)|G(z)\rangle = |G(e^{i\phi}z)\rangle$.

Another important class of Gaussian states are the mixed diagonal states

$$(4) \quad \Phi(r) = (1 - r) \sum_{k=0}^{\infty} r^k |k\rangle\langle k|, \quad 0 < r < 1$$

which are also called thermal states; cf. Section 3.3 in [44]. The corresponding Wigner function is a centred Gaussian

$$W_{\Phi(r)}(q, p) = \frac{1}{2\pi\sigma^2(r)} \exp\left(-\frac{q^2 + p^2}{2\sigma^2(r)}\right)$$

with covariance matrix $V = \sigma^2(r) \cdot I_2$ where $\sigma^2(r) = \frac{1}{2} \frac{1+r}{1-r}$.

PROPOSITION 2.2. *Consider the family of coherent states $\{|G(z)\rangle\langle G(z)|, z \in \mathbb{C}\}$, with random displacement (location) z distributed according to $\Pi(dz)$, having a Gaussian law with covariance matrix $\sigma^2 \cdot I_2$. Then the mixed state $\Phi = \int |G(z)\rangle\langle G(z)|\Pi(dz)$ is the thermal state $\Phi(r)$ with $r = \frac{2\sigma^2}{2\sigma^2+1}$.*

PROOF. Consider the corresponding Wigner function

$$(5) \quad \begin{aligned} W_{\Phi}(q, p) &= \int W_{|G(z)\rangle}(q, p) \exp\left(-\frac{1}{2\sigma^2}(x^2 + y^2)\right) \frac{1}{2\pi\sigma^2} dx dy \\ &= \frac{1}{\pi\sigma^2} \int \exp\left(-\frac{(q - \sqrt{2}x)^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}\right) \frac{dx}{\sqrt{2\pi}} \\ &\quad \times \int \exp\left(-\frac{(p - \sqrt{2}y)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2}\right) \frac{dy}{\sqrt{2\pi}} \\ &= \frac{1}{\pi(4\sigma^2 + 1)} \exp\left(-\frac{q^2 + p^2}{2(2\sigma^2 + 1/2)}\right). \end{aligned}$$

Therefore, the state Φ is identical to the thermal state $\Phi(r)$ with $2\sigma^2 + \frac{1}{2} = \frac{1}{2} \frac{1+r}{1-r}$, or equivalently $r = \frac{2\sigma^2}{1+2\sigma^2}$. \square

This fact will be used later on in in Section 5 in applications to functional estimation and testing.

2.2.2. Fock spaces and second quantisation. The above construction can be generalised to multimode systems by tensoring several one-mode systems. Thus, the Hilbert space of a k -mode system is $L_2(\mathbb{R})^{\otimes k} \cong L_2(\mathbb{R}^k)$, upon which we define “canonical pairs” (Q_i, P_i) acting on the i th tensor as above, and as identity on the other tensors. Similarly, we define the one-mode operators a_i, a_i^*, N_i . The number basis consists now of tensor products $|n\rangle := \otimes_{i=1}^k |n_i\rangle$ indexed by the sequences

of integers $\mathbf{n} = (n_1, \dots, n_k)$. A multi-mode coherent state is a tensor product of one-mode coherent states

$$\begin{aligned}
 |G(\mathbf{z})\rangle &= \bigotimes_{i=1}^k |G(z_i)\rangle \\
 (6) \qquad &= \exp(\mathbf{z}\mathbf{a}^\dagger - \mathbf{a}\mathbf{z}^\dagger)|0\rangle \\
 &= \exp(-|\mathbf{z}|^2/2) \sum_{\mathbf{n}=0}^\infty \left(\prod_{i=1}^k \frac{z_i^{n_i}}{\sqrt{n_i!}} \right) |\mathbf{n}\rangle \in L_2(\mathbb{R})^{\otimes k},
 \end{aligned}$$

where $\mathbf{z} = (z_1, \dots, z_k)$ is the vector of means, $\mathbf{a} = (a_1, \dots, a_k)$, and \dagger denotes the transposition and adjoint (complex conjugation) of individual entries.

We will now extend this construction to systems with infinitely many modes. One way to do this is by defining an infinite tensor product of one-mode spaces, as completion of the space spanned by tensors in which all but a finite number of modes are in the vacuum state. Instead, we will present an equivalent but more elegant construction called *second quantisation* which will be useful for later considerations.

DEFINITION 6. Let \mathcal{K} be a Hilbert space. The *Fock space* over \mathcal{K} is the Hilbert space

$$(7) \qquad \mathcal{F}(\mathcal{K}) = \bigoplus_{n \geq 0} \mathcal{K}^{\otimes_s n},$$

where $\mathcal{K}^{\otimes_s n}$ denotes the n -fold symmetric tensor product, that is, the subspace of $\mathcal{K}^{\otimes n}$ consisting of vectors which are symmetric under permutations of the tensors. The term $\mathcal{K}^{\otimes_s 0} =: \mathbb{C}|0\rangle$ is called the vacuum state.

In this definition, the space \mathcal{K} should be regarded as the “space of modes” rather than physical states. As we will see below, by fixing an orthonormal basis in \mathcal{K} , we can establish an isomorphism between the Fock space $\mathcal{F}(\mathcal{K})$ and a tensor product of one-mode cv spaces, one for each basis vector. In particular, if $\mathcal{K} = \mathbb{C}$, then $\mathcal{F}(\mathbb{C}) \cong L^2(\mathbb{R})$ so that the one-dimensional subspaces in the direct sum in (7) correspond to the number basis vectors $|0\rangle, |1\rangle, \dots \in L^2(\mathbb{R})$ of a one-mode cv system.

We now introduce the general notion of coherent state on a Fock space.

DEFINITION 7. Let $\mathcal{F}(\mathcal{K})$ be the Fock space over \mathcal{K} . For each $|v\rangle \in \mathcal{K}$, we define an associated coherent state

$$|G(v)\rangle := e^{-\|v\|^2/2} \bigoplus_{n \geq 0} \frac{1}{\sqrt{n!}} |v\rangle^{\otimes n} \in \mathcal{F}(\mathcal{K}).$$

The coherent vectors form a dense subspace of $\mathcal{F}(\mathcal{K})$. This fact can be used to prove the following factorisation property, and to define the annihilation operators below. Let $\mathcal{K} = \mathcal{K}_0 \oplus \mathcal{K}_1$ be a direct sum decomposition of \mathcal{K} into orthogonal subspaces, and let $|v\rangle = |v_0\rangle \oplus |v_1\rangle$ be the decomposition of a generic vector $|v\rangle \in \mathcal{K}$. Then the map

$$U : \mathcal{F}(\mathcal{K}) \rightarrow \mathcal{F}(\mathcal{K}_0) \otimes \mathcal{F}(\mathcal{K}_1),$$

$$U : |G(v)\rangle \mapsto |G(v_0)\rangle \otimes |G(v_1)\rangle$$

is unitary. We will use this correspondence to identify $\mathcal{F}(\mathcal{K})$ with the tensor product $\mathcal{F}(\mathcal{K}_0) \otimes \mathcal{F}(\mathcal{K}_1)$. By the same argument, for any orthonormal basis $\{|e_1\rangle, |e_2\rangle, \dots\}$ of \mathcal{K} , the Fock space $\mathcal{F}(\mathcal{K})$ is isomorphic with the tensor product of one mode spaces $\mathcal{F}_i := \mathcal{F}(\mathbb{C}|e_i\rangle)$ and the coherent states factorise as

$$(8) \quad \mathcal{F}(\mathcal{K}) \cong \bigotimes_i \mathcal{F}_i,$$

$$|G(u)\rangle \cong \bigotimes_i |G(u_i)\rangle, \quad u_i = \langle e_i | u \rangle$$

so that we recover the formula (6).

We define the *annihilation operators* through their action on coherent states as follows: for each mode $|u\rangle \in \mathcal{K}$ the associated annihilator $a(u) : \mathcal{F}(\mathcal{K}) \rightarrow \mathcal{F}(\mathcal{K})$ is given by

$$a(u) : |G(v)\rangle = \langle u | v \rangle |G(v)\rangle, \quad |v\rangle \in \mathcal{K}.$$

Then the annihilation and (their adjoint) the creation operators satisfy the commutation relations

$$a(u)a^*(w) - a^*(w)a(u) = \langle u | v \rangle \mathbf{1}.$$

For each mode, we can also define the canonical operators $Q(u)$, $P(u)$ and the number operator $N(u)$ in terms of $a(u)$, $a^*(u)$ as in the one-mode case. Moreover, if $|u\rangle = |u_0\rangle \oplus |u_1\rangle$ is the decomposition of $|u\rangle$ as above, then $a(u_0)$ acts as $a(u_0) \otimes \mathbf{1}_{\mathcal{F}(\mathcal{K}_1)}$, when the Fock space is represented in the tensor product form. Similar decompositions hold for $a^*(u_0)$, $N(u_0)$, $a(u_1)$, $a^*(u_1)$, $N(u_1)$.

