On asymptotic quantum statistical inference

Richard D. Gill[1],[*] and Mădălin I. Guţă[2],[†]

Leiden University, University of Nottingham

Abstract: We study asymptotically optimal statistical inference concerning the unknown state of N identical quantum systems, using two complementary approaches: a "poor man's approach" based on the van Trees inequality, and a rather more sophisticated approach using the recently developed quantum form of LeCam's theory of Local Asymptotic Normality.

1. Introduction

The aim of this paper is to show the rich possibilities for asymptotically optimal statistical inference for "quantum i.i.d. models". Despite the possibly exotic context, mathematical statistics has much to offer, and much that we have learned—in particular through Jon Wellner's work in semiparametric models and nonparametric maximum likelihood estimation—can be put to extremely good use. Exotic? In today's quantum information engineering, measurement and estimation schemes are put to work to recover the state of a small number of quantum states, engineered by the physicist in his or her laboratory. New technologies are winking at us on the horizon. So far, the physicists are largely re-inventing statistical wheels themselves. We think it is a pity statisticians are not more involved. If Jon is looking for some new challenges... ?

In this paper we do theory. We suppose that one has N copies of a quantum system each in the same state depending on an unknown vector of parameters θ, and one wishes to estimate θ, or more generally a vector function of the parameters $\psi(\theta)$, by making some measurement on the N systems together. This yields data whose distribution depends on θ and on the choice of the measurement. Given the measurement, we therefore have a classical parametric statistical model, though not necessarily an i.i.d. model, since we are allowed to bring the N systems together before measuring the resulting joint system as one quantum object. In that case the resulting data need not consist of (a function of) N i.i.d. observations, and a key quantum feature is that we can generally extract more information about θ using such "collective" or "joint" measurements than when we measure the systems

separately. What is the best we can do as $N \to \infty$, when we are allowed to optimize both over the measurement and over the ensuing data-processing?

A statistically motivated, approach to deriving methods with good properties for large N is to choose the measurement to optimize the Fisher information in the data, leaving it to the statistician to process the data efficiently, using for instance maximum likelihood or related methods, including Bayesian. This heuristic principle has already been shown to work in a number of special cases in quantum statistics. Since the measurement maximizing the Fisher information typically depends on the unknown parameter value this often has to be implemented in a two-step approach, first using a small fraction of the N systems to get a first approximation to the true parameter, and then optimizing on the remaining systems using this rough guess.

The approach favoured by many physicists, on the other hand, is to choose a prior distribution and loss function on grounds of symmetry and physical interpretation, and then to *exactly* optimize the Bayes risk over all measurements and estimators, for any given N. This approach succeeds in producing attractive methods on those rare occasions when a felicitous combination of all the mathematical ingredients leads to an analytically tractable solution.

Now it has been observed in a number of problems that the two approaches result in asymptotically equivalent estimators, though the measurement schemes can be strikingly different. Heuristically, this can be understood to follow from the fact that, in the physicists' approach, for large N the prior distribution should become increasingly irrelevant and the Bayes optimal estimator close to the maximum likelihood estimator. Moreover, we expect those estimators to be asymptotically normal with variances corresponding to inverse Fisher information.

Here we link the two approaches by deriving an asymptotic lower bound on the Bayes risk of the physicists' approach, in terms of the optimal Fisher information of the statisticians' approach. Sometimes one can find in this way asymptotically optimal solutions which are much easier to implement than the exactly optimal solution of the physicists' approach. On the other hand, it also suggests that the physicists' approach, when successful, leads to procedures which are *asymptotically* optimal for other prior distributions, and other loss functions, than those used in the computation. It also suggests that these solutions are asymptotically optimal in a pointwise rather than a Bayesian sense.

In the first part of our paper, we derive our new bound by combining an existing quantum Cramér-Rao bound [22] with the van Trees inequality, a Bayesian Cramér-Rao bound from classical statistics [7, 29]. The former can be interpreted as a bound on the Fisher information in an arbitrary measurement on a quantum system, the latter is a bound on the Bayes risk (for a quadratic loss function) in terms of the Fisher information in the data. This part of the paper can be understood without any familiarity with quantum statistics. Applications are given in an appendix to an eprint version of the paper at arXiv.org.

The paper contains only a brief summary of "what is a quantum statistical model"; for more information the reader is referred to the papers of Barndorff-Nielsen et al. [1], and Gill [4]. For an overview of the "state of the art" in quantum asymptotic statistics see Hayashi [19] which reprints papers of many authors together with introductions by the editor.

After this "simplistic" part of the paper we present some of the recently developed theory of quantum Local Asymptotic Normality (also mentioning a number of open problems). This provides an alternative but more sophisticated route to getting asymptotic optimality results, but at the end of the day it also explains "why"

our simplistic approach does indeed work. In classical statistics, we have learnt to understand asymptotic optimality of maximum likelihood estimation through the idea that an i.i.d. parametric model can be closely approximated, locally, by a Gaussian shift model with the same information matrix. To say the same thing in a deeper way, the two models have the same geometric structure of the score functions of one-dimensional sub-models; and in the i.i.d. case, after local rescaling, those score functions are asymptotically Gaussian.

Let us first develop enough notation to state the main result of the paper and compare it with the comparable result from classical statistics. Starting on familiar ground with the latter, suppose we want to estimate a function $\psi(\theta)$ of a parameter θ, both represented by real column vectors of possibly different dimension, based on N i.i.d. observations from a distribution with Fisher information matrix $I(\theta)$. Let π be a prior density on the parameter space and let $\widetilde{G}(\theta)$ be a symmetric positive-definite matrix defining a quadratic loss function $l(\widehat{\psi}^{(N)}, \theta) = (\widehat{\psi}^{(N)} - \psi(\theta))^\top \widetilde{G}(\theta)(\widehat{\psi}^{(N)} - \psi(\theta))$. (Later we will use $G(\theta)$, without the tilde, in the special case when ψ is θ itself). Define the mean square error matrix $V^{(N)}(\theta) = \mathbb{E}_\theta(\widehat{\psi}^{(N)} - \psi(\theta))(\widehat{\psi}^{(N)} - \psi(\theta))^\top$ so that the risk can be written $R^{(N)}(\theta) = \operatorname{trace} \widetilde{G}(\theta)V^{(N)}(\theta)$. The Bayes risk is $R^{(N)}(\pi) = \mathbb{E}_\pi \operatorname{trace} \widetilde{G}V^{(N)}$. Here, $\mathbb{E}_\theta$ denotes expectation over the data for given θ, $\mathbb{E}_\pi$ denotes averaging over θ with respect to the prior π. The estimator $\widehat{\psi}^{(N)}$ is completely arbitrary. We assume the prior density to be smooth, compactly supported and zero on the smooth boundary of its support. Furthermore a certain quantity roughly interpreted as "information in the prior" must be finite. Then it is very easy to show [7], using the van Trees inequality, that under minimal smoothness conditions on the statistical model,

$$
(1) \qquad \liminf_{N \to \infty} N R^{(N)}(\pi) \geq \mathbb{E}_\pi \operatorname{trace} G I^{-1},
$$

where $G = \psi' \widetilde{G} \psi'^\top$ and ψ' is the matrix of partial derivatives of elements of ψ with respect to those of θ.

Now in quantum statistics the data depends on the choice of measurement and the measurement should be tuned to the loss function. Given a measurement $M^{(N)}$ on N copies of the quantum system, denote by $\overline{I}_M^{(N)}$ the average Fisher information (i.e., Fisher information divided by N) in the data. The Holevo [22] quantum Cramér-Rao bound, as extended by Hayashi and Matsumoto [20] to the quantum i.i.d. model, can be expressed as saying that, for all θ, G, N and $M^{(N)}$,

$$
(2) \qquad \operatorname{trace} G(\theta)\big(\overline{I}_M^{(N)}(\theta)\big)^{-1} \geq \mathcal{C}_G(\theta)
$$

for a certain quantity $\mathcal{C}_G(\theta)$, which depends on the specification of the quantum statistical model (state of one copy, derivatives of the state with respect to parameters, and loss function G) *at the point* θ only, i.e., on local or pointwise model features (see (7) below).

We aim to prove that under minimal smoothness conditions on the quantum statistical model, and conditions on the prior similar to those needed in the classical case, but under essentially no conditions on the estimator-and-measurement sequence,

$$
(3) \qquad \liminf_{N \to \infty} N R^{(N)}(\pi) \geq \mathbb{E}_\pi \mathcal{C}_G,
$$

where, as before, $G = \psi' \widetilde{G} \psi'^\top$. The main result (3) is exactly the bound one would hope for, from heuristic statistical principles. In specific models of interest, the right

hand side is often easy to calculate. Various specific measurement-and-estimator sequences, motivated by a variety of approaches, can also be shown in interesting examples to achieve the bound, see the appendix to the eprint version of this paper.

It was also shown in [7], how—in the classical statistical context—one can replace a fixed prior π by a sequence of priors indexed by N, concentrating more and more on a fixed parameter value θ_0, at rate $1/\sqrt{N}$. Following their approach would, in the quantum context, lead to the pointwise asymptotic lower bounds

$$(4) \qquad \liminf_{N \to \infty} N R^{(N)}(\theta) \geq \mathcal{C}_G(\theta)$$

for each θ, for *regular* estimators, and to local asymptotic minimax bounds

$$(5) \qquad \lim_{M \to \infty} \liminf_{N \to \infty} \sup_{\|\theta - \theta_0\| \leq N^{-1/2}M} N R^{(N)}(\theta) \geq \mathcal{C}_G(\theta_0)$$

for *all* estimators, but we do not further develop that theory here. In classical statistics the theory of Local Asymptotic Normality is the way to unify, generalise, and understand this kind of result. In the last section of this paper we introduce the now emerging quantum generalization of this theory.

The basic tools used in the first part of this paper have now all been mentioned, but as we shall see, the proof is not a routine application of the van Trees inequality. The missing ingredient will be provided by the following new *dual* bound to (2): for all θ, K, N and $M^{(N)}$,

$$(6) \qquad \operatorname{trace} K(\theta) \overline{I}_M^{(N)}(\theta) \leq \mathcal{C}^K(\theta),$$

where $\mathcal{C}^K(\theta)$ actually equals $\mathcal{C}_G(\theta)$ for a certain G defined in terms of K (as explained in Theorem 2 below). This is an *upper* bound on Fisher information, in contrast to (2) which is a lower bound on inverse Fisher information. The new inequality (6) follows from the convexity of the sets of information matrices and of inverse information matrices for arbitrary measurements on a quantum system, and these convexity properties have a simple statistical explanation. Such dual bounds have cropped up incidentally in quantum statistics, for instance in [8], but this is the first time a connection is established.

