An invitation to quantum tomography II

An invitation to quantum tomography

L.M. Artiles

Eurandom, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, artiles@eurandom.tue.nl, http://euridice.tue.nl/∼lartiles/

R. Gill

Mathematical Institute, University of Utrecht, Box 80010, 3508 TA Utrecht, The Netherlands, gill@math.uu.nl, http://www.math.uu.nl/people/gill

M.I. Guţă

Eurandom, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, guta@eurandom.tue.nl, http://euridice.tue.nl/∼mguta/

Summary. The quantum state of a light beam can be represented as an infinite dimensional density matrix or equivalently as a density on the plane called the Wigner function. We describe quantum tomography as an inverse statistical problem in which the state is the unknown parameter and the data is given by results of measurements performed on identical quantum systems. We present consistency results for Pattern Function Projection Estimators as well as for Sieve Maximum Likelihood Estimators for both the density matrix of the quantum state and its Wigner function. Finally we illustrate via simulated data the performance of the estimators. An EM algorithm is proposed for practical implementation. There remain many open problems, e.g. rates of convergence, adaptation, studying other estimators, etc., and a main purpose of the paper is to bring these to the attention of the statistical community.

1. Introduction

It took more than eighty years from its discovery till it was possible to experimentally determine and visualize the most fundamental object in quantum mechanics, the wave function. The forward route from quantum state to probability distribution of measurement results has been the basic stuff of quantum mechanics textbooks for decennia. That the corresponding mathematical inverse problem had a solution, provided (speaking metaphorically) that the quantum state has been probed from a sufficiently rich set of directions, had also been known for many years. However it was only with Smithey et al. (1993), that it became feasible to actually carry out the corresponding measurements on one particular quantum system—in that case, the state of one mode of electromagnetic radiation (a pulse of laser light at a given frequency). Experimentalists have used the technique to establish that they have succeeded in creating non-classical forms of laser light such as squeezed light and Schrödinger cats. The experimental technique we are referring to here is called quantum homodyne tomography: the word homodyne referring to a comparison between the light being measured with a reference light beam at the same frequency. We will explain the word tomography in a moment.

The quantum state can be represented mathematically in many different but equivalent ways, all of them linear transformations on one another. One favorite is as the Wigner function W: a real function of two variables, integrating to plus one over the whole plane, but not necessarily nonnegative. It can be thought of as a "generalized joint probability density" of the electric and magnetic

∗*Key words and phrases.* Quantum tomography, Wigner function, Density matrix, Pattern Functions estimation, Sieve Maximum Likelihood estimation, E.M. algorithm.

fields, q and p. However one cannot measure both fields at the same time and in quantum mechanics it makes no sense to talk about the values of both electric and magnetic fields simultaneously. It does, however, make sense to talk about the value of any linear combination of the two fields, say $\cos(\phi)q + \sin(\phi)p$. And one way to think about the statistical problem is as follows: the unknown parameter is a joint probability density W of two variables Q and P. The data consists of independent samples from the distribution of $(X, \Phi) = (\cos(\Phi)Q + \sin(\Phi)P, \Phi)$, where Φ is chosen independently of (Q, P), and uniformly in the interval $[0, \pi]$. Write down the mathematical model expressing the joint density of (X, Φ) in terms of that of (Q, P). Now just allow that latter joint density, W, to take negative as well as positive values (subject to certain restrictions which we will mention later). And that is the statistical problem of this paper.

This is indeed a classical tomography problem: we take observations from all possible one-dimensional projections of a two-dimensional density. The non-classical feature is that though all these one-dimensional projections are indeed bona-fide probability densities, the underlying two-dimensional "joint density" need not itself be a bona-fide joint probability density, but can have small patches of "negative probability".

Though the parameter to be estimated may look strange from some points of view, it is mathematically very nice from others. For instance, one can also represent it by a matrix of (a kind of) Fourier coefficients: one speaks then of the "density matrix" ρ. This is an infinite dimensional matrix of complex numbers, but it is a positive and selfadjoint matrix with trace one. The diagonal elements are real numbers summing up to one, and forming the probability distribution of the number of photons found in the light beam (if one could do that measurement). Conversely, any such matrix ρ corresponds to a physically possible Wigner function W, so we have here a concise mathematical characterization of precisely which "generalized joint probability densities" can occur.

The initial reconstructions were done by borrowing analytic techniques from classical tomography – the data was binned and smoothed, the inverse Radon transform carried out, followed by some Fourier transformations. At each of a number of steps, there are numerical discretization and truncation errors. The histogram of the data will not lie in the range of the forward transformation (from quantum state to density of the data). Thus the result of blindly applying an inverse will not be a bona-fide Wigner function or density matrix. Moreover the various numerical approximations all involve arbitrary choices of smoothing, binning or truncation parameters. Consequently the final picture can look just how the experimenter would like it to look and there is no way to statistically evaluate the reliability of the result. On the other hand the various numerical approximations tend to destroy the interesting "quantum" features the experimenter is looking for, so this method lost in popularity after the initial enthousiasm.

So far there has been little attention paid to this problem by statisticians, although on the one hand it is an important statistical problem coming up in modern physics, and on the other hand it is "just" a classical nonparametric statistical inverse problem. The unknown parameter is some object ρ, or if you prefer W, lying in an infinite dimensional linear space (the space of density matrices, or the space of Wigner functions; these are just two concrete representations in which the experimenter has particular interest). The data has a probability distribution which is a linear transform of the parameter. Considered as an analytical problem, we have an ill-posed inverse problem, but one which has a lot of beautiful mathematical structure and about which a lot is known (for instance, close connection to the Radon transform). Moreover it has features in common with nonparametric missing data problems (the projections from bivariate to univariate, for instance, and there are more connections we will mention later) and with nonparametric density and regression estimation. Thus we think that the time is ripe for this problem to be "cracked" by mathematical and computational statisticians. In this paper we will present some first steps in that direction.

Our main theoretical results are consistency theorems for two estimators. Both estimators are based on approximating the infinite dimensional parameter ρ by a finite dimensional parameter, in fact, thinking of ρ as an infinite dimensional matrix, we simply truncate it to an $N \times N$ matrix where the truncation level N will be allowed to grow with the number of observations n. The first estimator employs some analytical inverse formulas expressing the elements of ρ as mean values of certain functions, called pattern functions, of the observations (X, Φ). Simply replace the theoretical means by empirical averages and one has unbiased estimators of the elements of ρ, with moreover finite variance. If one applies this technique without truncation the estimate of the matrix ρ as a whole will typically not satisfy the nonnegativity constraints. The resulting estimator will not be consistent either, with respect to natural distance measures. But provided the truncation level grows with n slowly enough, the overall estimator will be consistent. The effect of truncating the density matrix ρ is to project on the subspace generated by the first elements of the corresponding basis, we shall call it the Pattern Function Projection estimator (PFP).

The second estimator we study is sieve maximum likelihood (SML) based on the same truncation to a finite dimensional problem. The truncation level N has to depend on sample size n in order to balance bias and variance. We prove consistency of the SML estimator under an appropriate choice of $N(n)$ by applying a general theorem of van de Geer (2000). To verify the conditions we need to bound certain metric entropy integrals (with bracketing) which express the size of the statistical model under study.

This turns out to be feasible, and indeed to have an elegant solution, by exploiting features of the mapping from parameters (density matrices) to distributions of the data. Various distances between probability distributions possess analogues as distances between density matrices, the mapping from parameter to data turns out to be a contraction, so we can bound metric entropies for the statistical model for the data with quantum metric entropies for the class of density matrices. And the latter can be calculated quite conveniently.

Our results form just a first attempt at studying the statistical properties of estimators which are already being used by experimental physicists, but they show that the basic problem is both rich in interesting features and tractable to analysis. The main results so far are consistency theorems for PFP and SML estimators, of both the density matrix and the Wigner function. These results depend on an assumption of the rate at which a truncated density matrix approximates the true one. It seems that the assumption is satisfied for the kinds of states which are met with in practice. However, further work is needed here to describe in physically interpretable terms, when the estimators work. Secondly, we need to obtain rates of consistency and to further optimize the construction of the estimator. Thirdly, one should explore the properties of penalized maximum likelihood. This will make the truncation level data driven. Fourthly, one should try to make the estimators adaptive with respect to the smoothness of the parameter. We largely restrict attention to an ideal case of the problem where there is no further noise in the measurements. In practice, the observations have added to them Gaussian disturbances of known variance. There are some indications that when the variance is larger than a threshold of $1/2$, reconstruction becomes qualitatively much more difficult. This needs to be researched from the point of view of optimal rates of convergence.

We also only considered one particular though quite convenient way of sieving the model, i.e., one particular class of finite dimensional approximations. There are many other possibilities and some of them might allow easier analysis and easier computation. For instance, instead of truncating the matrix ρ in a given basis, one could truncate in an arbitrary basis, so that the finite dimensional approximations would correspond to specifying N arbitrary state vectors and a probability distribution over these "pure states". Now the problem has become a missing data problem, where the "full data" would assign to each observation also the label of the pure state from which it came. In the full data problem we need to reconstruct not a matrix but a set of vectors, together with an

ordinary probability distribution over the set, so the "full data" problem is statistically speaking a much easier problem that the missing data problem. We shall use a version of this to derive an Expectation-Maximization algorithm (EM) for the practical implementation of the SML estimator, see Section 5. One could imagine that Bayesian reconstruction methods could also exploit this structure.

The analogy with density estimation could suggest new statistical approaches here. Finally, it is most important to add to the estimated parameter, estimates of its accuracy. This is absolutely vital for applications, but so far no valid approach is available.

In Section 2 we introduce first, very briefly, the statistical problems we are concerned with. We then give a short review of the basic notions of quantum mechanics which are needed in this paper. Concepts such as observables, states, measurements and quantum state tomography are explained by using finite dimensional complex matrices. At the end of the section we show how to generalize this to the infinite dimensional case and describe the experimental set-up of Quantum Homodyne Detection pointing out the relation with computerized tomography.

In Section 3 we present results on consistency of density matrix estimators: projection estimator based on pattern functions, and sieve maximum likelihood estimator. The last subsection extends the results of the previous ones to estimating the Wigner function.

Section 4 deals with the detection losses occurring in experiments due in part to the inefficiency of the detectors. This adds a deconvolution problem on the top of our tomography estimation.

Section 5 shows experimental results. We illustrate the behavior of the studied estimators and propose some practical tools for the implementation — EM algorithm. The last section finishes with some concluding remarks to the whole paper and open problems. The main purpose of the paper is to bring the attention of the statistical community to these problem; thus, some proofs are just sketched. They fully appear in a complementary paper, Guţă (2004).

2. Physical background

Our statistical problem is to reconstruct the quantum state of light by using data obtained from measuring identical pulses of light through a technique called Quantum Homodyne Detection (QHD). In particular we will estimate the quantum state in two different representations or parameterizations: the density matrix and the Wigner function. The physics behind this statistical problem is presented in subsection 2.2 which serves as introduction to basic notions of quantum theory, and subsection 2.3 which describes the model of quantum homodyne detection from quantum optics. The relations between the different parameters of the problem are summarized in the diagram at the end of subsection 2.3 followed by Table 1 containing some examples of quantum states. For the reader who is not interested in the quantum physics background we state the statistical problem in the next subsection and we will return to it in Section 3.