The second quantisation has the following functorial properties which will be used later on.

DEFINITION 8. Let $W : \mathcal{K} \rightarrow \mathcal{K}$ be a unitary operator. The quantisation operator $\Gamma(W)$ is the unitary defined by $\Gamma(W) : \mathcal{F}(\mathcal{K}) \rightarrow \mathcal{F}(\mathcal{K})$ by

$$\Gamma(W) := \bigoplus_{n \geq 0} W^{\otimes n},$$

where $W^{\otimes n}$ acts on the n th level of the Fock space $\mathcal{K}^{\otimes_s n}$.

From the definition, it follows that the action of $\Gamma(W)$ on coherent states is covariant in the sense that

$$\begin{aligned} \Gamma(W) &: \mathcal{F}(\mathcal{K}) \rightarrow \mathcal{F}(\mathcal{K}), \\ \Gamma(W) &: |G(v)\rangle \mapsto |G(Wv)\rangle. \end{aligned}$$

In particular, it follows from the definitions that $\Gamma(e^{i\phi}\mathbf{1}) = \exp(i\phi N)$, where N is the total number operator, whose action on the n th level of the Fock space is $N|v\rangle^{\otimes n} = n|v\rangle^{\otimes n}$. Note that while $|v\rangle$ and $e^{i\phi}|v\rangle$ differ only by a phase, and hence represent the same state, the corresponding coherent states $|G(v)\rangle$ and $\Gamma(e^{i\phi})|G(v)\rangle = |G(e^{i\phi}v)\rangle$ are linearly independent and represent different states.

As in the single mode case, the coherent states can be obtained by acting with the unitary displacement (or Weyl) operators onto the vacuum

$$|G(u)\rangle = \exp(a^*(u) - a(u))|0\rangle.$$

Moreover, the coherent states $|G(u)\rangle$ are Gaussian with respect to all coordinates. The means of annihilation operators are given by $\langle G(u)|a(w)|G(u)\rangle = \langle w|v\rangle$, from which we can deduce that the the coordinates $(Q(w), P(w))$ have means $(\sqrt{2}\operatorname{Re}\langle w|u\rangle, \sqrt{2}\operatorname{Im}\langle w|u\rangle)$. The covariance of coherent states is constant (independent of the displacement u), and is given by $\langle 0|a(w)a^*(v)|0\rangle = \langle w|v\rangle$. This implies that orthogonal modes (i.e., $\langle w|v\rangle = 0$) have independent pairs of coordinates.

2.3. Metrics on the space of states. For future reference, we review here the states space metrics used in the paper. Recall that the space of states $\mathcal{S}(\mathcal{H})$ on a Hilbert space \mathcal{H} is the cone of positive, trace one operators in $\mathcal{T}_1(\mathcal{H})$. The *norm-one* (or trace-norm) distance between two states $\rho_0, \rho_1 \in \mathcal{S}(\mathcal{H})$ is given by

$$\|\rho_0 - \rho_1\|_1 := \operatorname{Tr}(|\rho_0 - \rho_1|),$$

where $|\tau| := \sqrt{\tau^*\tau}$ denotes the absolute value of τ . The norm-one distance can be interpreted in terms of the maximum difference between expectations of bounded observables

$$\|\rho_0 - \rho_1\|_1 = 2 \sup_{A: \|A\| \leq 1} |\operatorname{Tr}(\rho_0 A) - \operatorname{Tr}(\rho_1 A)|.$$

Another interpretation is in terms of quantum testing. Let $M = (M_0, M_1)$ be a binary POVM used to test between hypotheses $H_0 := \{\text{measured state is } \rho_0\}$ and $H_1 := \{\text{measured state is } \rho_1\}$. The sum of error probabilities is

$$\mathbb{P}_e^M = \operatorname{Tr}(M_0\rho_1) + \operatorname{Tr}(M_1\rho_0).$$

By optimizing over all possible POVM, we obtain [32] the optimal error probability sum

$$(9) \quad \mathbb{P}_e^* := \inf_M \mathbb{P}_e^M = 1 - \frac{1}{2} \|\rho_0 - \rho_1\|_1.$$

In the special case of pure states, the norm-one distance is given by

$$(10) \quad \|\psi_0\rangle\langle\psi_0| - |\psi_1\rangle\langle\psi_1|\|_1 = 2\sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2},$$

as proven, for example, in [39]. The previous formula becomes for coherent states

$$\| |G(\psi_0)\rangle\langle G(\psi_0)| - |G(\psi_1)\rangle\langle G(\psi_1)| \|_1 = 2\sqrt{1 - \exp(-\|\psi_0 - \psi_1\|^2)}.$$

The second important metric is the *Bures distance* whose square is given by

$$d_b^2(\rho_0, \rho_1) := 2(1 - \text{Tr}(\sqrt{\sqrt{\rho_0}\rho_1\sqrt{\rho_0}}))$$

and is a quantum extension of the Hellinger distance. In the case of pure states, the Bures distance becomes

$$(11) \quad d_b^2(|\psi_0\rangle\langle\psi_0|, |\psi_1\rangle\langle\psi_1|) = 2(1 - |\langle\psi_0|\psi_1\rangle|)$$

so for coherent states it is given by

$$d_b^2(|G(\psi_0)\rangle\langle G(\psi_0)|, |G(\psi_1)\rangle\langle G(\psi_1)|) := 2\left(1 - \exp\left(-\frac{1}{2}\|\psi_0 - \psi_1\|^2\right)\right).$$

Similar to the classical case, the following inequality holds for arbitrary states [20]:

$$(12) \quad d_b^2(\rho_0, \rho_1) \leq \|\rho_0 - \rho_1\|_1 \leq 2d_b(\rho_0, \rho_1).$$

Moreover, since $|\langle\psi_0|\psi_1\rangle|^2 \leq |\langle\psi_0|\psi_1\rangle|$, the additional inequality holds for *pure* states

$$(13) \quad \|\rho_0 - \rho_1\|_1 \geq \sqrt{2}d_b(\rho_0, \rho_1).$$

This means that for pure states, the trace and Bures distances are equivalent (up to constants).

Finally, we will be using the fact that both the norm-one and the Bures distance are contractive under quantum channels. $T : \mathcal{T}_1(\mathcal{H}) \rightarrow \mathcal{T}_1(\mathcal{H}')$, that is,

$$\|T(\rho_0) - T(\rho_1)\|_1 \leq \|\rho_0 - \rho_1\|_1, \quad d_b^2(T(\rho_0), T(\rho_1)) \leq d_b^2(\rho_0, \rho_1).$$

3. Quantum statistical models. In this section, we review key elements of quantum statistics, and introduce the quantum statistical models which will be analysed later on. For comparison, we briefly review certain asymptotic equivalence results for related classical statistical models.

The classical density model consists of n observations X_1, \dots, X_n which are independent, identically distributed (i.i.d.) with common probability density f . In the Gaussian white noise model, a function $g \in \mathbb{L}_2[0, 1]$ is observed with Gaussian white noise of variance n^{-1} , that is,

$$(14) \quad dY_t = g(t) dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1].$$

This model is equivalent to the Gaussian sequence model, where we observe a sequence of Gaussian random variables with means equal to the coefficients θ_j of g in some orthonormal basis of $\mathbb{L}_2[0, 1]$

$$(15) \quad y_j = \theta_j + \frac{1}{\sqrt{n}} \xi_j, \quad i = 1, 2, \dots,$$

where $\{\xi_i\}_{i \geq 1}$ are Gaussian i.i.d. random variables.

In [48], it was shown that for densities f on $[0, 1]$, the i.i.d. model is asymptotically equivalent to the white noise model (14) for $g = f^{1/2}$, in the sense that the Le Cam distance of the models converges to zero as $n \rightarrow \infty$ when f varies in a certain smoothness class of functions. For recent related results and extensions (cf. [53]; in Appendix A.1 [12]), we present a more detailed review of asymptotic equivalence results for classical statistical models.