The argument for (6), and given that, for (3), is based on some general structural features of quantum statistics, and hence it is not necessary to be familiar with the technical details of the set-up.

In the next section we will summarize the i.i.d. model in quantum statistics, focussing on the key facts which will be used in the proof of the dual Holevo bound (6) and of our main result, the asymptotic lower bound (3).

These proofs are given in a subsequent section, where no further "quantum" arguments will be used.

In the final section we will show how the bounds correspond to recent results in the theory of Q-LAN, according to which the i.i.d. model converges to a quantum Gaussian shift experiment, with the same Holevo bounds, which are actually attainable in the Gaussian case. An eprint version of this paper, Gill and Guţă (2012) includes an appendix with some worked examples.

2. Quantum statistics: The i.i.d. parametric case

The basic objects in quantum statistics are *states* and *measurements*, defined in terms of certain operators on a complex Hilbert space. To avoid technical complications we restrict attention to the finite-dimensional case, already rich in structure and applications, when operators are represented by ordinary (complex) matrices.

States and measurement

The state of a d-dimensional system is represented by a $d \times d$ matrix ρ, called the *density matrix* of the state, having the following properties: $\rho^* = \rho$ (self-adjoint or Hermitian), $\rho \geq \mathbf{0}$ (non-negative), trace(ρ) = 1 (normalized). "Non-negative" actually implies "self-adjoint" but it does no harm to emphasize both properties. $\mathbf{0}$ denotes the zero matrix; $\mathbf{1}$ will denote the identity matrix.

Example. When $d = 2$, every density matrix can be written in the form $\rho = \frac{1}{2}(\mathbf{1} + \theta_1\sigma_1 + \theta_2\sigma_2 + \theta_3\sigma_3)$ where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the three Pauli matrices and where $\theta_1^2 + \theta_2^2 + \theta_3^2 \leq 1$.

"Quantum statistics" concerns the situation when the state of the system $\rho(\theta)$ depends on a (column) vector θ of p unknown (real) parameters.

Example. A completely unknown two-dimensional quantum state depends on a vector of three real parameters, $\theta = (\theta_1, \theta_2, \theta_3)^\top$, known to lie in the unit ball. Various interesting submodels can be described geometrically: e.g., the equatorial plane; the surface of the ball; a straight line through the origin. More generally, a completely unknown d-dimensional state depends on $p = d^2 - 1$ real parameters.

Example. In the previous example the two-parameter case obtained by demanding that $\theta_1^2 + \theta_2^2 + \theta_3^2 = 1$ is called the case of a two-dimensional pure state. In general, a state is called pure if $\rho^2 = \rho$ or equivalently ρ has rank one. A completely unknown pure d-dimensional state depends on $p = 2(d-1)$ real parameters.

A measurement on a quantum system is characterized by the outcome space, which is just a measurable space $(\mathcal{X}, \mathcal{B})$, and a *positive operator valued measure* (POVM) M on this space. This means that for each $B \in \mathcal{B}$ there corresponds a $d \times d$ non-negative self-adjoint matrix $M(B)$, together having the usual properties of an ordinary (real) measure (sigma-additive), with moreover $M(\mathcal{X}) = \mathbf{1}$. The probability distribution of the outcome of doing measurement M on state $\rho(\theta)$ is given by the Born law, or trace rule: Pr(outcome $\in B$) = trace($\rho(\theta)M(B)$). It can be seen that this is indeed a bona-fide probability distribution on the sample space $(\mathcal{X}, \mathcal{B})$. Moreover it has a density with respect to the finite real measure trace($M(B)$).

Example. The most simple measurement is defined by choosing an orthonormal basis of $\mathbb{C}^d$, say $\psi_1, \ldots, \psi_d$, taking the outcome space to be the discrete space $\mathcal{X} = \{1, \ldots, d\}$, and defining $M(\{x\}) = \psi_x\psi_x^*$ for $x \in \mathcal{X}$; or in physicists' notation, $M(\{x\}) = |\psi_x\rangle\langle\psi_x|$. One computes that Pr(outcome $= x$) $= \psi_x^*\rho(\theta)\psi_x = \langle\psi_x|\rho|\psi_x\rangle$. If the state is pure then $\rho = \phi\phi^* = |\phi\rangle\langle\phi|$ for some $\phi = \phi(\theta) \in \mathbb{C}^d$ of length 1 and depending on the parameter θ. One finds that Pr(outcome $= x$) $= |\psi_x^*\phi|^2 = |\langle\psi_x|\phi\rangle|^2$.

So far we have discussed state and measurement for a single quantum system. This encompasses also the case of N copies of the system, via a tensor product construction, which we will now summarize. The joint state of N identical copies of a single system having state $\rho(\theta)$ is $\rho(\theta)^{\otimes N}$, a density matrix on a space of dimension d^N. A joint or collective measurement on these systems is specified by a POVM on this large tensor product Hilbert space. An important point is that joint

measurements give many more possibilities than measuring the separate systems independently, or even measuring the separate systems adaptively.

Fact to remember 1. State plus measurement determines probability distribution of data.

Quantum Cramér-Rao bound

Our main input is going to be the Holevo [22] quantum Cramér-Rao bound, with its extension to the i.i.d. case due to Hayashi and Matsumoto [20].

Precisely because of quantum phenomena, different measurements, incompatible with one another, are appropriate when we are interested in different components of our parameter, or more generally, in different loss functions. The bound concerns estimation of θ itself rather than a function thereof, and depends on a quadratic loss function defined by a symmetric real non-negative matrix $G(\theta)$ which may depend on the actual parameter value θ. For a given estimator $\widehat{\theta}^{(N)}$ computed from the outcome of some measurement $M^{(N)}$ on N copies of our system, define its mean square error matrix $V^{(N)}(\theta) = \mathbb{E}_\theta(\widehat{\theta}^{(N)} - \theta)(\widehat{\theta}^{(N)} - \theta)^\top$. The risk function when using the quadratic loss determined by G is $R^{(N)}(\theta) = \mathbb{E}_\theta(\widehat{\theta}^{(N)} - \theta)^\top G(\theta)(\widehat{\theta}^{(N)} - \theta) = \text{trace}(G(\theta)V^{(N)}(\theta))$.

One may expect the risk of good measurements-and-estimators to decrease like N^{-1} as $N \to \infty$. The quantum Cramér-Rao bound confirms that this is the best rate to hope for: it states that for unbiased estimators of a p-dimensional parameter θ, based on arbitrary joint measurements on N copies,

$$(7) \qquad NR^{(N)}(\theta) \geq \mathcal{C}_G(\theta) = \inf_{\vec{X}, V : V \geq Z(\vec{X})} \text{trace}(G(\theta)V),$$

where $\vec{X} = (X_1, \ldots, X_p)$, the X_i are $d \times d$ self-adjoint matrices satisfying

$$(8) \qquad \partial/\partial\theta_i \, \text{trace}(\rho(\theta)X_j) = \delta_{ij},$$

Z is the $p \times p$ self-adjoint matrix with elements $\text{trace}(\rho(\theta)X_iX_j)$, and V is a real symmetric matrix. It is possible to solve the optimization over V for given $\vec{X}$ leading to the formula

$$(9) \qquad \mathcal{C}_G(\theta) = \inf_{\vec{X}} \text{trace}(\Re(G^{1/2}Z(\vec{X})G^{1/2}) + \text{abs}\Im(G^{1/2}Z(\vec{X})G^{1/2})),$$

where $G = G(\theta)$. The absolute value of a matrix is found by diagonalising it and taking absolute values of the eigenvalues. We'll assume that the bound is finite, i.e., there exists $\vec{X}$ satisfying the constraints. A sufficient condition for this is that the Helstrom quantum information matrix H introduced in ([21]) below is nonsingular.

For specific interesting models, it often turns out not difficult to compute the bound $\mathcal{C}_G(\theta)$. Note, it is a bound which depends only on the density matrix of one system ($N = 1$) and its derivative with the respect to the parameter, and on the loss function, both at the given point θ. It can be found by solving a finite-dimensional optimization problem.

We will not be concerned with the specific form of the bound. What we are going to need, are just two key properties.

Firstly: the bound is local, and applies to the larger class of *locally unbiased estimators*. This means to say that *at the given point θ*, $\mathbb{E}_\theta\widehat{\theta}^{(N)} = \theta$, and at this

point also $\partial/\partial\theta_i\,\mathbb{E}_\theta\widehat{\theta}_j^{(N)} = \delta_{ij}$. Now, it is well known that the "estimator" $\theta_0 + I(\theta_0)^{-1}S(\theta_0)$, where $I(\theta)$ is Fisher information and $S(\theta)$ is score function, is locally unbiased at $\theta = \theta_0$ *and achieves the Cramér-Rao bound there.* Thus the Cramér-Rao bound for *locally* unbiased estimators is sharp. Consequently, we can rewrite the bound (7) in the form (2) announced above, where $\overline{I}_M^{(N)}(\theta)$ is the *average* (divided by N) Fisher information in the outcome of an arbitrary measurement $M = M^{(N)}$ on N copies and the right hand side is defined in (7) or (9).

Fact to remember 2. We have a family of computable lower bounds on the inverse average Fisher information matrix for an arbitrary measurement on N copies, given by (2) and (7) or (9),

Secondly, for given θ, define the following two sets of positive-definite symmetric real matrices, in one-to-one correspondence with one another through the mapping "matrix inverse". The matrices G occurring in the definition are also taken to be positive-definite symmetric real.

$$(10) \qquad\qquad \mathcal{V} = \big\{V : \mathrm{trace}(GV) \geq \mathcal{C}_G \; \forall G\big\},$$

$$(11) \qquad\qquad \mathcal{I} = \big\{I : \mathrm{trace}\big(GI^{-1}\big) \geq \mathcal{C}_G \; \forall G\big\}.$$

Elsewhere (Gill, 2005) we have given a proof by matrix algebra that that the set $\mathcal{I}$ is convex (for $\mathcal{V}$, convexity is obvious), and that the inequalities defining $\mathcal{V}$ define supporting hyperplanes to that convex set, i.e., all the inequalities are achievable in $\mathcal{V}$, or equivalently $\mathcal{C}_G = \inf_{V \in \mathcal{V}} \mathrm{trace}(GV)$. But now, with the tools of Q-LAN behind us (well—ahead of us—see the last section of this paper), we can give a short, statistical, explanation which is simultaneously a short, complete, proof.

