The Annals of Statistics
2022, Vol. 50, No. 1, 430–459
https://doi.org/10.1214/21-AOS2115

MINIMAX NONPARAMETRIC ESTIMATION OF PURE QUANTUM STATES

BY SAMRIDDHA LAHIRY[1] AND MICHAEL NUSSBAUM[2]

[1] *Department of Statistics and Data Science, Cornell University, sl2938@cornell.edu*
[2] *Department of Mathematics, Cornell University, nussbaum@math.cornell.edu*

In classical statistics, Pinsker's theorem provides an exact asymptotic minimax bound in nonparametric estimation, improving upon optimal rates of convergence results. We obtain a quantum version of the theorem by establishing asymptotic minimax results for estimation of the displacement vector in a quantum Gaussian white noise model, given by a sequence of shifted vacuum states. Analogous results are then obtained for estimation of a general pure state from an ensemble of identically prepared, independent quantum systems, using the recently established local asymptotic equivalence to the quantum Gaussian white noise model. Optimality holds with respect to the most fundamental distance measure for quantum states, that is, trace norm distance, and in a true quantum sense, allowing for all possible measurements. Adaptive estimators are also obtained for the above cases. As an application, we obtain asymptotic minimax adaptive estimators for Wigner functions of pure states.

1. Introduction. Problems of *quantum statistical inference* arise in connection with detection and processing of quantum signals. The need for a rigorous theory in this regard has been confirmed by recent breakthroughs in quantum technology, such as quantum computing, communication and metrology [41]. Many questions related to quantum measurements can be formulated in the language of mathematical statistics and can be answered using the tools familiar to "classical" statisticians.

It is well known that quantum theory yields many counterintuitive predictions, and that quantum probability modeled after the laws of quantum mechanics violates some of the classical "Kolmogorovian" laws of probability. It is no surprise that inference in quantum statistics dictated by the laws of quantum probability (which in turn is based on quantum mechanical axioms) deviates from inference as seen in classical statistics. The main differences are as follows:

1. The probabilistic nature of outcomes is not due to ignorance of the experimenter or error made during measurement, but because randomness is a fundamental feature of physical systems at the microscopic level (we elaborate on this in the next section).
2. The task of a classical statistician consists of processing the data obtained by an experimenter and then performing inference at the population level. In contrast, in quantum statistics one has to "choose the measurement" (and generate the data) in addition to processing the data generated by it.
3. There are many physical properties that cannot be jointly measured (in fact measurement of one property changes the physical state), and hence in many cases there is no concept of joint distributions.

Received June 2020; revised April 2021.
MSC2020 subject classifications. Primary 62G05; secondary 62G20, 62C20, 81P50.
Key words and phrases. Pinsker's theorem, pure quantum state, Hermite–Sobolev class, trace norm distance, quantum Gaussian sequence model, quantum Bayes estimation, local asymptotic equivalence, adaptive estimation, block Stein method, Wigner function.

One of the fundamental problems in quantum statistics is state estimation, where given a collection of identically prepared states indexed by a parameter, one designs a measurement and then provides an estimate of the parameter based on the outcome of the measurement. In analogy to the classical decision theoretic approach, one can develop a *quantum decision theoretic* framework for parameter estimation where the "best estimate" also involves optimizing over the measurements. In this direction, a *quantum Cramer–Rao bound* has been obtained in [8, 18, 19] for the covariance matrix of unbiased estimators. In the same spirit, one can devise optimal testing procedures for discrimination between multiple quantum states. In the asymptotic setup, a *quantum Chernoff bound* has been established in [4, 5, 29, 36].

In classical statistics, a fundamental paradigm is approximating complicated experiments (families of laws, or models) by simpler ones (see [27] for details). In particular, one establishes asymptotic equivalence between i.i.d. models indexed by a local parameter and a Gaussian shift model (with the shift given by the same local parameter). This approximation is called local asymptotic normality (LAN) and allows one to construct an estimator from a procedure in the Gaussian model with similar risk bounds. Since the approximation is only local in nature, one constructs a preliminary estimate of the parameter first and then uses LAN in the neighborhood of the estimated value. Global equivalence has also been shown between many nonparametric estimation problems like nonparametric regression [9], density estimation [34] and the Gaussian white noise model. In the quantum setup, *quantum LAN theory* for parametric models, established in [15, 16, 23, 42], shows that a model given by a large collection of identically prepared finite dimensional states can be approximated by a quantum Gaussian shift model in a local neighborhood. A recent paper [11] extends quantum LAN theory toward a quantum version of local asymptotic equivalence, in particular, it is shown that an ensemble of pure states in infinite dimensional Hilbert space can be approximated by coherent states, which are a quantum counterpart of the Gaussian white noise model.

However, the risk bounds for estimating pure states, obtained in [11] (see Section 5), establish only optimal rates of convergence and the estimator is not adaptive to the smoothness parameters. We refer to estimators (classical or quantum) as being *sharp minimax* if they attain the optimal rate and the best possible constant in the asymptotic minimax risk. Such "exact" asymptotically minimax and adaptive estimators can be envisaged in the abovementioned quantum models, as analogs of their classical counterparts given by Pinsker's theorem [38, 40]. We use a Bayes estimator, constructed by Holevo in [20], to show that sharp minimax and adaptive estimation can also be established in the quantum Gaussian sequence model, and the same holds for i.i.d. pure state models due to local asymptotic equivalence.

1.1. *Outline of results*. The paper is organized as follows. In Section 2, we review the basic quantum mechanical concepts of states, measurements, observables and quantum channels. Section 3 reviews some classical results in nonparametric estimation due to Pinsker [38] and others and also contains the two main theorems of the paper. Theorem 3.2 states sharp and adaptive minimax estimation in the Gaussian case and Theorem 3.3 states the same for i.i.d. pure states. We use a truncated version of the quadratic loss for both the quantum Gaussian sequence and the pure state model (with the truncation going to infinity) since a bounded loss is needed (or a loss function which grows slower than the asymptotic equivalence rate) for the transfer of risks.

Section 4 describes the Bayes estimator from [20, 21], which will be useful in our construction of the optimal estimator and also in establishing the minimax lower bound. We also observe a shrinkage phenomenon for this Bayes estimator—an observation crucial for establishing the minimax result in the quantum white noise model or the equivalent quantum Gaussian sequence model.

For minimaxity, we consider two classes of sequences of displacements of the vacuum, characterized as a Sobolev ellipsoid and an exponential ellipsoid. In Section 5, we outline the proof of the nonadaptive part of Theorem 3.2, that is, a sharp minimax result in the quantum Gaussian sequence model. We prove the upper bound in Theorem 5.1, as mentioned before, the quantum Bayes estimator described in Section 4 is crucially used here. The minimax lower bound in the Sobolev case (proved in Theorem 5.3) is established via the Bayes risk and a concentration property of the prior on the ellipsoid. For the exponential ellipsoid case, we use a quantum version of the van Trees inequality from [14]. In the quantum context, for joint estimation of several parameters, the Fisher information appearing in the inequality would have to be replaced by the Holevo bound; in fact, however, one needs a dual Holevo bound (proved in [14]) to obtain an upper bound on the trace of Fisher information matrices from all possible measurements. In both the Sobolev and exponential cases, we observe an inflation of the minimax risk compared to the classical case, and we argue that this is due to the noncommutativity aspect of quantum models.

In Section 6, proceeding from the quantum Gaussian sequence model to product models of pure states, we consider two types of classes: Hermite–Sobolev classes and Hermite-exponential classes of pure states. These are analogs of the Sobolev ellipsoids and exponential ellipsoids, respectively, for sequences of displacements. We prove the nonadaptive part of Theorem 3.3, that is, a sharp minimax result in the i.i.d. pure state model. The upper bound proved in Theorem 6.2 is obtained by constructing a preliminary estimator (from Theorem 5.1 of [11]), which is in a neighborhood of the true state with high probability. We then use local asymptotic equivalence (a stronger version than the one given in [11]) to transfer risks between the quantum Gaussian sequence model and the i.i.d. pure state model. The lower bound is proved in Theorem 6.3 with a similar risk transfer.

Adaptive versions of the estimators are described in Section 7. Inspired by techniques from [40], we replace the Pinsker type estimators of Sections 5 and 6 by weakly geometric block Stein estimators, which are adaptive to the nuisance parameters while preserving the risk asymptotics.

An application of these results to Wigner function estimation is discussed in Section 8. In that context, a method called quantum homodyne tomography is generally applied; [31] and [32] discuss nonparametric estimation of the purity and nonparametric goodness-of-fit testing of a quantum state, respectively, via this method. Nonparametric estimation of the Wigner function in the Bayesian framework is treated in [33]; the minimax risk (when the Wigner function is restricted to a certain smoothness class) is discussed in [1–3, 10, 30]. The minimax rate under L^2-loss obtained by these authors is generally of the order n^{-1} (up to logarithmic corrections) in the case of noiseless measurements and slower than any polynomial in noisy setups. We restrict ourselves to the estimation of Wigner functions for pure states in the absence of noise; under a certain smoothness restriction (a derived Hermite-exponential class) we obtain sharp asymptotic risk bounds. These results are not directly comparable to those cited above, in view of the restriction to pure states; Section 8 discusses this relationship in more detail, that is, to estimation of Wigner functions of general mixed states under exponential smoothness. We also observe that while previous minimax risk bounds for quantum homodyne tomography (such as [30]) assume a fixed measurement, and thus rely on methods of classical nonparametric statistics, our bound is of true quantum type in the sense that it extends over all possible measurements.

The remaining proofs (of Sections 5, 6 and 7) are deferred to Appendix C in [26].

1.2. *Notation.* In physics, the vectors of a Hilbert space $\mathcal{H}$ (assumed separable) are written as "ket" $|v\rangle$, v^* as "bra" $\langle v|$ and the inner product of two vectors as the "bra-ket" $\langle u|v\rangle \in \mathbb{C}$, which is linear with respect to the right entry and antilinear with respect to the left entry. Similarly, $M := |u\rangle\langle v|$ is the rank one operator acting as $M : |w\rangle \mapsto M|w\rangle = \langle v|w\rangle|u\rangle$. For an operator A, the expression $\langle u|Av\rangle$ will sometimes be denoted as $\langle u|A|v\rangle$. The space of bounded linear operators on $\mathcal{H}$ is denoted by $\mathcal{L}(\mathcal{H})$. Of particular interest are the following two subspaces of $\mathcal{L}(\mathcal{H})$.

1. $\mathcal{T}_1(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$—the trace class defined by $\mathcal{T}_1(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} : \mathrm{Tr}(A^*A)^{1/2} < \infty\}$. Operators in $\mathcal{T}_1(\mathcal{H})$ are equipped with the norm $\mathrm{Tr}(A^*A)^{1/2}$.
2. $\mathcal{T}_2(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$—the Hilbert–Schmidt operators defined by $\mathcal{T}_2(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} : \mathrm{Tr}(A^*A) < \infty\}$. Operators in $\mathcal{T}_2(\mathcal{H})$ are equipped with the norm $(\mathrm{Tr}(A^*A))^{1/2}$. The class $\mathcal{T}_2(\mathcal{H})$ is also a Hilbert space with respect to the inner product $(A, B) := \mathrm{Tr}(A^*B)$.

For any Hilbert space, the usual norm will be denoted by $\|\cdot\|$ where the particular space will be understood from the context. We will denote by $\|\mu - \nu\|_{\mathrm{TV}}$ the total variation norm between two measures μ and ν. By $a \vee b$ and $a \wedge b$, we will denote $\max(a, b)$ and $\min(a, b)$, respectively, and a_+ will be used to denote $a \vee 0$. By $\lfloor a \rfloor$ and $\lceil a \rceil$, we will denote the largest integer less than or equal to a and smallest integer greater than or equal to a, respectively. We will use the notation $a_n \asymp b_n$ whenever $c < \liminf_n (a_n/b_n) \le \limsup_n (a_n/b_n) < C$ for some constants $c, C > 0$. Throughout the paper, c and C will denote arbitrary constants.

2. Quantum mechanics preliminaries.

2.1. *States, measurements and observables*. A *state of a quantum system* is described by a self-adjoint operator ρ on a complex Hilbert space $\mathcal{H}$, which is positive ($\rho \ge 0$) and normalized to $\mathrm{Tr}(\rho) = 1$ (a density operator). A state is called *pure* if it is of the form $\rho = |\psi\rangle\langle\psi|$, otherwise it is called a *mixed state*. We denote the set of states by $\mathcal{S}(\mathcal{H})$. It can be shown that $\mathcal{S}(\mathcal{H}) \subset \mathcal{T}_1(\mathcal{H})$.

Data on a quantum system are obtained from *observables*, which are self-adjoint operators in the Hilbert space $\mathcal{H}$. If S is a self-adjoint operator in $\mathcal{H}$ with spectral decomposition $S = \sum_j \lambda_j \Pi_j$ where Π_j are projectors, then a measurement generates a *discrete random variable* X_S taking values in the set of eigenvalues $\{\lambda_1, \lambda_2, \dots\}$ with probabilities $p_j = \mathrm{Tr}(\rho \cdot \Pi_j)$. The expectation of X_S under the state ρ is then given by the *Born–von Neumann postulate*:

$$E_\rho X_S = \sum_j \lambda_j \mathrm{Tr}[\rho \Pi_j] = \mathrm{Tr}[\rho S]. \tag{1}$$

In quantum mechanics, one needs generalized versions of the above definitions of observables and measurements because the spectral decomposition of self-adjoint operators in the form of a weighted sum of projectors may fail to hold when the Hilbert space is infinite dimensional. If a measurement has outcomes in a measurable space $(\Omega, \mathfrak{B})$, it is determined by a positive operator valued measure.

DEFINITION 1. A positive operator valued measure (POVM) is a map $M : \mathfrak{B} \to \mathcal{L}(\mathcal{H})$ having the following properties:

(i) positivity: $M(B) \ge 0$ for all events $B \in \mathfrak{B}$ (hence M(B) is self-adjoint)
(ii) σ-additivity: $M(\bigcup_i B_i) = \sum_i M(B_i)$ for any countable set of mutually disjoint events B_i (here the convergence is in the weak operator topology of $\mathcal{L}(\mathcal{H})$)
(iii) normalization: $M(\Omega) = \mathbf{1}$.