3.1. Quantum models, randomisations and convergence. In this subsection, we introduce the basic notions of a theory of quantum statistical models which is currently still in its early stages; cf. [21, 26] for more details. We will focus on the notions of quantum-to-classical randomisation carried out through measurements, and quantum-to-quantum randomisations implemented by quantum channels, which allow us to define the equivalence and the Le Cam distance between models.

In analogy to the classical case, we make the following definition.

DEFINITION 9. A quantum statistical model over a parameter space Θ consists of a family of quantum states $\mathcal{Q} = \{\rho_\theta : \theta \in \Theta\}$ on a Hilbert space \mathcal{H} , indexed by an unknown parameter $\theta \in \Theta$.

A simple example is a family of pure states $\{\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta| : \theta \in \mathbb{R}\}$ with $|\psi_\theta\rangle := \exp(i\theta H)|\psi\rangle$, where H is a self-adjoint operator generating the one-dimensional family of unitaries $\exp(i\theta H)$, and $|\psi\rangle \in \mathcal{H}$ is a fixed vector. Physically, the parameter θ could be for instance time, a phase, or an external magnetic field. Another example is that of a completely unknown state of a finite dimensional system, which can be parametrised in terms of its density matrix elements, or the eigenvalues and eigenvectors. In order to increase the estimation precision, one typically prepares a number n of identical and independent copies of the state ρ_θ , in which case the corresponding model is $\mathcal{Q}_n := \{\rho_\theta^{\otimes n} : \theta \in \Theta\}$. Our work deals with *nonparametric* quantum statistical models for which the underlying Hilbert space is infinite dimensional, as we will detail below.

In order to obtain information about the parameter θ , we need to perform measurements on the system prepared in ρ_θ . Using the random measurement data, we then employ statistical methods to solve specific decision problems. For instance, the task of estimating an unknown quantum state (also known as quantum tomography) is a key component of quantum engineering experiments [29]. In

particular, the estimation of large dimensional states has received significant attention in the context of compressed sensing [19, 24], and estimation of low rank states [11]. Suppose that we perform a measurement M on the system in state ρ_θ , and obtain a random outcome $O \in \Omega$ with distribution $\mathbb{P}_\theta^M(E) := \text{Tr}(\rho_\theta M(E))$; cf. Section 2. The measurement data is therefore described by the classical model $\mathcal{P}^M := \{\mathbb{P}_\theta^M : \theta \in \Theta\}$, and the estimation problem can be treated using “classical” statistical methods. The measurement map

$$\begin{aligned} \mathcal{M} : \mathcal{T}_1 &\rightarrow L_1(\Omega, \Sigma, \mathbb{P}), \\ \mathcal{M} : \rho_\theta &\mapsto p_\theta := \frac{d\mathbb{P}_\theta}{d\mathbb{P}} \end{aligned}$$

can be seen as a randomisation from a quantum to a classical model, which intuitively means that \mathcal{Q} is more informative than \mathcal{P}^M for any measurement M . Here, \mathbb{P} can be chosen to be the distribution corresponding to an arbitrary full rank (strictly positive) state ρ which insures the existence of all probability densities p_θ . One of the distinguishing features of quantum statistics is the possibility to choose appropriate measurements for specific statistical problems (e.g., estimation, testing) and the fact that optimal measurements for different problems may be incompatible with each other. In the applications section, we will discuss specific instances of this phenomenon.

Beside measurements, the quantum model \mathcal{Q} can be transformed into another quantum model $\mathcal{Q}' := \{\rho'_\theta : \theta \in \Theta\}$ on a Hilbert space \mathcal{H}' by means of a *quantum randomisation*, that is, by applying a quantum channel

$$\begin{aligned} T : \mathcal{T}_1(\mathcal{H}) &\rightarrow \mathcal{T}_1(\mathcal{H}'), \\ T : \rho_\theta &\mapsto \rho'_\theta. \end{aligned}$$

The model \mathcal{Q}' is less informative than \mathcal{Q} in the sense that for any measurement M' on \mathcal{H}' one can construct the measurement $M := M' \circ T$ on \mathcal{H} such that $\mathbb{P}_\theta^{M'} = \mathbb{P}_\theta^M$ for all θ . If there exists another channel S such that $S(\rho'_\theta) = \rho_\theta$ for all θ , we say (in analogy to the classical case) that the models \mathcal{Q} and \mathcal{Q}' are *equivalent*; in particular, for any statistical decision problem, one can match a procedure for one model with a procedure with the same risk, for the other model. A closely related concept is that of *quantum sufficiency* whose theory was developed in [37]. More generally, we define the Le Cam distance in analogy to the classical case [43].

DEFINITION 10. Let \mathcal{Q} and \mathcal{Q}' be two quantum models over Θ . The deficiency between \mathcal{Q} and \mathcal{Q}' is defined by

$$\delta(\mathcal{Q}, \mathcal{Q}') := \inf_T \sup_{\theta \in \Theta} \|T(\rho_\theta) - \rho'_\theta\|_1,$$

where the infimum is taken over all channels T . The *Le Cam* distance between \mathcal{Q} and \mathcal{Q}' is defined as

$$(16) \quad \Delta(\mathcal{Q}, \mathcal{Q}') := \max(\delta(\mathcal{Q}, \mathcal{Q}'), \delta(\mathcal{Q}', \mathcal{Q})).$$

Its interpretation is that models which are “close” in the Le Cam distance have similar statistical properties. In practice, this metric is often used to approximate a sequence of models by another sequence of simpler models, providing a method to establish asymptotic minimax risks. In particular, the approximation of i.i.d. quantum statistical models by quantum Gaussian ones has been investigated in [26, 27, 38], in the case of finite dimensional systems with arbitrary mixed states. Our goal is to extend these results to nonparametric models consisting of pure states on infinite dimensional Hilbert spaces. The following lemma will be used later on.

LEMMA 3.1. *Let $\mathcal{Q}, \mathcal{Q}'$ be two quantum models as defined above. Let $\rho_i = \sum_j \mu_{i,j} \rho_{\theta_{i,j}}$ be two arbitrary mixtures ($i = 1, 2$) of states in \mathcal{Q} and let $\rho'_i = \sum_j \mu_{i,j} \rho'_{\theta_{i,j}}$ be their counterparts in \mathcal{Q}' . Then*

$$\|\rho'_1 - \rho'_2\|_1 - 2\Delta(\mathcal{Q}, \mathcal{Q}') \leq \|\rho_1 - \rho_2\|_1 \leq \|\rho'_1 - \rho'_2\|_1 + 2\Delta(\mathcal{Q}, \mathcal{Q}').$$

PROOF. Since quantum channels are contractive with respect to the norm-one

$$\|S(\rho'_1) - S(\rho'_2)\|_1 \leq \|\rho'_1 - \rho'_2\|_1$$

and by the triangle inequality we get

$$\begin{aligned} \|\rho_1 - \rho_2\|_1 &\leq \|\rho_1 - S(\rho'_1)\|_1 + \|S(\rho'_1) - S(\rho'_2)\|_1 + \|S(\rho'_2) - \rho_2\|_1 \\ &\leq 2\Delta(\mathcal{Q}, \mathcal{Q}') + \|\rho'_1 - \rho'_2\|_1 \end{aligned}$$

The second inequality can be shown in a similar way. \square

3.2. *The i.i.d. and the quantum white noise models.* We now introduce the nonparametric quantum models investigated in the paper. Let \mathcal{H} be an infinite dimensional Hilbert space and let $B := \{|e_0\rangle, |e_1\rangle, \dots\}$ be a fixed orthonormal basis in \mathcal{H} . The Fourier decomposition of an arbitrary vector is written as $|\psi\rangle = \sum_{j=0}^\infty \psi_j |e_j\rangle$. Since most of the models will consist of pure states, we will sometimes define them in terms of the Hilbert space vectors rather than the density matrices, but keep in mind that the vectors are uniquely defined only up to a complex phase.