The quantum statistical problem of collective measurements on N identical quantum systems, when rescaled at the proper $\sqrt{N}$-rate, approaches a quantum Gaussian problem as $N \to \infty$, as we will see the last section of this paper. In this problem, $\mathcal{V}$ consists precisely of all the covariance matrices of locally unbiased estimators achievable (by suitable choice of measurement) in the limiting p-parameter quantum Gaussian statistical model. The inequalities defining $\mathcal{V}$ are exactly the Holevo bounds for that model, and each of those bounds, as we show in Section 4, is attainable. Thus, for each G, there exists a $V \in \mathcal{V}$ achieving equality in $\mathrm{trace}(GV) \geq \mathcal{C}_G$. It follows from this that $\mathcal{I}$ consists of all non-singular information matrices (augmented with all non-singular matrices smaller than an information matrix) achievable by choice of measurement on the same quantum Gaussian model. Consider the set of information matrices attainable by some measurement, together with all smaller matrices; and consider the set of variance matrices of locally unbiased estimators based on arbitrary measurements, together with all larger matrices. Adding zero mean noise to a locally unbiased estimator preserves its local unbiasedness, so adding larger matrices to the latter set does not change it, by the mathematical definition of measurement, which includes addition of outcomes of arbitrary auxiliary randomization. The set of information matrices is convex: choosing measurement 1 with probability p and measurement 2 with probability q while remembering your choice, gives a measurement whose Fisher information is the convex combination of the informations of measurements 1 and 2. Augmenting the set with all matrices smaller than something in the set, preserves convexity. The set of variances of locally unbiased estimators is convex, by a similar randomization argument. Putting this together, we obtain

Fact to remember 3. For given θ, both $\mathcal{V}$ and $\mathcal{I}$ defined in (10) and (11) are convex, and all the inequalities defining these sets are achieved by points in the sets.

3. An asymptotic Bayesian information bound

We will now introduce the van Trees inequality, a Bayesian Cramér-Rao bound, and combine it with the Holevo bound (2) via derivation of a dual bound following from the convexity of the sets (7) and (9). We return to the problem of estimating the (real, column) vector function $\psi(\theta)$ of the (real, column) vector parameter θ of a state $\rho(\theta)$ based on collective measurements of N identical copies. The dimensions of ψ and of θ need not be the same. The sample size N is largely suppressed from the notation. Let V be the mean square error matrix of an arbitrary estimator $\widehat{\psi}$, thus $V(\theta) = \mathbb{E}_\theta(\widehat{\psi} - \psi(\theta))(\widehat{\psi} - \psi(\theta))^\top$. Often, but not necessarily, we'll have $\widehat{\psi} = \psi(\widehat{\theta})$ for some estimator of θ. Suppose we have a quadratic loss function $(\widehat{\psi} - \psi(\theta))^\top \widetilde{G}(\theta)(\widehat{\psi} - \psi(\theta))$ where $\widetilde{G}$ is a positive-definite matrix function of θ, then the Bayes risk with respect to a given prior π can be written $R(\pi) = \mathbb{E}_\pi \text{trace}\, \widetilde{G}V$. We are going to prove the following theorem:

Theorem 1. *Suppose $\rho(\theta) : \theta \in \Theta \subseteq \mathbb{R}^p$ is a smooth quantum statistical model and suppose π is a smooth prior density on a compact subset $\Theta_0 \subseteq \Theta$, such that Θ_0 has a piecewise smooth boundary, on which π is zero. Suppose moreover the quantity $\mathcal{J}(\pi)$ defined in (16) below, is finite. Then*

$$(12) \qquad \liminf_{N \to \infty} NR^{(N)}(\pi) \geq \mathbb{E}_\pi \mathcal{C}_{G_0},$$

where $G_0 = \psi' \widetilde{G} \psi'^\top$ (and assumed to be positive-definite), ψ' is the matrix of partial derivatives of elements of ψ with respect to those of θ, and $\mathcal{C}_{G_0}$ is defined by (7) or (9).

"Once continuously differentiable" is enough smoothness. Smoothness of the quantum statistical model implies smoothness of the classical statistical model following from applying an arbitrary measurement to N copies of the quantum state. Slightly weaker but more elaborate smoothness conditions on the statistical model and prior are spelled out in [7]. The restriction that G_0 be non-singular can probably be avoided by a more detailed analysis.

Let $\overline{I}_M$ denote the average Fisher information matrix for θ based on a given collective measurement on the N copies. Then the van Trees inequality states that for all matrix functions C of θ, of size $\dim(\psi) \times \dim(\theta)$,

$$(13) \qquad N\mathbb{E}_\pi \text{trace}\, \widetilde{G}V \geq \frac{(\mathbb{E}_\pi \text{trace}\, C\psi'^\top)^2}{\mathbb{E}_\pi \text{trace}\, \widetilde{G}^{-1} C\overline{I}_M C^\top + \frac{1}{N}\mathbb{E}_\pi \frac{(C\pi)'^\top \widetilde{G}^{-1}(C\pi)'}{\pi^2}},$$

where the primes in ψ' and in $(C\pi)'$ both denote differentiation, but in the first case converting the vector ψ into the matrix of partial derivatives of elements of ψ with respect to elements of θ, of size $\dim(\psi) \times \dim(\theta)$, in the second case converting the matrix $C\pi$ into the column vector, of the same length as ψ, with row elements $\sum_j (\partial/\partial\theta_j)(C\pi)_{ij}$. To get an optimal bound we need to choose $C(\theta)$ cleverly.

First though, note that the Fisher information appears in the denominator of the van Trees bound. This is a nuisance since we have a Holevo's *lower* bound (2) to the *inverse* Fisher information. We would like to have an *upper* bound on the information itself, say of the form (6), together with a recipe for computing $\mathcal{C}^K$.

All this can be obtained from the convexity of the sets $\mathcal{J}$ and $\mathcal{V}$ defined in (11) and (10) and the non-redundancy of the inequalities appearing in their definitions. Suppose V_0 is a boundary point of $\mathcal{V}$. Define $I_0 = V_0^{-1}$. Thus I_0 (though not necessarily

an attainable average information matrix $\overline{I}_M^{(N)})$ satisfies the Holevo bound for each positive-definite G, and attains equality in one of them, say with $G = G_0$. In the language of convex sets, and "in the V-picture", trace $G_0 V = \mathcal{C}_{G_0}$ is a supporting hyperplane to $\mathcal{V}$ at $V = V_0$.

Under the mapping "matrix-inverse" the hyperplane trace $G_0 V = \mathcal{C}_{G_0}$ in the V-picture maps to the smooth surface trace $G_0 I^{-1} = \mathcal{C}_{G_0}$ touching the set $\mathcal{I}$ at I_0 in the I-picture. Since $\mathcal{I}$ is convex, the tangent plane to the smooth surface at $I = I_0$ must be a supporting hyperplane to $\mathcal{I}$ at this point. The matrix derivative of the operation of matrix inversion can be written $\mathrm{d}A^{-1}/\mathrm{d}x = -A^{-1}(\mathrm{d}A/\mathrm{d}x)A^{-1}$. This tells us that the equation of the tangent plane is trace $G_0 I_0^{-1} I I_0^{-1} = \text{trace } G_0 I_0^{-1} = \mathcal{C}_{G_0}$. Since this is simultaneously a supporting hyperplane to $\mathcal{I}$ we deduce that for all $I \in \mathcal{I}$, trace $G_0 I_0^{-1} I I_0^{-1} \leq \mathcal{C}_{G_0}$. Defining $K_0 = I_0^{-1} G_0 I_0^{-1}$ and $\mathcal{C}^{K_0} = \mathcal{C}_{G_0}$ we rewrite this inequality as trace $K_0 I \leq \mathcal{C}^{K_0}$.

A similar story can be told when we start in the I-picture with a supporting hyperplane (at $I = I_0$) to $\mathcal{I}$ of the form trace $K_0 I = \mathcal{C}^{K_0}$ for some symmetric positive-definite K_0. It maps to the smooth surface trace $K_0 V^{-1} = \mathcal{C}^{K_0}$, with tangent plane trace $K_0 V_0^{-1} I V_0^{-1} = \mathcal{C}^{K_0}$ at $V = V_0 = I_0^{-1}$. By strict convexity of the function "matrix inverse", the tangent plane touches the smooth surface only at the point V_0. Moreover, the smooth surface lies above the tangent plane, but below $\mathcal{V}$. This makes V_0 the unique minimizer of trace $K_0 V_0^{-1} I V_0^{-1}$ in $\mathcal{V}$.

It would be useful to extend these computations to allow singular I, G and K. Anyway, we summarize what we have so far in a theorem.

Theorem 2. *Dual to the Holevo family of lower bounds on average inverse information, trace $G \overline{I}_M^{-1} \geq \mathcal{C}_G$ for each positive-definite G, we have a family of upper bounds on information,*

(14) $$\text{trace } K \overline{I}_M \leq \mathcal{C}^K \quad \text{for each } K.$$

If $I_0 \in \mathcal{I}$ satisfies trace $G_0 I_0^{-1} = \mathcal{C}_{G_0}$ then with $K_0 = I_0^{-1} G_0 I_0^{-1}$, $\mathcal{C}^{K_0} = \mathcal{C}_{G_0}$. Conversely if $I_0 \in \mathcal{I}$ satisfies trace $K_0 I_0 = \mathcal{C}^{K_0}$ then with $G_0 = I_0 K_0 I_0$, $\mathcal{C}_{G_0} = \mathcal{C}^{K_0}$. Moreover, none of the bounds is redundant, in the sense that for all positive-definite G and K, $\mathcal{C}_G = \inf_{V \in \mathcal{V}} \text{trace}(GV)$ and $\mathcal{C}^K = \sup_{I \in \mathcal{I}} \text{trace}(KI)$. The minimizer in the first equation is unique.