If the operators $M(B)$ are also orthogonal projections, that is, $M(A)^2 = M(A)$ and $M(B)M(A) = 0$ when $A \cap B = \varnothing$, then it is called a *simple measurement*. The collection of projectors $\{\Pi_j\}$ in the spectral decomposition $S = \sum_j \lambda_j \Pi_j$ is an example of a simple measurement. The outcome of the measurement has probability distribution

$$P_\rho(B) = \mathrm{Tr}\big(\rho M(B)\big), \quad B \in \mathfrak{B}. \tag{2}$$

The spectral theorem shows that any self-adjoint operator $S : \mathcal{H} \to \mathcal{H}$ can be diagonalized as follows:

$$S = \int_{\sigma(S)} x M(dx), \tag{3}$$

where $\sigma(S)$ is the spectrum of S and M is a POVM, also called spectral measure associated with the operator S. When S is an observable with a continuous spectrum, it generates a *continuous random variable* X_S with probabilities given by (2). Also it easily follows that

$$E[X_S] = \mathrm{Tr}(S\rho).$$

The expected value of an observable S is often denoted as $\langle S\rangle$, when the state dependence is not explicitly shown. There are POVMs (called *generalized measurements*) where the orthogonality does not hold, but it can be extended to a POVM in a larger Hilbert space where the extended version is orthogonal. Let $\mathrm{POVM}(\Omega, \mathcal{H})$ be the set of POVMs with values in $\mathcal{L}(\mathcal{H})$ and outcome space Ω and let $\mathcal{H}_0$ be another Hilbert space with a density operator ρ_0. Then any simple measurement M' in $\mathrm{POVM}(\Omega, \mathcal{H} \otimes \mathcal{H}_0)$ induces a measurement M in $\mathrm{POVM}(\Omega, \mathcal{H})$, which is determined by

$$\mathrm{Tr}\big(\rho M(B)\big) = \mathrm{Tr}\big((\rho \otimes \rho_0) M'(B)\big), \quad B \in \mathfrak{B}$$

for all states ρ on $\mathcal{H}$. The pair $(\mathcal{H}_0, \rho_0)$ is called an *ancilla* and Holevo ([19], Section 2.5) proved that given any measurement M in $\mathcal{H}$ there exist an ancilla $(\mathcal{H}_0, \rho_0)$ and a simple measurement M' such that the above equation holds. The triple $(\mathcal{H}_0, \rho_0, M')$ is called a *realization* of the measurement M and the notion of adding an ancilla before taking simple measurements is called *quantum randomization* in [6].

In many cases, it is convenient to perform measurement after "changing" the state of the original system by interacting with other systems. The maps describing such transformations are called quantum channels.

DEFINITION 2. A quantum channel between systems with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ is a mapping T, which assigns to every state ρ on $\mathcal{H}_1$ the state $T(\rho)$ on $\mathcal{H}_2$ given by

$$T(\rho) = \sum_{i=1}^{\infty} K_i \rho K_i^*, \tag{4}$$

where $\{K_i\}$ are bounded operators $K_i : \mathcal{H}_1 \to \mathcal{H}_2$ such that $\sum_{i=1}^{\infty} K_i^* K_i = \mathbf{1}$ (the series converging in the strong operator topology of $\mathcal{L}(\mathcal{H})$).

It can be shown that the map T is trace preserving and *completely positive*, that is, $\mathrm{Id}_m \otimes T$ is positive for all $m \geq 1$, where Id_m is the identity map on the space of m dimensional matrices. 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

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

is a quantum channel where $\mathrm{Tr}_{\mathcal{K}}$ is the partial trace over $\mathcal{K}$ (with respect to an orthonormal basis $\{|f_i\rangle\}_{i=1}^{\dim \mathcal{K}}$). If we define operators K_i on $\mathcal{H}$ such that $\langle\psi|K_i|\psi'\rangle := \langle\psi \otimes f_i|V|\psi' \otimes \varphi\rangle$, then it can be seen that $T(\rho)$ can be written as in the form given in Definition 2. We define a dual map T^* of a quantum channel as follows:

$$T^* : \mathrm{POVM}(\Omega, \mathcal{H}_2) \to \mathrm{POVM}(\Omega, \mathcal{H}_1),$$

$$T^*(M)(B) = \sum_{i=1}^{\infty} K_i^* M(B) K_i,$$

where the $\sum_{i=1}^{\infty} K_i^* M(B) K_i$ is a strongly convergent sum. From the definition, it can be easily verified that $T^*(M)$ is indeed an element of $\mathrm{POVM}(\Omega, \mathcal{H}_1)$ (i.e., a POVM satisfying properties i, ii and iii of Definition 1) and that it satisfies the following duality relation:

$$\mathrm{Tr}\big(\rho T^*(M)(B)\big) = \mathrm{Tr}\big(T(\rho)M(B)\big), \quad \forall B \in \mathfrak{B}$$

and all states ρ on $\mathcal{H}_1$ (cf. 29.9 of [37]).

For estimation purposes, we will need to define distances between two quantum states. The *trace-norm* distance between two states $\rho_0, \rho_1 \in \mathcal{S}(\mathcal{H})$ is given by

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

where $|\tau| := \sqrt{\tau^*\tau}$ denotes the absolute value of τ. An interpretation of this metric in terms of quantum testing can be found in [17]. In the special case of pure states, the trace-norm distance is given by

$$\big\| |\psi_0\rangle\langle\psi_0| - |\psi_1\rangle\langle\psi_1| \big\|_1 = 2\sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2}. \tag{5}$$

2.2. *Gaussian states*, *Fock space*, *quantum Gaussian sequence model*. To obtain Gaussian random variables, in the space $\mathcal{H} = L^2(\mathbb{R})$ one considers two special observables Q, P with continuous spectrum:

$$(Qf)(x) = xf(x), \qquad (Pf)(x) = -i\frac{df}{dx}(x), \quad f \in D \subset L^2(\mathbb{R})$$

(defined on an appropriate domain D) often associated to position (Q) and momentum (P) of a particle. It can be shown that $Z_u := u_1 Q + u_2 P$, $u \in \mathbb{R}^2$ are observables (called the *canonical observables*). In this context, we define the *quantum characteristic function* as $\widetilde{W}_\rho(u_1, u_2) = \mathrm{Tr}(\rho \exp(iZ_u))$. The inverse Fourier transform of $\widetilde{W}_\rho$ with respect to both variables is called *Wigner function* W_ρ, or quasidistribution associated to ρ:

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

A well-known identity relating Wigner functions and the states is the overlap formula

$$\mathrm{Tr}[\rho_1\rho_2] = 2\pi \int\int W_{\rho_1}(q, p) W_{\rho_2}(q, p)\, dq\, dp \tag{6}$$

a proof of which (along with other properties of Wigner functions) can be found in Chapter 3 of [28]. If the following relation holds:

$$E_\rho \exp(iZ_u) = \mathrm{Tr}\big(\rho \exp(iZ_u)\big) = \exp\Big(iu'\mu - \frac{1}{2}u'\Sigma u\Big), \quad u \in \mathbb{R}^2,$$

then ρ is called a Gaussian state with mean μ and covariance matrix Σ. It trivially follows that the Wigner function of ρ is the bivariate Gaussian density with mean μ and covariance matrix Σ. For such quantum Gaussian states in $L^2(\mathbb{R})$, we adopt a compact notation, resembling the one for the 2-variate normal law:

$$\rho = \mathbb{N}_2(\mu, \Sigma). \tag{7}$$

To define the simplest Gaussian state, let $\psi_0 = \sqrt{\varphi_{1/2}}$ be the square root of the density function of the normal $N(0, 1/2)$ distribution and consider the operator ρ_0 acting by $\rho_0 f = \psi_0 \langle \psi_0, f \rangle$, $f \in L^2(\mathbb{R})$. Since ψ_0 is a unit vector in $L^2(\mathbb{R})$, the operator ρ_0 (henceforth called the vacuum state) is a projection (written $\rho_0 = |\psi_0\rangle\langle\psi_0|$ in Dirac notation) and it can be shown that $\rho_0 = \mathbb{N}_2(0, \frac{1}{2} I)$ in the notation described above.

An important class is the collection of *coherent states* $\mathbb{N}_2(\mu, I/2)$; these are pure states which can be interpreted as a vacuum shifted by $\mu \in \mathbb{R}^2$ (similar to the Gaussian shift model in classical statistics). Consider the operators $a^* = (Q - iP)/\sqrt{2}$ (the *creation operator*), $a = (Q + iP)/\sqrt{2}$ (the *annihilation operator*) and $N = a^* a$ (the number operator). It is well known that the *Hermite basis* $\{|0\rangle, |1\rangle \ldots\}$ forms an eigenbasis of the number operator, that is, $N|k\rangle = k|k\rangle$. For any $z \in \mathbb{C}$, define the displacement operator as

$$D(z) = \exp\bigl(z a^* - \bar{z} a\bigr)$$

and the *coherent state* as

$$\bigl|G(z)\bigr\rangle = D(z)|0\rangle = \exp\bigl(-|z|^2/2\bigr) \sum_{k=0}^{\infty} \frac{z^k}{\sqrt{k!}} |k\rangle. \tag{8}$$

In the density operator notation, this pure Gaussian state is $|G(z)\rangle\langle G(z)|$. The expectations of the canonical observables Q and P under the state $|G(z)\rangle\langle G(z)|$ are

$$\langle Q \rangle = \sqrt{2}\,\mathrm{Re}\, z, \qquad \langle P \rangle = \sqrt{2}\,\mathrm{Im}\, z$$

and the characteristic function of $|G(z)\rangle\langle G(z)|$ is

$$\varphi(t) = \exp\biggl(i\bigl(t_1\sqrt{2}\,\mathrm{Re}\, z + t_2\sqrt{2}\,\mathrm{Im}\, z\bigr) - \frac{1}{4}\bigl(t_1^2 + t_2^2\bigr)\biggr), \quad t \in \mathbb{R}^2. \tag{9}$$

The presence of the factor $\sqrt{2}$ motivates us to adopt a modified notation for the coherent vector: setting $\mu = (\sqrt{2}\,\mathrm{Re}\, z, \sqrt{2}\,\mathrm{Im}\, z)$, we will write $|G(z)\rangle = |\psi_\mu\rangle$ so that now the expectations are $(\langle Q \rangle, \langle P \rangle) = \mu$. The characteristic function $\varphi(t)$ is that of $N_2(\mu, I/2)$, and hence in the notation of (7)

$$\bigl|G(z)\bigr\rangle\bigl\langle G(z)\bigr| = |\psi_\mu\rangle\langle\psi_\mu| = \mathbb{N}_2(\mu, I/2). \tag{10}$$

This setting describes Gaussian states in *a single mode*; to extend this model to an "infinite mode" system, one uses Fock spaces. We adopt the definition of Fock space and the general notion of coherent state from [37].

DEFINITION 3. Let $\mathcal{K}$ be a Hilbert space. The *Fock space* over $\mathcal{K}$ is the Hilbert space

$$\mathcal{F}(\mathcal{K}) = \bigoplus_{n \geq 0} \mathcal{K}^{\otimes_s n}, \tag{11}$$

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 space.

DEFINITION 4. 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

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

When $v = z \in \mathbb{C}$, we notice the similarity of the RHS of the above equation and that of (8). It can be shown that if one fixes an orthonormal basis in $\mathcal{K}$, an isomorphism between the Fock space $\mathcal{F}(\mathcal{K})$ and a tensor product of one-mode spaces (one for each basis vector) can be established. In particular, if $\mathcal{K} = \mathbb{C}$, then $\mathcal{F}(\mathbb{C}) \cong L^2(\mathbb{R})$ and 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 factorize as

$$
\begin{aligned}
\mathcal{F}(\mathcal{K}) &\cong \bigotimes_i \mathcal{F}_i, \\
|G(\psi)\rangle &\cong \bigotimes_i |G(\psi_i)\rangle, \qquad \psi_i = \langle e_i|\psi\rangle.
\end{aligned} \tag{12}
$$

Note that in the above expression $|G(\psi_i|e_i\rangle)\rangle$ is written as $|G(\psi_i)\rangle$ since the span of $|e_i\rangle$ and $\mathbb{C}$ are isomorphic. Now let $\mathcal{K} = L^2(\mathbb{R})$ and consider $\psi = \sum_{i=1}^{\infty} \psi_i |e_i\rangle$ for any orthonormal basis $\{|e_1\rangle, |e_2\rangle, \dots\}$ of $L^2(\mathbb{R})$. For $z = \psi_i$, we set

$$
\mu_i = (\sqrt{2}\,\mathrm{Re}\,\psi_i, \sqrt{2}\,\mathrm{Im}\,\psi_i), \tag{13}
$$

and also write $|G(z)\rangle = |\psi_{\mu_i}\rangle = |\bar{G}(\mu_i)\rangle$ so that now the expectations are $(\langle Q_i\rangle, \langle P_i\rangle) = \mu_i$ where Q_i and P_i are position and momentum operators, respectively, for the ith mode. To clarify notation, we now write $|G(z)\rangle$ when z is complex and $|\bar{G}(\mu)\rangle = |G(z)\rangle$ for $\mu = (\sqrt{2}\,\mathrm{Re}\,z, \sqrt{2}\,\mathrm{Im}\,z)$. We now factorize the coherent state as follows:

$$
\begin{aligned}
|G(\sqrt{n}\psi)\rangle\langle G(\sqrt{n}\psi)| &\cong \bigotimes_{i=1}^{\infty} |G(\sqrt{n}\psi_i)\rangle\langle G(\sqrt{n}\psi_i)| = \bigotimes_{i=1}^{\infty} |\bar{G}(\sqrt{n}\mu_i)\rangle\langle \bar{G}(\sqrt{n}\mu_i)| \\
&= \bigotimes_{i=1}^{\infty} \mathbb{N}_2(\sqrt{n}\mu_i, I/2),
\end{aligned} \tag{14}
$$

where we have used the notation from (10) in the last step. The last expression of (14) is the analog of the classical Gaussian sequence model $y_j \sim N(\theta_j, n^{-1})$, $j \geq 1$. We will see in Section 5 that the random variables generated using certain measurements (called covariant measurements) indeed form a Gaussian sequence model.

