Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data

Cristina Butucea, Madalin Guta, Luis Artiles

Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data

Cristina BUTUCEA[1], Mădălin GUŢĂ[2] and Luis ARTILES[3]

[1]Université Paris X and Université Paris VI
[2]Mathematisch Instituut, Universiteit Utrecht
[3]Eurandom, Technische Universiteit Eindhoven

March 30, 2005

Abstract

We estimate the quantum state of a light beam from results of quantum homodyne measurements performed on identically prepared quantum systems. The state is represented through the Wigner function, a density on $\mathbb{R}^2$ which may take negative values but must respect intrinsic positivity constraints imposed by quantum physics. The effect of the losses due to detection inefficiencies which are always present in a real experiment is the addition to the tomographic data of independent Gaussian noise.

We construct a kernel estimator for the Wigner function and prove that it is minimax efficient for the pointwise risk over a class of infinitely differentiable functions. For the $\mathbb{L}_2$ risk, we compute the upper bounds of a truncated kernel estimator over the same classes, restricted to functions with sub-Gaussian asymptotic behaviour. We construct adaptive estimators, i.e. which do not depend on the smoothness parameters, and prove that in some set-ups they attain the minimax rates for the corresponding smoothness class.

Mathematics Subject Classifications 2000: 62G05, 62G20, 81V80

Key Words: Adaptive estimation, deconvolution, nonparametric estimation, infinitely differentiable functions, exact constants in nonparametric smoothing, minimax risk, quantum state, quantum homodyne tomography, Radon transform, Wigner function.

Short title: QHT with noisy data

1 Introduction

When measuring a quantum system, some particular aspect of its wave function is revealed through the probability distribution of the measurement results. For instance, a picture of an electron cloud shows the relative positioning of atoms in a molecule while the emission spectrum indicates the transitions between the different energy levels. The mathematical inverse problem is to reconstruct the entire wave function from the distribution of measurement results. It has been known for many years that a solution exists provided (speaking metaphorically) that the quantum state has been probed from a sufficiently rich set of directions. However it was only with Smithey *et al.* [15], that it became feasible to 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 favourite 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 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 $\cos(\phi)q + \sin(\phi)p$.

One way to think about quantum tomography as a statistical problem is as follows: the unknown parameter is a joint probability density W of two variables Q and P. Consider the random variable $(X, \Phi) = (\cos(\Phi)Q + \sin(\Phi)P, \Phi)$ where Φ is chosen independently of (Q, P), and uniformly in the interval $[0, \pi]$. The joint density of (X, Φ) can be expressed mathematically in terms of the joint density W of (Q, P) which is allowed to take negative as well as positive values, subject to certain restrictions which guarantee that (X, Φ) does have a proper probability density. In an ideal situation the statistical problem would be to estimate W from independent samples of (X, Φ). In the context of PET tomography this problem has been addressed in Cavalier [4] which provides minimax rates for the pointwise risk on a class of "very smooth" probability densities. The quantum tomography version is treated along similar lines in Guţă and Artiles [9] with the important difference that the proof of the lower bound requires the construction of a "worst parametric family" of Wigner functions rather than probability densities.

In this paper we consider a statistical problem which is more relevant for the experimentalist confronted with various noise sources corrupting the ideal data (X, Φ). It turns out that a good model for a realistic quantum tomography measurement amounts to replacing (X, Φ) by the noisy observations (Y, Φ), where $Y := \sqrt{\eta}X + \sqrt{(1-\eta)/2}\,\xi$ with ξ a standard Gaussian random variable independent of (X, Φ). The parameter $0 < \eta < 1$ is called the detection efficiency and represents the proportion of photons which are not detected due to various losses in the measurement process. This is the statistical problem of this paper, a combination of two classical problems: noise deconvolution and PET tomography. The non-classical feature is that though all the one-dimensional projections of W 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". Moreover, it lives on the whole plane, while in classical tomography, the object to be reconstructed globally (with $\mathbb{L}_2$ risk) lives on a bounded region, e.g. Johnstone and Silverman [12].

Though the parameter to be estimated looks strange from some points of view, it is mathematically very nice from others. One can 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 nonnegative real numbers summing to one. They are the probability distribution of the number of photons found in the laser pulse (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.