Now we are ready to apply the van Trees inequality. First we make a guess for what the left hand side of (13) should look like, at its best. Suppose we use an estimator $\widehat{\psi} = \psi(\widehat{\theta})$ where $\widehat{\theta}$ makes optimal use of the information in the measurement M. Denote now by I_M the asymptotic normalized Fisher information of a sequence of measurements. Then we expect that the asymptotic normalized covariance matrix V of $\widehat{\psi}$ is equal to $\psi' I_M^{-1} \psi'^{\top}$ and therefore the asymptotic normalized Bayes risk should be $\mathbb{E}_\pi \text{trace} \widetilde{G} \psi' I_M^{-1} \psi'^{\top} = \mathbb{E}_\pi \text{trace} \psi'^{\top} \widetilde{G} \psi' I_M^{-1}$. This is bounded below by the integrated Holevo bound $\mathbb{E}_\pi \mathcal{C}_{G_0}$ with $G_0 = \psi'^{\top} \widetilde{G} \psi'$. Let $I_0 \in \mathcal{I}$ satisfy trace $G_0 I_0^{-1} = \mathcal{C}_{G_0}$; its existence and uniqueness are given by Theorem 2. (Heuristically we expect that I_0 is asymptotically attainable). By the same theorem, with $K_0 = I_0^{-1} G_0 I_0^{-1}$, $\mathcal{C}^{K_0} = \mathcal{C}_{G_0} = \text{trace } G_0 I_0^{-1} = \text{trace } \psi'^{\top} \widetilde{G} \psi' I_0^{-1}$.

Though these calculations are informal, they lead us to try the matrix function $C = \widetilde{G} \psi' I_0^{-1}$. Define $V_0 = I_0^{-1}$. With this choice, in the numerator of the van Trees inequality, we find the square of trace $C \psi'^{\top} = \text{trace } \widetilde{G} \psi' I_0^{-1} \psi'^{\top} = \text{trace } G_0 V_0 = \mathcal{C}_{G_0}$. In the main term of the denominator, we find trace $\widetilde{G}^{-1} \widetilde{G} \psi' I_0^{-1} \overline{I}_M I_0^{-1} \psi'^{\top} \widetilde{G} = \text{trace } I_0^{-1} G_0 I_0^{-1} \overline{I}_M = \text{trace } K_0 \overline{I}_M \leq \mathcal{C}^{K_0} = \mathcal{C}_{G_0}$ by the dual Holevo bound (14).

This makes the numerator of the van Trees bound equal to the square of this part of the denominator, and using the inequality $a^2/(a+b) \geq a - b$ we find

$$(15) \qquad N\mathbb{E}_\pi \text{trace}\, GV \geq \mathbb{E}_\pi \mathcal{C}_{G_0} - \frac{1}{N}\mathcal{J}(\pi),$$

where

$$(16) \qquad \mathcal{J}(\pi) = \mathbb{E}_\pi \frac{(C\pi)'^\top \widetilde{G}^{-1}(C\pi)'}{\pi^2}$$

with $C = \widetilde{G}\psi'V_0$ and V_0 uniquely achieving in $\mathcal{V}$ the bound $\text{trace}\, G_0V \geq \mathcal{C}_{G_0}$, where $G_0 = \psi'^\top \widetilde{G}\psi'$. Finally, provided $\mathcal{J}(\pi)$ is finite (which depends on the prior distribution and on properties of the model), we obtain the asymptotic lower bound

$$(17) \qquad \liminf_{N\to\infty} N\mathbb{E}_\pi \text{trace}\, \widetilde{G}V \geq \mathbb{E}_\pi \mathcal{C}_{G_0}.$$

4. Q-LAN for i.i.d. models

In this section we sketch some elements of a theory of comparison and convergence of quantum statistical models, which is currently being developed in analogy to the LeCam theory of classical statistical models. We illustrate the theory with the example of local asymptotic normality for (finite dimesional) i.i.d. quantum states, which provides a route to proving that the Holevo bound is asymptotically achievable. For more details we refer to the papers [12–14, 16], for the i.i.d. case and to [10] for the case of mixing quantum Markov chains.

The Q-LAN theory surveyed here concerns *strong* local asymptotic normality. Just as in the classical case, the "strong" version of the theory enables us not only to derive asymptotic bounds, but also to actually construct asymptotically optimal statistical procedures, by explicitly lifting the optimal solution of the asymptotic problem back to the finite N situation, where it is approximately optimal. It will be useful to build up theory and applications of the corresponding *weak* local asymptotic normality concept. A start has been made by [13]. Such a theory would be easier to apply, and would be sufficient to obtain rigorous asymptotic bounds, but would not contain recipes for how to attain them. At present there are some situations (involving degeneracy) where stong local asymptotic normality is conjectured but not yet proven. It would be interesting to study these analytically tricky problems first using the simpler tools of weak Q-LAN.

4.1. Convergence of classical statistical models

To facilitate the comparison between classical and quantum, we will start with a brief summary of some basic notions from the classical theory of convergence of statistical models, specialised to the case of dominated models.

Recall that if $\mathbb{P}_\theta$ is a probability distribution on (Ω, Σ) with $\theta \in \Theta$ unknown, then model $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ is called dominated if $\mathbb{P}_\theta \ll \mathbb{P}$ for some measure $\mathbb{P}$. We will denote by p_θ the probability density of $\mathbb{P}_\theta$ with respect to $\mathbb{P}$. Similarly, let $\mathcal{P}' := \{\mathbb{P}'_\theta : \theta \in \Theta\}$ be another model on (Ω', Σ') with densities $p'_\theta = d\mathbb{P}'_\theta/d\mathbb{P}'$. Then we say that $\mathcal{P}$ and $\mathcal{P}'$ are statistically equivalent (denoted $\mathcal{P} \sim \mathcal{P}'$) if their distributions can be transformed into each other via randomisations, i.e., if there exists a linear transformation

$$R : L^1(\Omega, \Sigma, \mathbb{P}) \to L^1(\Omega', \Sigma', \mathbb{P}')$$

mapping probability densities into probability densities, such that for all $\theta \in \Theta$

$$R(p_\theta) = p'_\theta,$$

and similarly in the opposite direction. In particular, $S : \Omega \to \Omega'$ is a sufficient statistic for $\mathcal{P}$ if and only if $\mathcal{P} \sim \mathcal{P}'$ where $\mathbb{P}'_\theta := \mathbb{P}_\theta \circ S^{-1}$.

In asymptotics one often needs to show that a sequence of models converges to a limit model without being statistically equivalent to it at any point. This can be formulated by using LeCam's notion of deficiency and the associated distance on the space of statistical models. The deficiency of $\mathcal{P}$ with respect to $\mathcal{P}'$ (expressed here in L^1 rather than total variation norm) is

$$\delta(\mathcal{P}, \mathcal{P}') := \inf_R \sup_{\theta \in \Theta} \|R(p_\theta) - p'_\theta\|_1,$$

where the infimum is taken over all randomisations R. The LeCam distance between $\mathcal{P}$ and $\mathcal{P}'$ is defined as

$$\Delta(\mathcal{P}, \mathcal{P}') := \max(\delta(\mathcal{P}, \mathcal{P}'), \delta(\mathcal{P}', \mathcal{P})),$$

and is equal to zero if and only if the models are equivalent. A sequence of models $\mathcal{P}^{(n)}$ converges strongly to $\mathcal{P}$ if

$$\lim_{n \to \infty} \Delta(\mathcal{P}^{(n)}, \mathcal{P}) = 0.$$

This can be used to prove the convergence of optimal procedures and risks for statistical decision problems. We illustrate this with the example of local asymptotic normality (LAN) for i.i.d. parametric models, whose quantum extension provides an alternative route to optimal estimation in quantum statistics. Suppose that $\mathcal{P}$ is a model over an open set $\Theta \subset \mathbb{R}^k$ and that p_θ depends sufficiently smoothly on θ (e.g., $p_\theta^{1/2}$ is differentiable in quadratic mean), and consider the local i.i.d. models around θ_0 with local parameter $h \in \mathbb{R}^k$

$$\mathcal{P}^{(n)} := \{\mathbb{P}^n_{\theta_0 + h/\sqrt{n}} : \|h\| \leq C\}.$$

LAN means that $\mathcal{P}^{(n)}$ converges strongly to the Gaussian shift model consisting of a single sample from an k-variate normal distribution with mean h and variance equal to the inverse Fisher information matrix of the original model at θ_0

$$\mathcal{N} := \{N(h, I_{\theta_0}^{-1}) : \|h\| \leq C\}.$$

4.2. Convergence of quantum statistical models

As we have seen, an important problem in quantum statistics is to find the most informative measurement for a given quantum statistical model and a given decision problem. A partial solution to this problem is provided by the quantum Cramér-Rao theory which aims to construct lower bounds to the quadratic risk of any estimator, expressed solely in terms of the properties of the quantum states. Classical mathematical statistics suggests that rather than searching for optimal decisions, more insight could be gained by analysing the structure of the quantum statistical models themselves, beyond the notion of quantum Fisher information. Therefore we will start by addressing a more basic question of how to decide whether two quantum models over a parameter space Θ are statistically equivalent, or close to

each other in a statistical sense. To answer this question we will introduce the notion of quantum channel, which is a transformation of quantum states that could—in principle—be physically implemented in a lab, and should be seen as the analog of a classical randomisation which defines a particular data processing procedure. The simplest example of such transformation is a unitary channel which rotates a state ($d \times d$ density matrix ρ) by means of a $d \times d$ unitary matrix U, i.e.,

$$\mathcal{U} : \rho \mapsto U\rho U^*.$$

Since $\mathcal{U}$ can be reversed by applying the inverse unitary U^{-1}, we anticipate that it will map any quantum model into an equivalent one. More generally, a quantum channel $C : M(\mathbb{C}^d) \to M(\mathbb{C}^k)$ must satisfy the minimal requirement of being positive and trace preserving linear map, i.e., it must transform quantum states into quantum states in an affine way, similarly to the action of a classical randomisation. However, unlike the classical case, it turns out that this condition needs to be strengthened to the requirement that C is *completely positive*, i.e., the ampliated maps

$$C \otimes \text{Id}_n : M(\mathbb{C}^d) \otimes M(\mathbb{C}^n) \to M(\mathbb{C}^k) \otimes M(\mathbb{C}^n)$$

must be positive for all $n \geq 0$, where Id_n is the identity transformation on $M(\mathbb{C}^n)$. An example of a positive but not completely positive, and hence unphysical transformation, is the transposition $tr : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ with respect to a given basis. Indeed, the reader can verify that applying $tr \otimes \text{Id}_d$ to any *pure entangled* state (i.e., not a product state $|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|$) produces a matrix which is not positive, hence not a state.