Let $\psi_0, \psi_1 \in \mathcal{K}$. The trace-norm distance between the corresponding coherent states is given by

$$
\big\| |G(\psi_0)\rangle\langle G(\psi_0)| - |G(\psi_1)\rangle\langle G(\psi_1)| \big\|_1 = 2\sqrt{1 - \exp(-\|\psi_0 - \psi_1\|^2)}. \tag{15}
$$

If $\psi_0, \psi_1 \in \mathbb{C}$, we write the above equation, in our present notation, as

$$
\big\| |\bar{G}(\mu_0)\rangle\langle \bar{G}(\mu_0)| - |\bar{G}(\mu_1)\rangle\langle \bar{G}(\mu_1)| \big\|_1 = 2\sqrt{1 - \exp\left(-\frac{1}{2}\|\mu_0 - \mu_1\|^2\right)}.
$$

2.3. *Quantum statistical inference*. In this section, we formalize the quantum counterparts of the basic notions of optimality in classical statistical inference. In classical statistics, an experiment is defined to be a family of probability measures on a sample space and denoted by $E = \{P_\theta, \theta \in \Theta\}$ where Θ is the parameter space.

DEFINITION 5. 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 $\theta \in \Theta$.

Inference in quantum models generally involves two steps. In the first step, one performs a measurement on the state ρ_θ and generates data, while in the second step, one uses standard statistical tools to solve the specific decision problem using data from the first step. If one performs a measurement M on the system in state ρ_θ, a random outcome is obtained with distribution $\mathbb{P}_\theta^M(E) := \mathrm{Tr}(\rho_\theta M(E))$ (cf. Section 2.1). 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. However, in many scenarios the optimal estimators for individual components of a parameter are incompatible with each other and the optimal joint estimator for the two components can be entirely different from the optimal estimators of the individual components.

In the classical setup, a randomized decision function is given by a Markov kernel ν. If $L(\theta, u)$ is the loss function, then the risk is given by

$$\begin{aligned} R(\theta, \nu) &= \int\int L(\theta, u)\nu_x(du)\mu_\theta(dx) = \int L(\theta, u)\int \nu_x(du)\mu_\theta(dx) \\ &= \int L(\theta, u)\tilde{\nu}_\theta(du), \end{aligned} \tag{16}$$

where $\tilde{\nu}_\theta(A) = \int \nu_x(A)\mu_\theta(dx)$.

Section 2.2.4 of [22] discusses the quantum counterpart of this classical formulation. Let ρ_θ be the quantum state and $\mu_\theta^M(B) = \mathrm{Tr}(\rho_\theta M(B))$ be the probability measure generated by the POVM M. Then the risk is given by

$$R(\theta, M) = \int L(\theta, u)\mu_\theta^M(du). \tag{17}$$

By using the fact that every affine map $\rho_\theta \to \mu_\theta()$ can be associated with a POVM, we see that M is an analog of the classical randomized decision function ν given in (16). We can easily define the Bayes and minimax problems for quantum estimation.

Minimax problem

$$\inf_M \sup_{\theta\in\Theta} R(\theta, M) = \inf_M \sup_{\theta\in\Theta} \int L(\theta, u)\mu_\theta^M(du) = \inf_{\hat{m}} \sup_{\theta\in\Theta} E_\theta\big[L(\theta, \hat{m})\big].$$

Bayes problem

$$\inf_M \int_\Theta R(\theta, M)\pi(d\theta) = \inf_M \int_\Theta \int L(\theta, u)\mu_\theta^M(du)\pi(d\theta) = \inf_{\hat{m}} \int E_\theta\big[L(\theta, \hat{m})\big]\pi(d\theta).$$

The notation $\inf_{\hat{m}} \sup_{\theta\in\Theta} E_\theta[L(\theta, \hat{m})]$ and $\inf_{\hat{m}} \int E_\theta[L(\theta, \hat{m})]\pi(d\theta)$ will be called *condensed notation* henceforth and will be used in Section 8 and in the proofs in [26]. Note that the infimum is over all POVM and the notation $\hat{m}$ should not be confused with a deterministic estimator seen in the classical setup. We will also denote the Bayes risk as $\inf_{\hat{m}} E[L(\theta, \hat{m})]$ where the expectation is also taken over the parameter θ.

In classical statistics, a well-known paradigm is using asymptotic equivalence of experiments to transfer risk bounds from one experiment to another. Suppose we have two experiments $E = \{P_\theta, \theta \in \Theta\}$ on a sample space $(\Omega_1, \mathcal{A}_1)$ and $F = \{Q_\theta, \theta \in \Theta\}$ on a a sample space $(\Omega_2, \mathcal{A}_2)$. Also let the loss function satisfy the condition $0 \le L(\theta, u) \le 1$. If there exists a Markov kernel K such that

$$\sup_{\theta\in\Theta} \|KP_\theta - Q_\theta\|_{\mathrm{TV}} \le \epsilon,$$

then for any randomized decision function μ of θ in the model Q_θ, the randomized decision function $\nu = \mu \circ K$ (composition of two Markov kernels) satisfies

$$R^1(\theta, \nu) \le R^2(\theta, \mu) + \epsilon,$$

where $R^1(\theta, \nu)$ and $R^2(\theta, \mu)$ are the risks in the models E and F, respectively. We discuss the generalization of this paradigm to the quantum setup and also generalize it to the case of unbounded loss.

The quantum equivalent of a Markov kernel is the transformation by quantum channels. 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 applying a quantum channel

$$T : \mathcal{T}_1(\mathcal{H}) \to \mathcal{T}_1(\mathcal{H}'),$$
$$T : \rho_\theta \mapsto \rho'_\theta.$$

In this context, we define the *quantum Le Cam distance* between two models from [11].

DEFINITION 6. 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

$$\Delta(\mathcal{Q}, \mathcal{Q}') := \max\big(\delta(\mathcal{Q}, \mathcal{Q}'), \delta(\mathcal{Q}', \mathcal{Q})\big). \tag{18}$$

Its interpretation is that models which are "close" in the Le Cam distance have similar risk bounds. Now, suppose we have two sequences of quantum models (or experiments) $E^{(n)} = \{\rho_\theta^{(1,n)} : \theta \in \Theta\}$ and $F^{(n)} = \{\rho_\theta^{(2,n)} : \theta \in \Theta\}$. Assume that $\Delta(E^{(n)}, F^{(n)}) \to 0$ with the associated sequence of Hilbert spaces $\mathcal{H}^{1,n}$ and $\mathcal{H}^{2,n}$. This implies $\delta(E^{(n)}, F^{(n)}) \to 0$ and in particular there exists a sequence of quantum channels T_n, such that

$$\big\|T_n\big(\rho_\theta^{(1,n)}\big) - \rho_\theta^{(2,n)}\big\|_1 = o(1).$$

Let the loss function also change with n and satisfy the relation $0 \le L_n(\theta, u) \le c_n$. Also assume that the sequence of quantum channels T_n is such that

$$c_n \sup_{\theta\in\Theta} \big\|T_n\big(\rho_\theta^{(1,n)}\big) - \rho_\theta^{(2,n)}\big\|_1 = o(1). \tag{19}$$

Recall the dual map T^* of a quantum channel T. It follows that for any $M \in \mathrm{POVM}(\Omega, \mathcal{H}^{2,n})$,

$$\begin{aligned}
R_n^1(\theta, T_n^*(M)) &= \int L_n(\theta, u) \operatorname{Tr}\big(\rho_\theta^{(1,n)} T_n^*(M(du))\big) \\
&= \int L_n(\theta, u) \operatorname{Tr}\big(\rho_\theta^{(2,n)} M(du)\big) \qquad (20)\\
&\quad + \int L_n(\theta, u) \big[\operatorname{Tr}\big(\rho_\theta^{(1,n)} T_n^*(M(du))\big) - \operatorname{Tr}\big(\rho_\theta^{(2,n)} M(du)\big)\big] \\
&= R_n^2(\theta, M) + \int L_n(\theta, u)\big[\operatorname{Tr}\big(T_n\big(\rho_\theta^{(1,n)}\big)(M(du))\big) - \operatorname{Tr}\big(\rho_\theta^{(2,n)} M(du)\big)\big] \\
&\le R_n^2(\theta, M) + c_n \big\|T_n\big(\rho_\theta^{(1,n)}\big) - \rho_\theta^{(2,n)}\big\|_1 \\
&\le R_n^2(\theta, M) + o(1) \qquad (21)
\end{aligned}$$

the term $o(1)$ going to 0 uniformly over all θ. Thus we can compare the risks of the two models $E^{(n)}$ and $F^{(n)}$ if (19) holds. Note that we have similar relations for minimax risks and

Bayes risks, by taking a supremum over Θ or integrating with respect to a prior, respectively, and then taking an infimum over all estimators:

$$\inf_M \sup_{\theta\in\Theta} R_n^1(\theta, M) \le \inf_M \sup_{\theta\in\Theta} R_n^2(\theta, M) + o(1). \tag{22}$$

$$\inf_M \int_\Theta R_n^1(\theta, M)\pi(d\theta) \le \inf_M \int_\Theta R_n^2(\theta, M)\pi(d\theta) + o(1). \tag{23}$$

3. Main theorems.

Consider the classical Gaussian white noise model

$$dY(t) = f(t)\,dt + \frac{1}{\sqrt{n}} dW(t), \quad t \in [0, 1], \tag{24}$$

where $W(t)$ is the standard Brownian motion and the equivalent Gaussian sequence model

$$y_j = \theta_j + \frac{1}{\sqrt{n}}\xi_j, \quad j = 1, 2, \dots. \tag{25}$$

Here, the ξ_j's are i.i.d. $N(0, 1)$. The second model can be obtained from the first by the following transformations:

$$y_j = \int_0^1 \phi_j(t)\,dY(t), \qquad \theta_j = \int_0^1 \phi_j(t) f(t)\,dt, \qquad \xi_j = \int_0^1 \phi_j(t)\,dW(t),$$

where $\phi_j(t)$ is the trigonometric basis of $L^2[0, 1]$ given by

$$\phi_1(t) = 1, \qquad \phi_{2k}(t) = \sqrt{2}\cos(2\pi kt), \qquad \phi_{2k+1}(t) = \sqrt{2}\sin(2\pi kt), \quad k = 1, 2, \dots.$$

Consider the Sobolev ellipsoid

$$\Theta(\beta, Q) = \left\{\theta = \{\theta_j\} \in \ell^2(\mathbb{N}) : \sum_{j=1}^\infty \alpha_j \theta_j^2 \le Q\right\},$$

where $\alpha_j = j^{2\beta}$ or $(j-1)^{2\beta}$ for even and odd j, respectively.

Also consider the class of functions,

$$W(\beta, Q) = \left\{f \in L^2[0, 1] : \theta = \{\theta_j\} \in \Theta(\beta, Q), \text{ where } \theta_j = \int_0^1 \phi_j(t) f(t)\,dt\right\}.$$

Pinsker's theorem ([38], cf. also [7, 35, 40]) gives sharp estimates of the asymptotic minimax risk in the above models, where one can be derived from the other using Parseval's identity.

THEOREM 3.1 (Pinsker's theorem). *For $\beta > 0$, $Q > 0$, we have*

$$\lim_{n\to\infty} \inf_{\hat{f}_n} \sup_{f\in W(\beta,Q)} n^{2\beta/(2\beta+1)} E_f \|f - \hat{f}_n\|^2 = C^*,$$

$$\lim_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta(\beta,Q)} n^{2\beta/(2\beta+1)} E_\theta \|\theta - \hat{\theta}_n\|^2 = C^*,$$

where

$$C^* = Q^{1/(2\beta+1)} c(\beta), \quad c(\beta) = \left(\frac{\beta}{\beta+1}\right)^{2\beta/(2\beta+1)} (2\beta+1)^{1/(2\beta+1)}$$

and $\inf_{\hat{f}_n}$, $\inf_{\hat{\theta}_n}$ extend over all estimators of f, θ in the classical models (24) *and* (25), *respectively.*

Now consider the following norm in $\ell^2(\mathbb{R}^2)$ (square summable sequences in $\mathbb{R}^2$): for $m = (\mu_1, \mu_2, \dots)$,

$$\|m\|^2 = \sum_{j=1}^{\infty} \|\mu_j\|^2$$

and consider the quantum Gaussian sequence model given in (14). In particular, consider the following form of the tensor product of coherent states:

$$\rho_m = \bigotimes_{j=1}^{\infty} |\bar{G}(n^{1/2}\mu_j)\rangle\langle\bar{G}(n^{1/2}\mu_j)|, \quad \|m\|^2 < \infty. \tag{26}$$

In the Sobolev case, we will assume that the sequence $m = (\mu_1, \mu_2, \dots)$ is in an ellipsoid (for $\beta > 0$, $R > 0$) described by the following equation:

$$\Sigma(\beta, R) := \left\{m : \sum_{j=1}^{\infty} j^{2\beta} \|\mu_j\|^2 \le R\right\} \tag{27}$$

and in the exponential case we will assume that $m = (\mu_1, \mu_2, \dots)$ is in an ellipsoid (for $\beta > 0$, $r \in (0, 1]$, $R > 0$) described by the following equation:

$$\mathcal{E}(\beta, r, R) := \left\{m : \sum_{j=1}^{\infty} \exp(2\beta j^r) \|\mu_j\|^2 \le R\right\}. \tag{28}$$

In [11], the authors prove an optimal rate of convergence result for estimation of m under the Sobolev ellipsoid in terms of the norm $\|m\|^2$. Using a Bayesian framework of estimation of a single component μ_j (discussed in the next section), we can sharpen that result to obtain the following analog of Pinsker's theorem in the quantum setup. We also extend our theorem to the exponential ellipsoid case, the classical analog of which can be found in [7].

THEOREM 3.2. Case A: Sobolev ellipsoid
Consider the model given in (26) *with* $m = (\mu_1, \mu_2 \dots)$. *Let* $\Sigma(\beta, R)$ *as given in* (27) *and* $w_n(x) = x \wedge \log\log n$. *Then the following result holds*:

$$\lim_{n\to\infty} \inf_{M} \sup_{m\in\Sigma(\beta,R)} \int w_n\bigl(\eta_n(\beta, R)\|m - u\|^2\bigr) \operatorname{Tr}\bigl(\rho_m M(du)\bigr) = 1, \tag{29}$$

where

$$\eta_n(\beta, R) = (n/2)^{2\beta/(2\beta+1)} R^{-1/(2\beta+1)} c^{-1}(\beta)$$

and $c(\beta)$ *is the constant defined in Theorem* 3.1. *An estimator attaining this bound asymptotically can be chosen independent of* β *and* R *as long as* $\beta, R > 0$.