Let us consider the general problem of estimating an unknown pure quantum state in \mathcal{H} . For *finite* dimensional systems, the risk with respect to typical rotation invariant loss functions scales linearly with the number of parameters [22], hence with the dimension of the space. Therefore, since \mathcal{H} is infinite dimensional, it is not possible to develop a meaningful estimation theory without any prior information about the state. Motivated by physical principles and statistical methodology, we introduce the following *Hermite–Sobolev classes* [8] and [7] of pure states characterised by an appropriate decay of the coefficients with respect to the basis B :

$$(17) \quad S^\alpha(L) := \left\{ |\psi\rangle\langle\psi| : \sum_{j=0}^\infty |\psi_j|^2 j^{2\alpha} \leq L, \text{ and } \|\psi\| = 1 \right\}, \quad \alpha > 0, L > 0.$$

To gain some intuition about the meaning of this class, let us assume that B is the Fock basis of a one-mode cv system. Then the constraint translates into the moment condition for the number operator $\langle \psi | N^{2\alpha} | \psi \rangle \leq L$; this is a mild assumption considering that all experimentally feasible states have finite moments to all orders. Even more, the coefficients of typical states such as coherent, squeezed and Fock states decay exponentially with the photon number.

Our first model describes n identical copies of a pure state belonging to the Sobolev class

$$(18) \quad \mathcal{Q}_n := \{ |\psi\rangle\langle\psi|^{\otimes n} : |\psi\rangle\langle\psi| \in S^\alpha(L) \}.$$

In Section 5.1, we show that the minimax rate of \mathcal{Q}_n for the norm-one and Bures distance loss functions is $n^{-\alpha/(2\alpha+1)}$. This is identical to the minimax rate of the classical i.i.d. model described in Appendix A.1 [12].

We now introduce the corresponding quantum Gaussian model. Let $\mathcal{F} := \mathcal{F}(\mathcal{H})$ be the Fock space over \mathcal{H} , and let $|G(\sqrt{n}\psi)\rangle \in \mathcal{F}$ be the coherent state with “displacement” vector $\sqrt{n}\psi$. As discussed in Section 2.2.2, the vector $\sqrt{n}\psi$ should be seen now as the *expectation* of the infinite dimensional Gaussian state rather than a quantum state in itself, for which reason we have omitted the ket notation. We define the coherent states model:

$$(19) \quad \mathcal{G}_n = \{ |G(\sqrt{n}\psi)\rangle\langle G(\sqrt{n}\psi)| : |\psi\rangle \in \mathcal{H}, \text{ such that } |\psi\rangle\langle\psi| \in S^\alpha(L) \}.$$

Using the factorisation property (8) with respect to the orthonormal basis B , we see that the model is equivalent to the product of independent one-mode coherent Gaussian states of mean $\sqrt{n}\psi_i$

$$|G(\sqrt{n}\psi)\rangle \cong \bigotimes_{i=1}^{\infty} |G(\sqrt{n}\psi_i)\rangle$$

which is analogous to the classical Gaussian sequence model \mathcal{N}_n defined in equation (15).

Similarly, we can draw an analogy with the white noise model \mathcal{F}_n by realising \mathcal{H} as $L^2([0, 1])$. Let us define the *quantum stochastic process* [52] on $\mathcal{F}(L^2([0, 1]))$

$$B(t) := a(\chi_{[0,t]}) + a^*(\chi_{[0,t]})$$

and note that $[B(t), B(s)] = 0$ for all $t, s \in [0, 1]$ so that $\{B(t) : t \in [0, 1]\}$ is a commutative family of operators. This implies that $\{B(t) : t \in [0, 1]\}$ have a joint probability distribution which is uniquely determined by the quantum state, and can be regarded as a classical stochastic process. If the state is the vacuum $|0\rangle$, the process is Gaussian and has the same distribution as the Brownian motion. Consider now the process $X(t) := W(\sqrt{n}\psi)^* B(t) W(\sqrt{n}\psi)$, which is obtained by applying a unitary Weyl transformation to $B(t)$. In physics terms, we work here in the “Heisenberg picture” where the transformation acts on operators while the

state is fixed. Using quantum stochastic calculus, one can derive the following differential equation for $X(t)/\sqrt{n}$:

$$\frac{1}{\sqrt{n}} dX(t) = \psi(t) dt + \frac{1}{\sqrt{n}} dB(t).$$

Therefore, $X(t)/\sqrt{n}$ is similar to the process (14) with the exception that it has a complex rather than real valued drift function. Note that in this correspondence $\psi(t)$ plays the role of $f^{1/2}$, which agrees with the intuitive interpretation of the wave function as square root of the state $|\psi\rangle\langle\psi|$. Alternatively, one can use the Schrödinger picture, where the state is $|\sqrt{n}\psi\rangle = W(\sqrt{n}\psi)|0\rangle$, such that the process $B(t)$ has the same law as $X(t)$ under the vacuum state.

In Section 5.1, we show that the minimax rate of \mathcal{G}_n for loss functions based on the norm-one and the Bures distance is $n^{-\alpha/(2\alpha+1)}$. Although the rate is identical to that of the corresponding classical model, the result does not follow from the classical case but relies on an explicit measurement strategy for the upper bounds, and on the quantum local asymptotic equivalence Theorem 4.1 for the lower bound. Furthermore, the minimax rate for the estimation of certain quadratic functionals are established in Section 5.2, and the minimax testing rates are derived in Section 5.3. While the former are similar to the classical ones, the quantum testing rates are parametric as opposed to nonparametric in the classical case. This reflects the fact that in the quantum case, the optimal measurements for different statistical problems are in general incompatible with each other and in some cases they differ significantly from what is expected on classical basis.

4. Local asymptotic equivalence for quantum models. In this section, we prove that the sequence (18) of nonparametric pure states models is locally asymptotically equivalent (LAE) with the sequence (19) of quantum Gaussian models, in the sense of the Le Cam distance. This is one of the main results of the paper and will be subsequently used in the applications. Throughout the section, $|\psi_0\rangle$ is a fixed but arbitrary state in an infinite dimensional Hilbert space \mathcal{H} . We let $\mathcal{H}_0 := \{|\psi\rangle \in \mathcal{H} : \langle\psi_0|\psi\rangle = 0\}$ denote the orthogonal complement of $\mathbb{C}|\psi_0\rangle$. Any vector state $|\psi\rangle \in \mathcal{H}$ decomposes uniquely as

$$(20) \quad |\psi\rangle = |\psi_u\rangle := \sqrt{1 - \|u\|^2}|\psi_0\rangle + |u\rangle, \quad |u\rangle \in \mathcal{H}_0,$$

where the phase has been chosen such that the overlap $\langle\psi|\psi_0\rangle$ is real and positive. Therefore, the pure states are uniquely parametrised by vectors $|u\rangle \in \mathcal{H}_0$.

Further to the i.i.d. and Gaussian models \mathcal{Q}_n and \mathcal{G}_n defined in (18), and respectively (19), we now introduce their local counterparts which are parametrised by the local parameter $|u\rangle$ rather than by $|\psi\rangle$. Let γ_n be a sequence such that $\gamma_n = o(1)$, and define the pure state models:

$$(21) \quad \mathcal{Q}_n(\psi_0, \gamma_n) := \{|\psi_u^{\otimes n}\rangle \in \mathcal{H}^{\otimes n} : |u\rangle \in \mathcal{H}_0, \|u\| \leq \gamma_n\},$$

$$(22) \quad \mathcal{G}_n(\psi_0, \gamma_n) := \{|G(\sqrt{n}u)\rangle \in \mathcal{F}(\mathcal{H}_0) : |u\rangle \in \mathcal{H}_0, \|u\| \leq \gamma_n\}.$$

The LAE theorem below shows that these local models are asymptotically equivalent. An interesting fact is that LAE holds without imposing global restrictions such as defined by the Sobolev classes, rather it suffices that the local balls shrink at an arbitrary slow rate $\gamma_n = o(1)$. This contrasts with the classical case where both types of conditions are needed, as explained in Appendix A.1 [12]. However, since the state cannot be “localised” without any prior knowledge, in applications we need to make additional assumptions which allow us to work in a small neighbourhood and make use of local asymptotic equivalence. In particular, the convergence holds for the restricted models where the Sobolev condition is imposed on top of the local one. This will be used in establishing the estimation, testing and functional estimation results.

THEOREM 4.1. *Let $\mathcal{Q}_n(\psi_0, \gamma_n)$ and $\mathcal{G}_n(\psi_0, \gamma_n)$ be the models defined in (21), and respectively (22) where $\gamma_n = o(1)$. Then the following convergence holds uniformly over states $|\psi_0\rangle$:*

$$(23) \quad \limsup_{n \rightarrow \infty} \sup_{|\psi_0\rangle \in \mathcal{H}} \Delta(\mathcal{Q}_n(\psi_0, \gamma_n), \mathcal{G}_n(\psi_0, \gamma_n)) = 0,$$

where $\Delta(\cdot, \cdot)$ is the quantum Le Cam distance defined in equation (16).