Definition 1. A linear map $C : M(\mathbb{C}^d) \to M(\mathbb{C}^k)$ which is completely positive and trace preserving is called a quantum channel.

The Stinespring-Kraus Theorem [26] says a linear map $C : M(\mathbb{C}^d) \to M(\mathbb{C}^k)$ is completely positive map if and only if it is of the form

$$C(\rho) = \sum_{i=1}^{dk} K_i \rho K_i^*,$$

with K_i linear transformations from $\mathbb{C}^d$ to $\mathbb{C}^k$, some of which may be equal to zero. Moreover, C is trace preserving if and only if $\sum_i K_i^* K_i = \mathbf{1}_d$. In particular, if the sum consists of a single non-zero term $V\rho V^*$, the action of the channel C is to embed the state ρ isometrically into a the d-dimensional subspace $\text{Ran}(V) \subset \mathbb{C}^k$. As in the unitary case, it is easy to see that this action is reversible (hence noiseless) and maps any statistical model into an equivalent one. We are now ready to define the notion of equivalence of statistical models, as an extension of the classical characterisation.

Definition 2. Let $\mathcal{Q} := \{\rho(\theta) \in M(\mathbb{C}^d) : \theta \in \Theta\}$ and $\mathcal{R} := \{\varphi(\theta) \in M(\mathbb{C}^k) : \theta \in \Theta\}$ be two quantum statistical models over Θ. Then $\mathcal{Q}$ is statistically equivalent to $\mathcal{R}$ if there exist quantum channels $T : M(\mathbb{C}^d) \to M(\mathbb{C}^k)$ and $S : M(\mathbb{C}^k) \to M(\mathbb{C}^d)$ such that for all $\theta \in \Theta$

$$T(\rho(\theta)) = \varphi(\theta) \qquad \text{and} \qquad S(\varphi(\theta)) = \rho(\theta).$$

The interpretation of this definition is immediate. Suppose that we want to solve a statistical decision problem concerning the model $\mathcal{R}$, e.g., estimating θ, and we perform a measurement M on the state φ_θ whose outcome is the estimator $\hat{\theta}$ with distribution $\mathbb{P}_\theta^M = M(\rho(\theta))$ and risk $R_\theta^M := \mathbb{E}_\theta(d(\hat{\theta}, \theta)^2)$. Consider now the same

problem for the model $\mathcal{Q}$, and define the measurement $N = M \circ R$ realised by first mapping the quantum states $\rho(\theta)$ through the channel T into $\varphi(\theta)$, and then performing the measurement M. Clearly, the distribution of the obtained outcome is again $\mathbb{P}_\theta^M$ and the risk is R_θ^M, so we can say that $\mathcal{Q}$ is at least as informative as $\mathcal{P}$ from a statistical point of view. By repeating the argument in the opposite direction we conclude that any statistical decision problem is equally difficult for the two models, and hence they are equivalent in this sense. However, unlike the classical case the opposite implication is not true. For instance, models whose states are each other's transpose have the same set of risks for any decision problem but are usually not equivalent in the sense of being connected by quantum channels. It turns out that a full statistical interpretation of Definition 2 is possible if one considers a larger set of *quantum decision problems*, which do not involve measurements, but quantum channels as statistical procedures.

Until this point we have tacitly assumed that any (finite dimensional) quantum model is built upon the algebra of square matrices of a certain dimension. However this setting is too restrictive as it excludes the possibility of considering hybrid classical-quantum models, as well as the development of a theory of quantum sufficiency. We motivate this extension through the following example. We throw a coin whose outcome X has probabilities $p_\theta(1) = \theta$ and $p_\theta(0) = 1 - \theta$, and subsequently we prepare a quantum system in the state $\rho_\theta(X) \in M(\mathbb{C}^d)$ which depends on X and the parameter θ. What is the corresponding statistical model? Since the "data" is both classical and quantum, the "state" is a matrix valued density on $\{0, 1\}$

$$\varrho_\theta(i) = p_\theta(i)\rho_\theta(i), \qquad i \in \{0, 1\}$$

or equivalently, a block-diagonal density matrix $\varrho_\theta(1) \oplus \varrho_\theta(2) \in M(\mathbb{C}^d) \oplus M(\mathbb{C}^d)$ which is positive and normalised in the usual sense. While this can be seen as a state on the full matrix algebra $M(\mathbb{C}^{2d})$, it is clear that since the off-diagonal blocks have expectation zero for all θ, we can restrict ϱ_θ to the block diagonal sub-algebra $M(\mathbb{C}^d) \oplus M(\mathbb{C}^d)$ without loosing any statistical information. In other words, the latter is a sufficient algebra of our quantum statistical model. In general, for a model defined on some matrix algebra, one can ask what is the smallest sub-algebra to which we can restrict without loosing statistical information, i.e., such that the restricted model is equivalent to the original one in the sense of definition 2. The theory of quantum sufficiency was developed in [28] where a number of classical results were extended to the quantum set-up, in particular the fact that the minimal sufficient algebra is generated by the likelihood ratio statistic.

We now make a step further and characterise the "closeness" rather than equivalence of quantum statistical models, by generalising LeCam's notion of deficiency between models.

Definition 3. Let $\mathcal{Q} := \{\rho(\theta) \in M(\mathbb{C}^d) : \theta \in \Theta\}$ and $\mathcal{R} := \{\varphi(\theta) \in M(\mathbb{C}^k) : \theta \in \Theta\}$ be two quantum statistical models over Θ. The deficiency of $\mathcal{R}$ with respect to $\mathcal{Q}$ is defined as

$$(18) \qquad \delta(\mathcal{R}, \mathcal{Q}) = \inf_T \sup_{\theta \in \Theta} \big\| \varphi(\theta) - T\big(\rho(\theta)\big) \big\|_1,$$

where the infimum is taken over all channels $T : M(\mathbb{C}^d) \to M(\mathbb{C}^k)$. The LeCam distance between $\mathcal{Q}$ and $\mathcal{R}$ is

$$\Delta(\mathcal{Q}, \mathcal{R}) = \max(\delta(\mathcal{R}, \mathcal{Q}), \delta(\mathcal{Q}, \mathcal{R})).$$

This is an extension of the classical definition of deficiency for dominated statistical models. We will use the LeCam distance to formulate the concept of local asymptotic normality for quantum states and find asymptotically optimal measurement procedures.

4.3. Continuous variables systems and quantum Gaussian states

In this section we introduce the basic concepts associated to continuous variables (cv) quantum systems, and then analyse the problem of optimal estimation for simple quantum Gaussian shifts models.

Firstly we will restrict our attention to the elementary "building block" cv system which physically may be a particle moving on the real line, or a mono-chromatic light pulse. Then we will show how more complex cv systems can be reduced to a tensor product of such "building blocks" by a standard "diagonalisation" procedure.

The Hilbert space of the system is $\mathcal{H} = L^2(\mathbb{R})$ and its quantum states are given by density matrices, i.e., positive operators of trace one. Unlike the finite dimensional case, their linear span, called the space of trace-class operators $\mathcal{T}_1(\mathcal{H})$, is a proper subspace of all bounded operators on $\mathcal{H}$, which is a Banach space with respect to the trace-norm

$$\|\tau\|_1 := \mathrm{Tr}(|\tau|) = \sum_{i=1}^{\infty} s_i,$$

where s_i are the singular values of τ. The key observables are two "canonical coordinates" $\mathbf{Q}$ and $\mathbf{P}$ representing the position and momentum of the particle, or the electric and magnetic field of the light pluse, and are defined as follows

$$(19) \qquad (\mathbf{Q}f)(x) = xf(x), \qquad (\mathbf{P}f)(x) = -i\frac{df}{dx}(x).$$

Although they do not commute with each other, they satisfy Heisenberg's commutation relation which essentially captures the entire algebraic properties of the system:

$$\mathbf{QP} - \mathbf{PQ} = i\mathbf{1}.$$

The label "continuous variables" stems from the fact that the probability distributions of $\mathbf{Q}$ and $\mathbf{P}$ are always absolutely continuous with respect to the Lebesgue measure. Indeed since any state is a mixture of pure states, it suffices to prove this for a pure state $|\psi\rangle\langle\psi|$. If Q and P denote the real valued random variables representing the outcomes of measuring $\mathbf{Q}$ and respectively $\mathbf{P}$ then using (19) one can verify that

$$\mathbb{E}\big(e^{iuQ}\big) = \langle \psi, e^{iu\mathbf{Q}}\psi \rangle = \int e^{iuq}|\psi(q)|^2 \, dq,$$

$$\mathbb{E}\big(e^{ivP}\big) = \langle \psi, e^{iv\mathbf{P}}\psi \rangle = \int e^{ivp}|\widehat{\psi}(p)|^2 \, dp.$$

where $\widehat{\psi}$ is the Fourier transform of ψ. This means that Q and P have probability densities $|\psi(q)|^2$ and respectively $|\widehat{\psi}(p)|^2$, and suggests that the cv system should be seen as the non-commutative analogue of an $\mathbb{R}^2$ valued random variable. Following up on this idea we define the "quantum characteristic function" of a state ρ

$$\widetilde{W}_\rho(u, v) := \mathrm{Tr}(\rho e^{-i(u\mathbf{Q}+v\mathbf{P})})$$

and the Wigner or "quasidistribution" function

$$W_\rho(q,p) = \frac{1}{(2\pi)^2} \int\int e^{i(uq+vp)} \widetilde{W}_\rho(u,v) \, du \, dv.$$

These functions have a number of interesting and useful properties, which make them into important tools in visualising and analysing states of cv quantum systems.