2.1. Summary of statistical problem

We observe $(X_1, \Phi_1), \ldots, (X_n, \Phi_n)$, i.i.d. random variables with values in $\mathbb{R} \times [0, \pi]$ and distribution P_ρ depending on the unknown parameter ρ which is an infinite dimensional matrix $\rho = [\rho_{j,k}]_{j,k=0,\ldots,\infty}$ such that $\mathrm{Tr}\rho = 1$ (trace one) and $\rho \geq 0$ (positive definite). The probability density of P_ρ is

$$p_\rho(x, \phi) = \frac{1}{\pi} \sum_{j,k=0}^{\infty} \rho_{j,k} \psi_k(x) \psi_j(x) e^{-i(j-k)\phi}, \tag{1}$$

where the functions $\{\psi_n\}$ to be specified later, form a orthonormal basis of the space of complex square integrable functions on $\mathbb{R}$. Because ρ is positive definite and has trace 1, this is a probability density: real, nonnegative, integrates to 1. The data (X_ℓ, Φ_ℓ) comes from independent QHD measurements on identically prepared pulses of light whose properties or state are completely encoded in the matrix ρ called a density matrix. We will consider the problem of estimating ρ from a given sample.

Previous attempts by physicists to estimate the density matrix ρ have focused mainly on the estimation of the individual matrix elements without considering the accuracy of the estimated density matrix with respect to natural distances of the underlying parameter space. In Section 3 we will present consistency results in the space of density matrices w.r.t. L_1 and L_2-norms using two different types of estimators, namely projection and sieve maximum likelihood estimators.

An alternative representation of the quantum state is through the Wigner function $W_\rho : \mathbb{R}^2 \to \mathbb{R}$ whose estimation is close to a classical computer tomography problem namely, Positron Emission Tomography (PET), Vardi et al. (1985). In PET one would like to estimate a probability density f on $\mathbb{R}^2$ from i.i.d. observations $(X_1, \Phi_1), \ldots, (X_n, \Phi_n)$, with probability density equal to the Radon transform of f:

$$\mathcal{R}[f](x, \phi) = \int_{-\infty}^{\infty} f(x\cos\phi + t\sin\phi, x\sin\phi - t\cos\phi)dt.$$

Although the Wigner function is in general *not* positive, its Radon transform is always a probability density, in fact $\mathcal{R}[W_\rho](x, \phi) = p_\rho(x, \phi)$. As the Wigner function W_ρ is in one-to-one correspondence with the density matrix ρ, our state reconstruction problem can be stated as to estimate the Wigner function W_ρ. This is an ill posed inverse problem as seen from the formula for the inverse of the Radon transform

$$W_\rho(q, p) = \frac{1}{2\pi^2}\int_0^\pi \int_{-\infty}^{+\infty} p_\rho(x, \phi)K(q\cos\phi + p\sin\phi - x)\, dx d\phi, \tag{2}$$

where

$$K(x) = \frac{1}{2}\int_{-\infty}^{+\infty} |\xi| \exp(i\xi x)d\xi, \tag{3}$$

makes sense only as a generalized function. To correct this one usually makes a cut-off in the range of the above integral and gets a well behaved kernel function $K_c(x) = \frac{1}{2}\int_{-c}^{c} |\xi| \exp(i\xi x)d\xi$. Then the tomographic estimator of W_ρ is the average sampled kernel

$$\widehat{W}_{c,n}(q, p) = \frac{1}{2\pi^2 n}\sum_{\ell=1}^{n} K_c(q\cos\Phi_\ell + p\sin\Phi_\ell - X_\ell).$$

For consistency one needs to let the 'bandwidth' $h = 1/c$ depend on the sample size n and $h_n \to 0$ as $n \to \infty$ at an appropriate rate.

In this paper we will not follow this approach, which will be treated separately in future work. Instead, we use a plug-in type estimator based on the property

$$W_\rho(q, p) = \sum_{k,j=0}^{\infty} \rho_{k,j} W_{k,j}(q, p), \tag{4}$$

where $W_{k,j}$'s are known functions and we replace ρ by its above mentioned estimators. We shall prove consistency of the proposed estimators of the Wigner function w. r. t. L_2 and supremum norms in the corresponding space.

2.2. Quantum systems and measurements

This subsection serves as a short introduction to the basic notions of quantum mechanics which will be needed in this paper. For simplicity we will deal first with finite dimensional quantum systems and leave the infinite dimensional case for the next subsection. For further details on quantum statistical inference we refer to the review Barndorff-Nielsen et al. (2003) and the classic textbooks Helstrom (1976) and Holevo (1982).

In classical mechanics the state of macroscopic systems like billiard balls, pendulums or stellar systems is described by points on a manifold or "phase space", each of the point's coordinates corresponding to an attribute which we can measure such as position and momentum. Therefore the functions on the phase space are called observables. When there exists uncertainty about the exact point in the phase space, or we deal with a statistical ensemble, the state is modelled by a probability distribution on the phase space, and the observables become random variables.

Quantum mechanics also deals with observables such as position and momentum of a particle, spin of an electron, number of photons in a cavity, but breaks from classical mechanics in that these are no longer represented by functions on a phase space but by Hermitian matrices, that is, complex valued matrices which are invariant under transposition followed by complex conjugation. For example, the components in different directions of the spin of an electron are certain 2×2 complex Hermitian matrices $\sigma_x, \sigma_y, \sigma_z$.

Any d-dimensional complex Hermitian matrix $\mathbf{X}$ can be diagonalized by changing the standard basis of $\mathbb{C}^d$ to another orthonormal basis $\{e_1, \ldots, e_d\}$ such that $\mathbf{X}e_i = x_i e_i$ for $i = 1, \ldots, d$, with $x_i \in \mathbb{R}$. The vectors e_i and numbers x_i are called eigenvectors and respectively eigenvalues of $\mathbf{X}$. With respect to the new basis we can write

$$\mathbf{X} = \begin{pmatrix} x_1 & 0 & 0 & \ldots & 0 \\ 0 & x_2 & 0 & \ldots & 0 \\ 0 & 0 & x_3 & \ldots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \ldots & x_d \end{pmatrix}$$

The physical interpretation of the eigenvalues is that when measuring the observable $\mathbf{X}$ we obtain (randomly) one of the values x_i according to a probability distribution depending on the state of the system before measurement and on the observable $\mathbf{X}$. This probability measure is degenerate if and only if the system before measurement was prepared in a special state called an eigenstate of $\mathbf{X}$. We represent such a state mathematically by the projection $\mathbf{P}_i$ onto the one dimensional space generated by the vector e_i in $\mathbb{C}^d$. Given a probability distribution $\{p_1, \ldots, p_d\}$ over the finite set $\{x_1, \ldots, x_d\}$, we describe a statistical ensemble in which a proportion p_i of systems is prepared in the state $\mathbf{P}_i$ by the convex combination $\rho = \sum_i p_i \mathbf{P}_i$. The expected value of the random result X when measuring the observable $\mathbf{X}$ for this particular state is equal to $\sum_i p_i x_i$ which can be written shortly

$$\mathbb{E}_\rho(X) := \mathrm{Tr}(\rho \mathbf{X}). \tag{5}$$

Similarly, the probability distribution can be recovered as

$$p_i = \mathrm{Tr}(\rho \mathbf{P}_i) \tag{6}$$

thanks to the orthogonality property $\mathrm{Tr}(\mathbf{P}_i \mathbf{P}_j) = \delta_{ij}$.

Now, let $\mathbf{Y}$ be a different observable and suppose that $\mathbf{Y}$ does not commute with $\mathbf{X}$, that is $\mathbf{XY} \neq \mathbf{YX}$, then the two observables cannot be diagonalized in the same basis, their eigenvectors are different. Consequently, states which are mixtures of eigenvectors of $\mathbf{X}$ typically will not be

mixtures of eigenvectors of $\mathbf{Y}$ and vice-versa. This leads to an expanded formulation of the notion of state in quantum mechanics independent of any basis associated to a particular observable, and the recipe for calculating expectations and distributions of measurement results.

Any preparation procedure results in an statistical ensemble, or state, which is described mathematically by a matrix ρ with the following properties

(a) $\rho \geq 0$ (positive definite matrix),
(b) $\mathrm{Tr}(\rho) = 1$ (normalization).

In physics ρ is called a *density matrix*, and is for a quantum mechanical system an analogue of a probability density. Notice that the special state $\sum_i p_i \mathbf{P}_i$ defined above is a particular case of density matrix, since it is a mixture of eigenstates of the observable $\mathbf{X}$. The density matrices of dimension d form a convex set $\mathcal{S}_d$, whose extremals are the *pure* or *vector* states, represented by orthogonal projections $\mathbf{P}(\psi)$ onto one dimensional spaces spanned by *arbitrary* vectors $\psi \in \mathbb{C}^d$. Any state can be represented as a mixture of pure states which are not necessarily eigenstates of a particular observable.

When measuring an observable, for example $\mathbf{X}$, of a quantum system prepared in the state ρ we obtain a random result $X \in \{x_1, \ldots, x_d\}$ with probability distribution given by equation (6), expectation as in equation (5), and characteristic function

$$G(t) := \mathbb{E}_\rho\big(\exp(itX)\big) = \mathrm{Tr}\big(\rho \exp(it\mathbf{X})\big). \tag{7}$$

In order to avoid confusion we stress the important difference between $\mathbf{X}$ which is a matrix and X which is a real-valued random variable. More concretely, if we write ρ in the basis of eigenvectors of $\mathbf{X}$ then we obtain the map $\mathbf{M}$ from states to probability distributions $P_\rho^{(\mathbf{X})}$ over results $\{x_1, \ldots, x_d\}$

$$\mathbf{M} : \rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2d} \\ & & \ddots & \\ \rho_{d1} & \rho_{d2} & \cdots & \rho_{dd} \end{pmatrix} \longmapsto P_\rho^{(\mathbf{X})} = \begin{pmatrix} \rho_{11} \\ \rho_{22} \\ \vdots \\ \rho_{dd} \end{pmatrix}.$$

Notice that $P_\rho^{(\mathbf{X})}$ is indeed a probability distribution as a consequence of the defining properties of states, and it does not contain information about the off-diagonal elements of ρ, meaning that measuring only the observable $\mathbf{X}$ is not enough to identify the unknown state. Roughly speaking, as $\dim(\mathcal{S}_d) = d^2 - 1 = (d-1)(d+1)$, one has to measure on many identical systems each one of a number of $d+1$ mutually non-commuting observables in order to have a one-to-one map between states and probability distributions of results. The probing of identically prepared quantum systems from different 'angles' in order to reconstruct their state is broadly named *quantum state tomography* in the physics literature.

Let us suppose that we have at our disposal n systems identically prepared in an unknown state $\rho \in \mathcal{S}_d$, and for each of the systems we can measure one of the fixed observables $\mathbf{X}(1), \ldots, \mathbf{X}(d+1)$. We write the observables in diagonal form

$$\mathbf{X}(i) = \sum_{a=1}^{d} x_{i,a} \mathbf{P}_{i,a} \tag{8}$$

where $x_{i,a}$ eigenvalues and $\mathbf{P}_{i,a}$ eigenstates. We will perform a randomized experiment, i.e. for each system we will choose the observable to be measured by randomly selecting its index according to

a probability distribution P^{Φ} over $\{1, \ldots, d+1\}$. The results of the measurement on the k^{th} system are the pair $Y_k = (X_k, \Phi_k)$ where $\Phi_1, \ldots, \Phi_n$ are i.i.d. with probability distribution $P^{(\Phi)}$ and X_k is the result of measuring the observable $\mathbf{X}(\Phi_k)$ whose conditional distribution is given by

$$P^{\mathbf{M}}(X_k = x_{i,a} | \Phi_k = i) = \mathrm{Tr}(\rho \mathbf{P}_{i,a}) \tag{9}$$

The statistical problem is now to estimate the parameter ρ from the data $Y_1, Y_2, \ldots, Y_n$. In the next subsection we will describe quantum homodyne tomography as an analogue of this problem for infinite dimensional systems.