The proof is given in [12].

5. Applications. In this section, we discuss three major applications of the local asymptotic equivalence result in Theorem 4.1, namely to the estimation of pure states, estimation of a physically meaningful quadratic functional and finally to testing between pure states. We stress that local asymptotic equivalence allows us to translate these problems into similar but easier ones involving Gaussian states. This strategy has already been successfully employed [27] in finding asymptotically optimal estimation procedures for *finite dimensional mixed states*, which otherwise appeared to be a difficult problem due to the complexity of the set of possible measurements.

As discussed in Section 3.2, we will assume that we are given n independent systems, each prepared in a state $|\psi\rangle \in \mathcal{H}$ belonging to the Sobolev ellipsoid $S^\alpha(L)$ defined in equation (17). The corresponding quantum statistical model \mathcal{Q}_n was defined in equation (18), and the Gaussian counterpart model \mathcal{G}_n was defined in equation (19).

Here is a summary of the results. In Theorem 5.2, we show that the estimation rates over such ellipsoids are $n^{-\alpha/(2\alpha+1)}$; this is similar to the well-known rates, for example, for density estimation, in nonparametric statistics (see [57]). The estimation of the quadratic functional

$$F(\psi) = \sum_{j \geq 0} |\psi_j|^2 j^{2\beta} \quad \text{for some fixed } \beta > 0$$

of the unknown pure state presents two regimes: a parametric rate n^{-1} for the MSE is attained when the unknown state has enough “smoothness” (i.e., $\alpha \geq 2\beta$), whereas a nonparametric rate $n^{-2(1-\beta/\alpha)}$ is obtained when $\beta < \alpha < 2\beta$. This double regime is known in nonparametric estimation for the density model, with different values for both the rates and the values of the parameters where the phase-transition occurs; cf. [14, 40] and references therein.

Parametric rates and sharp asymptotic constants are obtained for the testing problem of a pure state against an alternative described by the Sobolev-type ellipsoid with an L_2 -ball removed. In the classical density model, only nonparametric rates for testing of order $n^{-2\alpha/(4\alpha+1)}$ can be obtained for the L_2 norm. In our quantum i.i.d. model, the parametric rate $n^{-1/2}$ is shown to be minimax for testing $H_0 : \psi = \psi_0$, for some ψ_0 in $S^\alpha(L)$ over the nonparametric set of alternatives:

$$H_1 : \psi \in S^\alpha(L) \quad \text{is such that } \|\lvert\psi\rangle\langle\psi\rvert - \lvert\psi_0\rangle\langle\psi_0\rvert\|_1 \geq cn^{-1/2}.$$

The sharp asymptotic constant we obtain for testing is specific for ensembles of pure states. As we discuss in the sequel, quantum testing of states allows us to optimize over the measurements, and thus to obtain the most distinguishable likelihoods for the underlying unknown quantum state.

5.1. Estimation. We consider the problem of estimating an unknown pure state belonging to the Hermite–Sobolev class $S^\alpha(L)$ given an ensemble of n independent, identically prepared systems. The corresponding sequence of statistical models \mathcal{Q}_n was defined in equation (18). We first describe a specific measurement procedure which provides an estimator whose risk attains the nonparametric rate $n^{-2\alpha/(2\alpha+1)}$. We prove the lower bounds for estimating a Gaussian state in the model \mathcal{G}_n defined in (19). Subsequently, we use LAE to establish a lower bound showing that the rate is optimal in the i.i.d. model as well.

Before deriving the bounds, we briefly review the definitions of the loss functions used here and the relations between them; cf. Section 2.3. Recall that the trace norm distance between states ρ and ρ' is given by $\|\rho - \rho'\|_1 := \text{Tr}(|\rho - \rho'|)$, and is the quantum analogue of the norm-one distance between probability densities. The square of the Bures distance is given by $d_b^2 := 2(1 - \text{Tr}(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}))$, and is a quantum extension of the Hellinger distance. These distances satisfy the inequalities (12).

In the case of pure states (i.e., $\rho = \lvert\psi\rangle\langle\psi\rvert$ and $\rho' = \lvert\psi'\rangle\langle\psi'\rvert$), these metrics become [cf. (10) and (11)]

$$\|\rho - \rho'\|_1 = 2\sqrt{1 - |\langle\psi|\psi'\rangle|^2}, \quad d_b^2(\rho, \rho') = 2(1 - |\langle\psi|\psi'\rangle|).$$

Since vectors are not uniquely defined by the states, the distances cannot be expressed directly in terms of the length $\|\psi - \psi'\|$. However, if we consider a reference vector $\lvert\psi_0\rangle$ and define the representative vector $\lvert\psi\rangle$ such that $\langle\psi_0|\psi\rangle \geq 0$,

then we can write (as in Section 4)

$$\begin{aligned}
 |\psi_u\rangle &= \sqrt{1 - \|u\|^2}|\psi_0\rangle + |u\rangle, \\
 |\psi_{u'}\rangle &= \sqrt{1 - \|u'\|^2}|\psi_0\rangle + |u'\rangle, \quad |u\rangle, |u'\rangle \perp |\psi_0\rangle
 \end{aligned}$$

and the distances have the same (up to a constant) quadratic approximation

$$\begin{aligned}
 \|\rho_u - \rho_{u'}\|_1^2 &= 4\|u - u'\|^2 + O(\max(\|u\|, \|u'\|\|^4), \\
 d_b^2(\rho_u, \rho_{u'}) &= \|u - u'\|^2 + O(\max(\|u\|, \|u'\|\|^4),
 \end{aligned}
 \tag{24}$$

where the correction terms are of order 4 as $\|u\|$ and $\|u'\|$ tend to 0. Below we show that asymptotically with n the estimation risk for norm-one square and Bures distance square will have the same rate as that of estimating the local parameter u with respect to the Hilbert space distance.

5.1.1. *Upper bounds.* We first describe a two steps measurement procedure, which provides an estimator whose risk has rate $n^{-2\alpha/(2\alpha+1)}$.

THEOREM 5.1. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18). There exists an estimator $\hat{\rho}_n := |\hat{\psi}_n\rangle\langle\hat{\psi}_n|$ such that*

$$\limsup_{n \rightarrow \infty} \sup_{|\psi\rangle \in S^\alpha(L)} n^{2\alpha/(2\alpha+1)} \mathbb{E}_\rho [d^2(\hat{\rho}_n, \rho)] \leq C,$$

where $\rho := |\psi\rangle\langle\psi|$, $d(\hat{\rho}_n, \rho)$ denotes either the trace-norm distance, or the Bures distance, and $C > 0$ is a constant depending only on $\alpha > 0$ and $L > 0$.

The proof is given in [12].

5.1.2. *Lower bounds—unimprovable rates.* We will first consider the Gaussian model \mathcal{G}_n given by equation (19) which is indexed by Hilbert space vectors $\psi \in \mathcal{H}$ in the Sobolev class $S^\alpha(L)$, playing the role of means of quantum Gaussian states $|G(\sqrt{n}\psi)\rangle$. In Theorem 5.2, we find a lower bound for the mean square error of any estimator $\hat{\psi}$. This is then used in conjunction with the local asymptotic equivalence Theorem 4.1 to obtain a lower bound for the risk of the i.i.d. model \mathcal{Q}_n , with respect to the norm-one and Bures distances.

THEOREM 5.2. *Consider the quantum Gaussian model \mathcal{G}_n given by equation (19). There exists some constant $c > 0$ depending only on α and L such that*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\psi}_n} \sup_{\psi \in S^\alpha(L)} n^{2\alpha/(2\alpha+1)} \mathbb{E}_\psi [\|\hat{\psi}_n - \psi\|_2^2] \geq c,$$

where the infimum is taken over all estimators $\hat{\psi}_n$, understood as combination of measurements and classical estimators.

The proof is given in [12].

We now proceed to consider the i.i.d. model \mathcal{Q}_n defined in (18). We are given n copies of an unknown pure state $|\psi\rangle\langle\psi|$, with ψ in the Sobolev class $S^\alpha(L)$. The goal is to find an asymptotic lower bound for the estimation risk (with respect to the Bures or norm-one loss functions) which matches the upper bound derived in Section 5.1.1. Since both loss functions satisfy the triangle inequality, it can be shown that by choosing estimators which are mixed states, rather than pure states, one can improve the risk by at most a constant factor 2. Therefore, we consider estimators which are pure states. In order to fix the phase of the vector representing the true and the estimated state, we will assume that $\langle\psi|e_0\rangle \geq 0$ and $\langle\hat{\psi}|e_0\rangle \geq 0$.