Case B: Exponential ellipsoid
Let $\mathcal{E}(\beta, r, R)$ *as given in* (28) *and* $v_n(x) = x \wedge (n^{1/2}/\log\log n)$. *Then the following result holds*:

$$\lim_{n\to\infty} \inf_{M} \sup_{m\in\mathcal{E}(\beta,r,R)} \int v_n\bigl(\tau_n(\beta, r)\|m - u\|^2\bigr) \operatorname{Tr}\bigl(\rho_m M(du)\bigr) = 1, \tag{30}$$

where

$$\tau_n(\beta, r) = \frac{n}{2}\left(\frac{\log n}{2\beta}\right)^{-1/r}.$$

For every $0 < r_0 < 1$, *there exists an estimator, not depending on* β, R *and* r, *attaining this bound asymptotically, whenever* $\beta, R > 0$ *and* $r_0 < r \le 1$.

In both the cases, the POVMs are maps described in general as follows: $M : \mathcal{B}(\ell^2(\mathbb{R}^2)) \to \mathcal{L}(\mathcal{F}(L^2(\mathbb{R})))$ *where* $\mathcal{B}(\ell^2(\mathbb{R}^2))$ *is the Borel sigma algebra of* $\ell^2(\mathbb{R}^2)$ *and* $\mathcal{L}(\mathcal{F}(L^2(\mathbb{R})))$ *is the space of bounded operators on the Fock space* $\mathcal{F}(L^2(\mathbb{R}))$.

Note: *In the exponential case, the minimax constant does not depend on* R.

We note that the risk functions here are derived from normalized truncated quadratic $\ell^2(\mathbb{R}^2)$-loss, with truncation effected by functions $w_n(x)$ or $v_n(x)$, and where the normalizer sequences $\eta_n(\beta, r)$ or $\tau_n(\beta, r)$ encode the sharp risk asymptotics. This variant of Pinsker type bounds has first been stated in [39], using fixed, possibly bounded functions $w(x)$; it is particularly suited for transfer by asymptotic equivalence.

Indeed, the quantum asymptotic equivalence result of [11] now ensures that a similar sharp risk bound can be established in the product model of pure states $|\psi\rangle\langle\psi|^{\otimes n}$ with squared trace norm loss. It can be shown that the rate of approximation of the models in the asymptotic equivalence sense can be chosen to be faster than the truncation rates, that is, the condition (19) is satisfied with $c_n = \log\log n$ or $n^{1/2}/\log\log n$ for the Sobolev and the exponential cases, respectively. So we can transfer the risk bound of the quantum Gaussian sequence model to a risk bound in the quantum pure product state model.

Let $\mathcal{H} = L^2(\mathbb{R})$ and let $|e_0\rangle, |e_1\rangle, \ldots$ be the standard ONB (Hermite functions). Introduce the following *Hermite–Sobolev class* and *Hermite-exponential class* of pure states characterized by ellipsoid conditions:

$$S^\beta(L) := \left\{ |\psi\rangle\langle\psi| : \|\psi\| = 1 \text{ and } \sum_{j=0}^{\infty} \left|\langle\psi|e_j\rangle\right|^2 j^{2\beta} \le L \right\}, \quad \beta > 0, L > 0, \tag{31}$$

$$E^{\beta,r}(L) := \left\{ |\psi\rangle\langle\psi| : \|\psi\| = 1 \text{ and } \sum_{j=0}^{\infty} \left|\langle\psi|e_j\rangle\right|^2 \exp(2\beta j^r) \le L \right\}, \\ \beta > 0, L > 1, r \in (0, 1]. \tag{32}$$

THEOREM 3.3. *Assume measurements on states* $|\psi\rangle\langle\psi|^{\otimes n}$. *Then we have the following results*.

Case A: Hermite–Sobolev class

Let $S^\beta(L)$ *be the Hermite–Sobolev class defined above and* $|\psi\rangle\langle\psi| \in S^\beta(L)$. *Then the following result holds*:

$$\lim_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi| \in S^\beta(L)} \int w_n\left(\frac{\tilde{\eta}_n(\beta, L)}{2} \|\rho - |\psi\rangle\langle\psi|\|_1^2\right) \operatorname{Tr}\left(|\psi\rangle\langle\psi|^{\otimes n} M(d\rho)\right) = 1, \tag{33}$$

where

$$\tilde{\eta}_n(\beta, L) = \frac{1}{2} n^{2\beta/(2\beta+1)} L^{-1/(2\beta+1)} c^{-1}(\beta).$$

For every $\beta_0 > 1/2$, *there exists an estimator, not depending on* β *and* L, *attaining this bound asymptotically, whenever* $\beta > \beta_0$ *and* $L > 0$.

Case B: Hermite-exponential class

Let $E^{\beta,r}(L)$ *be the Hermite-exponential class defined above and* $|\psi\rangle\langle\psi| \in E^{\beta,r}(L)$. *Then the following holds*:

$$\lim_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi| \in E^{\beta,r}(L)} \int v_n\left(\frac{\tau_n(\beta, r)}{2} \|\rho - |\psi\rangle\langle\psi|\|_1^2\right) \operatorname{Tr}\left(|\psi\rangle\langle\psi|^{\otimes n} M(d\rho)\right) = 1. \tag{34}$$

For every $0 < r_0 < 1$ *and* $\beta_0 > 0$, *there exists an estimator, not depending on* β, R *and* r, *attaining this bound asymptotically, whenever* $\beta > \beta_0$, $L > 1$ *and* $r_0 < r \le 1$.

In both the cases, the POVMs are maps described in general as follows: $M : \mathcal{B} \to \mathcal{L}(\mathcal{H}^{\otimes n})$ *where* $\mathcal{B}$ *is the Borel sigma algebra of* $\mathcal{T}_1(L^2(\mathbb{R}))$ *restricted to density operators of rank* 1 *and* $\mathcal{H} = L^2(\mathbb{R})$.

Note: *In the exponential case*, *the minimax constant does not depend on L.*

REMARK. Henceforth, "Case A" will be used to denote the Sobolev ellipsoid and the Hermite– Sobolev class for the Gaussian sequence model and the i.i.d. model, respectively. Similarly, "Case B" will be used to denote the exponential ellipsoid and the Hermite-exponential class for the Gaussian sequence model and the i.i.d. model, respectively. While discussing the POVMs in the following sections, the spaces on which they act will be those described in Theorems 3.2 and 3.3.

4. Measurement of the shift parameter. In this section, we consider measurement of the shift parameter μ of the model $\rho = \mathbb{N}_2(\mu, I/2)$. In the first part, we describe the particular generalized measurement (the *covariant measurement*) that is used to measure the shift parameter and mention its optimality as a pointwise estimator. A Bayes estimator constructed by appropriately "shrinking" the covariant measurement is discussed in the following subsection. Since the details of this measurement scheme are thoroughly discussed in Chapters 3, 4 and 5 of [19], we refer the reader to the relevant sections for proofs and details. Appendix A in [26] also contains some discussion of the observables generating the covariant measurement.

4.1. *Covariant measurement*. An essential fact is that the coherent vectors $\{|\psi_m\rangle, m \in \mathbb{R}^2\}$ form an "overcomplete system" (if multiplied by a factor $1/\sqrt{2\pi}$), that is, fulfill

$$\frac{1}{2\pi}\int_{\mathbb{R}^2} |\psi_m\rangle\langle\psi_m|\, dm = I, \tag{35}$$

where I is the identity operator in the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$ (see equation (3.5.45), p. 101 of [19], with proof after Proposition 3.5.1). A complete orthonormal system $\{|\psi_m\rangle\}$ is an example of an overcomplete system (with integration replaced by summation); however, in the general case the vectors ψ_m can be nonorthogonal and linearly dependent. The system of coherent vectors $\{|\psi_m\rangle, m \in \mathbb{R}^2\}$ generates a *resolution of the identity*, that is, a normalized POVM M on the measurable space $(\mathbb{R}^2, \mathfrak{B})$ (with $\mathfrak{B}$ being the Borel sigma-algebra) according to

$$M(B) = \frac{1}{2\pi}\int_B |\psi_m\rangle\langle\psi_m|\, dm, \qquad B \in \mathfrak{B}. \tag{36}$$

The POVM M then generates a (generalized) observable X_M with values in $\mathbb{R}^2$, which under the state ρ has probability distribution

$$P(X_M \in B) = \operatorname{Tr} \rho M(B), \qquad B \in \mathfrak{B}. \tag{37}$$

This is called the *canonical covariant measurement* in Section 3.6 of [19], covariance referring to the action of the Weyl unitaries (or the displacement operator). There also optimality properties are proved, as well as equivalence to simple measurements on an extended system (corresponding to an *orthogonal* resolution of the identity there). When the covariant measurement is clear from the context we will write often write X_M as X; in Appendix A in [26] we show that $X \sim N_2(\mu, I_2)$ if $\rho = |\psi_\mu\rangle\langle\psi_\mu|$.

An equivalent description can be given as follows; cf. Proposition 3.6.1 of [19] and also relation (3.18) in [21]. Let $\mathcal{H} = L^2(\mathbb{R})$ and let $\mathcal{H}_0$ be an identical Hilbert space with canonical observables Q_0 and P_0. In the tensor product, $\mathcal{H} \otimes \mathcal{H}_0$ consider the operators

$$\tilde{Q} = Q \otimes I_0 + I \otimes Q_0, \qquad \tilde{P} = P \otimes I_0 - I \otimes P_0, \tag{38}$$

where I_0 is the identity operator in $\mathcal{H}_0$. Let ρ be the state in $\mathcal{H}$ to be measured and ρ_0 be an auxiliary state in $\mathcal{H}_0$ to be chosen; then a simple measurement of $\rho \otimes \rho_0$ can be understood as a "randomized" measurement of ρ. These randomized measurements correspond to nonorthogonal resolutions like (35).

It can be shown that $\tilde{Q}$ and $\tilde{P}$ commute (see Appendix A in [26]) which means that the observables $\tilde{Q}$, $\tilde{P}$ are jointly measurable in the system given by $\mathcal{H} \otimes \mathcal{H}_0$. The operators $\tilde{Q}$, $\tilde{P}$ are self-adjoint, and thus generate jointly distributed real valued random variables $X_{\tilde{Q}}$, $X_{\tilde{P}}$. We define $\tilde{X}$ as follows:

$$\tilde{X} = (X_{\tilde{Q}}, X_{\tilde{P}}) \tag{39}$$

and it can be checked that if $\rho = |\psi_\mu\rangle\langle\psi_\mu|$ and the auxiliary state ρ_0 is the vacuum $\rho_0 = |\psi_0\rangle\langle\psi_0|$ then the distribution of $\tilde{X}$ coincides with the distribution of X (obtained with the covariant measurement), that is, with $N_2(\mu, I_2)$ (see Appendix A in [26]). Suppose $\mu_1 = \operatorname{Tr} Q|\psi_\mu\rangle\langle\psi_\mu|$ is to be estimated, then Q turns out to be the uniformly best unbiased measurement of μ_1 and the random variable Y_Q corresponding to the canonical observable Q has distribution $N(\mu_1, 1/2)$. Similar results hold for the parameter μ_2 and the observable P. We have

$$E_\mu(Y_Q - \mu_1)^2 = \operatorname{Var}_\mu(Y_Q) = \frac{1}{2} = E_\mu(Y_P - \mu_2)^2$$

so that

$$E_\mu \|(Y_Q, Y_P) - \mu\|^2 = 1. \tag{40}$$

But Y_Q, Y_P are not jointly observable, so the bound 1 for the MSE of estimating μ cannot actually be attained. Indeed Sections 6.5 and 6.6 of [19] discuss two different Cramer–Rao type bounds for the MSE, one of which is attainable for one-dimensional parameters only (the bound achievable for higher dimensional parameters is known as the Holevo bound in the literature [14, 42]). Since the unbiased estimator X of μ (based on the canonical measurement (36)) has distribution $N_2(\mu, I_2)$ under $|\psi_\mu\rangle\langle\psi_\mu|$, we have

$$E_\mu \|X - \mu\|^2 = 2.$$

It is noted in [19] that the higher bound 2 compared to 1 in (40) is an expression of the fact that Y_Q, Y_P are not jointly observable. Also a scheme is presented where in a continuum of quantum models, depending on a parameter $\hbar$ expressing the "degree of noncommutativity," the higher bound tends to the lower one as $\hbar \to 0$.

4.2. *Bayes estimation.* Consider the problem of Bayes estimation with quadratic loss of the parameter $\mu \in \mathbb{R}^2$ under a normal prior $\mu \sim N_2(0, \sigma_0^2 I_2)$. The solution for a quadratic risk and given in [22], p. 55, with details and proofs in [21]. Consider the loss function

$$L(\hat{\mu}, \mu) = g_1(\hat{\mu}_1 - \mu_1)^2 + g_2(\hat{\mu}_2 - \mu_2)^2.$$

The form of the solution (the Bayes estimator) crucially depends on the ratio g_1/g_2. Define $s^2 = \sigma_0^2 + 1/2$; we have to distinguish three cases: $g_1/g_2 \geq (2s^2)^2$, $g_1/g_2 \leq (2s^2)^{-2}$ and $(2s^2)^{-2} < g_1/g_2 < (2s^2)^2$ and discuss them separately. The solutions overlap at the boundaries.

Case A1. If $g_1/g_2 \geq (2s^2)^2$, then

$$\hat{\mu}_1 = \left(\frac{\sigma_0^2}{\sigma_0^2 + 1/2}\right) Y_Q, \qquad \hat{\mu}_2 = 0 \tag{41}$$

Case A2. If $g_1/g_2 \le (2s^2)^{-2}$, then

$$\hat{\mu}_1 = 0, \qquad \hat{\mu}_2 = \left(\frac{\sigma_0^2}{\sigma_0^2 + 1/2}\right) Y_P, \tag{42}$$

where Y_Q is the r.v. generated by Q and Y_P is the r.v. generated by P. For a point on the presentation in [22], see the technical remark in Appendix A in [26].