2.3. Quantum homodyne tomography

Although correct and sufficient when describing certain quantum properties such as the spin of a particle, the model presented above needs to be enlarged in order to cope with quantum systems with 'continuous variables' which will be central in our statistical problem. This technical point can be summarized as follows: we replace $\mathbb{C}^d$ by an infinite dimensional complex Hilbert space $\mathcal{H}$, the Hermitian matrices becoming *selfadjoint operators* acting on $\mathcal{H}$. The spectral theorem tells us that selfadjoint operators can be 'diagonalized' in the spirit of (2.2) but the spectrum (the set of 'eigenvalues') can have a more complicated structure, for example it can be continuous as we will see below. The density matrices are *positive* selfadjoint operators ρ such that $\mathrm{Tr}(\rho) = 1$ and can be regarded as infinite dimensional matrices with elements $\rho_{k,j} := \langle \psi_j, \rho \psi_k \rangle$ for a given orthonormal basis $\{\psi_1, \psi_2, \ldots\}$ in $\mathcal{H}$.

The central example of a system with continuous variables in this paper is the quantum particle. Its basic observables position and momentum, are two unbounded selfadjoint operators $\mathbf{Q}$ and $\mathbf{P}$ respectively, acting on $L^2(\mathbb{R})$, the space of square integrable complex valued functions on $\mathbb{R}$

$$(\mathbf{Q}\psi_1)(x) = x\psi_1(x),$$
$$(\mathbf{P}\psi_2)(x) = -i\frac{d\psi_2(x)}{dx},$$

for ψ_1, ψ_2 arbitrary functions. The operators satisfy Heisenberg's *commutation relations* $\mathbf{QP} - \mathbf{PQ} = i\mathbf{1}$ which implies that they cannot be measured simultaneously. The problem of (separately) measuring such observables has been elusive until ten years ago when pioneering experiments in quantum optics by Smithey et al. (1993), led to a powerful measurement technique called *quantum homodyne detection*. This technique is the basis of a continuous analogue of the measurement scheme presented at the end of the previous subsection where $d+1$ observables were measured in the case of a d-dimensional quantum system.

The quantum system to be measured is a beam of light with a fixed frequency whose observables are the electric and magnetic field amplitudes which satisfy commutation relations identical to those characterizing the quantum particle, with which they will be identified from now on. Their linear combinations $\mathbf{X}_\phi = \cos\phi\mathbf{Q} + \sin\phi\mathbf{P}$ are called *quadratures*, and homodyne detection is about measuring the quadratures for *all* phases $\phi \in [0, \pi]$. The experimental setup shown in Figure 1 contains an additional laser called a local oscillator (LO) of high intensity $|z| \gg 1$ and relative phase ϕ with respect to the beam in the unknown state ρ. The two beams are combined through a fifty-fifty beam splitter, and the two emerging beams are then measured by two photon detectors. A simple quantum optics computation (see Leonhardt (1997)) shows that in the limit of big LO intensity the difference of the measurement results (countings) of the two detectors rescaled by the LO intensity $X = \frac{I_1 - I_2}{|z|}$ has the probability distribution corresponding to the measurement of the quadrature $\mathbf{X}_\phi$. The result X takes values in $\mathbb{R}$ and its probability distribution $P_\rho(\cdot|\phi)$ has a density

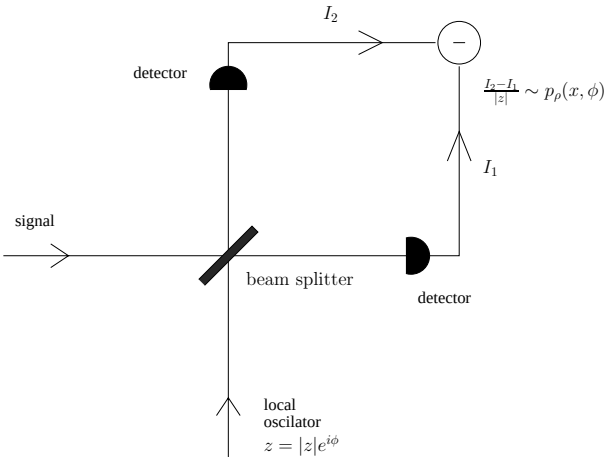


Fig. 1. Quantum Homodyne Tomography measurement system

$p_\rho(x|\phi)$ and characteristic function (see equation (7))

$$G_{\rho,\mathbf{X}_\phi}(t) = \mathrm{Tr}\big(\rho \exp(it\mathbf{X}_\phi)\big). \tag{10}$$

The phase ϕ can be controlled by the experimenter by adjusting a parameter of the local oscillator. We assume that he chooses it randomly uniformly distributed over the interval $[0, \pi]$. Then the joint probability distribution P_ρ for the pair consisting in measurement result and phase $Y := (X, \Phi)$ has density $p_\rho(x, \phi)$ equal to $\frac{1}{\pi} p_\rho(x|\phi)$ with respect to the measure $dx \times d\phi$ on $\mathbb{R} \times [0, \pi]$. An attractive feature of the homodyne detection scheme is the invertibility of the map $\mathbf{T}$ that associates P_ρ to ρ, making it possible to asymptotically infer the unknown parameter ρ from the i.i.d. results $Y_1, \ldots, Y_n$ of homodyne measurements on n systems prepared in the state ρ.

We will see now why this state estimation method is called *quantum homodyne tomography* by drawing a parallel with computerized tomography used in the hospitals. In quantum optics it is common to represent the state of a quantum system by a certain function on $\mathbb{R}^2$ called the *Wigner function* $W_\rho(q, p)$ which is much like a joint probability density for $\mathbf{Q}$ and $\mathbf{P}$, for instance its marginals are the probability densities for measuring $\mathbf{Q}$ and respectively $\mathbf{P}$. The Wigner function of the state ρ is defined by demanding that its Fourier transform $\mathcal{F}_2$ with respect to *both* variables has the following property

$$\widetilde{W}_\rho(u, v) := \mathcal{F}_2[W_\rho](u, v) = \mathrm{Tr}\big(\rho \exp(-iu\mathbf{Q} - iv\mathbf{P})\big). \tag{11}$$

We see from this equation that if $\mathbf{Q}$ and $\mathbf{P}$ were commuting operators then $W(q, p)$ would indeed be the joint probability density of outcomes of their measurement. As the two observables cannot be measured simultaneously, we cannot speak of a joint density, in fact the Wigner function need not be positive, but many interesting features of the quantum state can be visualized in this way. Let $(u, v) = (t\cos\phi, t\sin\phi)$, then

$$\widetilde{W}(u, v) = \mathcal{F}_1[p(\cdot, \phi)](t) = \mathrm{Tr}\big(\rho \exp(-it\mathbf{X}_\phi)\big) \tag{12}$$

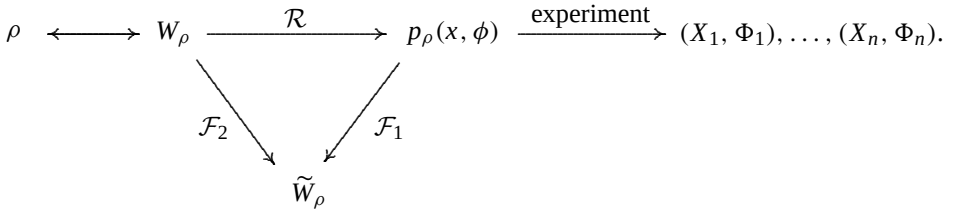
where the Fourier transform $\mathcal{F}_1$ in the last term is with respect to the first variable, keeping ϕ fixed.

The equations (11) and (12) are well known in the theory of Radon transform $\mathcal{R}$ and imply that for each fixed ϕ, the probability density $p_\rho(x,\phi)$ is the marginal of the Wigner function with respect to the direction ϕ in the plane,

$$p_\rho(x,\phi) = \mathcal{R}[W_\rho](x,\phi) = \int_{-\infty}^{\infty} W_\rho(x\cos\phi + y\sin\phi, x\sin\phi - y\cos\phi)dy, \tag{13}$$

adding quantum homodyne tomography to a number of applications ranging from computerized tomography to astronomy and geophysics, Deans (1983). In computerized tomography one reconstructs an image of the tissue distribution in a cross-section of the human body by recording events whereby pairs of positrons emitted by an injected radioactive substance hit detectors placed in a ring around the body after flying in opposite directions along an axis determined by an angle $\phi \in [0,\pi]$. In quantum homodyne tomography the role of the unknown distribution is played by the Wigner function which is in general not positive, but has a probability density $p_\rho(x|\phi)$ as marginal along any direction ϕ.

The following diagram summarizes the relations between the various objects in our problem:



Finally in Table 1 we give some examples of density matrices and their corresponding Wigner function representations for different states. The matrix elements $\rho_{k,j}$ are calculated with respect to the orthonormal base corresponding to the wave functions of k photons states

$$\psi_k(x) = H_k(x)e^{-x^2/2} \tag{14}$$

where H_k are the Hermite polynomials normalized such that $\int \psi_k^2 = 1$. A few graphical representations can be seen in Figure 3.

State	Density matrix $\rho_{k,j}$	Wigner function $W(q,p)$
Vacuum state	$\rho_{0,0} = 1$, rest zero	$\frac{1}{\pi}\exp(-q^2-p^2)$
Single Photon state	$\rho_{1,1} = 1$, rest zero	$\frac{1}{\pi}(2q^2+2p^2-1)\exp(-q^2-p^2)$
Thermal state $\beta > 0$	$\delta_k^j(1-e^{-\beta})e^{-\beta k}$	$\frac{1}{\pi}\tanh(\beta/2)\exp[-(q^2+p^2)\tanh(\beta/2)]$
Coherent state, $N \in \mathbb{R}_+$	$\exp(-N)\frac{N^{k+j}}{\sqrt{k!j!}}$	$\frac{1}{\pi}\exp(-(q-\sqrt{N})^2-p^2)$
Squeezed state $N \in \mathbb{R}_+$, $\xi \in \mathbb{R}$,	$C(N,\xi)(\frac{1}{2}\tanh(\xi))^{k+j}\times$ $H_j(\gamma)H_k(\gamma)/\sqrt{j!k!}$	$\frac{1}{\pi}\exp(-e^{2\xi}(q-\alpha)^2-e^{-2\xi}p^2)$

Table 1: Density matrix and Wigner function of some quantum states

The vacuum is the pure state of zero photons, notice that in this case the distributions of **Q** and **P** are Gaussian. The thermal state is a mixed state describing equilibrium at temperature $T = 1/\beta$,

having Gaussian Wigner function with variance increasing with the temperature. The coherent state is pure and characterizes the laser pulse. The photon number is Poisson distributed with an average of N photons. The squeezed states have Gaussian Wigner functions whose variances for the two directions are different but have a fixed product. The parameters N and ξ satisfy the condition $N \geq \sinh^2(\xi)$, $C(N, \xi)$ is a normalization constant, $\alpha = \frac{(N-\sinh^2(\xi))^{1/2}}{\cosh(\xi)-\sinh(\xi)}$, and $\gamma = (\frac{\alpha}{\sinh(2\xi)})^{1/2}$.