1. there is a one-to-one correspondence between ρ and W_ρ;
2. the Wigner function may take negative values, but its marginal along any direction ϕ is a bona-fide probability density corresponding to the measurement of the quadrature observable $\mathbf{X}_\phi := \mathbf{Q}\cos\phi + \mathbf{P}\sin\phi$;
3. Both W_ρ and $\widetilde{W}_\rho$ belong to $L^2(\mathbb{R}^2)$ and the following isometry holds between the space of Hilbert-Schmidt operators $\mathcal{T}_2(L^2(\mathbb{R}))$ and $L^2(\mathbb{R}^2)$

$$\mathrm{Tr}(\rho A) = \int\int W_\rho(q,p) W_A(q,p) \, dq \, dp.$$

We can now introduce the class of quantum Gaussian states by analogy to the classical definition.

Definition 4. Let ρ be a state with mean $(q,p) = (\mathrm{Tr}(\rho\mathbf{Q}), \mathrm{Tr}(\rho\mathbf{Q}))$ and covariance matrix

$$V := \begin{pmatrix} \mathrm{Tr}(\rho(Q-q)^2) & \mathrm{Tr}(\rho(Q-q)\circ(P-p)) \\ \mathrm{Tr}(\rho(Q-q)\circ(P-p)) & \mathrm{Tr}(\rho(P-p)^2) \end{pmatrix}.$$

Then ρ is called Gaussian if its characteristic function is

$$\mathrm{Tr}(\rho e^{-i(u\mathbf{Q}+v\mathbf{P})}) = e^{-itx^t} \cdot e^{-tVt^t/2}, \qquad t = (u,v), \quad x = (q,p),$$

in particular the Wigner function W_ρ is equal to the probability density of $N(x,V)$.

While the definition looks deceptively similar to that of a classical normal distribution, there are a couple of important differences. The first one is that the covariance matrix V cannot be arbitrary but must satisfy the *uncertainty principle*

(20) $$\mathrm{Det}(V) \geq \frac{1}{4}.$$

This restriction can be traced back to the commutation relations $[\mathbf{Q},\mathbf{P}] = i\mathbf{1}$ which says that we cannon assign classical values to $\mathbf{Q}$ and $\mathbf{P}$ simultaneously. Which leads us to the second point, and the problem of optimal estimation: since $\mathbf{Q}$ and $\mathbf{P}$ cannot be measured simultaneously, their covariance matrix V is not "achievable" by any measurement aimed at estimating the means (q,p) and the experimenter needs to make a trade-off between measuring $\mathbf{Q}$ with high accuracy but ignoring $\mathbf{P}$, and vice-versa. In the last part of this section we look at this problem in more detail and explain the optimal measurement procedure.

Definition 5. A quantum Gaussian shift model is family of Gaussian states

$$\mathcal{G} := \{\Phi(x,V) : x \in \mathbb{R}^2\}$$

with unknown mean x and fixed and known covariance matrix V. If G is a 2×2 positive real weight matrix, the optimal estimation problem is to find the measurement M with outcome $\hat{x} = (\hat{q},\hat{p})$ which minimises the maximum quadratic risk

(21) $$R(M) = \sup_x \mathbb{E}_x\big((\hat{x}-x)G(\hat{x}-x)^t\big).$$

This is a provisional definition only: a definitive version follows as Definition 6 below. Finding the optimal measurement, relies on the equivariance (or covariance in physics terminology) of the problem with respect to the action of the translations (or displacements) group $\mathbb{R}^2$ on the states

$$\mathcal{D}(y) : \Phi(x, V) \mapsto \Phi(x + y, V), \qquad y \in \mathbb{R}^2.$$

This action is implemented by a unitary channel

$$\Phi(x + y, V) = D(y)\Phi(x, V)D(y)^*, \qquad y = (u, v),$$

where $D(y) = \exp(iv\mathbf{Q} - iu\mathbf{P})$ are called the displacement or Weyl operators. Since $R(M)$ is invariant under the transformation $[x, \hat{x}] \mapsto [x + y, \hat{x} + y]$, a standard equivariance argument shows that the infimum risk is achieved on the special subset of *covariant* measurements, defined by the property

$$\mathbb{P}^{(M)}_{\Phi(x+y,V)}(d\hat{x} + y) = \mathbb{P}^{(M)}_{\Phi(x,V)}(d\hat{x}).$$

Such measurements, and the more general class of covariant quantum channels, have a simple description in terms of linear transformation on the space of coordinates of the system together with an auxiliary system, [25]. More specifically, consider an independent quantum cv system with coordinates $(\mathbf{Q}', \mathbf{P}')$, prepared in a state τ with zero mean and covariace matrix Y. By the commutation relations, the observables $\mathbf{Q} + \mathbf{Q}'$ and $\mathbf{P} - \mathbf{P}'$ commute with each other and hence can be measured simultaneously. Since the joint state of the two independent systems is $\Phi(x, V) \otimes \tau$, the outcome $(\hat{q}, \hat{p})$ of the measurement is an unbiased estimator of (q, p) with covariance matrix $V + Y$, and the risk is

$$R(M) = \text{Tr}(G(V + Y)) = \text{Tr}(GV) + \text{Tr}(GY),$$

where the first term is the risk of the corresponding classical problem, and the second is the non-vanishing contribution due to the auxiliary "noisy" system. To find the optimum, it remains to minimise the above expression over all possible covariance matrices of the auxiliary system which must satisfy the constraint $\text{Det}(Y) \geq 1/4$. If G has the form $G = O \text{Diag}(g_1, g_2) O^t$ with O orthogonal, then it can be easily verified that the optimal Y is the matrix

$$Y_0 = \frac{1}{2}O \left(\begin{matrix} \sqrt{g_2/g_1} & 0 \\ 0 & \sqrt{g_1/g_2} \end{matrix} \right) O^t.$$

Moreover, the unique state with such "minimum uncertainty" is the Gaussian state $\tau = \Phi(0, Y_0)$. In conclusion, the minimax risk is

$$R_{minmax} = \inf_M R(M) = \text{Tr}(GV) + \sqrt{\text{Det}(G)}.$$

4.4. General Gaussian shift models and optimal estimation

We now extend the findings of the previous section from the "building block" system to a multidimensional setting. In essence, we show that the Holevo bound is *achievable* for general Gaussian shift models, a result which has been known—in various degrees of generality—since the pioneering work of V.P. Belavkin and of A.S. Holevo in the 70's.

Let us consider a system composed of $p \geq 1$ mutually commuting pairs of canonical coordinates $(\mathbf{Q}_i, \mathbf{P}_i)$, so that the commutation relations hold

$$[\mathbf{Q}_i, \mathbf{P}_j] = i\delta_{i,j}\mathbf{1}, \qquad i,j = 1, \ldots, p.$$

The joint system can be represented on the Hilbert space $L^2(\mathbb{R})^{\otimes p}$ such that the pair $(\mathbf{Q}_i, \mathbf{P}_i)$ acts on i-th copy of the tensor product as in (19), and as identity on the other spaces. Additionally, we allow for a number l of "classical variables" $\mathbf{C}_k$ which commute with each other and with all $(\mathbf{Q}_i, \mathbf{P}_i)$, and can be represented separately as position observables on k additional copies of $L^2(\mathbb{R})$. For simplicity we will denote all variables as

$$(\mathbf{X}_1, \ldots, \mathbf{X}_m) \equiv (\mathbf{Q}_1, \mathbf{P}_1, \ldots, \mathbf{Q}_p, \mathbf{P}_p, \mathbf{C}_1, \ldots, \mathbf{C}_l), \quad m = 2p + l,$$

and write their commutation relations as

$$[\mathbf{X}_i, \mathbf{X}_j] = iS_{i,j}\mathbf{1},$$

where S is the $m \times m$ block diagonal symplectic matrix of the form $S = \mathrm{Diag}(\Omega, \ldots, \Omega, 0, \ldots 0)$ with

$$\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that while this may seem to be rather special cv system, it actually captures the general situation since any symplectic (bilinear antisymmetric) can be transformed into the above one by a change of basis.

The states of this hybrid quantum-classical system are described by positive normalised densities in $\mathcal{T}_1(L^2(\mathbb{R}^p)) \otimes L^1(\mathbb{R}^l)$, e.g., if the quantum and classical variables are independent the state is of the form $\rho \otimes p$ with ρ a density matrix and p a probability density. In general the classical and quantum parts may be correlated, and the state is a positive operator valued density $\varrho : \mathbb{R}^l \to \mathcal{T}_1(L^2(\mathbb{R}^p))$, whose characteristic function can be computed as

$$\mathbb{E}_\varrho(e^{i\sum_{i=1}^m u_i \mathbf{X}_i}) = \int \ldots \int \mathrm{Tr}(\varrho(y)e^{\sum_{i=1}^{2p} u_i \mathbf{X}_i}) e^{i\sum_{j=1}^l u_{2p+j} y_j} \, dy_1 \ldots dy_l.$$

Definition 6. A state $\Phi(x, V)$ with mean $x \in \mathbb{R}^m$ and $m \times m$ covariance matrix V is Gaussian if

$$\mathbb{E}_{\Phi(x,V)}(e^{i\sum_{i=1}^m u_i \mathbf{X}_i}) = e^{iux^t} e^{-uVu^t/2}.$$

A Gaussian shift model over the parameter space $\Theta := \mathbb{R}^k$ is a family

$$\mathcal{G} := \{\Phi(Lh, V) : h \in \mathbb{R}^k\},$$

where $L : \mathbb{R}^k \to \mathbb{R}^m$ is a linear map.