THEOREM 5.3. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18). There exists some constant $c > 0$ depending only on $\alpha > 0$ and $L > 0$ such that*

$$\liminf_{n \rightarrow \infty} \inf_{|\hat{\psi}_n\rangle | \psi\rangle \in S^\alpha(L)} \sup n^{2\alpha/(2\alpha+1)} \mathbb{E}_\rho [d^2(\hat{\rho}_n, \rho)] \geq c,$$

where $\rho := |\psi\rangle\langle\psi|$, the infimum is taken over all estimators $\hat{\rho}_n := |\hat{\psi}_n\rangle\langle\hat{\psi}_n|$ (defined by a combination of measurement and a classical estimator), and the loss function $d(\hat{\rho}, \rho)$ is either the norm-one or the Bures distance.

The proof is given in [12].

5.2. Quadratic functionals. This section deals with the estimation of the quadratic functional

$$F(\psi) = \sum_{j \geq 0} |\psi_j|^2 \cdot j^{2\beta} \quad \text{for some fixed } 0 < \beta < \alpha,$$

which is well defined for all pure states $|\psi\rangle$ in the ellipsoid $S^\alpha(L)$. If the Hilbert space \mathcal{H} is represented as $L_2(\mathbb{R})$ and $\{|j\rangle : j \geq 0\}$ is the Fock basis (cf. Section 2.2.1), then $F(\psi)$ is the moment of order 2β of the number operator N :

$$F(\psi) = \text{Tr}(|\psi\rangle\langle\psi| \cdot N^{2\beta}).$$

Below we derive upper and lower bounds for the rate of the quadratic risk for estimating $F(\psi)$, which is of order n^{-1} if $\alpha \geq 2\beta$, and $n^{-2(1-\beta/\alpha)}$ if $\beta < \alpha < 2\beta$.

5.2.1. Upper bounds. Let us describe an estimator \hat{F}_n of $F(\psi)$ in the quantum i.i.d. model. We consider the measurement of the number operator with projections $\{|j\rangle\langle j|\}_{j \geq 0}$. For a pure state $|\psi\rangle = \sum_{j \geq 0} \psi_j |j\rangle$, we obtain an outcome X taking values $j \in \mathbb{N}$ with probabilities $p_j := \mathbb{P}_\psi(X = j) = |\psi_j|^2$, for $j \geq 0$. By measuring each quantum sample $|\psi\rangle$ separately, we obtain i.i.d. copies X_1, \dots, X_n of X , allowing us to estimate each p_j empirically, by

$$\hat{p}_j = \frac{1}{n} \sum_{k=1}^n I(X_k = j), \quad j \geq 0$$

which is an unbiased estimator of p_j with variance $p_j(1 - p_j)/n$. The estimator of the quadratic functional is defined as

$$(25) \quad \hat{F}_n = \sum_{j=1}^N \hat{p}_j \cdot j^{2\beta}$$

for an appropriately chosen truncation parameter N defined below. The next theorem shows that a parametric rate can be attained for estimating the quadratic functional $F(\psi)$ if $\alpha \geq 2\beta$, whereas a nonparametric rate is attained if $\beta < \alpha < 2\beta$.

THEOREM 5.4. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18). Let \hat{F}_n be the estimator (25) of $F(\psi)$ with $N \asymp n^{1/4(\alpha-\beta)}$, for $\alpha \geq 2\beta$, respectively, $N \asymp n^{1/2\alpha}$, for $\beta < \alpha < 2\beta$. Then*

$$(26) \quad \sup_{\psi \in S^\alpha(L)} \mathbb{E}_\psi (\hat{F}_n - F(\psi))^2 = O(\eta_n^2),$$

$$\text{where } \eta_n^2 = \begin{cases} n^{-1} & \text{if } \alpha \geq 2\beta, \\ n^{-2(1-\beta/\alpha)} & \text{if } \beta < \alpha < 2\beta. \end{cases}$$

The proof is given in [12].

5.2.2. Lower bounds. The next theorem proves the optimality of the previously attained rate for the estimation of quadratic functionals.

THEOREM 5.5. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18). Then there exists some constant $c > 0$ depending only on α, β (with $\alpha > \beta > 0$), and $L > 0$ such that*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{F}_n} \sup_{\psi \in S^\alpha(L)} \eta_n^{-2} \cdot \mathbb{E}_\psi (\hat{F}_n - F(\psi))^2 \geq c,$$

where the infimum is taken over all measurements and resulting estimators \hat{F}_n of $F(\psi)$.

Further discussion on quadratic functionals can be found in Appendix A.2 [12]; proofs are presented in Appendix B.

5.3. Testing. In the problem of testing for signal in classical Gaussian white noise, over a smoothness class with an L_2 -ball removed, minimax rates of convergences (separation rates) are well known [36]; they are expressed in the rate of the ball radius tending to zero along with noise intensity, such that a nontrivial asymptotic power is possible. We will consider an analogous testing problem here for pure states. Accordingly, let $\rho = |\psi\rangle\langle\psi|$ denote pure states, let $\rho_0 = |\psi_0\rangle\langle\psi_0|$ be a fixed pure state to serve as the null hypothesis and let

$$(27) \quad B(\varphi) = \{\|\rho - \rho_0\|_1 \geq \varphi\}$$

be the complement of a trace norm ball around ρ_0 . We want to test in the i.i.d. quantum model \mathcal{Q}_n given by equation (18) the following hypotheses about a pure state ρ :

$$(28) \quad \begin{aligned} H_0 &: \rho = \rho_0, \\ H_1(\varphi_n) &: \rho \in S^\alpha(L) \cap B(\varphi_n) \end{aligned}$$

for $\{\varphi_n\}_{n \geq 1}$ a decreasing sequence of positive real numbers. Consider a binary POVM $M = (M_0, M_1)$, acting on the product states $\rho^{\otimes n}$; cf. Definition 2. We denote the testing risk between two fixed hypotheses by the sum of the two error probabilities

$$R_n^T(M) = R_n^T(\rho_0^{\otimes n}, \rho^{\otimes n}, M) = \text{Tr}(\rho_0^{\otimes n} \cdot M_1) + \text{Tr}(\rho^{\otimes n} \cdot M_0).$$

In the minimax α -testing approach which dominates the literature on the classical Gaussian white noise case, one would require $\text{Tr}(\rho_0^{\otimes n} \cdot M_1) \leq \alpha$ while trying to minimize the worst case type 2 error $\sup_{\rho \in S^\alpha(L) \cap B(\varphi_n)} \text{Tr}(\rho^{\otimes n} \cdot M_0)$. However, we will consider here the so-called detection problem [35] where the target is the worst case total error probability

$$\begin{aligned} \mathbb{P}_e^M(\varphi_n) &= \sup_{\rho \in S^\alpha(L) \cap B(\varphi_n)} R_n^T(\rho_0^{\otimes n}, \rho^{\otimes n}, M) \\ &= \text{Tr}(\rho_0^{\otimes n} \cdot M_1) + \sup_{\rho \in S^\alpha(L) \cap B(\varphi_n)} \text{Tr}(\rho^{\otimes n} \cdot M_0). \end{aligned}$$

The minimax total error probability is then obtained by optimizing over T :

$$\mathbb{P}_e^*(\varphi_n) = \inf_{M \text{ binary POVM}} \mathbb{P}_e^M(\varphi_n).$$

5.3.1. *Separation rate.* A sequence $\{\varphi_n^*\}_{n \geq 1}$ is called a *minimax separation rate* if any other sequence $\{\varphi_n\}_{n \geq 1}$ fulfills

$$(29) \quad \mathbb{P}_e^*(\varphi_n) \rightarrow 1 \quad \text{if } \varphi_n/\varphi_n^* \rightarrow 0 \quad \text{and} \quad \mathbb{P}_e^*(\varphi_n) \rightarrow 0 \quad \text{if } \varphi_n/\varphi_n^* \rightarrow \infty.$$

Below we establish that $\varphi_n^* = n^{-1/2}$ is a separation rate in the current problem, even though the alternative $H_1(\cdot)$ in (28) is a nonparametric set of pure states. Recall relations (9), (10) describing the total optimal error for testing between simple hypotheses given by two pure states.