3. Density matrix estimation

D'Ariano et al. (1994) presented the density matrix analogue of formula (2) of the Wigner function as inverse Radon transform of the probability density p_ρ

$$\rho = \int_{-\infty}^{\infty} dx \int_0^{\pi} \frac{d\phi}{\pi} p_\rho(x, \phi) K(\mathbf{X}_\phi - x\mathbf{1}), \tag{15}$$

where K is the generalized function given in equation (3) whose argument is a selfadjoint operator $\mathbf{X}_\phi - x\mathbf{1}$. The method has been further analyzed in D'Ariano (1995), Leonhardt et al. (1995), see also D'Ariano et al. (1995). We recall that in the case of the Wigner function we needed to regularize the kernel K by introducing a cut-off in the integral (3). For density matrices the philosophy will be rather to project on a finite dimensional subspace of $L_2(\mathbb{R})$ whose dimension N will play the role of the cut-off. In fact all the matrix elements of the density matrix ρ with respect to the orthonormal basis $\{\psi_k\}_{k=0}^{\infty}$ defined in (14), can be expressed as kernel integrals

$$\rho_{k,j} = \int_{-\infty}^{\infty} dx \int_0^{\pi} \frac{d\phi}{\pi} p_\rho(x, \phi)\, f_{k,j}(x) e^{-i(j-k)\phi}, \tag{16}$$

with $f_{k,j} = f_{j,k}$ bounded real valued functions which in the quantum tomography literature are called *pattern functions*. The singularity of the kernel K is reflected in the asymptotic behavior of $f_{k,j}$ as $k, j \to \infty$. A first formula for $f_{k,j}$ was found in Leonhardt et al. (1995) and uses Laguerre polynomials. This was followed by a more transparent one due to Leonhardt et al. (1996),

$$f_{k,j}(x) = \frac{d}{dx}(\psi_k(x)\varphi_j(x)), \tag{17}$$

for $j \geq k$, where ψ_k and φ_j represent the square integrable and respectively the unbounded solutions of the Schrödinger equation,

$$\left[-\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2\right]\psi = \omega\,\psi, \qquad \omega \in \mathbb{R}. \tag{18}$$

Figure 2 shows pattern functions for different values of k and j. We notice that the oscillatory part is concentrated in an interval centered at zero whose length increase with k and j, the number of oscillations increases with k and j and the functions become more irregular as we move away from the diagonal. It can be shown that tails of the pattern function decay like $x^{-2-|k-j|}$. More properties of the pattern function can be found in Leonhardt et al. (1996) and Guţă (2004).

3.1. Pattern function projection estimation

Equation (16) suggests the *unbiased estimator* $\hat{\rho}^{(n)}$ of ρ, based on n i.i.d. observations of (X, Φ), whose matrix elements are:

$$\hat{\rho}^{(n)}_{k,j} = \frac{1}{n}\sum_{\ell=1}^{n} F_{k,j}(X_\ell, \Phi_\ell), \tag{19}$$

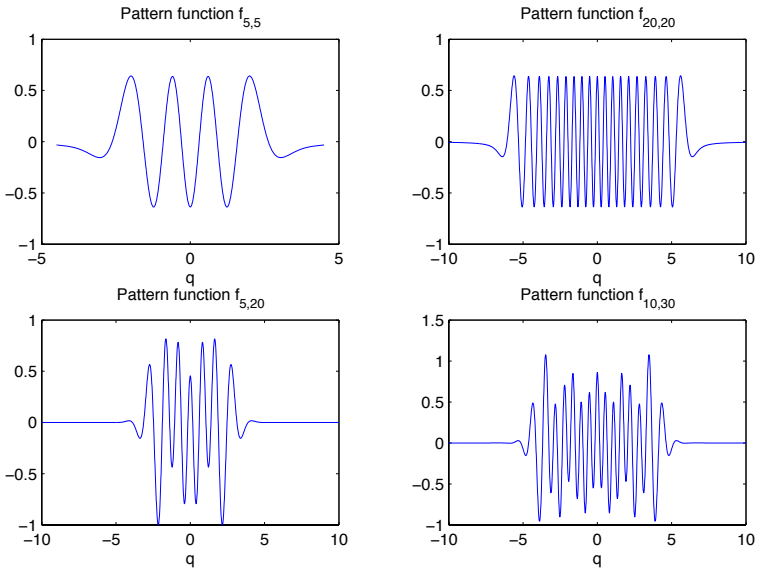


Fig. 2. Pattern functions

where $F_{k,j}(x,\phi) = f_{k,j}(x)e^{-i(j-k)\phi}$, see D'Ariano et al. (1995), Leonhardt et al. (1995, 1996). By the strong law of large numbers the individual matrix elements of this estimator converge to the matrix elements of the true parameter ρ. However the infinite matrix $\hat{\rho}^{(n)}$ need not be positive, normalized, or even selfadjoint, thus it cannot be interpreted as a quantum state. These problems are similar to those encountered when trying to estimate an unknown probability density by using unbiased estimators for all its Fourier coefficients. The remedy is to estimate only a finite number of coefficients at any moment, obtaining a projection estimator onto the subspace generated by linear combinations of a finite subset of the basis vectors. In our case we will project onto the space of matrices of dimension $N = N(n)$ with respect to the basis $\{\psi_k\}_{k=0}^{\infty}$,

$$\hat{\rho}_{k,j}^{(N,n)} = \frac{1}{n}\sum_{\ell=1}^{n} F_{k,j}(X_\ell, \Phi_\ell), \qquad \text{for } 0 \le k,\, j \le N-1,$$

and $\hat{\rho}_{k,j}^{(n)} = 0$ for $(k \wedge j) \ge N$.

In order to test the performance of our estimators we introduce the L_1 and L_2 distances on the space of density matrices. Let ρ and τ be two density matrices with $\rho - \tau = \sum_i \lambda_i \mathbf{P}_i$ the diagonal form of their difference, and notice that some of the eigenvalues are positive and some negative such that their sum is zero due to the normalization of the density matrices. We define the absolute value $|\rho - \tau| := \sum_i |\lambda_i| \mathbf{P}_i$ and the norms

$$\|\rho - \tau\|_1 := \mathrm{Tr}(|\rho - \tau|), \tag{20}$$

$$\|\rho - \tau\|_2 := \left[\mathrm{Tr}(|\rho - \tau|^2)\right]^{1/2} = \left[\sum_{j,k=0}^{\infty} |\rho_{k,j} - \tau_{k,j}|^2\right]^{1/2}. \tag{21}$$

Let us consider now the mean integrated square error (MISE) and split it into the bias and variance parts:

$$\mathbb{E}\Big(\sum_{j,k=0}^{\infty}|\hat{\rho}_{k,j}^{(N,n)}-\rho_{k,j}|^2\Big)=\sum_{(k\wedge j)\geq N}|\rho_{k,j}|^2+\mathbb{E}\Big(\sum_{j,k=0}^{N-1}|\hat{\rho}_{k,j}^{(N,n)}-\rho_{k,j}|^2\Big)=b^2(n)+\sigma^2(n). \quad (22)$$

By choosing $N(n)\to\infty$ as $n\to\infty$ the bias $b^2(n)$ converges to zero. For the variance we have the upper bound

$$\begin{aligned}\sigma^2(n) &= \frac{1}{n}\sum_{j,k=0}^{N-1}\int_0^{\pi}\int_{\mathbb{R}}|F_{k,j}(x,\phi)-\rho_{k,j}|^2 p_\rho(x,\phi)d\phi dx\\ &\leq \frac{1}{n}\sum_{j,k=0}^{N-1}\int_0^{\pi}\int_{\mathbb{R}}|F_{k,j}(x,\phi)|^2 p_\rho(x,\phi)d\phi dx\\ &= \frac{1}{n}\sum_{j,k=0}^{N-1}\int_{\mathbb{R}}dx\, f_{k,j}(x)^2\int_0^{\pi}p_\rho(x,\phi)d\phi. \quad (23)\end{aligned}$$

The proof of the following lemma on the norms of the pattern functions can be found in Guţă (2004).

LEMMA 3.1. *There exist constants C_1, C_2, C_3 such that*

$$\sum_{k,j\geq 0}^{N}\|f_{k,j}\|_\infty^2\leq C_1N^{7/3}, \quad (24)$$

$$\sum_{k,j\geq 0}^{N}\|f_{k,j}\|_2^2\leq C_2N^3, \quad (25)$$

$$\int_0^{2\pi}p_\rho(x,\phi)d\phi\leq C_3,\qquad \text{for all } x\in\mathbb{R}. \quad (26)$$

By applying the lemma to equation (23) we conclude that the estimator $\hat{\rho}^{(N,n)}$ is consistent with respect to the L_2 distance if we choose $N(n)\to\infty$ as $n\to\infty$ such that $N(n)=o(n^{3/7})$. Based on the property $\|\rho^{(N,n)}\|_1\leq\sqrt{N}\|\rho^{(N,n)}\|_2$ we can prove a similar result concerning L_1-consistency, see Guţă (2004).

THEOREM 3.2. *Let $N(n)\to\infty$ be the dimension of the pattern function projection estimator. If $N(n)=o(n^{3/7})$, then*

$$\mathbb{E}\big(\|\hat{\rho}^{(N,n)}-\rho\|_2^2\big)\to 0,\qquad \text{as } n\to\infty. \quad (27)$$

If $N(n)=o(n^{3/10})$ then

$$\mathbb{E}\big(\|\hat{\rho}^{(N,n)}-\rho\|_1\big)\to 0,\qquad \text{as } n\to\infty. \quad (28)$$

Rates of consistency can be obtained by assuming that the state belongs to a given class for which upper bounds of the bias can be calculated and $N(n)$ is chosen such as to balance bias and variance. This problem will be attacked in future work within the minimax framework. In Section 5 we present a data-dependent way of selecting the dimension of the projection estimator based on the minimization of the empirical L_2-risk using a cross-validation technique.

3.2. Sieve maximum likelihood estimation

We will consider now a maximum likelihood approach to the estimation of the state ρ. Let us recall the terms of the problem: we are given a sequence $Y_1, Y_2 \ldots, Y_n$ of i.i.d. random variables $Y_i = (X_i, \Phi_i)$ with values in $\mathbb{R} \times [0, \pi]$ and probability density p_ρ depending on the parameter ρ which is an infinite dimensional density matrix. We would like to find an estimator of ρ

$$\bar{\rho}^{(n)} = \bar{\rho}^{(n)}(Y_1, Y_2, \ldots, Y_n).$$

Let ρ and τ be density matrices and denote by

$$h(P_\rho, P_\tau) := \left(\int (\sqrt{p_\rho} - \sqrt{p_\tau})^2 \, dx d\varphi \right)^{1/2}, \tag{29}$$

the Hellinger distance between the two probability distributions. The following relations are well known

$$d_{\mathrm{tv}}(P_\rho, P_\tau) = \frac{1}{2} \| p_\rho - p_\tau \|_1 \leq h(P_\rho, P_\tau) \leq \sqrt{\| p_\rho - p_\tau \|_1}, \tag{30}$$

An important property which is true for any measurement is the following inequality between the classical and quantum distances, cf. Guţă (2004),

$$\| p_\rho - p_\tau \|_1 \leq \| \rho - \tau \|_1. \tag{31}$$

From (30) and (31) we obtain

$$h(P_\rho, P_\tau) \leq \sqrt{\| \rho - \tau \|_1}, \tag{32}$$

for arbitrary states ρ, τ. As we have seen previously, the inverse map from P_ρ to ρ is unbounded thus we do not have an inequality in the opposite direction to the one above. However we can prove the continuity of the inverse map by using a matrix analogue of the Scheffé's lemma from classical probability (see Williams (1991)) stating that if a sequence of probability densities converges pointwise almost everywhere to a probability density, then they also converge in total variation norm. The matrix Scheffé's lemma which can be found in Simon (1979) says that if $\rho, \rho^{(1)}, \rho^{(2)}, \ldots$ are density matrices (positive and trace one) and if the coefficients $\rho_{i,j}^{(n)}$ converge to $\rho_{i,j}$ as n goes to infinity, for any fixed indices i, j, then $\| \rho^{(n)} - \rho \|_1 \to 0$. But by equation (16) if the sequence $P_{\rho^{(n)}}$ converges to P_ρ as $n \to \infty$ with respect to the d_{tv}-distance then $\rho_{i,j}^{(n)}$ converges to $\rho_{i,j}$ and thus $\| \rho^{(n)} - \rho \|_1 \to 0$, completing the proof of the continuity of the map from P_ρ to ρ. In particular we have $\| \rho^{(n)} - \rho \|_2 \to 0$ due to the inequality between the L_1 and L_2 norms