Note that the dimension of the parameter h may be smaller than the dimension of mean value x. One may distinguish *full* and *partial* quantum Gaussian shift models: in the full model case, the dimensions are equal (and the matrix L invertible). A non-classical feature of the general quantum Gaussian shift is that *a linear submodel of a full Gaussian shift model is not, in general, equivalent to a full model with lower-dimensional mean vector.*

The analogue of the uncertainty principle (20) for general cv systems is the (complex) matrix inequality

$$(22) \qquad\qquad\qquad V \geq \frac{i}{2}S.$$

The statistical decision problem is to find the measurement which optimally estimates the parameter h of the Gaussian state $\Phi(Lh, V)$, for a mean square error risk with a given $k \times k$ weight matrix G, cf. (21). As before, we can restrict our attention to covariant measurements, i.e., to measuring *mutually commuting* variables of the form

$$\mathbf{W}^{(i)} = \mathbf{Y}^{(i)} + \tilde{\mathbf{Y}}^{(i)},$$

where

$$\mathbf{Y}^{(i)} = \sum_{j=1}^{m} y_j^{(i)} \mathbf{X}_j, \qquad \mathbb{E}_{\Phi(Lh,V)}(\mathbf{Y}^{(i)}) = h_i$$

and

$$\tilde{\mathbf{Y}}^{(i)} = \sum_{j=1}^{\tilde{m}} \tilde{y}_j^{(i)} \tilde{\mathbf{X}}_j, \qquad \mathbb{E}_\varrho(\tilde{\mathbf{Y}}^{(i)}) = 0.$$

Here $(\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_{\tilde{m}})$ are the coordinates of an independent, auxiliary system with symplectic matrix $\tilde{S}$, prepared in a state ϱ with mean zero and covariance matrix $\tilde{V}$. Let $V^{(\mathbf{Y})}$ and $V^{(\tilde{\mathbf{Y}})}$ denote the covariance matrices of the independent systems $(\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(k)})$ and $(\tilde{\mathbf{Y}}^{(1)}, \ldots, \tilde{\mathbf{Y}}^{(k)})$. Then the risk of the $(\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(k)})$ measurement is

$$R(\mathbf{W}) = \mathrm{Tr}(G V^{(\mathbf{Y})}) + \mathrm{Tr}(G V^{(\tilde{\mathbf{Y}})}).$$

On the other hand, since all $\mathbf{W}^{(i)}$ must commute with each other, we have

$$[\tilde{\mathbf{Y}}^{(i)}, \tilde{\mathbf{Y}}^{(j)}] = -[\mathbf{Y}^{(i)}, \mathbf{Y}^{(j)}] := -i S_{i,j}^{(\mathbf{Y})} \mathbf{1}.$$

The uncertainty principle (22) applied to to the auxiliary variables $\tilde{\mathbf{Y}}^{(i)}$ gives the constraint

$$V^{(\tilde{\mathbf{Y}})} \geq \pm \frac{i}{2} S^{(\mathbf{Y})}.$$

Lemma 1. *Let V and S be real symmetric and respectively anti-symmetric $k \times k$ matrices, such that $V \geq iS/2$. Then $\mathrm{Tr}(V) \geq \mathrm{Tr}(|S|)/2$, with equality for $V = |S|/2$.*

By optimising $V^{(\tilde{\mathbf{Y}})}$'s contribution to the risk and applying the above lemma with a fixed choice of $\mathbf{Y}^{(i)}$ we obtain

$$\inf_{\tilde{\mathbf{Y}}^{(i)}} \mathrm{Tr}(G V^{(\tilde{\mathbf{Y}})}) = \inf_{\tilde{\mathbf{Y}}^{(i)}} \mathrm{Tr}(\sqrt{G} V^{(\tilde{\mathbf{Y}})} \sqrt{G}) = \frac{1}{2} \mathrm{Tr}(\sqrt{G} |S^{(\mathbf{Y})}| \sqrt{G}).$$

and the infimum is achieved for the covariance matrix $V^{(\tilde{\mathbf{Y}})} = |S^{(\tilde{\mathbf{Y}})}|/2$, which is only possible if the auxiliary system is prepared in the Gaussian state $\Phi(0, V^{(\tilde{\mathbf{Y}})})$, [23].

It remains now to optimise the risk over all unbiased $(\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(k)})$ i.e., which satisfy the condition (8) from the formulation of the Holevo bound:

$$(23) \qquad \qquad \frac{\partial}{\partial h_j} \mathbb{E}_{\Phi_{h,V}}(\mathbf{Y}^{(i)}) = \delta_{i,j}.$$

The minimax risk is then

$$R_{minmax}(\mathcal{G}, G) = \inf_{\{\mathbf{Y}^{(i)}\}} \mathrm{Tr}(\sqrt{G} V^{(\mathbf{Y})} \sqrt{G}) + \frac{1}{2} \mathrm{Tr}(\sqrt{G} |S^{(\mathbf{Y})}| \sqrt{G})$$

which is equal to the Holevo bound (9) if we consider that

$$V_{i,j}^{\mathbf{Y}} = \Re \, \mathbb{E}_{\Phi(0,V)}(\mathbf{Y}^{(i)} \mathbf{Y}^{(j)}), \quad \text{and} \quad \frac{1}{2} S^{(\mathbf{Y})} = \Im \, \mathbb{E}_{\Phi(0,V)}(\mathbf{Y}^{(i)} \mathbf{Y}^{(j)}).$$

4.5. Local asymptotic normality for i.i.d. states

In this section we show how the general Gaussian shift models discussed above emerge from i.i.d. models through local asymptotic normality.

Suppose that we are given N independent quantum systems prepared identically in an unknown state $\rho \in M(\mathbb{C}^d)$. For large N we can sacrifice a small part of the systems (e.g., $\tilde{N} = N^{1-\epsilon}$) and use them to construct an estimator ρ_0 of the state, by means of a quantum tomography procedure. Using standard concentration inequalities it can be shown that ρ belongs to a neighbourhood of size $N^{-1/2+\epsilon}$ centred at ρ_0, with probability converging to one. Therefore, the asymptotic behaviour of parameter estimation problems is determined by the structure of local quantum models around a fixed state ρ_0, and from now on we will restrict our attention to such models. By choosing the eigenvectors of ρ_0 as the standard basis, and assuming that the eigenvalues satisfy $\mu_1 > \ldots \mu_d > 0$, we have $\rho_0 = \mathrm{Diag}(\mu_1, \ldots, \mu_d)$ and an arbitrary state in its neighbourhood is of the form

$$
(24) \qquad \rho_h := \begin{bmatrix} \mu_1 + u_1 & \zeta_{1,2}^* & \cdots & \zeta_{1,d}^* \\ \zeta_{1,2} & \mu_2 + u_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \zeta_{d-1,d}^* \\ \zeta_{1,d} & \cdots & \zeta_{d-1,d} & \mu_d - \sum_{i=1}^{d-1} u_i \end{bmatrix}, \qquad u_i \in \mathbb{R}, \ \zeta_{j,k} \in \mathbb{C}.
$$

with local parameter $h = (\vec{u}, \vec{\zeta}) \in \mathbb{R}^{d-1} \times \mathbb{C}^{d(d-1)/2} \cong \mathbb{R}^{d^2-1}$. The local i.i.d. quantum model around ρ_0 is then defined as

$$
(25) \qquad \mathcal{Q}_N := \left\{ \rho_h^N := \rho_{h/\sqrt{N}}^{\otimes N} : \|h\| \le N^\epsilon \right\}.
$$

If some eigenvalues μ_i are equal to one another or to zero, degeneracies occur which are tricky to deal with. Completing the theory for such situations is a topic of ongoing research. In the rest of this section we give an intuitive argument for the emergence of the limit Gaussian model and finish with the precise formulation of LAN, restricting attention to the nondegenerate situation.

We define $m = d^2 - 1$ operators whose expectation with respect to the state ρ_0 is zero, and together with the identity form a basis of the space of selfadjoint $d \times d$ matrices

$$
\{X_1, \ldots, X_m\} = \{Q_{1,2}, P_{1,2}, \ldots, Q_{d-1,d}, P_{d-1,d}, C_1, \ldots, C_{d-1}\},
$$

where

$$
Q_{j,k} := \frac{|j\rangle\langle k| + |j\rangle\langle k|}{\sqrt{2(\mu_j - \mu_k)}}, \quad P_{j,k} := \frac{i(|k\rangle\langle j| - |j\rangle\langle k|)}{\sqrt{2(\mu_j - \mu_k)}}, \quad C_i := |i\rangle\langle i| - \mu_i \mathbf{1}.
$$

Let $Q_{j,k}(N) \in M(\mathbb{C}^d)^{\otimes N}$ denote the corresponding collective observables

$$
Q_{j,k}(N) := \sum_{s=1}^{N} Q_{j,k}^{(s)}, \qquad Q_{jk}^{(s)} := \mathbf{1} \otimes \cdots \otimes Q_{j,k} \otimes \cdots \otimes \mathbf{1},
$$

with $Q_{j,k}^{(s)}$ acting on the position s of the tensor product; similar definitions hold for $P_{j,k}(N), C_i(N)$. The collective observables play the role of sufficient statistic for our i.i.d. model, and we would like to understand their asymptotic behaviour.

Since all systems are independent and identically prepared, and the terms in each collective observable commute, we can apply classical Central Limit techniques to show that, under the state ρ_h^n, we have

$$\frac{C_i(N)}{\sqrt{N}} \xrightarrow{\mathcal{L}} N\big(u_i, \mu_i(1-\mu_i)\big), \quad 1 \le i \le d-1;$$

$$\frac{Q_{j,k}(N)}{\sqrt{N}} \xrightarrow{\mathcal{L}} N(\Re\tilde{\zeta}_{j,k}, v_{j,k}), \qquad 1 \le j < k \le d;$$

$$\frac{P_{j,k}(N)}{\sqrt{N}} \xrightarrow{\mathcal{L}} N(\Im\tilde{\zeta}_{j,k}, v_{j,k}), \qquad 1 \le j < k \le d,$$

where $\tilde{\zeta}_{j,k} = \zeta_{j,k}/\sqrt{(\mu_j - \mu_k)/2}$ and $v_{j,k} = 1/(2(\mu_j - \mu_k))$. This indicates that the model converges to a Gaussian shift model, but does not tell us what the *covariance* and *commutation relations* of the different limit variables are. For this, we need a quantum CLT, that is a multivariate CLT which takes into account the fact that the collective variables do not commute with each other. Its precise formulation can be found in [27], but for our purposes it is enough to give the following recipe. The limit is a general cv system as described in Section 4.4, with $m = d^2 - 1$ coordinates $(\mathbf{X}_1, \ldots, \mathbf{X_m}) = (\mathbf{Q}_{j,k}, \mathbf{P}_{j,k}, \mathbf{C}_i)$ having the commutation relations

$$[\mathbf{X}_a, \mathbf{X}_b] = \mathrm{Tr}(\rho_0[X_a, X_b])\mathbf{1} = 2i\Im\mathrm{Tr}(\rho_0 X_a X_b)\mathbf{1},$$

whose state is Gaussian with covariance matrix

$$V_{a,b} = \mathrm{Tr}\big(\rho_0(X_a X_b + X_b X_a)/2\big) = \Re\mathrm{Tr}(\rho_0 X_a X_b)\mathbf{1}.$$