These estimators are called "deterministic" in [22] since they are based on an simple observable (not a generalized one), that is, they do not involve additional "randomization" by an auxiliary system (equivalently, do not involve a nonorthogonal resolution like (35)). We compute the risk for case A1 noting that the calculations for the case A2 is analogous. Recall Y_Q, Y_P defined in the last page; the risk of this estimator of $\hat{\mu}_1$ for fixed (nonrandom) μ is

$$\begin{aligned} &E_\mu g_1(\hat{\mu}_1 - \mu_1)^2 + g_2\mu_2^2 \\ &\quad = g_1\left(1 - \frac{\sigma_0^2}{\sigma_0^2 + 1/2}\right)^2 \mu_1^2 + g_1\left(\frac{\sigma_0^2}{\sigma_0^2 + 1/2}\right)^2 \frac{1}{2} + g_2\mu_2^2 \end{aligned}$$

and if π denotes the prior distribution $\pi = N_2(0, \sigma_0^2 I_2)$ for μ then the overall (Bayes) risk is the expectation of the above under π, that is,

$$\begin{aligned} R(\hat{\mu}, \pi) &= g_1\left(\frac{1/2}{\sigma_0^2 + 1/2}\right)^2 \sigma_0^2 + g_1\left(\frac{\sigma_0^2}{\sigma_0^2 + 1/2}\right)^2 \frac{1}{2} + g_2\sigma_0^2 \\ &= g_1 \frac{\sigma_0^2(1/2)}{\sigma_0^2 + 1/2} + g_2\sigma_0^2. \end{aligned} \tag{43}$$

We note that if $g_2 = 0$ and $g_1 = 1$, that is, we are only trying to estimate μ_1 with quadratic loss, then the Bayes risk and the estimator coincide with those in the classical model where we only observe a random variable $Y_Q \sim N(\mu_1, 1/2)$.

Case B. If $(2s^2)^{-2} < g_1/g_2 < (2s^2)^2$, then the Bayes estimator is given by a nonorthogonal resolution obtained from (35) by a linear change of variables. Note that $2s^2 = 2\sigma_0^2 + 1 > 1$, so if $g_1/g_2 = 1$ then this case necessarily obtains. Henceforth, we will focus on the case $g_1 = g_2 = 1$ and we state the "randomized" observables and the corresponding POVM where the optimal risk is obtained. Although the general case ($(2s^2)^{-2} < g_1/g_2 < (2s^2)^2$) is worked out in [21], Chapter III, Sections 5 and 6; the special case ($g_1 = g_2 = 1$) is explained in [20]. In addition to the observables, [20] describes the equivalent POVM which is optimal in the Bayes sense. The problem is discussed in Section 3 of [20]; by Proposition 3 there, the optimal POVM is given by (denoting $m = (x, y)$)

$$M_c(B) = \frac{1}{2\pi c^2} \int_B |\psi_{m/c}\rangle\langle\psi_{m/c}|\, dm, \qquad B \in \mathfrak{B}.$$

From (35) and a change of variables, it can be easily verified that

$$\frac{1}{2\pi c^2} \int_{\mathbb{R}^2} |\psi_{m/c}\rangle\langle\psi_{m/c}|\, dm = I,$$

and hence M_c is a resolution of identity (the other properties of POVM are also trivial to check). Here, $c = 2\sigma_0^2/(2\sigma_0^2 + 2\sigma^2 + 1)$ and in the special case $\sigma^2 = 1/2$, we have $c = \frac{\sigma_0^2}{\sigma_0^2 + 1}$. The minimal Bayes risk according to (17) in [20] thus becomes

$$\inf_M R(M, \pi) = \frac{2\sigma_0^2}{\sigma_0^2 + 1}. \tag{44}$$

The randomized measurements are then given by

$$\tilde{Q}_c = c(Q \otimes I_0 + I \otimes Q_0), \qquad \tilde{P}_c = c(P \otimes I_0 - I \otimes P_0). \tag{45}$$

The above expressions are given by equation (12) of [20] and the fact that these randomized measurements and M_c generate the same random variables follows by an argument analogous to the one given using characteristic functions in Appendix A in [26].

From (45), it is clear that $\tilde{X}_c = (X_{\tilde{Q}_c}, X_{\tilde{P}_c})$ satisfies $\tilde{X}_c = c\tilde{X}$ where $\tilde{X}$ is given in (39). As $c = \frac{\sigma_0^2}{\sigma_0^2+1} < 1$, we note that Bayes estimation in the quantum case exhibits the same shrinkage phenomenon as witnessed in the classical counterpart.

5. Minimax nonparametric estimation of a signal in quantum Gaussian white noise. In this section, we consider the estimation of parameters of the quantum Gaussian sequence model given in (26) under the ellipsoid restrictions (27) and (28). We will also consider the following variant of the previous model:

$$\bigotimes_{j=0}^{\infty} |\bar{G}(n^{1/2}\mu_j)\rangle\langle\bar{G}(n^{1/2}\mu_j)|. \tag{46}$$

This model, with the index starting from 0 instead of 1 and modified ellipsoids described below, arises from an approximate model, different from the one given in [11] (for a motivation see Lemma B.5 in [26]). In the Sobolev case, we will assume that the sequence $m = (\mu_0, \mu_1, \mu_2, \dots)$ is in an ellipsoid described as follows:

$$\Sigma_1(\beta, R) := \left\{ m : \sum_{j=0}^{\infty} j^{2\beta} \|\mu_j\|^2 \le R \right\} \tag{47}$$

(the above ellipsoid is related to the Hermite–Sobolev class) and in the exponential case we will assume that $m = (\mu_0, \mu_1, \mu_2, \dots)$ is in an ellipsoid (associated to the Hermite-exponential class) described by the following equation:

$$\mathcal{E}_1(\beta, r, R) := \left\{ m : \sum_{j=0}^{\infty} \exp(2\beta j^r) \|\mu_j\|^2 \le R \right\}. \tag{48}$$

In this case, we consider the following norm:

$$\|m\|^2 = \sum_{j=0}^{\infty} \|\mu_j\|^2.$$

5.1. *Upper asymptotic risk bound.* In this subsection, we consider the upper bound of minimax risk in estimating the parameter m over the ellipsoids $\Sigma(\beta, R)$, $\Sigma_1(\beta, R)$, $\mathcal{E}(\beta, r, R)$ and $\mathcal{E}_1(\beta, r, R)$. For each component state $|\bar{G}(n^{1/2}\mu_j)\rangle\langle\bar{G}(n^{1/2}\mu_j)|$, take a generalized observable $\tilde{X}_j$ constructed as $\tilde{X}$ in (39), and set $Y_j = n^{-1/2}\tilde{X}_j$. This gives a sequence of independent random vectors

$$Y_j \sim N_2(\mu_j, n^{-1} I_2), \quad j = 0, 1, 2, \dots, \tag{49}$$

which is a classical Gaussian sequence model, written with $\mathbb{R}^2$-valued Gaussian sequence elements Y_j. Then there is an estimator $\hat{m}_n$ such that we have the following theorem.

THEOREM 5.1. *(i)* Case A

Let

$$\Sigma(\beta, R) = \left\{ m = (\mu_j)_{j=1}^{\infty} : \sum_{j=1}^{\infty} \|\mu_j\|^2 j^{2\beta} \le R \right\}$$

and let $w_n(x) = x \wedge \log\log n$. *The following result holds*:

$$\limsup_{n\to\infty} \sup_{m\in\Sigma(\beta,R)} E_m\big[w_n\big(\eta_n(\beta, R)\|m - \hat{m}_n\|^2\big)\big] \le 1, \tag{50}$$

where

$$\eta_n(\beta, R) = (n/2)^{2\beta/(2\beta+1)} R^{-1/(2\beta+1)} c^{-1}(\beta),$$

$$c(\beta) = \left(\frac{\beta}{\beta+1}\right)^{2\beta/(2\beta+1)} (2\beta+1)^{1/(2\beta+1)}.$$

Case B

Let

$$\mathcal{E}(\beta, r, R) = \left\{ m = (\mu_j)_{j=1}^{\infty} : \sum_{j=1}^{\infty} \|\mu_j\|^2 \exp(2\beta j^r) \le R \right\}$$

and $v_n(x) = x \wedge (n^{1/2}/\log\log n)$. *The following result holds*:

$$\limsup_{n\to\infty} \sup_{m\in\mathcal{E}(\beta,r,R)} E_m\big[v_n\big(\tau_n(\beta, r)\|m - \hat{m}_n\|^2\big)\big] \le 1, \tag{51}$$

where

$$\tau_n(\beta, r) = \frac{n}{2}\left(\frac{\log n}{2\beta}\right)^{-1/r}.$$

The optimal estimator in the general form in terms of coefficients α_j *is given below* (*the estimators in the Sobolev case and the exponential case can be found by letting* α_j *be* $j^{2\beta}$ *or* $\exp(2\beta j^r)$, *resp.*). *The optimal estimator* $\hat{m}_n$ *is linear in* Y_j:

$$\hat{m}_{n,j} = \hat{\mu}_j = l_j Y_j, \quad l_j = \big(1 - \kappa(\alpha_j)^{1/2}\big)_+, \tag{52}$$

where $\kappa > 0$ *is found from the equation*,

$$\frac{(n/2)^{-1}}{\kappa} \sum_{j=1}^{\infty} \alpha_j^{1/2}\big(1 - \kappa(\alpha_j)^{1/2}\big)_+ = R. \tag{53}$$

(*ii*) *Relation* (50) *holds with* $\Sigma(\beta, R)$ *replaced by* $\Sigma_1(\beta, R)$ *with* $l_0 = 1$ *and* l_j *the same as in* (52) *for* $j \ge 1$. *Similarly*, *relation* (51) *holds with* $\mathcal{E}(\beta, r, R)$ *replaced by* $\mathcal{E}_1(\beta, r, R)$ *with* $l_0 = 1$ *and* l_j *the same as in* (52) *for* $j \ge 1$.

Note that we have used the first of the following equivalent expressions:

$$E_m\big[w_n\big(\eta_n(\beta, R)\|m - \hat{m}_n\|^2\big)\big] = \int w_n\big(\eta_n(\beta, R)\|m - u\|^2\big) \operatorname{Tr}\big(\rho_m M(du)\big).$$

Here, $\hat{m}_n$ was constructed from $Y_j = n^{-1/2}\tilde{X}_j$ and if M_j was the POVM which generates $\tilde{X}_j$ by acting on the component $|\bar{G}(n^{1/2}\mu_j)\rangle\langle\bar{G}(n^{1/2}\mu_j)|$, then we can easily associate M with $\bigotimes_{j=1}^{\infty} M_j$ (adjusting for the change in scale due to the multipliers l_j). Similarly, we use the equivalent notation with the expectation for the exponential case and (ii). Henceforth for the upper bounds, we will always use the $E_m(\cdot)$ notation keeping in mind that the associated POVM is implied. Also we will use η_n instead of $\eta_n(\beta, R)$ and τ_n instead of $\tau_n(\beta, R)$ for the rest of the paper. First, we state the main lemma (proved in Appendix C in [26]), which will help proving Theorem 5.1.

LEMMA 5.2. (*i*) Case A

Let $c(\beta)$ and $\hat{m}_n$ (for the Sobolev case) be as given in Theorem 5.1. As $n \to \infty$, we have

$$(n/2)^{2\beta/(2\beta+1)} \sup_{m\in\Sigma(\beta,R)} E_m \|\hat{m}_n - m\|^2 \le R^{1/(2\beta+1)} c(\beta)\big(1+o(1)\big). \tag{54}$$

Case B

Let $\hat{m}_n$ (for the exponential case) be as given in Theorem 5.1. As $n \to \infty$, we have

$$\left(\frac{n}{2}(\log n)^{-1/r}\right) \sup_{m\in\mathcal{E}(\beta,r,R)} E_m \|\hat{m}_n - m\|^2 \le (2\beta)^{-1/r}\big(1+o(1)\big) \text{ as } n \to \infty . \tag{55}$$

(*ii*) *Equation* (54) *holds with $\Sigma(\beta, R)$ replaced by $\Sigma_1(\beta, R)$ with $l_0 = 1$ and l_j the same as in* (52) *for $j \ge 1$. Similarly, equation* (55) *holds with $\mathcal{E}(\beta, r, R)$ replaced by $\mathcal{E}_1(\beta, r, R)$ with $l_0 = 1$ and l_j the same as in* (52) *for $j \ge 1$.*

PROOF OF THEOREM 5.1. To prove part (i), we note that due to Lemma 5.2,

$$\limsup_{n\to\infty} \sup_{m\in\Sigma(\beta,R)} E_m\big[\eta_n \|m - \hat{m}_n\|^2\big] \le 1.$$

Since the truncated loss is bounded above by the untruncated one, that is,

$$w_n\big(\eta_n\|m-\hat{m}_n\|^2\big) = \eta_n\|m-\hat{m}_n\|^2 \wedge \log\log n \le \eta_n\|m-\hat{m}_n\|^2,$$

the result follows.

Similarly, observing that the truncated loss is less than the quadratic loss and the exponential case of Lemma 5.2 (i.e., equation (55)) holds, one infers the result for the exponential case (i.e., equation (51)) for Theorem 5.1. Part (ii) of the theorem can be proved similarly using part (ii) of the above lemma. □

5.2. *Lower asymptotic risk bound.* The minimax lower bound can be obtained from the Bayes risk computed for a suitable prior.

THEOREM 5.3. Case A

With the sequence $\eta_n(\beta, R)$, the constant $c(\beta)$ and the function $w_n(x)$ as defined in Theorem 5.1, the following result holds:

$$\liminf_{n\to\infty} \inf_M \sup_{m\in\Sigma(\beta,R)} \int w_n\big(\eta_n(\beta, R)\|m-u\|^2\big) \operatorname{Tr}\big(\rho_m M(du)\big) \ge 1, \tag{56}$$

where the infimum is taken over all POVMs.

Case B

With the sequence $\tau_n(\beta, R)$ and the function $v_n(x)$ as defined in Theorem 5.1, the following result holds:

$$\liminf_{n\to\infty} \inf_M \sup_{m\in\mathcal{E}(\beta,r,R)} \int v_n\big(\tau_n(\beta, r)\|m-u\|^2\big) \operatorname{Tr}\big(\rho_m M(du)\big) \ge 1, \tag{57}$$

where the infimum is taken over all POVMs.