$$\| \tau - \rho \|_2 \leq \| \tau - \rho \|_1. \tag{33}$$

The maximum likelihood estimator of ρ is defined as

$$\arg\max_\tau \sum_{\ell=1}^{n} \log p_\tau(X_\ell, \Phi_\ell) \tag{34}$$

where the maximum is taken over all density matrices in on the space $L_2(\mathbb{R})$. However there exist density matrices τ such that the probability density p_τ takes arbitrarily high values at all the points (X_ℓ, Φ_ℓ). To see this let us first remind the reader that any density matrix is a convex combination

of "pure states" which are projections $\mathbf{P}(\psi)$ on one dimensional sub-spaces of $L_2(\mathbb{R})$ generated by vectors ψ which can be written as a Fourier sum

$$\psi(x) = \sum_{k=0}^{\infty} \alpha_k \psi_k(x) \tag{35}$$

in the basis $\{\psi_k\}$ given in equation (14), with $\int |\psi(x)|^2 dx = \sum_k |\alpha_k|^2 = 1$. For any such state the corresponding probability density is

$$p_\psi(x,\phi) = \Big| \sum_{k=0}^{\infty} \alpha_k e^{ik\phi} \psi_k(x) \Big|^2 = |\psi^\phi|^2(x), \tag{36}$$

where $\psi^\phi(x)$ is the square integrable function with Fourier coefficients $e^{ik\phi}\alpha_k$. It is clear that there exists a one-to-one relation between ψ and ψ^ϕ which preserves the L_2-norms, thus we can choose vectors $\varphi_1, \dots \varphi_n$ such that $\int |\varphi_\ell(x)|^2 dx = 1$ and $|\varphi_\ell^{\Phi_\ell}(X_\ell)|^2 > C$ for all $\ell = 1, \dots n$ and arbitrary $C > 0$. Then the density matrix

$$\rho = \frac{1}{n} \sum_{k=1}^{n} \mathbf{P}(\varphi_k) \tag{37}$$

representing a statistical mixture of the pure states leads to the likelihood

$$p_\rho(x,\phi) = \prod_{\ell}^{n} \Big(\frac{1}{n} \sum_{k=1}^{n} |\varphi_k^{\Phi_\ell}(X_\ell)|^2 \Big) \geq \Big(\frac{C}{n} \Big)^n. \tag{38}$$

which can be arbitrarily high for any fixed n. This drawback can be corrected by using for example penalized maximum likelihood estimators or by restricting the state space to some subspace $\mathcal{Q}(n)$ of density matrices such that for any amount of data the maximum of the likelihood over $\mathcal{Q}(n)$ exists and $\cup_n \mathcal{Q}(n)$ is dense in the space of all density matrices $\mathcal{S}$ with respect to some chosen distance function. Such a method is called sieve maximum likelihood and we refer to van de Geer (2000) and Wong and Shen (1995) for the general theory. The choice of the sieves $\mathcal{Q}(n)$ should be tailored according to the problem one wants to solve, the class of states one is interested in, etc. Here we will use the same subspaces as for the projection estimator of the previous subsection, that is $\mathcal{Q}(n)$ consists of those states with maximally $N(n) - 1$ photons described by density matrices over the subspace spanned by the basis vectors $\psi_0, \dots, \psi_{N(n)-1}$. We will call $\{\mathcal{Q}(n)\}_{n\geq 0}$ the *number states sieves* and the dimension $N(n)$ will be an increasing function of n which will be fixed later so as to guarantee consistency.

$$\mathcal{Q}(n) = \left\{ \tau \in \mathcal{S} \ : \ \tau_{j,k} = 0 \text{ for all } j \geq N(n) \text{ or } k \geq N(n) \right\}. \tag{39}$$

Notice that the dimension of the space $\mathcal{Q}(n)$ is $N(n)^2 - 1$. Let now the estimator be

$$\bar{\rho}^{(n)} := \arg \max_{\tau \in \mathcal{Q}(n)} \sum_{\ell=1}^{n} \log p_\tau(X_\ell, \Phi_\ell), \tag{40}$$

where the maximum can be shown to exist for example by using compactness arguments. We will denote the corresponding sieve in the space of probability densities by

$$\mathcal{P}(n) = \{ p_\tau \ : \ \tau \in \mathcal{Q}(n) \}. \tag{41}$$

The theory of M-estimators van de Geer (2000) tells us that the consistency of ML estimators depends essentially on the "size" of the parameter space, in our case the sieves $\mathcal{Q}(n)$ or $\mathcal{P}(n)$, which is measured by entropy numbers with respect to some distance, for example the L_1-norm on density matrices or the Hellinger distance between probability distributions.

DEFINITION 3.3. *Let $\mathcal{G}$ be a class of density matrices. Let $N_{B,1}(\delta, \mathcal{G})$ be the smallest value of $p \in \mathbb{N}$ for which there exist pairs of Hermitian matrices (not necessarily density matrices) $[\tau_j^L, \tau_j^U]$ with $j = 1, \ldots, p$ such that $\|\tau_j^L - \tau_j^U\|_1 \leq \delta$ for all j, and such that for each $\tau \in \mathcal{G}$ there is a $j = j(\tau) \in \{1, \ldots, p\}$ satisfying*

$$\tau_j^L \leq \tau \leq \tau_j^U.$$

Then $H_{B,1}(\delta, \mathcal{G}) = \log N_{B,1}(\delta, \mathcal{G})$ is called δ-entropy with bracketing of $\mathcal{G}$.

We note that this definition relies on the concept of positivity of matrices and the existence of the L_1-distance between states. But the same notions exist for the space of integrable functions thus by replacing density matrices with probability densities and selfadjoint operators with functions we obtain the definition of the δ-entropy with bracketing $H_{B,1}(\delta, \mathcal{F})$ for some space of probability densities $\mathcal{F}$, see van de Geer (2000). Moreover by using equation (31) and the fact that the linear extension of the map from density matrices to probability densities sends a positive matrix to a positive function, we get that for any δ-bracketing $[\tau_j^L, \tau_j^U]$ for $\mathcal{Q}(n)$, the corresponding functions $[p_j^L, p_j^U]$ form a δ-bracketing for $\mathcal{P}(n)$, i.e. they satisfy $\|p_j^L - p_j^U\|_1 \leq \delta$ and for any $p \in \mathcal{P}(n)$ there exists a $j = j(p)$ such that $p_j^L \leq p \leq p_j^U$. Thus

$$H_{B,1}(\delta, \mathcal{P}(n)) \leq H_{B,1}(\delta, \mathcal{Q}(n)).$$

The following proposition gives an upper bound of the "quantum" bracketing entropy and in consequence for $H_{B,1}(\delta, \mathcal{P}(n))$. Its proof can be found in Guţă (2004) and relies on choosing a maximal number of nonintersecting balls centered in $\mathcal{Q}(n)$ having radius $\frac{\delta}{2N(n)}$ and then providing a pair of brackets for each ball.

PROPOSITION 3.4. *Let $\mathcal{Q}(n)$ be the class of density matrices of dimension $N(n)$. Then*

$$H_{B,1}(\delta, \mathcal{Q}(n)) \leq CN(n)^2 \log \frac{N(n)}{\delta}\,. \tag{42}$$

for some constant C independent of n and δ.

By combining the previous inequalities with equation (32) we get the following bound for the bracketing entropy of the class of square-root densities

$$\mathcal{P}^{1/2}(n) = \left\{\sqrt{p}_\rho \;:\; p_\rho \in \mathcal{P}(n)\right\}, \tag{43}$$

with respect to the L_2-distance

$$H_{B,2}(\delta, \mathcal{P}^{1/2}(n)) \leq C_2 N(n)^2 \log \frac{N(n)}{\delta}\,. \tag{44}$$

We will concentrate now on the Hellinger consistency of the sieve maximum likelihood estimator $\bar{P}_n$. We will appeal to a theorem from van de Geer (2000), which is similar to other results in the literature on non-parametric M-estimation (see for example Wong and Shen (1995)). There are two competing factors which contribute to the convergence of $h(\bar{P}_n, P_\rho)$. The first is related to the

approximation properties of the sieves with respect to the whole parameter space. Such a distance from ρ to the sieve $\mathcal{Q}(n)$ can take different expressions, for example in terms of the χ^2-distance between the corresponding probability measures

$$\tau_n := \operatorname*{argmin}_{p_n \in \mathcal{P}_n} \chi^2(p_\rho, p_n), \tag{45}$$

where the χ^2-distance between two probability distributions is given by

$$\chi^2(P_1, P_2) := \begin{cases} \int (\frac{dP_1}{dP_2} - 1)^2 dP_2 & P_1 \ll P_2, \\ \infty & \text{otherwise.} \end{cases} \tag{46}$$

Notice that τ_n depends on n through the growth rate of the sieve $N(n)$. The second factor influencing the convergence of $h(\bar{P}_n, P_\rho)$ is the size of the sieves which is expressed by the bracketing entropy. The non-parametric sieve maximum likelihood estimation theory shows that consistency holds if there exists a sequence $\delta_n \to 0$ such that the following *entropy integral inequalities* are satisfied for all n

$$J_B(\delta_n, \mathcal{P}^{1/2}(n)) := \int_{\delta_n^2/c}^{\delta_n} H_B^{1/2}(u, \mathcal{P}(n)^{1/2}) du \ \vee \ \delta_n \leq \sqrt{n}\delta_n^2/c. \tag{47}$$

where c is some universal constant, van de Geer (2000). From (44) we get

$$J_B(\delta_n, \mathcal{P}^{1/2}(n)) = O\left[N(n)\delta_n \left(\log \frac{N(n)}{\delta_n} \right)^{1/2} \right], \tag{48}$$

which implies the following constraint for $N(n) \to \infty$ and $\delta_n \to 0$,

$$\frac{N(n)}{\delta_n} = O\left(\sqrt{\frac{n}{\log n}} \right). \tag{49}$$

THEOREM 3.5. *Suppose that the state ρ satisfies $\tau_n \to 0$. Let $\bar{\rho}^{(n)}$ be the sieve MLE with $N(n)$ and δ_n satisfying (47), then*

$$\mathbb{P}(h(\bar{P}_n, P_\rho) \geq \delta_n + \tau_n) \leq c \exp(-n\delta_n^2/c^2). \tag{50}$$

Proof. Details can be found in Guţă (2004) based on Theorem 10.13 of van de Geer (2000). □

From the physical point of view, we are interested in the convergence of the state estimator $\bar{\rho}^{(n)}$ with respect to the L_1 and L_2-norms on the space of density matrices. Clearly the rates of convergence for such estimators are slower than those of their corresponding probability densities. As shown in the beginning of this subsection the map sending probability densities p_ρ to density matrices ρ is continuous, thus an estimator $\bar{\rho}_n$ taking values in the space of density matrices $\mathcal{S}$ is consistent in the L_1 or L_2-norms if and only if $\bar{P}_n$ converges to P_ρ almost surely with respect to the Hellinger distance.