THEOREM 5.6. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18), and the testing problem (28). Assume that ρ_0 is in the interior of $S^\alpha(L)$, that is, $\rho_0 \in S^\alpha(L')$ for some $L' < L$. Then $\varphi_n^* = n^{-1/2}$ is a minimax separation rate.*

The proof is given in [12].

5.3.2. *Sharp asymptotics.* Having identified the optimal rate of convergence in the testing problem, we will go a step further and aim at a sharp asymptotics for the minimax testing error. We will adopt the approach of [17], extended in [36], where testing analogs of the Pinsker-type sharp risk asymptotics in nonparametric estimation were obtained. The result will be framed as follows: if the radius is chosen $\varphi_n \sim cn^{-1/2}$ for a certain $c > 0$, then the minimax testing error behaves as $\mathbb{P}_e^*(\varphi_n) \sim \exp(-c^2/4)$. Thus the sharp asymptotics is expressed as a type of scaling result: a choice of constant c in the radius implies a certain minimax error asymptotics depending on c .

To outline the problem, consider the upper and lower error bounds obtained in the proof of the separation rate, that is, the proof of Theorem 5.6 in [12]. The upper risk bound obtained is

$$(30) \quad \mathbb{P}_e^{M_n}(\varphi_n) \leq \exp(-c_n^2/4)$$

if $\varphi_n = c_n n^{-1/2}$, where M_n is the sequence of projection tests $M_n = (\rho_0^{\otimes n}, I - \rho_0^{\otimes n})$. The corresponding lower risk bound is

$$\inf_{M \text{ binary POVM}} \mathbb{P}_e^M(\varphi_n) \geq 1 - \sqrt{1 - (1 - c_n^2 n^{-1}/4)^n}.$$

If $c_n = c$, we can summarize this as

$$1 - \sqrt{1 - \exp(-c^2/4)} + o(1) \leq \mathbb{P}_e^*(\varphi_n) \leq \exp(-c^2/4).$$

Our result will be that the upper bound is sharp and represents the minimax risk asymptotics.

THEOREM 5.7. *Consider the i.i.d. quantum model \mathcal{Q}_n given by equation (18), and the testing problem (28). Assume that $\rho_0 \in S^\alpha(L')$ for some $L' < L$. At the minimax separation rate for the radius, that is, for $\varphi_n \asymp n^{-1/2}$ we have*

$$\lim_n n^{-1} \varphi_n^{-2} \log \mathbb{P}_e^*(\varphi_n) = -1/4.$$

Further discussion on nonparametric testing can be found in Appendix A.3 [12]; proofs are presented in Appendix B.

5.4. *Discussion: State estimation.*

Tomography and optimal rates. Consider a model where the Sobolev-type assumption $\rho \in S^\alpha(L)$ about the pure state $\rho = |\psi\rangle\langle\psi|$ [cf. (17)] is replaced by a finite dimensionality assumption: $\rho \in \mathcal{H}_d$ where

$$\mathcal{H}_d = \{|\psi\rangle\langle\psi| : \psi_j = 0, j \geq d\}$$

and d is known. One observes n identical copies of the pure state $\rho = |\psi\rangle\langle\psi|$, with possibly $d = d_n \rightarrow \infty$, that is, the model \mathcal{Q}_n of (18) is replaced by

$$\mathcal{Q}_n := \{\rho^{\otimes n} : \rho \in \mathcal{H}_d\}.$$

Since \mathcal{H}_d can be written $\mathcal{H}_d = \mathcal{S}_{1,d}$ where

$$\mathcal{S}_{r,d} := \{ \rho : \langle e_i | \rho | e_j \rangle = 0, i, j \geq d, \text{rank}(\rho) = r \},$$

the model is effectively a special case of the $d \times d$ density matrices of $\text{rank}(\rho) = r$ considered in [42]. In [42], however, it is not known in advance that $r = 1$ but ρ is a density matrix of possibly low rank r , and the aim is estimation of ρ using quantum state tomography performed on n identical copies of ρ . Data are obtained by defining an observable $\otimes_{i=1}^n E_i$ where E_1, \dots, E_n are i.i.d. uniformly selected elements of the Pauli basis of the linear space of $d \times d$ Hermitian matrices, and applying the corresponding measurement to $\rho^{\otimes n}$. Let $\hat{\rho}_n^*$ denote an arbitrary estimator of ρ based on that measurement. A lower asymptotic risk bound for norm-one risk is established; in the special case $d^2 r^2 = o(n)$ it reads as

$$(31) \quad \inf_{\hat{\rho}_n^*} \sup_{\rho \in \mathcal{S}_{r,d}} \mathbb{E}_\rho [\| \hat{\rho}_n^* - \rho \|_1^2] \geq c \frac{r^2 d^2}{n}$$

for some $c > 0$ (Theorem 10 in [42]). It is also shown in [42] that (31) is attained, up to a different constant and logarithmic terms, by an entropy penalized least squares type estimator based on measurement of $\otimes_{i=1}^n E_i$, even when the rank r is unknown. Analogous optimal rates for $d \times d$ mixed states ρ with Pauli measurements, but under sparsity assumptions on the entries of the matrix ρ have been obtained in [13].

Returning to our setting of pure states, where $r = 1$ is known, with an infimum over all measurements of $\rho^{\otimes n}$ and corresponding estimators $\hat{\rho}_n$, according to [30] one has

$$(32) \quad \inf_{\hat{\rho}_n} \sup_{\rho \in \mathcal{S}_{1,d}} \mathbb{E}_\rho [\| \hat{\rho}_n - \rho \|_1^2] = \frac{4(d-1)}{d+n}$$

and the bound is attained by an estimator of the pure state ρ based on the covariant measurement; cf. equation (B.8) [12]. Comparing (31) for $r = 1$ and $d_n \rightarrow \infty$, $d_n = o(n)$ with (32), we find that the latter bound is of order d_n/n whereas the former is of order d_n^2/n . It means that for estimation of finite dimensional pure states, estimators based on the Pauli type measurement $\otimes_{i=1}^n E_i$ do not attain the optimal rate when $d_n \rightarrow \infty$. It may be conjectured that the same holds for the optimal rate over $\rho \in S^\alpha(L)$, that is, our rate of Theorem 5.1. We emphasize again that our results establish lower asymptotic risk bounds over all quantum measurements and estimators, whereas lower risk bounds within one specific measurement scheme [13, 41, 42] are essentially results of classical (nonquantum) statistics.

Separate measurements. A notable fact is also that $\otimes_{i=1}^n E_i$ is a *separate* (or local) measurement, that is, produces independent random variables (or random elements) Y_1, \dots, Y_n each based on a measurement of a copy of ρ , whereas the covariant measurement (cp. equation (B.8) [12]) we used for attainment our risk

bound of Theorem 5.1 is of *collective* (or joint)-type with regard to the product $\rho^{\otimes n}$. Separate measurements are of interest from a practical point of view since collective measurements of large quantum systems may be unfeasible in implementations [46]. In [5], it is shown that for fixed $d = 2$, the bound (32) can be attained asymptotically as $n \rightarrow \infty$ [up to a factor $1 + o(1)$] by a separate measurement of $\rho^{\otimes n}$; it is an open question whether in our infinite dimensional setting, the optimal rate of Theorem 5.1 can be attained by a separate measurement. For mixed qubits ($d = 2$), an asymptotic efficiency gap between separate and collective measurements is known to exist [4].

SUPPLEMENTARY MATERIAL

Supplement to “Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise” (DOI: [10.1214/17-AOS1672SUPP](https://doi.org/10.1214/17-AOS1672SUPP); .pdf). A more detailed overview of asymptotic equivalence for classical models is provided in Appendix A.1. The results on quadratic functionals and nonparametric testing are further discussed in Appendix A.2 and A.3. Proofs of all results are given in Appendix B.