For the Sobolev case, let us set up a prior distribution π on μ_j in the quantum model (26) as independent random 2-vectors,

$$\mu_j \sim N_2\left(0, \frac{1}{2}\tilde{\theta}_j^2 I_2\right), \quad j = 1, 2, \dots, \tag{58}$$

where $\tilde{\theta}_j^2 = (n/2)^{-1}(c_n \wedge (\kappa_\epsilon \alpha_j^{1/2})^{-1} - 1)_+$, $c_n = \log\log\log n$ and also assuming that $\tilde{\theta}_j^2 = 0$ means μ_j is simply set 0. Here, κ_ϵ is the solution of the equation

$$\frac{(n/2)^{-1}}{\kappa_\epsilon} \sum_{j=1}^{\infty} \alpha_j^{1/2}(1 - \kappa_\epsilon \alpha_j^{1/2})_+ = R(1-\epsilon).$$

Define the a ball as follows:

$$B_n = \left\{ m : \sum_{j=1}^{\infty} \|\mu_j\|^2 \leq \delta_n(1+\epsilon_0) \right\},$$

where $\delta_n = E(\sum_{j=1}^{\infty} \|\mu_j\|^2)$ (this term can be shown to be $O(n^{-2\beta/(2\beta+1)} \log\log\log n)$). The following lemma implies that the above prior concentrates over the ball B_n and the ellipsoid $\Sigma(\beta, R)$.

LEMMA 5.4. *Consider the prior given in* (58). *Then the following concentration properties hold*:

$$P\big(m \notin \Sigma(\beta, R)\big) \to 0, \tag{59}$$

$$P(m \notin B_n) \to 0. \tag{60}$$

Some standard reasoning (see Appendix C in [26]) in connection with Pinsker's lower risk bound and using the previous lemma yields

LEMMA 5.5.

$$\begin{aligned}\inf_M \sup_{m\in\Sigma(\beta,R)} &\int w_n\big(\eta_n(\beta, R)\|m-u\|^2\big) \operatorname{Tr}\big(\rho_m M(du)\big) \\ &\geq \inf_M \int\int \eta_n(\beta, R)\|m-u\|^2 \operatorname{Tr}\big(\rho_m M(du)\big)\pi(dm) + o(1),\end{aligned}$$

where the infimum is taken over all POVMs.

We note that by Lemma 5.5, it is enough to show that

$$\lim_{\varepsilon\downarrow 0} \liminf_n \inf_M \int\int \eta_n(\beta, R)\|m-u\|^2 \operatorname{Tr}\big(\rho_m M(du)\big)\pi(dm) \geq 1 \tag{61}$$

to prove Theorem 5.3 for the Sobolev case.

The above lemmas are proved in Appendix C and the main theorem is proved in Appendix B in [26]. For the lower bound in the exponential case, we can set up a Gaussian prior similarly, but the asymptotic concentration properties like those in Lemma 5.4 do not hold. Instead we compute the Bayes risk by a different approach, similar to [7], to give a one-shot lower bound using the van Trees inequality. However, since we are in the quantum setup we need a suitable generalization of the inequality. Since the arguments involving the quantum van Trees inequality are rather technical, we defer the details to the main proof in Appendix B in [26].

6. Minimax nonparametric estimation of pure quantum states. Consider the framework of [11]. 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 $\mathcal{H}$. Motivated by physical principles and statistical methodology, we introduce the following *Hermite–Sobolev classes* of pure states characterized by an appropriate decay of the coefficients with respect to the basis B:

$$S^{\beta}(L) := \left\{ |\psi\rangle\langle\psi| : \sum_{j=0}^{\infty} |\psi_j|^2 j^{2\beta} \le L, \text{ and } \|\psi\| = 1 \right\}, \quad \beta > 0, L > 0. \tag{62}$$

To gain some intuition about the meaning of this class, let us assume that B is the Fock basis of a one-mode system. Then the constraint translates into a moment condition for the number operator $\langle\psi|N^{2\beta}|\psi\rangle \le L$; 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. We also consider the following *Hermite-exponential class*, which is *smoother* than the previous one. The relation of this class to other classes of exponential smoothness is discussed in Section 8 in the context of Wigner function estimation:

$$\begin{aligned} E^{\beta,r}(L) := \left\{ |\psi\rangle\langle\psi| : \sum_{j=0}^{\infty} |\psi_j|^2 \exp(2\beta j^r) \le L \text{ and } \|\psi\| = 1 \right\}, \\ \beta > 0, L > 0, r \in (0, 1]. \end{aligned} \tag{63}$$

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

$$\mathcal{Q}_n^S := \{ |\psi\rangle\langle\psi|^{\otimes n} : |\psi\rangle\langle\psi| \in S^{\beta}(L) \}. \tag{64}$$

Similarly, define the i.i.d. model when the pure state belongs to the exponential class,

$$\mathcal{Q}_n^E := \{ |\psi\rangle\langle\psi|^{\otimes n} : |\psi\rangle\langle\psi| \in E^{\beta,r}(L) \}. \tag{65}$$

We now introduce the corresponding quantum Gaussian models. 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 $\sqrt{n}\psi$. We define the coherent states models,

$$\mathcal{G}_n^S = \{ |G(\sqrt{n}\psi)\rangle\langle G(\sqrt{n}\psi)| : \psi \in \mathcal{H}, \text{ such that } |\psi\rangle\langle\psi| \in S^{\beta}(L) \}, \tag{66}$$

$$\mathcal{G}_n^E = \{ |G(\sqrt{n}\psi)\rangle\langle G(\sqrt{n}\psi)| : \psi \in \mathcal{H}, \text{ such that } |\psi\rangle\langle\psi| \in E^{\beta,r}(L) \}. \tag{67}$$

Using the factorization property (12) with respect to the orthonormal basis B, we see that the Gaussian states in both models defined above can be factored into a product of independent one-mode coherent Gaussian states of mean $\sqrt{n}\psi_i$,

$$|G(\sqrt{n}\psi)\rangle \cong \bigotimes_{i=0}^{\infty} |G(\sqrt{n}\psi_i)\rangle.$$

Note that the above model was discussed as a variant of the quantum Gaussian white noise model in (46) and the optimal estimation procedures established in Section 5 can be transferred to the i.i.d. model using the local asymptotic equivalence described below.

Let $|\psi_0\rangle$ be a fixed state in an infinite dimensional Hilbert space $\mathcal{H}$. We let $\mathcal{H}_0 := \{\psi \in \mathcal{H} : \langle\psi_0|\psi\rangle = 0\}$ denote the orthogonal complement of $\mathbb{C}\psi_0$. Any vector $\psi \in \mathcal{H}$ "close" to ψ_0 decomposes uniquely (see step 2 of the proof of Theorem 6.2 in Appendix B for further details) as

$$\psi = \psi_u := \sqrt{1 - \|u\|^2}\psi_0 + u, \quad u \in \mathcal{H}_0, \tag{68}$$

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 parametrized by vectors $u \in \mathcal{H}_0$.

In [11], $|\psi_0\rangle$ is arbitrary, but for the lower bound we select $|\psi_0\rangle$ as the basis element for index $j = 0$ in (62), that is, $|e_0\rangle$. For the upper bound, we will use the decomposition (68) but with $|\psi_0\rangle = |\hat{\psi}_1\rangle$—a preliminary estimator (see step 1 of the proof of Theorem 6.2 in Appendix B in [26] for further details). In addition to the i.i.d. and Gaussian models, we now introduce their local counterparts (without the Sobolev or exponential restrictions), which are parametrized by the local parameter u rather than by ψ. Let γ_n be a sequence such that $\gamma_n = o(1)$, and define the pure state models:

$$\mathcal{Q}_n(\psi_0, \gamma_n) := \{|\psi_u^{\otimes n}\rangle \in \mathcal{H}^{\otimes n} : u \in \mathcal{H}_0, \|u\| \le \gamma_n\}, \tag{69}$$

$$\mathcal{G}_n(\psi_0, \gamma_n) := \{|G(\sqrt{n}u)\rangle \in \mathcal{F}(\mathcal{H}_0) : u \in \mathcal{H}_0, \|u\| \le \gamma_n\}. \tag{70}$$

We state the LAE (Local Asymptotic Equivalence) theorem (proved in [11]) below, which 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 or exponential classes, rather it suffices that the local balls shrink at rate $\gamma_n = o(1)$.

THEOREM 6.1 (LAE). *Let $\mathcal{Q}_n(\psi_0, \gamma_n)$ and $\mathcal{G}_n(\psi_0, \gamma_n)$ be the models defined in* (69) *and* (70), *respectively, and $\gamma_n = o(1)$. Then the following convergence holds*:

$$\limsup_{n\to\infty} \sup_{|\psi_0\rangle\in\mathcal{H}} \Delta(\mathcal{Q}_n(\psi_0, \gamma_n), \mathcal{G}_n(\psi_0, \gamma_n)) = 0, \tag{71}$$

where $\Delta(\cdot, \cdot)$ is the quantum Le Cam distance defined in equation (18).

We will consider the rates of convergence for specific channels T_n and S_n where

$$\sup_{u:\|u\|\le\gamma_n} \|T_n(|\psi_u\rangle\langle\psi_u|^{\otimes n}) - |G(\sqrt{n}u)\rangle\langle G(\sqrt{n}u)|\|_1 \to 0,$$

$$\sup_{u:\|u\|\le\gamma_n} \|S_n(|G(\sqrt{n}u)\rangle\langle G(\sqrt{n}u)|) - |\psi_u\rangle\langle\psi_u|^{\otimes n}\|_1 \to 0.$$

In particular, we will show that the rates of convergence of the above channels will satisfy equation (19) with $c_n = \log\log n$ for the Sobolev case and $n^{1/2}/\log\log n$ for the exponential case and then we will use (22) to transfer risks between the two models.

6.1. *Upper asymptotic risk bound.* For the upper bound, we first split the sample and use a part of it to form a crude estimate $\hat{\rho}_1 = |\hat{\psi}_1\rangle\langle\hat{\psi}_1|$ of the pure state $\rho = |\psi\rangle\langle\psi|$ such that the estimate lies within a γ_n-neighborhood of ρ with high probability. Then we use local asymptotic equivalence to transfer this problem to the estimation of a Gaussian state. The final step is to transfer the bounds for the Gaussian state (using the estimator in Section 5) to the pure state model. We have the following theorem whose proof is given in Appendix B of [26].

THEOREM 6.2. *Consider the i.i.d. pure state model, that is, states are $\rho^{\otimes n}$ where $\rho = |\psi\rangle\langle\psi|$.*

Case A

Let $S^\beta(L)$ be the Hermite–Sobolev class defined in (62) *and let*

$$\tilde{\eta}_n = \frac{1}{2} n^{2\beta/(2\beta+1)} L^{-1/(2\beta+1)} c^{-1}(\beta). \tag{72}$$

Then there exists an estimator $|\hat{\psi}_n^{(S)}\rangle\langle\hat{\psi}_n^{(S)}|$ *of* $|\psi\rangle\langle\psi|$ *such that for all* $\beta > 1/2$ *and* $L > 0$ *the following holds*:

$$\limsup_{n\to\infty} \sup_{|\psi\rangle\langle\psi|\in S^\beta(L)} E_\psi\left[w_n\left(\frac{\tilde{\eta}_n}{2}\big\||\hat{\psi}_n^{(S)}\rangle\langle\hat{\psi}_n^{(S)}| - |\psi\rangle\langle\psi|\big\|_1^2\right)\right] \le 1. \tag{73}$$

Case B

Let $E^{\beta,r}(L)$ *be the Hermite-exponential class defined in* (63); *then there exists an estimator* $|\hat{\psi}_n^{(E)}\rangle\langle\hat{\psi}_n^{(E)}|$ *of* $|\psi\rangle\langle\psi|$ *such that for all* $\beta > 0$, $L > 1$ *and* $r \in (0, 1]$ *the following holds*:

$$\limsup_{n\to\infty} \sup_{|\psi\rangle\langle\psi|\in E^{\beta,r}(L)} E_\psi\left[v_n\left(\frac{\tau_n(\beta,r)}{2}\big\||\hat{\psi}_n^{(E)}\rangle\langle\hat{\psi}_n^{(E)}| - |\psi\rangle\langle\psi|\big\|_1^2\right)\right] \le 1. \tag{74}$$

Again in the upper bound, we have used the first of the following equivalent notation:

$$\begin{aligned} &E_\psi\left[w_n\left(\frac{\tilde{\eta}_n}{2}\big\||\hat{\psi}_n^{(S)}\rangle\langle\hat{\psi}_n^{(S)}| - |\psi\rangle\langle\psi|\big\|_1^2\right)\right] \\ &\quad = \int w_n\left(\frac{\tilde{\eta}_n(\beta,L)}{2}\big\|\rho - |\psi\rangle\langle\psi|\big\|_1^2\right)\operatorname{Tr}\left(|\psi\rangle\langle\psi|^{\otimes n} M'(d\rho)\right). \end{aligned}$$

M' being constructed from the POVM M (used for the Gaussian model) by applying the dual channel. Similarly, we use the condensed expectation notation in the Hermite-exponential class.

6.2. *Lower asymptotic risk bound.* For the lower bound, we restrict ψ to be of form (68) for $|u\rangle \in \mathcal{H}_0$, $|\psi_0\rangle = |e_0\rangle$ and $\|u\| \le \gamma_n = o(1)$. Then $\hat{\psi} = \psi_{\hat{u}}$ can be restricted similarly, without increasing the risk. By Lemma B.3 in [26],

$$\big\||\psi_{\hat{u}}\rangle\langle\psi_{\hat{u}}| - |\psi_u\rangle\langle\psi_u|\big\|_1^2 = 4\|\hat{u} - u\|^2 + O(\gamma_n^4). \tag{75}$$

Thus it suffices to obtain a lower bound for the estimation of the local parameter u. We already have a lower bound of the minimax risk (for the estimation of u) in the Gaussian model and we use the reverse channel S_n (defined in the proof of Theorem 4.1 in [12]) to give a lower bound to the minimax risk in the i.i.d. model. The lower bound is given in the following theorem whose proof is given in Appendix B in [26].

THEOREM 6.3. *Consider the i.i.d. pure state model, that is, states are* $\rho^{\otimes n}$ *where* $\rho = |\psi\rangle\langle\psi|$.