COROLLARY 3.6. *The Hellinger consistency of $\bar{P}_n$ is equivalent to the $\|\cdot\|_1$-consistency of $\bar{\rho}^{(n)}$. In particular, if $\sum_n \exp(-n\delta_n^2/c^2) < \infty$ and the assumptions of Theorem 3.5 hold, then we have $\|\bar{\rho}^{(n)} - \rho\|_1 \to 0$, a.s..*

3.3. Wigner function estimation

The Wigner function plays an important role in quantum optics as an alternative way of representing quantum states and calculating an observable's expectation: for any observable $\mathbf{X}$ there exists a function $W_{\mathbf{X}}$ from $\mathbb{R}^2$ to $\mathbb{R}$ such that

$$\mathrm{Tr}(\mathbf{X}\rho) = \iint W_{\mathbf{X}}(q, p) W_\rho(q, p) dq dp. \tag{51}$$

Besides, physicists are interested in estimating the Wigner function for the purpose of identifying features which can be easier visualized than read off from the density matrix, for example a "non-classic" state may be recognized by its patches of negative Wigner function, while "squeezing" is manifest through the oval shape of the support of the Wigner function, see Table 1 and Figure 3. As described in Subsection 2.3 the Wigner function should be seen formally as a joint density of the observables $\mathbf{Q}$ and $\mathbf{P}$ which may take non-negative values, reflecting the fact that the two observables cannot be measured simultaneously. However the Wigner function shares some common properties with probability densities, in particular their marginals $\int W_\rho(q, p) dq$ and $\int W_\rho(q, p) dp$ are probability densities on the line. In fact this is true for the marginals in any direction which are nothing else then the densities $p_\rho(x, \phi)$. On the other hand there exist probability densities which are not Wigner functions and vice-versa, for example the latter cannot be too "peaked":

$$|W_\rho(q, p)| \leq \frac{1}{\pi}, \qquad \text{for all } (q, p) \in \mathbb{R}^2,\ \rho \in \mathcal{S}. \tag{52}$$

As a corollary of this uniform boundedness we get

$$\| W_\rho - W_\tau \|_\infty \leq \frac{1}{\pi} \| \rho - \tau \|_1. \tag{53}$$

for any density matrices ρ and τ. Indeed we can write $\rho - \tau = \rho_+ - \rho_-$ where ρ_+ and $-\rho_-$ represent the positive and negative part of $\rho - \tau$. Then

$$\| W_\rho - W_\tau \|_\infty = \| W_{\rho_+} - W_{\rho_-} \|_\infty \leq \| W_{\rho_+} \| + \| W_{\rho_-} \|_\infty \leq$$
$$\frac{1}{\pi} (\| \rho_+ \|_1 + \| \rho_- \|_1) = \frac{1}{\pi} \| \rho - \tau \|_1.$$

Another important property is the fact that the linear span of the Wigner functions is dense in $L_2(\mathbb{R}^2)$, the space of real valued, square integrable functions on the plane, and there is an isometry (up to a constant) between the space of Wigner functions and that of density matrices with respect to the L_2-distances

$$\| W_\rho - W_\tau \|_2^2 =: \iint |W_\rho(q, p) - W_\tau(q, p)|^2 dp dq = \frac{1}{2\pi} \| \rho - \tau \|_2^2 . \tag{54}$$

In Section 2 we have described the standard estimation method employed in computerized tomography which used a regularized kernel K_c with bandwidth $h_n = 1/c$ converging to zero as $n \to \infty$ at an appropriate rate. This type of estimators for the Wigner function will be analyzed separately in future work in the minimax framework along the lines of Cavalier (2000). The estimators which we propose in this subsection are of a different type, they are based on estimators for ρ plugged into the following linearity equation

$$W_\rho(q, p) = \sum_{k,j} \rho_{k,j} W_{k,j}(q, p),$$

where $W_{k,j}$ are known functions corresponding to the matrix with the entry (k, j) equal to 1 and all the rest equal to zero, see Leonhardt (1997). The isometry (54) implies that the family $\{W_{k,j}\}_{k,j=0}^{\infty}$ forms an orthogonal basis of $L_2(\mathbb{R}^2)$. Following the same idea as in the previous section we consider the projection estimator

$$\widehat{W}^{(n)}(q, p) = \sum_{k,j=0}^{N(n)-1} \left(\frac{1}{n} \sum_{\ell=1}^{n} F_{k,j}(X_\ell, \Phi_\ell) \right) W_{k,j}(q, p).$$

COROLLARY 3.7. *Let $N(n)$ be such that $N(n) \to \infty$ and*

$$N(n) = o(n^{3/7}).$$

Then

$$\mathbb{E}\|\widehat{W}^{(n)} - W_\rho\|_2^2 = \|W_\rho^{(n)} - W_\rho\|_2^2 + O\left(\frac{N(n)^{7/3}}{n}\right).$$

Proof: Apply isometry property and Theorem 3.2.

□

Similarly we can extend the SML estimator of the density matrix to the Wigner function. Define the subspace

$$\mathcal{W}(n) = \{W_\rho : \rho \in \mathcal{Q}(n)\},$$

with $\mathcal{Q}(n)$ as in equation (39), and define the corresponding SML estimator as $\overline{W}^{(n)} = W_{\bar{\rho}^{(n)}}$ where $\bar{\rho}^{(n)}$ was defined in (40).

COROLLARY 3.8. *Suppose that ρ satisfies $\tau_n \to 0$. Let $\overline{W}_\rho^{(n)}$ be the SML estimator with $N(n)$ and δ_n satisfying (47) and $\sum_n \exp(-n\delta_n^2/c^2) < \infty$. Then we have*

$$\|\overline{W}^{(n)} - W_\rho\|_2 \to 0$$

almost surely. Under the same conditions

$$\|\overline{W}^{(n)} - W_\rho\|_\infty \to 0$$

almost surely.

Proof: Apply the inequalities (53, 54) and Corollary 3.6.

□

4. Noisy observations

The homodyne tomography measurement as presented up to now does not take into account various losses (mode mismatching, failure of detectors) in the detection process which modify the distribution of results in a real measurement compared with the idealized case. Fortunately, an analysis of such losses (see Leonhardt (1997)) shows that they can be quantified by a single *efficiency* coefficient $0 < \eta < 1$ and the change in the observations amounts replacing X_i by the noisy observations

$$X_i' := \sqrt{\eta} X_i + \sqrt{(1-\eta)/2}\xi_i \tag{55}$$

with ξ_i a sequence of i.i.d. standard Gaussians which are independent of all X_j. The problem is again to estimate the parameter ρ from $Y_i' = (X_i', \Phi_i)$, for $i = 1, \ldots, n$. The efficiency-corrected probability density is then the convolution

$$p_\rho(y, \phi; \eta) = (\pi(1-\eta))^{-1/2} \int_{-\infty}^{\infty} p(x, \phi) \exp\left[-\frac{\eta}{1-\eta} (x - \eta^{-1/2} y)^2 \right] dx. \tag{56}$$

The physics of the detection process detailed in Leonhardt (1997) offers an alternative route from the state to the probability density of the observations Y_i'. In a first step one performs a *Bernoulli transformation* $B(\eta)$ on the state ρ which is a quantum equivalent of the convolution with noise for probability densities, and obtains a new density matrix ρ^η. To understand the Bernoulli transformation let us consider first the diagonal elements $\{p_k = \rho_{k,k},\ k = 0, 1..\}$ and $\{q_j = \rho^\eta_{j,j},\ j = 0, 1..\}$ which are both probability distributions over $\mathbb{N}$ and represent the statistics of the number of photons in the two states. Let $b_k^{k+p} = \binom{k+p}{k} \eta^k (1-\eta)^p$ be the binomial distribution. Then

$$q_j = \sum_{k=j}^{\infty} b_j^k(\eta)\, p_k \tag{57}$$

which has a simple interpretation in terms of an "absorption" process by which each photon of the state ρ goes independently through a filter and is allowed to pass with probability η or absorbed with probability $1 - \eta$. The formula of the Bernoulli transformation for the whole matrix is

$$\rho^\eta_{j,k} = \sum_{p=0}^{\infty} \left[b_j^{j+p}(\eta) b_k^{k+p}(\eta) \right]^{1/2} \rho_{j+p,k+p}. \tag{58}$$

The second step is to perform the usual quantum tomography measurement with ideal detectors on the "noisy" state ρ^η obtaining the results Y_i' with density $p_\rho(x, \phi; \eta)$. It is noteworty that the transformations $B(\eta)$ form a semigroup, that is they can be composed as $B(\eta_1)B(\eta_2) = B(\eta_1\eta_2)$ and the inverse of $B(\eta)$ is simply obtained by replacing η with η^{-1} in equation (58). Notice however that if $\eta \leq 1/2$ the power series $(1 - \eta^{-1})^k$ appearing in the inverse transformation diverges, thus we need to take special care in this range of parameters.

A third way to compute the inverse map from $p_\rho(x, \phi; \eta)$ to ρ is by using pattern functions depending on η which incorporate the deconvolution map from $p_\rho(x, \phi; \eta)$ to $p_\rho(x, \phi; 1)$:

$$\rho_{k,j} = \int_{-\infty}^{\infty} dx \int_0^\pi \frac{d\phi}{\pi} p_\rho(x, \phi; \eta)\, f_{k,j}(x; \eta). \tag{59}$$

Such functions are analyzed in D'Ariano (1995); D'Ariano et al. (1995) where it is argued that the method has a fundamental limitation for $\eta \leq 1/2$ in which case the pattern functions are unbounded, while for $\eta > 1/2$ numerical calculations show that their range grows exponentially fast with both indices j, k.

The two estimation methods considered in Section 3 can be applied to the state estimation with noisy observations. The projection estimator has the same form as in Subsection 3.1 with a similar analysis of the mean L_2-risk taking into account the norms of the new pattern functions $f_{k,j}(x; \eta)$. The sieve maximum likelihood estimator follows the definition in Subsection 3.2 and a consistency result can be formulated on the lines of Corollary 3.6. We expect however that the rates of convergence will be dramatically slower and we will leave this analysis for a separate work.

5. Experimental results

In this section we study the performance of the Pattern Function Projection estimator and the Sieve Maximum Likelihood estimator using simulated data. In Table 1 we showed some examples of density matrices and Wigner functions of quantum states. In Figure 3, we display their corresponding graphical representation. For some of them the corresponding probability distribution can be expressed explicitly and it is possible to simulate data. In particular we shall simulate data from QHT measurements on squeezed states with efficiency $\eta = 1$.

5.1. Implementation

In order to implement the two estimators we need to compute the basis functions ψ_n and the functions ϕ_n, which are solutions of Schrödinger equation (18). For this, we use an appropriate set of recurrent equations, see Leonhardt (1997), Ch. 5. Pattern functions can then be calculated as

$$f_{k,j}(x) = 2x\psi_j(x)\phi_k(x) - \sqrt{2(j+1)}\psi_{j+1}(x)\phi_k(x) - \sqrt{2(k+1)}\psi_j(x)\phi_{k+1}(x),$$

for all $j \geq k$, otherwise $f_{k,j}(x) = f_{j,k}(x)$ and then $F_{k,j}(x,\theta) = f_{k,j}(x)e^{i(k-j)\theta}$.