It can be easily verified that thanks to our special choice of basis, $(\mathbf{Q}_{j,k}, \mathbf{P}_{j,k})$ are pairs of position and momentum operators, which commute with all other coordinates and $\mathbf{C}_i$ are "classical" variables, cf. Section 4.4. Moreover the covariance matrix is block diagonal, with each pair $(\mathbf{Q}_{j,k}, \mathbf{P}_{j,k})$ having a 2×2 the covariance matrix $V_{j,k}^{\mathrm{q}} = v_{j,k}\mathbf{1}$, and no correlation with the other coordinates, and the classical variables have covariance matrix

$$V_{ij}^{\mathrm{cl}} := \delta_{ij}\mu_i - \mu_i\mu_j, \qquad i, j = 1, \ldots d-1.$$

In summary, the limit Gaussian model consists of a tensor product between a Gaussian probability density and a density matrix of $d(d-1)/2$ independent quantum Gaussian states

$$(26) \qquad G(h, \mu) := \mathcal{N}\big(u, V^{\mathrm{cl}}\big) \otimes \bigotimes_{j<k} \Phi\big((\Re\tilde{\zeta}_{j,k}, \Im\tilde{\zeta}_{j,k}), V_{j,k}^{\mathrm{q}}\big).$$

We can now formulate the LAN Theorem which shows that the i.i.d. model can be asymptotically approximated by the Gaussian one, by means of quantum-classical randomisations, as discussed in Section 4.2. An alternative approach based on a generalisation of the notion of weak convergence of models, can be found in [13].

Theorem 3. *Let $\mathcal{Q}_N$ be the i.i.d. quantum model (25) and let*

$$\mathcal{G}_N := \{G(h, \mu) : \|h\| \le N^\epsilon\}.$$

be the Gaussian model with $G(h, \mu)$ defined in (26). *Then there exist channels (completely positive, normalised maps)*

$$T_N : M\left(\mathbb{C}^d\right)^{\otimes N} \to L^1\left(\mathbb{R}^{d-1}\right) \otimes \mathcal{T}_1\left(L^2(\mathbb{R})^{\otimes d(d-1)/2}\right),$$
$$S_N : L^1\left(\mathbb{R}^{d-1}\right) \otimes \mathcal{T}_1\left(L^2(\mathbb{R})^{\otimes d(d-1)/2}\right) \to M\left(\mathbb{C}^d\right)^{\otimes N},$$

such that

$$\lim_{N\to\infty} \Delta(\mathcal{Q}_N, \mathcal{G}_N) = 0,$$

where $\Delta(\cdot, \cdot)$ is the LeCam distance, cf. Definition 3.

Clearly, in the same i.i.d. setting, smooth lower-dimensional submodels of the model of a completely unknown state converge to a partial Gaussian shift model.

4.6. Asymptotic attainability of the Holevo bound

Besides its theoretical importance, local asymptotic normality has been used as a tool for solving various asymptotic problems such as optimal quantum learning [15], teleportation benchmarks [11], quantum state purification [2]. Here we give a short non-technical argument for the asymptotic attainability of the Holevo bound for i.i.d. models, using local asymptotic normality.

In Section 4.4 we showed that the Holevo bound is attained for arbitrary classical-quantum Gaussian shift models. We then saw that the model of N i.i.d. systems prepared in a completely unknown state converges locally to a Gaussian shift model with $(d^2 - 1)$ parameters. If some prior information about the state of the systems is available, we consider a lower dimensional model $\rho_\theta \in M(\mathbb{C}^d)$ with $\theta \in \Theta \subset \mathbb{R}^k$. By applying LAN to this sub-model of the "full" one, we find that it is approximated in the LeCam sense by a Gaussian shift of the form

$$\mathcal{G}' = \{G(Lh', \mu) : h' \in \mathbb{R}^k\},$$

where $L : \mathbb{R}^k \to \mathbb{R}^{d^2-1}$ is a linear map which depends only on the local behaviour of the restricted model around θ_0. To identify the linear transformation L we recall the correspondence between the collective variables and the limit continuous variables

$$(Lh')_a := \mathbb{E}_{G(h',\mu)}(\mathbf{X}_a) = \lim_{N\to\infty} \mathrm{Tr}(\rho_{h'}^N X_a(N)) = \sum_{i=1}^{k} h'_i \mathrm{Tr}\left(\frac{\partial \rho_{h'}}{\partial h'_i}\bigg|_{h=0} X_a\right)$$

from which we deduce

$$L_{i,a} = \mathrm{Tr}\left(\frac{\partial \rho_{h'}}{\partial h'_i}\bigg|_{h=0} X_a\right).$$

By a technical but otherwise rather standard argument, one can show that the asymptotic minimax risk for the problem of estimating the local parameter h' converges to the minimax risk for the same problem and the Gaussian model $\mathcal{G}'$, where in both cases the loss function is quadratic with weight matrix G

$$\lim_{N\to\infty} \inf_{M_N} \sup_{\|h'\|\leq N^\epsilon} NR(M_N, h') = R_{minmax}(\mathcal{G}', G).$$

The final step in proving the asymptotic attainability of the Holevo bound for finite dimensional systems it is to observe that its expression coincides with that of the minimax risk deduced in Section 4.4, applied to the Gaussian shift model $\mathcal{G}'$.

The optimisation (9) is performed over selfadjoint matrices satisfying the condition (8), which becomes (23) when translated into the cv language. Similarly, the real and imaginary parts of $Z(X)$ become the covariance and the symplectic matrices $V^{\mathbf{Y}}$ and respectively $S^{\mathbf{Y}}/2$.

References

[1] BARNDORFF-NIELSEN, O. E., GILL, R. D. AND JUPP, P. E. (2003). On quantum statistical inference. *J. Roy. Statist. Soc. (B)* **65** 775–816. With discussion and reply by the authors. Available at arXiv:quant-ph/0307191.

[2] BOWLES, P., GUŢĂ, M. AND ADESSO, G. (2011). Asymptotically optimal purification and dilution of mixed qubit and gaussian states. *Phys. Rev. A* **84** 022320.

[3] FUCHS, C. A. (1995). Distinguishability and accessible information in quantum theory. Ph.D. thesis, University of New Mexico. Available at arXiv:quant-ph/9601020.

[4] GILL, R. D. (2001). Asymptotics in quantum statistics. In *State of the Art in Probability and Statistics (Leiden, 1999)* (A. W. VAN DER VAART, M. DE GUNST AND C. A. J. KLAASSEN, eds.). *IMS Lecture notes Monogr. Ser.* **36** 255–285. Inst. Math. Statist., Beachwood, OH. Available at arXiv:math.ST/0405571.

[5] GILL, R. D. (2005). Conciliation of Bayes and pointwise quantum state estimation: Asymptotic information bounds in quantum statistics. Available at arXiv:math/0512443.

[6] GUŢĂ, M. AND KAHN, J. (2012). On asymptotic quantum statistical inference. Available at arXiv:abs/1112.2078.

[7] GILL, R. D. AND LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli* **1** 59–79.

[8] GILL, R. D. AND MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* **61** 042312–042335. Available at arXiv:quant-ph/9902063.

[9] GUŢĂ, M. AND KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* **73** 052108. Available at arXiv:quant-ph/0512075.

[10] GUŢĂ, M. (2011). Fisher information and asymptotic normality in system identification for quantum markov chains. *Phys. Rev. A* **83** 062324.

[11] GUŢĂ, M., BOWLES, P. AND ADESSO, G. (2010). Quantum-teleportation benchmarks for independent and identically distributed spin states and displaced thermal states. *Phys. Rev. A* **82** 042310.

[12] GUŢĂ, M., JANSSENS, B. AND KAHN, J. (2008). Optimal estimation of qubit states with continuous time measurements. *Commun. Math. Phys.* **277** 127–160.

[13] GUŢĂ, M. AND JENÇOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Commun. Math. Phys.* **276** 341–379.

[14] GUŢĂ, M. AND KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* **73** 052108.

[15] GUŢĂ, M. AND KOTLOWSKI, W. (2010). Quantum learning: asymptotically optimal classification of qubit states. *New J. Phys.* **12** 123032.

[16] KAHN, J. AND GUŢĂ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Commun. Math. Phys.* **289** 597–652.

[17] HAYASHI, M. (1998). Asymptotic estimation theory for a finite dimensional pure state model. *J. Phys. A* **31** 4633–4655. Available at arXiv:quant-ph/9704041.

[18] HAYASHI, M. (2003). Quantum estimation and quantum central limit theorem (in Japanese). *Sugaku* **55**(4) 368–391. New, English translation: arXiv:quant-ph/0608198.

[19] HAYASHI, M. (ed.). (2005). *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. World Scientific, Singapore.

[20] HAYASHI, M. AND MATSUMOTO, K. (2004). Asymptotic performance of optimal state estimation in quantum two level system. Available at arXiv:quant-ph/0411073.

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

[22] HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam. First appeared, in Russian, 1980.

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

[24] MATSUMOTO, K. (2002). A new approach to the Cramér-Rao-type bound of the pure state model. *J. Phys. A* **35** 3111–3123.

[25] NACHTERGAELE, B., SCHOLZ, V. B. AND WERNER, R. F. (2011). Local approximation of observables and commutator bounds. Available at arXiv:math-ph:1103.5663.

[26] NIELSEN, M. A. AND CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge University Press.

[27] OHYA, M. AND PETZ, D. (2004). *Quantum Entropy and its Use*. Springer Verlag, Berlin-Heidelberg.

[28] PETZ, D. AND JENCOVA, A. (2006). Sufficiency in quantum statistical inference. *Commun. Math. Phys.* **263** 259–276.

[29] VAN TREES, H. (1968). *Detection, Estimation and Modulation Theory, Part 1*. Wiley, New York.