Case A

Let $S^\beta(L)$ *be the Hermite–Sobolev class defined in* (62) *and let* $\tilde{\eta}_n$ *as defined in equation* (72). *Then for all* $\beta > 0$ *and* $L > 0$ *the following holds*:

$$\liminf_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi|\in S^\beta(L)} \int w_n\left(\frac{\tilde{\eta}_n(\beta,L)}{2}\big\|\rho - |\psi\rangle\langle\psi|\big\|_1^2\right)\operatorname{Tr}\left(|\psi\rangle\langle\psi|^{\otimes n} M(d\rho)\right) \ge 1. \tag{76}$$

Case B

Let $E^{\beta,r}(L)$ *be the Hermite-exponential class defined in* (63). *Then for all* $\beta > 0$, $L > 1$ *and* $r \in (0, 1]$ *the following holds*:

$$\liminf_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi|\in E^{\beta,r}(L)} \int v_n\left(\frac{\tau_n(\beta,r)}{2}\big\|\rho - |\psi\rangle\langle\psi|\big\|_1^2\right)\operatorname{Tr}\left(|\psi\rangle\langle\psi|^{\otimes n} M(d\rho)\right) \ge 1. \tag{77}$$

7. Adaptive estimators. We note that neither of the estimators obtained in Theorem 5.1 and Theorem 6.2 are adaptive since the weights $l_j = (1-\kappa(\alpha_j)^{1/2})_+ = (1-\kappa j^\beta)_+$ are dependent on β and R (or L in the i.i.d. model). Similarly, for the exponential ellipsoid the weights depend on β, r, R (or L). We can modify our estimators so that they are adaptive with respect to β and R (or L) for the Sobolev case, and β, r and R (or L) for the exponential case. We use weakly geometric block Stein estimators (see [40]), which are known to be adaptive to the Gaussian sequence models. We will define a natural number $N_{\max}$ separately for the Sobolev and exponential cases.

Partition the set $\{1, 2, \dots, N_{\max}\}$ in J blocks, that is,

$$\bigcup_{j=1}^{J} B_j = \{1, 2, \dots, N_{\max}\}$$

such that $B_i \cap B_j = \varnothing$ for $i \neq j$ and $\min\{k : k \in B_j\} > \max\{k : k \in B_{j-1}\}$. Let $T_j = |B_j|$ and define blockwise weights:

$$\breve{\lambda}_i = \begin{cases} \left(1 - \dfrac{(n/2)^{-1} T_j}{\sum_{k\in B_j} \|Y_k\|^2}\right)_+ & \text{if } i \in B_j, \\ 0 & \text{if } i > N_{\max}. \end{cases} \tag{78}$$

The block Stein estimator (in this context recall that each component is an element of $\mathbb{R}^2$) is given by $\breve{\mu}_i = \breve{\lambda}_i Y_i$. We will use a particular structure of blocks called weakly geometric blocks.

Case A

Define $\epsilon = \frac{1}{\sqrt{n}}$, $\rho_\epsilon = (\log(1/\epsilon))^{-1}$ and $N_{\max} = n$ and consider the following block structure:

$$\begin{cases} \tilde{T}_1^{(S)} = \lceil \rho_\epsilon^{-1} \rceil, \\ \tilde{T}_j^{(S)} = \lfloor \tilde{T}_1^{(S)} (1+\rho_\epsilon)^{j-1} \rfloor & \text{for } j \in \{2, \dots J-1\}, \\ \tilde{T}_J^{(S)} = N_{\max} - \displaystyle\sum_{j=1}^{J-1} \tilde{T}_j^{(S)}, \end{cases} \tag{79}$$

where

$$J = \min\left\{ m : \tilde{T}_1^{(S)} + \sum_{j=2}^{m} \lfloor \tilde{T}_1^{(S)} (1+\rho_\epsilon)^{j-1} \rfloor \geq N_{\max} \right\}.$$

Consider $\breve{\lambda}_i$ given in (78) with T_j replaced by $\tilde{T}_j^{(S)}$ (in this special case write $\breve{\lambda}_i$ as $\breve{\lambda}_i^{(S)}$) and define $\breve{\mu}_j^{(S)} = \breve{\lambda}_j^{(S)} Y_j$, $\breve{m}_n^{(S)} = (\breve{\mu}_1^{(S)}, \breve{\mu}_2^{(S)}, \dots)$. In the Sobolev case, $\breve{m}_n^{(S)}$ will be our adaptive estimator for m.

Case B

In the exponential case, we take $N_{\max} = \lfloor (\log n)^{1/r_0} \rfloor$ and $\epsilon' = (\log n)^{-1/2r_0}$. Define $\rho_{\epsilon'}$ (as in the Sobolev case) and the blocks $\tilde{T}_j^{(E)}$ (using $\rho_{\epsilon'}$ instead of ρ_ϵ).

Consider $\breve{\lambda}_i$ given in (78) with T_j replaced by $\tilde{T}_j^{(E)}$ (in this special case write $\breve{\lambda}_i$ as $\breve{\lambda}_i^{(E)}$) and define $\breve{\mu}_j^{(E)} = \breve{\lambda}_j^{(E)} Y_j$, $\breve{m}_n^{(E)} = (\breve{\mu}_1^{(E)}, \breve{\mu}_2^{(E)}, \dots)$. In the exponential case, $\breve{m}_n^{(E)}$ will be our adaptive estimator for m.

We have the following theorems which are adaptive versions of Theorem 5.1 and Theorem 6.2 (the proofs are deferred to Appendix B in [26]).

THEOREM 7.1. (*i*) Case A

Consider the estimator $\breve{m}_n^{(S)}$. *Then for any* $R > 0$, $\beta > 0$, *the following holds*:

(80) $$\limsup_{n\to\infty} \sup_{m\in\Sigma(\beta,R)} E_m w_n\big(\eta_n \|m - \breve{m}_n^{(S)}\|^2\big) \le 1.$$

Case B

Consider the estimator $\breve{m}_n^{(E)}$. *Then for any* $R > 0$, $\beta > 0$ *and* $0 < r_0 < r \le 1$, *the following holds*:

(81) $$\limsup_{n\to\infty} \sup_{m\in\mathcal{E}(\beta,r,R)} E_m v_n\big(\tau_n \|m - \breve{m}_n^{(E)}\|^2\big) \le 1.$$

(*ii*) *Equation* (80) *holds with* $\Sigma(\beta, R)$ *replaced by* $\Sigma_1(\beta, R)$ *with* $\breve{\lambda}_0^{(S)} = 1$ *and* $\breve{\lambda}_j^{(S)}$ *as in* (79) *for* $j \ge 1$. *Equation* (81) *holds with* $\mathcal{E}(\beta, r, R)$ *replaced by* $\mathcal{E}_1(\beta, r, R)$ *with* $\breve{\lambda}_0^{(E)} = 1$ *and* $\breve{\lambda}_j^{(E)}$ *as in* (78) (*specialized to the exponential case*) *for* $j \ge 1$.

THEOREM 7.2. Case A

Let $S^\beta(L)$ *be the Hermite–Sobolev class defined in* (31) *and let* $\beta_0 > 1/2$. *Then there exists an estimator* $|\breve{\psi}_n^{(S)}\rangle\langle\breve{\psi}_n^{(S)}|$ (*not depending on* β *and* L) *for which the following upper bound holds*:

(82) $$\limsup_{n\to\infty} \sup_{|\psi\rangle\langle\psi|\in S^\beta(L)} E_\psi\left[w_n\left(\frac{\tilde{\eta}_n}{2}\big\||\breve{\psi}_n^{(S)}\rangle\langle\breve{\psi}_n^{(S)}| - |\psi\rangle\langle\psi|\big\|_1^2\right)\right] \le 1$$

whenever $\beta > \beta_0$ *and* $L > 0$.

Case B

Let $E^{\beta,r}(L)$ *be the Hermite-exponential class defined in* (32) *and let* $\beta_0 > 0$ *and* $0 < r_0 < 1$. *Then there exists an estimator* $|\breve{\psi}_n^{E}\rangle\langle\breve{\psi}_n^{(E)}|$ (*not depending on* β, r *and* L) *for which the following upper bound holds*:

(83) $$\limsup_{n\to\infty} \sup_{|\psi\rangle\langle\psi|\in E^{\beta,r}(L)} E_\psi\left[v_n\left(\frac{\tau_n}{2}\big\||\breve{\psi}_n^{(E)}\rangle\langle\breve{\psi}_n^{(E)}| - |\psi\rangle\langle\psi|\big\|_1^2\right)\right] \le 1$$

whenever $\beta > \beta_0$, $L > 1$ *and* $r_0 < r \le 1$.

8. Wigner function estimation. For estimation of pure states, we assumed that the coefficients belong to the Sobolev ellipsoid, that is, if $|\psi\rangle\langle\psi|$ is the pure state and $|\psi\rangle = \sum_{i=0}^\infty \psi_i|e_i\rangle$ is the expansion with respect to the Hermite basis, then

$$\sum_j |\psi_j|^2 j^{2\beta} \le L.$$

Henceforth for this section, we will assume that the indices for the summation run from 0 to ∞ unless mentioned otherwise. Recall (5) by which we have $\|\rho_1 - \rho_2\|_1^2 = 4(1 - |\langle\psi_1|\psi_2\rangle|^2)$. Similarly, it can be verified that $\|\rho_1 - \rho_2\|_2^2 = 2(1 - |\langle\psi_1|\psi_2\rangle|^2)$ where $\|A\|_2^2 = \mathrm{Tr}(A^*A)$. On the other hand, by the overlap formula (6) we have

$$\|\rho_1 - \rho_2\|_2^2 = 2\pi\|W_{\rho_1} - W_{\rho_2}\|^2.$$

Thus we have the following identity for pure states ρ_1 and ρ_2:

(84) $$\|W_{\rho_1} - W_{\rho_2}\|^2 = \frac{1}{4\pi}\|\rho_1 - \rho_2\|_1^2.$$

Since Theorem 3.3 of the paper provides a scheme for sharp and adaptive minimax estimation of a pure state $\rho = |\psi\rangle\langle\psi|$ in the trace norm, we can use equation (84) to develop a sharp and adaptive minimax estimator for the Wigner function of pure states. To translate the smoothness condition on the pure states to a property of the Wigner function of a pure state, first note that

$$\rho = |\psi\rangle\langle\psi| = \sum_{j,k} \psi_j \bar{\psi}_k |e_j\rangle\langle e_k|.$$

By linearity of the quantum and ordinary Fourier transforms, we have

$$W_\rho = \sum_{j,k} \psi_j \bar{\psi}_k W_{jk}, \tag{85}$$

where

$$\begin{aligned} W_{ij}(q,p) &= \frac{1}{(2\pi)^2} \int\int \exp(-iuq - ivp) \operatorname{Tr}\big(\exp(iuQ + ivP)|e_i\rangle\langle e_j|\big)\, du\, dv \\ &= \frac{1}{(2\pi)^2} \int\int \exp(-iuq - ivp) \langle e_j| \exp(iuQ + ivP)|e_i\rangle\, du\, dv. \end{aligned}$$

It can be easily verified that $\bar{W}_{ij}(q,p) = W_{ji}(q,p)$. Since $|e_i\rangle$ is an orthonormal basis, an application of the overlap formula (it also holds for trace-class operators as shown in [28]) gives

$$2\pi \int\int \bar{W}_{ij}(q,p) W_{kl}(q,p)\, dq\, dp = \langle e_l|e_j\rangle\langle e_i|e_k\rangle = \delta_{jl}\delta_{ik},$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Thus $\{\sqrt{2\pi} W_{ij}\}$ is an orthonormal set of functions. Let $\mathcal{H}_W$ be the Hilbert space spanned by these functions. From (85), it can be seen that if W_ρ is the Wigner function of a pure state then $W_\rho \in \mathcal{H}_W$. By linearity, it can also be established that $W_\rho \in \mathcal{H}_W$ for any mixed state ρ. For a general state, we write the expansion as

$$W_\rho = \sum_{j,k} \rho_{jk} W_{jk}. \tag{86}$$

Define the following classes of Wigner functions:

$$S_W^\beta(L) = \left\{ W_\rho : \sum_{j,k} |\rho_{jk}|^2 j^{2\beta} k^{2\beta} \le L, \text{ where } W_\rho = \sum_{j,k} \rho_{jk} W_{jk} \right\}, \tag{87}$$

$$E_W^{\beta,r}(L) = \left\{ W_\rho : \sum_{j,k} |\rho_{jk}|^2 e^{2\beta(j^r + k^r)} \le L, \text{ where } W_\rho = \sum_{j,k} \rho_{jk} W_{jk} \right\}. \tag{88}$$

Observe that

$$\sum_{j,k} |\psi_j \bar{\psi}_k|^2 j^{2\beta} k^{2\beta} = \left(\sum_j |\psi_j|^2 j^{2\beta} \right)^2$$

and hence for a pure state $|\psi\rangle\langle\psi|$ we have

$$|\psi\rangle\langle\psi| \in S^\beta(L) \quad \text{iff} \quad W_{|\psi\rangle\langle\psi|} \in S_W^\beta(L^2).$$

Similarly,

$$|\psi\rangle\langle\psi| \in E^{\beta,r}(L) \quad \text{iff} \quad W_{|\psi\rangle\langle\psi|} \in E_W^{\beta,r}(L^2).$$

In order to facilitate comparison with other results in the literature, the theorem to follow is stated in condensed notation.

THEOREM 8.1. Hermite–Sobolev Wigner class

Let ρ be a pure state and $W_\rho \in S_W^\beta(L^2)$. Then the following result holds:

$$\lim_{n\to\infty} \inf_{W_{|\hat\psi\rangle\langle\hat\psi|}} \sup_{W_{|\psi\rangle\langle\psi|}\in S_W^\beta(L^2)} E_\psi\big[w_n\big(2\pi\tilde\eta_n\|W_{|\psi\rangle\langle\psi|}-W_{|\hat\psi\rangle\langle\hat\psi|}\|^2\big)\big]=1. \tag{89}$$

For every $\beta_0 > 1/2$, there exists an estimator, not depending on β and L, attaining this bound asymptotically, whenever $\beta > \beta_0$ and $L > 0$.