On the practical side, finding the maximum of the likelihood function over a set of density matrices is a more complicated problem due to the positivity and trace one constraints which must be taken into account. A solution was proposed in Banaszek et al. (1999), where the restriction on positivity of a density matrix ρ is satisfied by writing the Cholevski decomposition

$$\rho = T^*T \tag{60}$$

where T of *upper triangular* matrices of the same dimension as ρ with complex coefficients above the diagonal and reals on the diagonal. The normalization condition $\mathrm{Tr}\rho = 1$ translates into $\|T\|_2 = 1$ which defines a ball in the space of upper triangular matrices with the L_2-distance. We will denote by $\mathcal{T}(n)$ the set of such matrices having dimension $N(n)$. The sieve maximum likelihood estimator is the solution of the following optimization problem with $N = N(n)$

$$\begin{aligned}\widehat{T} &= \underset{T\in\mathcal{T}(n)}{\arg\max} \sum_{\ell=1}^{n} \log p_\rho(X_\ell, \Phi_\ell) \\ &= \underset{T\mathcal{T}(n)}{\arg\max} \sum_{\ell=1}^{n} \log \sum_{k=1}^{N-1} \left| \sum_{m=0}^{k} T_{k,m}\psi_m(X_\ell)e^{im\Phi_\ell} \right|^2.\end{aligned}$$

The numerical optimization was performed using a classical descendent method with constraints. Notice that we have an optimization problem on $N^2 - 1$ real variables. Given the problem of high dimensionality and computational cost we propose an alternative method to the procedure mentioned above. It exploits the mixing properties of our model. Any density matrix of dimension N can be written as convex combination of N *pure* states, i.e. $\rho = \sum_{r=0}^{N-1} p_r \rho_r$, where $p_r \geq 0$, $\sum_r p_r = 1$ and ρ_r is a one dimensional projection whose Cholevski decomposition is of the form $\rho_r = t_r^* t_r$ where t_r is the row vector of dimension N on which ρ_r projects, and t_r^* is the column vector of the complex conjugate of t_r. It should be noted that even though decomposition of our state in pure states is not unique this is not a problem given we are actually not interested in this representation but in the resulting convex combination, the state itself. Now we can state the problem as to find the maximizer of the loglikelihood

$$L(\{(X_\ell, \Phi_\ell)\}; p, t) = \sum_{\ell=1}^{n} \log \sum_{r=0}^{N-1} p_r\, p_{\rho_r}(X_\ell, \Phi_\ell) = \sum_{\ell=1}^{n} \log \sum_{r=0}^{N-1} p_r \left| \sum_{m=0}^{N(n)-1} t_{r,m}\psi_m(X_\ell)e^{im\Phi_\ell} \right|^2,$$

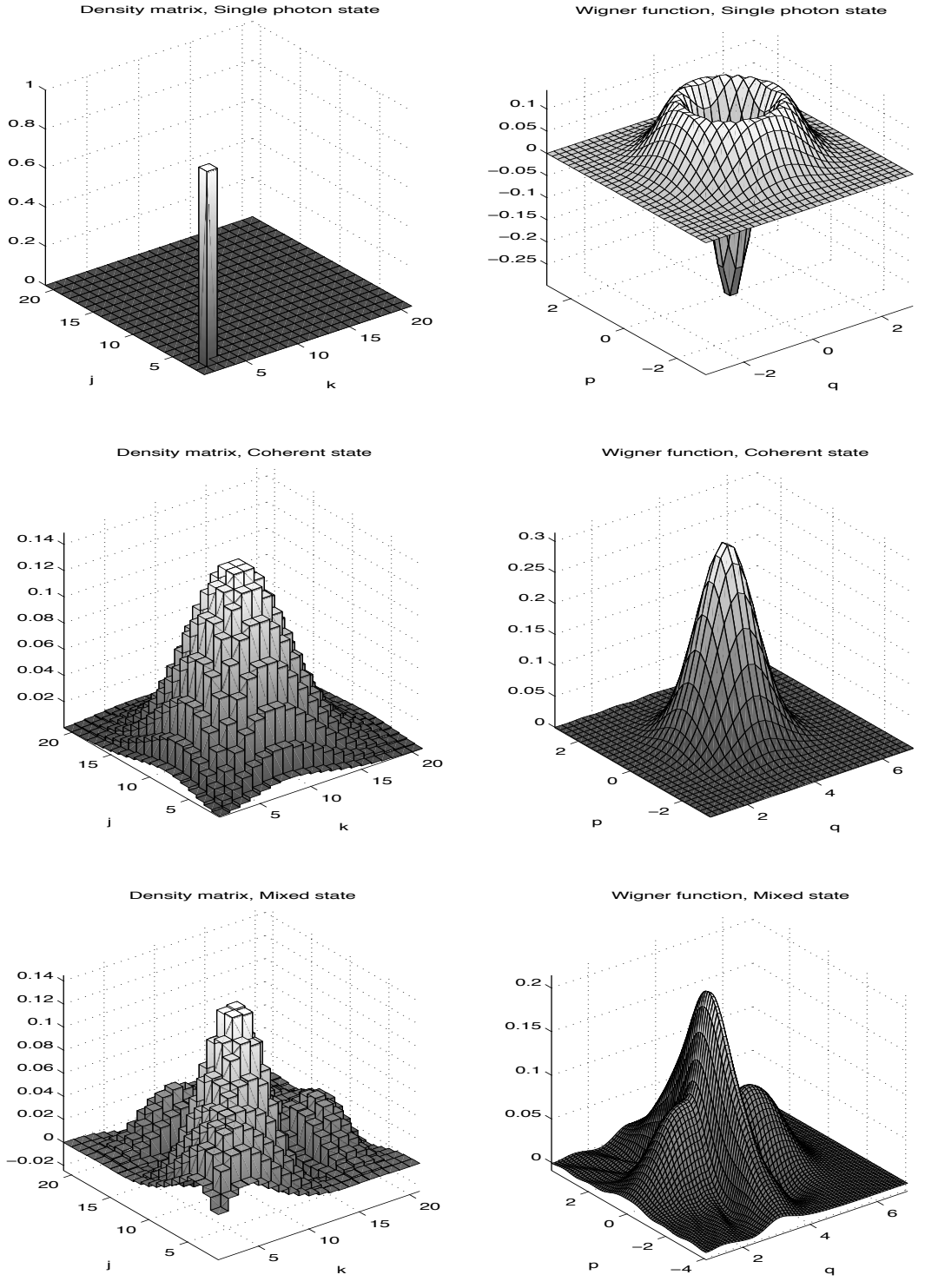


Fig. 3. Graphical representation of quantum states

where $t_{r,m}$ represents the m$^{\text{th}}$ coordinate of t_r. We now maximize over all $\beta \in \mathrm{B}$ where

$$\mathrm{B} = \left\{ \{p_r, t_r\}_{r=0}^{N-1} \ : \ p_r \geq 0, \sum_{r=0}^{N-1} p_r = 1, \|t_r\|_2 = 1, \text{ for all } r \right\}.$$

We propose an EM algorithm as an alternative method to the one presented in Banaszek et al. (1999). See, Dempster et al. (1977) for an exposition on the formulation of the EM algorithm for problems of ML estimation with mixtures of distributions. The iteration procedure is given then in the following steps.

1) Compute the expectation of the conditional likelihood:

$$Q(\beta|\beta^{\text{old}}) = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{r=0}^{N(n)-1} \frac{p_r^{\text{old}} p_{\rho_r}^{\text{old}}(X_\ell, \Phi_\ell)}{\sum_{j=0}^{N(n)-1} p_j^{\text{old}} p_{\rho_j}^{\text{old}}(X_\ell, \Phi_\ell)} \log p_r \, p_{\rho_r}(X_\ell, \Phi_\ell).$$

2) Maximize $Q(\beta|\beta^{\text{old}})$ over all $\beta \in \mathrm{B}$ and obtain β^{new} with components

$$p_r^{\text{new}} = \frac{1}{n} \sum_{\ell=1}^{n} \frac{p_r^{\text{old}} p_{\rho_r}^{\text{old}}(X_\ell, \Phi_\ell)}{\sum_{j=0}^{N(n)-1} p_j^{\text{old}} p_{\rho_j}^{\text{old}}(X_\ell, \Phi_\ell)},$$

$$\rho_r^{\text{new}} = t_r^* t_r, \qquad t_r = \arg \max_{\|t\|_2=1} \sum_{\ell=1}^{n} f_{r,\ell}^{\text{old}} \log p_{\rho_r}(X_\ell, \Phi_\ell),$$

where

$$f_{r,\ell}^{\text{old}} = \frac{p_r^{\text{old}} p_{\rho_r}^{\text{old}}(X_\ell, \Phi_\ell)}{\sum_{j=0}^{N(n)-1} p_j^{\text{old}} p_{\rho_j}^{\text{old}}(X_\ell, \Phi_\ell)}.$$

As initial condition one could take very simple ad hoc states. For example, take t_0 as the vector $(1, 0, \ldots, 0)$ and $p_0 = 1$ and t_r to be the null vector and $p_r = 0$ for $r > 0$. This corresponds to the *one photon* state. Another possible combination is to take $t_{r,m} = \delta_r^m$ and $p_r = \frac{1}{N}$. This corresponds to the state represented by a diagonal matrix, called a chaotic state. Another strategy is to consider a preliminary estimator, based on just few observations, diagonalize it and take t_r equal to its eigenstates and p_r the corresponding eigenvalues. In this way one hopes to start the iteration from a state closer to the optimum one. In terms of speed our simulations suggests to use the EM version as dimension grows. Establishing any objective comparison between direct optimization and EM algorithm has proven to be difficult given the dependency on initial conditions, and high dimensionality of the problem.

5.2. Analysis of results

In Figure 5 we show the result of estimating the squeezed state defined in Table 1 using samples of size 1600, for both Pattern Function and Maximum Likelihood estimators. At a first glance one can see that the Pattern Function estimator result is rougher when compared to the Maximum Likelihood estimator. This is due to the fact discussed in Subsection 3.1 that the variance of $F_{k,j}$ increases as a function of k and j as we move away from the diagonal. The relation between quality of estimation and dimension of the truncated estimator is seen more clearly in Figure 6 where the L_2-errors of estimating the coherent state is shown for both estimators at different sample sizes. The $*$-dot represents the point of minimum – and thus, optimum dimension $N^*(n)$ – for each curve. The

curves presented there are the mean L_2-error estimated using 15 simulations for each sample size. From there we can see that the optimum PFP estimator for the sample of size $n = 1600$ is the one corresponding to $N^*(n) = 15$ while the optimum SML would be obtained using the sieve of size $N^*(n) = 19$.

Let us first analyze the performance of PFP estimator. Notice that for $N > N^*$ the mean L_2-error increases quadratically with N due to contribution from the variance term. As n increases the variance decreases like n^{-1} for a fixed dimension N and consequently, the optimal dimension $N^*(n)$ increases. One can see that the minimum is attained rather sharply. This suggests that, in order to get a good result a refined method of guessing the optimum dimension becomes necessary, eg. BIC, AIC or cross-validation. Figure 4 shows a cross-validation estimator for the L_2-error of the PFP estimator for three simulations (continuous lines), each one based on 1600 observations with a squeezed state, and for comparison the expected L_2-error (dashed line). This represents only a first attempt to implement model selection procedures for this problem which should be investigated more thoroughly from the theoretical and practical point of view.

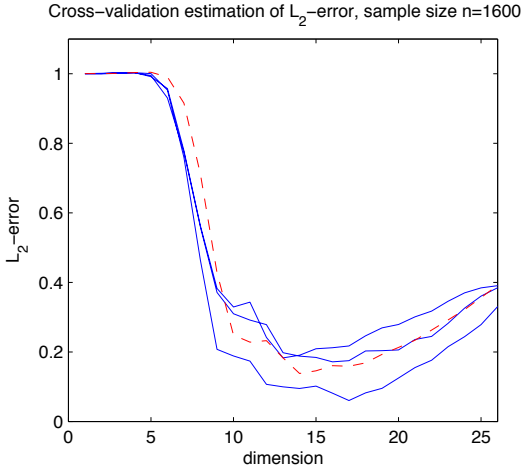


Fig. 4. Estimated optimal dimension for three simulations with 1600 observations

We pass now to the SML estimator. In Figure 6 we see that it has smaller L_2-error the the PFP estimator at its optimum sieve dimension. It is remarkable that the behavior of the L_2-error, for $N > N^*$ has a different behavior in this case, increasing much slower than the PFP estimator at the right side of its corresponding optimal dimension. This suggests that SML estimators could have a lower risk if the optimum dimension is overestimated. The bottom right pane of Figure 6 shows the optimum value of the L_2-risk in terms of sample size. Both axis are represented in a logarithmic scale. The observed linear pattern indicates that the L_2-risk decreases as $an^{-\tau}$. The slope of both curves correspond to $\tau \sim 0.4$, showing an almost parametric rate which is not surprising given the smoothness of the example that we consider. The value of τ for the PFP estimator is a bit smaller than for the SML estimator, confirming its worse performance. Notice also that the constant a is bigger for PFP than for the SML estimator. We expect that the contrast between the two estimators will be accentuated when $\eta < 1$.