REFERENCES

- [1] ARTILES, L. M., GILL, R. D. and GUȚĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 109–134. [MR2136642](#)
- [2] AUDENAERT, K. M. R., CALSAMIGLIA, J., MUÑOZ-TAPIA, R., BAGAN, E., MASANES, L., ACIN, A. and VERSTRAETE, F. (2007). Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501.
- [3] AUDENAERT, K. M. R., NUSSBAUM, M., SZKOLA, A. and VERSTRAETE, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** 251–283. [MR2377635](#)
- [4] BAGAN, E., BALLESTER, M. A., GILL, R. D., MUÑOZ-TAPIA, R. and ROMERO-ISART, O. (2006). Separable measurement estimation of density matrices and its fidelity gap with collective protocols. *Phys. Rev. Lett.* **97** 130501.
- [5] BAGAN, E., MONRAS, A. and MUÑOZ-TAPIA, R. (2005). Comprehensive analysis of quantum pure-state estimation for two-level systems. *Phys. Rev. A* **71** 062318.
- [6] BARNDORFF-NIELSEN, O. E., GILL, R. and JUPP, P. E. (2003). On quantum statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **65** 775–816. [MR2017871](#)
- [7] BONGIOANNI, B. (2011). Sobolev spaces diversification. *Rev. Un. Mat. Argentina* **52** 23–34. [MR2952949](#)
- [8] BONGIOANNI, B. and TORREA, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.* **116** 337–360. [MR2256010](#)
- [9] BRAUNSTEIN, S. L. and CAVES, C. M. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72** 3439–3443.
- [10] BUTUCEA, C., GUȚĂ, M. and ARTILES, L. (2007). Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.* **35** 465–494. [MR2336856](#)
- [11] BUTUCEA, C., GUȚĂ, M. and KYPRAIOS, T. (2015). Spectral thresholding quantum tomography for low rank states. *New J. Phys.* **17** 113050.

- [12] BUTUCEA, C., GUȚĂ, M. and NUSSBAUM, M. (2018). Supplement to “Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise”. DOI: [10.1214/17-AOS1672SUPP](https://doi.org/10.1214/17-AOS1672SUPP).
- [13] CAI, T., KIM, D., WANG, Y., YUAN, M. and ZHOU, H. H. (2016). Optimal large-scale quantum state tomography with Pauli measurements. *Ann. Statist.* **44** 682–712. [MR3476614](https://doi.org/10.1214/15-AOS1364)
- [14] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. [MR2253108](https://doi.org/10.1214/05-AN117)
- [15] CARPENTIER, A., EISERT, J., GROSS, D. and NICKL, R. Uncertainty quantification for matrix compressed sensing and quantum tomography problems. Available at [arXiv:1504.03234](https://arxiv.org/abs/1504.03234).
- [16] CHRISTANDL, M. and RENNER, R. (2012). Reliable quantum state tomography. *Phys. Rev. Lett.* **109** 120403.
- [17] ERMAKOV, M. S. (1990). Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 704–715. [MR1090496](https://doi.org/10.1070/1990/v35n04<704::1990>2.0.CO;2)
- [18] FAIST, P. and RENNER, R. (2016). Practical and reliable error bars in quantum tomography. *Phys. Rev. Lett.* **117** 010404.
- [19] FLAMMIA, S. T., GROSS, D., LIU, Y.-K. and EISERT, J. (2012). Quantum tomography via compressed sensing: Error bounds, sample complexity, and efficient estimators. *New J. Phys.* **14** 095022.
- [20] FUCHS, C. A. and VAN DE GRAAF, J. (1999). Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inform. Theory* **45** 1216–1227. [MR1686254](https://doi.org/10.1109/18.781763)
- [21] GILL, R. D. and GUȚĂ, M. I. (2013). On asymptotic quantum statistical inference. In *From Probability to Statistics and Back: High-Dimensional Models and Processes*. *Inst. Math. Stat. (IMS) Collect.* **9** 105–127. IMS, Beachwood, OH. [MR3186752](https://doi.org/10.1214/12-IMS-COLLECT-9)
- [22] GILL, R. D. and MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* **61** 042312.
- [23] GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. [MR2815834](https://doi.org/10.1109/18.1000186)
- [24] GROSS, D., LIU, Y.-K., FLAMMIA, S. T., BECKER, S. and EISERT, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401.
- [25] GUȚĂ, M., JANSSENS, B. and KAHN, J. (2008). Optimal estimation of qubit states with continuous time measurements. *Comm. Math. Phys.* **277** 127–160. [MR2357428](https://doi.org/10.1007/s00220-008-0500-0)
- [26] GUȚĂ, M. and JENČOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. [MR2346393](https://doi.org/10.1007/s00220-007-0239-0)
- [27] GUȚĂ, M. and KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* **73** 052108.
- [28] GUTA, M. and KAHN, J. (2008). Local asymptotic normality and optimal estimation for d-dimensional quantum systems. In *Quantum Stochastics and Information: Statistics, Filtering and Control* 300–322. World Scientific, Singapore. [MR2562438](https://doi.org/10.1142/9789812701111_017)
- [29] HÄFFNER, H., HÄNSEL, W., ROOS, C. F., BENHELM, J., CHEK-AL-KAR, D., CHWALLA, M., KÖRBER, T., RAPOL, U. D., RIEBE, M., SCHMIDT, P. O., BECHER, C., GÜHNE, O., DÜR, W. and BLATT, R. (2005). Scalable multiparticle entanglement of trapped ions. *Nature* **438** 643–646.
- [30] HAYASHI, M. (1998). Asymptotic estimation theory for a finite-dimensional pure state model. *J. Phys. A: Math. Gen.* **31** 4633. [MR1631191](https://doi.org/10.1088/0954-3899/31/17/1998)
- [31] HAYASHI, M. (2005). *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. World Scientific, Singapore.
- [32] HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.
- [33] HIAI, F. and PETZ, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.* **143** 99–114. [MR1139426](https://doi.org/10.1007/BF02046983)

- [34] HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland Series in Statistics and Probability **1**. North-Holland, Amsterdam. MR0681693
- [35] INGSTER, Y. and STEPANOVA, N. (2011). Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.* **5** 484–506. MR2813552
- [36] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446
- [37] JENČOVÁ, A. and PETZ, D. (2006). Sufficiency in quantum statistical inference. *Comm. Math. Phys.* **263** 259–276. MR2207329
- [38] KAHN, J. and GUŤÁ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764
- [39] KARGIN, V. (2005). On the Chernoff bound for efficiency of quantum hypothesis testing. *Ann. Statist.* **33** 959–976. MR2163165
- [40] KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. MR2214904
- [41] KOLTCHINSKII, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. MR3012397
- [42] KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798
- [43] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411
- [44] LEONHARDT, U. (1997). *Measuring the Quantum State of Light*. Cambridge Univ. Press, Cambridge.
- [45] LI, K. (2016). Discriminating quantum states: The multiple Chernoff distance. *Ann. Statist.* **44** 1661–1679. MR3519936
- [46] NAIR, R., GUHA, S. and TAN, S.-H. (2014). Realizable receivers for discriminating coherent and multicopy quantum states near the quantum limit. *Phys. Rev. A* **89** 032318.
- [47] NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge. MR1796805
- [48] NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959
- [49] NUSSBAUM, M. and SZKOLA, A. (2006). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. MR2502660
- [50] OGAWA, T. and NAGAOKA, H. (2000). Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory* **46** 2428–2433. MR1806811
- [51] PARIS, M. G. A. and ŘEHÁČEK, J. (2004). Quantum state estimation. In *Quantum State Estimation. Lecture Notes in Physics* **649**. Springer, Berlin.
- [52] PARTHASARATHY, K. R. (1992). *An Introduction to Quantum Stochastic Calculus. Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. MR1164866
- [53] RAY, K. and SCHMIDT-HIEBER, J. (2017). The Le Cam distance between density estimation, Poisson processes and Gaussian white noise. Available at arXiv:1608.01824 [math.ST].
- [54] STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin. MR812467
- [55] TEMME, K. and VERSTRAETE, F. (2015). Quantum chi-squared and goodness of fit testing. *J. Math. Phys.* **56** 012202. MR3390829
- [56] TORGENSEN, E. (1991). *Comparison of Statistical Experiments. Encyclopedia of Mathematics and Its Applications* **36**. Cambridge Univ. Press, Cambridge. MR1104437
- [57] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359
- [58] WALTER, M. and RENES, J. M. (2014). Lower bounds for quantum parameter estimation. *IEEE Trans. Inform. Theory* **60** 8007–8023. MR3285761

C. BUTUCEA
CREST, ENSAE
UNIVERSITÉ PARIS-SACLAY
5, AVENUE HENRY LE CHATELIER
91120 PALAISEAU
FRANCE
E-MAIL: crisrina.butucea@ensae.fr

M. GUȚĂ
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF NOTTINGHAM
UNIVERSITY PARK
NG7 2RD NOTTINGHAM
UNITED KINGDOM
E-MAIL: madalin.guta@nottingham.ac.uk

M. NUSSBAUM
DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
MALOTT HALL
ITHACA, NEW YORK 14853
USA
E-MAIL: nussbaum@math.cornell.edu