Hermite-exponential Wigner class

Let ρ be a pure state and $W_\rho \in E_W^{\beta,r}(L^2)$. Then the following result holds:

$$\lim_{n\to\infty} \inf_{W_{|\hat\psi\rangle\langle\hat\psi|}} \sup_{W_{|\psi\rangle\langle\psi|}\in E_W^{\beta,r}(L^2)} E_\psi\big[v_n\big(2\pi\tau_n\|W_{|\psi\rangle\langle\psi|}-W_{|\hat\psi\rangle\langle\hat\psi|}\|^2\big)\big]=1. \tag{90}$$

For every $0 < r_0 < 1$ and $\beta_0 > 0$, there exists an estimator, not depending on β, R and r, attaining this bound asymptotically, whenever $\beta > \beta_0$, $L > 1$ and $r_0 < r \le 1$.

In POVM notation, the statements of the theorems are as follows:

$$\lim_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi|\in S^\beta(L)} \int w_n\big(2\pi\tilde\eta_n\|W_{|\psi\rangle\langle\psi|}-W_\rho\|^2\big)\operatorname{Tr}\big(|\psi\rangle\langle\psi|^{\otimes n}M(d\rho)\big)=1. \tag{91}$$

$$\lim_{n\to\infty} \inf_M \sup_{|\psi\rangle\langle\psi|\in E^{\beta,r}(L)} \int v_n\big(2\pi\tau_n\|W_{|\psi\rangle\langle\psi|}-W_\rho\|^2\big)\operatorname{Tr}\big(|\psi\rangle\langle\psi|^{\otimes n}M(d\rho)\big)=1. \tag{92}$$

Here, the range and domain of the POVMs are same as in Theorem 3.3.

Comparison with other classes

In the literature, various authors consider estimation of Wigner functions over different smoothness classes. Only classes of infinitely differentiable functions are considered; we compare these to the exponential classes $E_W^{\beta,r}(L)$ defined in (88). In [3], the authors consider the following class:

$$\mathcal{R}(B,r)=\{\rho:|\rho_{jk}|\le\exp(-B(j+k)^{r/2})\}$$

and a related class has been considered in [1]:

$$\mathcal{R}(C,B,r)=\{\rho:|\rho_{jk}|\le C\exp(-B(j+k)^{r/2})\},$$

where $0 < r \le 2$. Replacing r with $2r$, we have $0 < r \le 1$ and since $2(j+k)^r > j^r + k^r$ we have

$$\sum_{j,k}|\rho_{jk}|^2e^{2\beta(j^r+k^r)}\le C\sum_{j,k}e^{-2B(j+k)^r}e^{2\beta(j^r+k^r)}\le C\sum_{j,k}e^{-(B-2\beta)(j^r+k^r)}.$$

For $B > 2\beta$, the above sum is finite and in that case

$$\mathcal{R}_W(C,B,r)\subset E_W^{\beta,r}(L)$$

for some $L > 0$ and $0 < r \le 1$, where

$$\mathcal{R}_W(C,B,r)=\{W_\rho:|\rho_{jk}|\le C\exp(-B(j+k)^r)\}.$$

Yet another class has been considered in [10] and [30]:

$$A(\beta,r,L)=\left\{W_\rho:\int\big|\tilde W_\rho(w)\big|^2e^{2\beta\|w\|^r}\,dw\le(2\pi)^2L^2\right\},$$

where $\tilde{W}$ is the Fourier transform of the Wigner function. In [10], it is mentioned (first paragraph page 13) that if $\mathrm{Tr}[\rho e^{aN^{r/2}}] < \infty$, then $W_\rho \in A(\beta, r, L)$ for some $\beta > 0$, $L > 0$. Here, $N = \frac{1}{2}(Q^2 + P^2 - 1)$ is the number operator. Before verifying this claim, we note that by replacing r with $2r$ for pure states, we have

$$\mathrm{Tr}\big[\rho e^{aN^r}\big] = \langle\psi| e^{aN^r} |\psi\rangle = \sum_j |\psi_j|^2 e^{aj^r}.$$

The last equality follows by expanding $|\psi\rangle$ in the Hermite basis $\{|e_j\rangle\}$ and noting that the $|e_j\rangle$ form an eigenbasis of the number operator N. But the condition that the RHS of the last equation is appropriately bounded is exactly the description of the class $E^{\beta,r}(L)$ for $a = 2\beta$. For a general state (mixed or pure) ρ, the condition $\mathrm{Tr}[\rho e^{aN^r}] \le L$ implies $W_\rho \in E_W^{\beta,r}(L^2)$ for $a = 2\beta$. We defer the proof of this statement to Appendix A in [26]. We also have the following lemma (also proved in Appendix A).

LEMMA 8.2.

$$E_W^{\beta,r}(L) \subset \tilde{A}(\beta_1, r, L_1),$$

where

$$\tilde{A}(\beta_1, r, L_1) = \left\{ W_\rho : \int \big|\tilde{W}_\rho(w)\big|^2 e^{2\beta_1 \|w\|^{2r}}\, dw \le (2\pi)^2 L_1^2 \right\}$$

for some $\beta_1 > 0$, $L_1 > 0$. *Note that in our case* $0 < r \le 1$.

We compare our result to those in [1, 3] and [30] (excluding [10] since it only considers pointwise estimation). We only consider the noiseless case ($\eta = 1$ or $\gamma = 0$) since our measurements did not include additional Gaussian noise like assumed in these papers. Theorem 1 in [3] describes the L^2 risk for $\eta = 1$ and the rate obtained is $\varphi_n^2 = n^{-1}(\log n)^{\frac{17}{3r}}$ or replacing r by $2r$ (to compare with our results) $\varphi_n^2 = n^{-1}(\log n)^{\frac{17}{6r}}$. We note that apart from sharp minimaxity our rate is also faster, that is, $n^{-1}(\log n))^{\frac{1}{r}}$ which should be expected since we are only maximizing over the smaller class of pure states. Theorem 4.1 of [3] also yields similar slow rates with a kernel density estimator. The statements of [1] are probabilistic in nature. Theorem 2.1 there (which directly computes the L^2 risk of estimating the states instead of estimating the Wigner function) states that with probability at least $1 - \epsilon$,

$$\|\rho - \tilde{\rho}^n\|_2^2 \le C_1 n^{-1}(\log n)^{\frac{20}{3r}}(\log\log n)\epsilon^{-1}).$$

Again if we replace r with $2r$ we obtain a slower rate compared to ours. Furthermore, we observe that while [1] and [3] only gives upper bounds for L^2 risk over the class $\mathcal{R}(B, r)$ or $\mathcal{R}(C, B, r)$, [30] gives a lower bound for the L^2 risk. Setting $\gamma = 0$ in Theorem 2 of [30] one obtains that the lower bound is of the order n^{-1}. Our result shows that, when restricted to a similar class (i.e., $E_W^{\beta,r}(L)$) and pure states, the logarithmic correction is unavoidable, which is of course also suggested by basic results about exponential ellipsoids [7, 38].

Finally, we stress again that our lower bound is over all quantum measurements, whereas the lower bound mentioned in [30] assumes that the measurement is quantum homodyne tomography. Risk bounds for quantum state estimation based on other specific measurement schemes (e.g., Pauli measurements) are obtained in [13, 24, 25].

Funding. Supported in part by NSF Grants DMS-1407600 and DMS-1915884.

SUPPLEMENTARY MATERIAL

Supplement to "Minimax nonparametric estimation of pure quantum states" (DOI: 10.1214/21-AOS2115SUPP; .pdf). Some details on the covariant measurement and the random variable generated by it are discussed in Appendix A.1. A comparison with the classical Gaussian sequence model can be found in Appendix A.2., while Appendix A.3 contains a discussion of Wigner function classes. Proofs of the main theorems are presented in Appendix B while those of some technical lemmas are given in Appendix C.

REFERENCES

[1] ALQUIER, P., MEZIANI, K. and PEYRÉ, G. (2013). Adaptive estimation of the density matrix in quantum homodyne tomography with noisy data. *Inverse Probl*. **29** 075017, 20. MR3080478 https://doi.org/10.1088/0266-5611/29/7/075017

[2] ARTILES, L. M., GILL, R. D. and GUŢĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **67** 109–134. MR2136642 https://doi.org/10.1111/j.1467-9868.2005.00491.x

[3] AUBRY, J.-M., BUTUCEA, C. and MEZIANI, K. (2009). State estimation in quantum homodyne tomography with noisy data. *Inverse Probl*. **25** 015003, 22. MR2465335 https://doi.org/10.1088/0266-5611/25/1/015003

[4] 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.

[5] AUDENAERT, K. M. R., NUSSBAUM, M., SZKOŁA, A. and VERSTRAETE, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys*. **279** 251–283. MR2377635 https://doi.org/10.1007/s00220-008-0417-5

[6] BARNDORFF-NIELSEN, O. E., GILL, R. D. and JUPP, P. E. (2003). On quantum statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **65** 775–816. With discussion and reply by the authors. MR2017871 https://doi.org/10.1111/1467-9868.00415

[7] BELITSER, E. N. and LEVIT, B. Y. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist*. **4** 259–273. MR1355248

[8] BRAUNSTEIN, S. L. and CAVES, C. M. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett*. **72** 3439–3443. MR1274631 https://doi.org/10.1103/PhysRevLett.72.3439

[9] BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist*. **24** 2384–2398. MR1425958 https://doi.org/10.1214/aos/1032181159

[10] BUTUCEA, C., GUŢĂ, 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 https://doi.org/10.1214/009053606000001488

[11] BUTUCEA, C., GUŢĂ, M. and NUSSBAUM, M. (2018). Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise. *Ann. Statist*. **46** 3676–3706. MR3852665 https://doi.org/10.1214/17-AOS1672

[12] BUTUCEA, C., GUŢĂ, M. and NUSSBAUM, M. (2018). Supplementary material to "Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise". *Ann. Statist*. **46**. MR3852665 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-AOS1382

[14] GILL, R. D. and GUŢĂ, 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-IMSCOLL909

[15] GUŢĂ, M. and JENČOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys*. **276** 341–379. MR2346393 https://doi.org/10.1007/s00220-007-0340-1

[16] GUŢĂ, M. and KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* (3) **73** 052108, 15. MR2229156 https://doi.org/10.1103/PhysRevA.73.052108

[17] HAYASHI, M. (2017). *Quantum Information Theory: Mathematical Foundation*, 2nd ed. *Graduate Texts in Physics*. Springer, Berlin. MR3558531 https://doi.org/10.1007/978-3-662-49725-8

[18] HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.

[19] HOLEVO, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. *Quaderni/Monographs* **1**. Edizioni della Normale, Pisa. With a foreword from the second Russian edition by K. A. Valiev. MR2797301 https://doi.org/10.1007/978-88-7642-378-9

[20] HOLEVO, A. S. (1973). Optimal quantum measurements. *Theoret. Math. Phys.* **17** 1172–1177. MR0468885 https://doi.org/10.1007/BF01037594

[21] HOLEVO, A. S. (1978). *Investigations in the General Theory of Statistical Decisions*. Amer. Math. Soc., Providence, RI. Translated from the Russian by Lisa Rosenblatt, Proc. Steklov Inst. Math., **1**978, no. 3. MR0652516

[22] HOLEVO, A. S. (2001). *Statistical Structure of Quantum Theory*. *Lecture Notes in Physics*. *Monographs* **67**. Springer, Berlin. MR1889193 https://doi.org/10.1007/3-540-44998-1

[23] KAHN, J. and GUŢĂ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764 https://doi.org/10.1007/s00220-009-0787-3

[24] KOLTCHINSKII, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. MR3012397 https://doi.org/10.1214/11-AOS926

[25] KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798

[26] LAHIRY, S. and NUSSBAUM, M. (2022). Supplement to "Minimax nonparametric estimation of pure quantum states." https://doi.org/10.1214/21-AOS2115SUPP

[27] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[28] LEONHARDT, U. (1997). *Measuring the Quantum State of Light*. Cambridge Univ. Press.

[29] LI, K. (2016). Discriminating quantum states: The multiple Chernoff distance. *Ann. Statist.* **44** 1661–1679. MR3519936 https://doi.org/10.1214/16-AOS1436

[30] LOUNICI, K., MEZIANI, K. and PEYRÉ, G. (2018). Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography. *Ann. Statist.* **46** 1318–1351. MR3798005 https://doi.org/10.1214/17-AOS1586

[31] MÉZIANI, K. (2007). Nonparametric estimation of the purity of a quantum state in quantum homodyne tomography with noisy data. *Math. Methods Statist.* **16** 354–368. MR2378280 https://doi.org/10.3103/S1066530707040047

[32] MEZIANI, K. (2008). Nonparametric goodness-of fit testing in quantum homodyne tomography with noisy data. *Electron. J. Stat.* **2** 1195–1223. MR2461899 https://doi.org/10.1214/08-EJS286

[33] NAULET, Z. and BARAT, É. (2017). Bayesian nonparametric estimation for quantum homodyne tomography. *Electron. J. Stat.* **11** 3595–3632. MR3709864 https://doi.org/10.1214/17-EJS1322

[34] NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959 https://doi.org/10.1214/aos/1032181160

[35] NUSSBAUM, M. (1999). Minimax risk: Pinsker bound. In *Encyclopedia of Statistical Sciences*, *Update Volume* **3** 451–460. Wiley, New York. MR1679440 https://doi.org/10.1002/0471667196.ess1098.pub2

[36] NUSSBAUM, M. and SZKOŁA, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. MR2502660 https://doi.org/10.1214/08-AOS593

[37] PARTHASARATHY, K. R. (1992). *An Introduction to Quantum Stochastic Calculus*. *Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. [2012 reprint of the 1992 original] [MR1164866]. MR3012668

[38] PINSKER, M. S.(1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** no. 2, 120-133; translated from *Problemy Peredachi Informatsii* **16** (1980), no. 2, 52-68 (Russian). MR0624591

[39] TSYBAKOV, A. B. (1997). Asymptotically efficient estimation of a signal in L_2 under general loss functions. *Problems Inform. Transmission* **33** no. 1, 78-88; translated from *Problemy Peredachi Informatsii* **33** (1997), no. 1, 94-106 (Russian). MR1441532

[40] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[41] WANG, Y. and SONG, X. (2020). Quantum science and quantum technology. *Statist. Sci.* **35** 51–74. MR4071358 https://doi.org/10.1214/19-STS745

[42] YAMAGATA, K., FUJIWARA, A. and GILL, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** 2197–2217. MR3127863 https://doi.org/10.1214/13-AOS1147