Finally, in Figure 7 we show the result of estimating the Wigner function of the squeezed state using both estimators. As explained in Section 3.3 the corresponding estimator can be obtained by

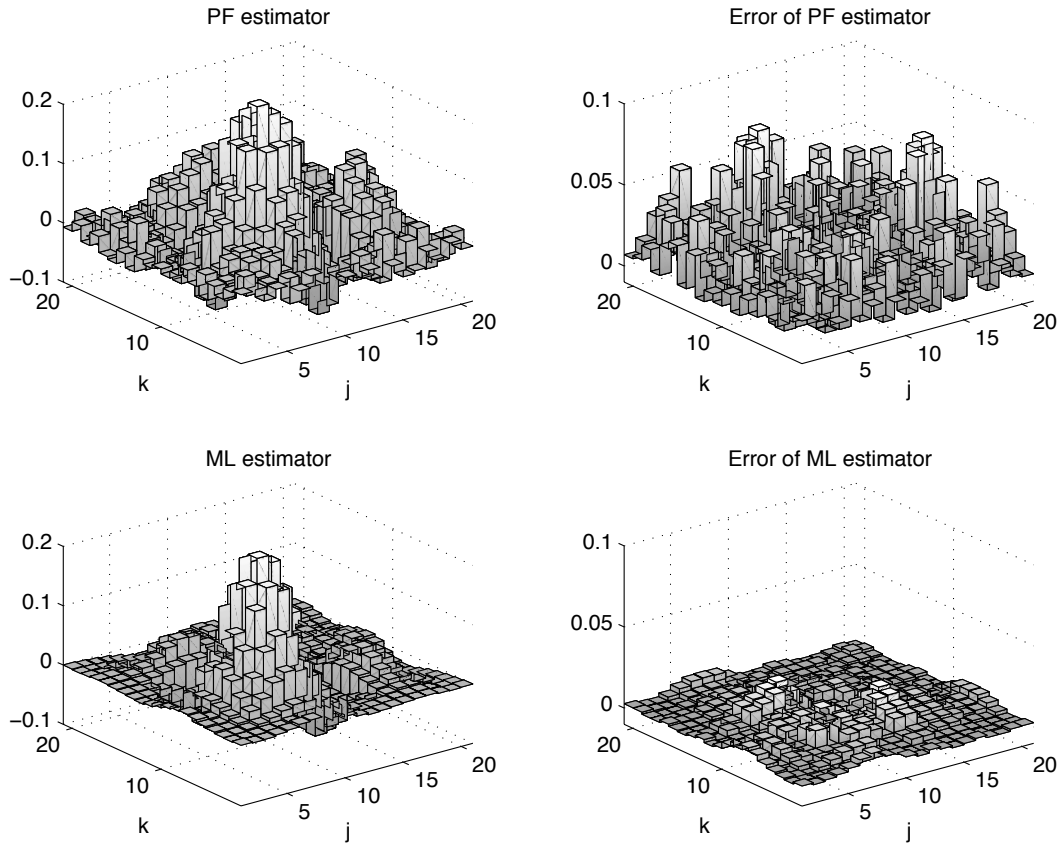


Fig. 5. Estimation of Squeezed state. First column, from top to bottom, PFP and SML estimators. Second column, corresponding errors

plugging-in the density matrix estimator in equation 4. The density matrix estimators for the same state are represented in Figure 5.

6. Concluding remarks

In this paper we have proposed a Pattern Function Projection estimator and a Sieve Maximum Likelihood estimator for the density matrix of the quantum state and its Wigner function. We proved they are consistent for different norms in their corresponding spaces. There are many open statistical questions related to quantum tomography and we would like to enumerate a few of them here.

- *Cross-validation.* For both types of estimators, a data dependent method is needed for selecting the optimal sieve dimension. We mention criteria such as unbiased cross-validation, hard thresholding or other types of minimum contrast estimators Barron et al. (1999).

- *Efficiency* $0 < \eta < 1$. A realistic detector has detection efficiency $0 < \eta < 1$ which introduces an additional noise in the homodyne data. From the statistical point of view we deal

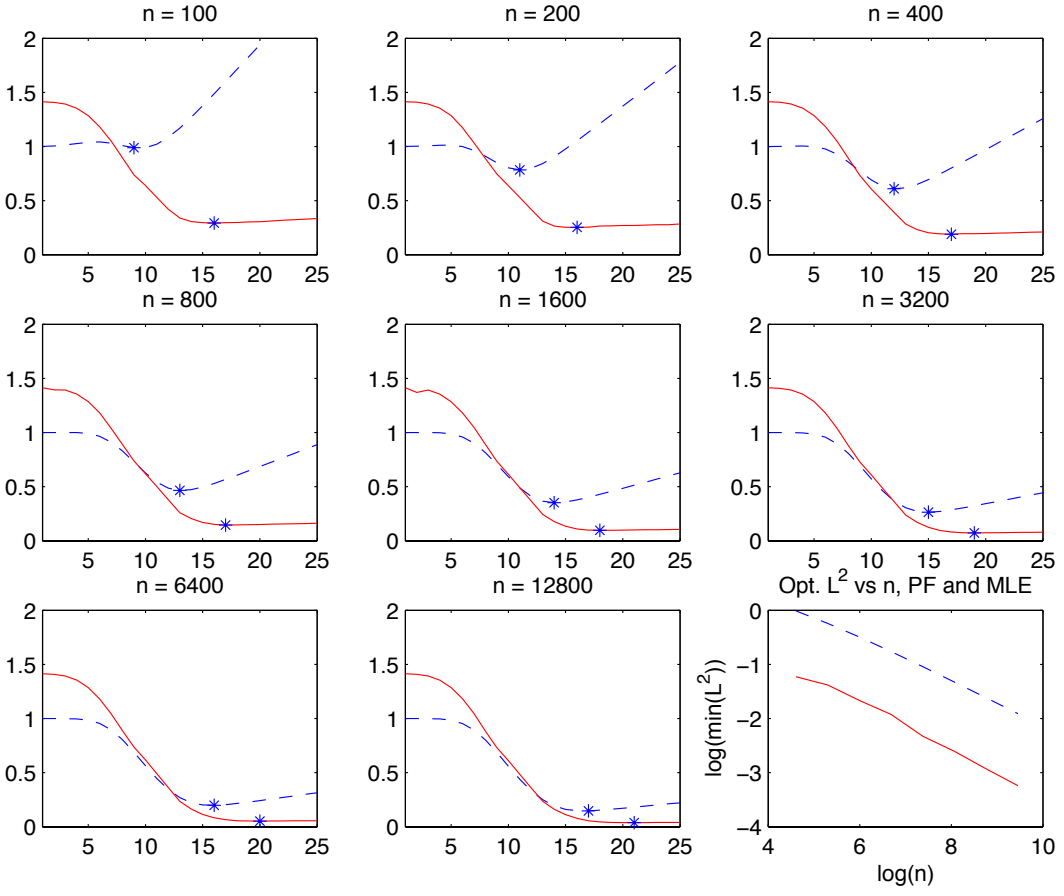


Fig. 6. L_2 error for Pattern Function Estimator and Sieve Maximum Likelihood Estimator and different sample sizes: $n = 100, \ldots, 12800$. Last graphic represents the optimum L_2 error for different sample size, using a logarithmic scale on both axis.

with a Gaussian deconvolution problem on top of the usual quantum tomography estimation.

- *Rates of convergence.* Going beyond consistency requires the selection of classes of states which are natural both from the physical, as well as statistical point of view. One should study optimal and achieved rates of convergence for given classes. For $0 < \eta < 1$ the rates are expected to be significantly lower than in the ideal case, so it becomes even more crucial to use optimal estimators. In applications, sometimes only the estimation of a functional of ρ such as average number of photons or entropy may be needed. This will require a separate analysis, cf. Shen (2001).

- *Kernel estimators for Wigner function.* When estimating the Wigner function it seems more natural to use a kernel estimator such as in Cavalier (2000) and to combine this analysis with the deconvolution problem in the case noisy observations $\eta < 1$, Butucea (2004).

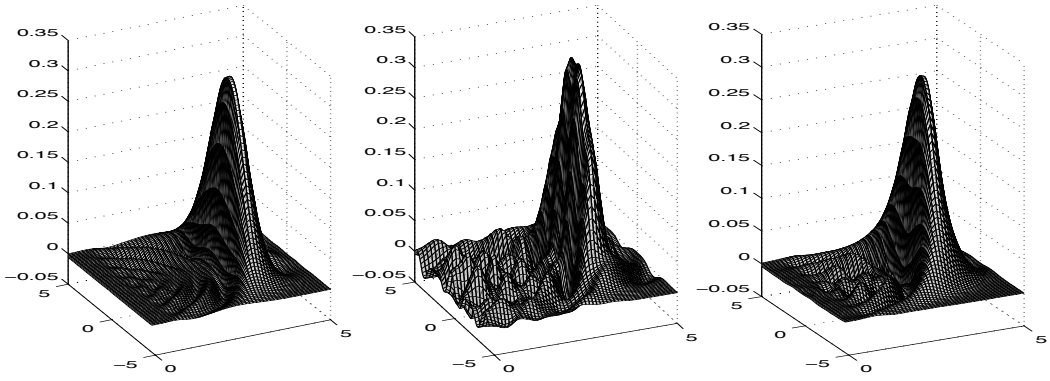


Fig. 7. Graphical representation of Wigner functions estimation. a) Original Wigner function of the squeezed state, b) Pattern Function estimation, c) Maximum Likelihood estimation

- *Other quantum estimation problems.* The methods used here for quantum tomography can be applied in other problems of quantum estimation, such as for example the calibration of measurement devices or the estimation of transformation of states under the action of quantum mechanical devices.

References

Banaszek, K., D'Ariano, G. M., Paris, M. G. A., and Sacchi, M. F. (1999). Maximum-likelihood estimation of the density matrix. *Phys. Rev. A 61*, R010304.

Barndorff-Nielsen, O. E., Gill, R., and Jupp, P. E. (2003). On quantum statistical inference (with discussion). *J. Royal Stat. Soc. B 65*, 775–816.

Barron, A. R., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields 113*, 301–413.

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics 32*. to appear.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Ann. Statist. 28*, 630–647.

D'Ariano, G. M. (1995). Tomographic measurement of the density matrix of the radiation field. *Quantum Semiclass. Optics 7*, 693–704.

D'Ariano, G. M., Leonhardt, U., and Paul, H. (1995). Homodyne detection of the density matrix of the radiation field. *Phys. Rev. A 52*, R1801–R1804.

D'Ariano, G. M., Macchiavello, C., and Paris, M. G. A. (1994). Detection of the density matrix through optical homodyne tomography without filtered back projection. *Phys. Rev. A 50*, 4298–4302.

Deans, S. R. (1983). *The Radon transform and some of its applications*. John Wiley & Sons, New York.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B 39*, 1–38.

Guţă, M. I. (2004). Quantum tomography. in preparation.

Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.

Holevo, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland.

Leonhardt, U. (1997). *Measuring the Quantum State of Light*. Cambridge University Press.

Leonhardt, U., Munroe, M., Kiss, T., Richter, Th., and Raymer, M. G. (1996). Sampling of photon statistics and density matrix using homodyne detection. *Optics Communications 127*, 144–160.

Leonhardt, U., Paul, H., and D'Ariano, G. M. (1995). Tomographic reconstruction of the density matrix via pattern functions. *Phys. Rev. A 52*, 4899–4907.

Shen, X. (2001). On Bahadur efficiency and maximum likelihood estimation in general parameter spaces. *Statistica Sinica 11*, 479–498.

Simon, B. (1979). *Trace Ideals and their Applications*. Cambridge University Press.

Smithey, D. T., Beck, M., Raymer, M. G., and Faridani, A. (1993). Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett. 70*, 1244–1247.

van de Geer, S. (2000). *Applications of Empirical Process Theory*. Cambridge University Press.

Vardi, Y., Shepp, L. A., and Kaufman, L. (1985). A statistical model for positron emission tomography. *J. Am. Stat. Assoc. 80*, 8–37.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

Wong, W. H. and X. Shen (1995). Probability inequalities for likelihood rations and convergence rates of sieve MLEs. *Ann. Statist. 23*(2), 339–362.