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. A first step in the direction of estimating ρ has been made in Artiles *et al.* [1] where consistency results are presented for linear and sieve maximum likelihood estimators. We recommend this paper as a complement to the present one.

Section 2 starts with a a short introduction to quantum mechanics followed by the particular problem of estimating the Wigner function in quantum homodyne tomography. In subsection 2.3 we describe some features of Wigner functions and show to what extent these functions differ from probability densities on the plane. The section ends with a concise diagram presenting the analytical relations between the different mathematical objects of quantum tomography with noisy observations.

Section 3 contains the main results of this paper. We assume that the unknown Wigner function belongs to a class $\mathcal{A}(\beta, r, L)$ of "very smooth" functions similar to those of Cavalier [4], Butucea and Tsybakov [3], Guţă and Artiles [9]. The estimator has a standard kernel-type form performing in one step the deconvolution and the inverse Radon transform. In Propositions 1 and 2 we compute upper bounds for the pointwise and the $\mathbb{L}_2$ risks respectively . Theorem 1 establishes the lower bound in the pointwise case and gives the minimax rate which is slower than any power of $1/n$ but faster than any power of $1/\log n$. Rates with a similar behavior have been obtained in [3] which inspired some of the results obtained in this paper. Theorem 2 computes upper bounds for the $\mathbb{L}_2$ risk of a truncated version of the previous estimator. Adaptive estimators can be derived in some cases (when $r \leq 1$), see Theorem 3, converging at the same rates as their non-adaptive correspondents.

Section 4 collects the proofs of Propositions 1 and 2 and a sketch of the proof of the adaptive upper bounds, in order to have a clearer presentation of results in the previous Section. These results can be easily adapted to other setups that practitioners may consider suitable, like other noise distributions or different asymptotic behaviours of Wigner functions in association with $\mathbb{L}_2$ risk.

Section 5 concentrates on the proof of the lower bound for the pointwise risk. For this we construct a pair of Wigner functions $W_{1,2}$ belonging to the class $\mathcal{A}(\beta, r, L)$ such that the distance between them is large enough and the χ^2 distance between the likelihoods of the corresponding models is small. It is now a well-known lower-bounds principle that the best rate of estimation can be viewed as the largest distance between parameters in order to detect the change in the statistical model. This construction is original as it relies on the positivity of the corresponding density matrices ρ_1 and ρ_2 rather than of the Wigner functions themselves.

2 Physical backgroung of the quantum tomography

In this section we present a short introduction to quantum mechanics in as far as it is needed for understanding the background of our statistical problem. The reader who is not interested in the physics can skip this section and continue with Section 3. Subsection 2.1 introduces the concept of state of a quantum system and that of quantum measurement. In Subsection 2.2 we describe the measurement technique called quantum homodyne tomography and show how this can be used to estimate the Wigner function which is a particular parametrization of the quantum state of a monochromatic pulse of light. The main issue tackled in this paper is the influence of noise due to the detection process onto the estimation of the Wigner function. Quantum homodyne tomography with noisy observations is discussed in Subsection 2.4.

For more background material we refer to the paper Gill *et al.* [1] which deals with the problem of quantum tomography, the review paper on quantum statistical inference Barndorff-Nielsen *et al.* [2] and the classic textbooks by Helstrom [10] and Holevo [11].

2.1 Short excursion into quantum mechanics

Quantum mechanics is the theory which describes the physical phenomena taking place at the microscopic level such as the emission and absorption of light by individual atoms, the detection of light photons. The power of quantum mechanics as a theory about physical reality, lies in its predictions concerning results of measurements which experimenters can perform in the lab. Such predictions are statistical in nature, in the sense that in general we cannot infer the result of a measurement on a single quantum system but we can compute the probability distribution of results for a given measurement performed on a statistical ensemble of identically prepared systems. Any such distribution is a function of the state in which the system is prepared, and of the performed measurement. Our statistical problem can then be briefly described as follows: estimate the state based on results of measurements on a number of identically prepared systems.

Mathematically, the main concepts of quantum mechanics are formulated in the language of selfadjoint operators acting on Hilbert spaces. To every quantum system one can associate a complex Hilbert space $\mathcal{H}$ whose vectors play represent the wave functions of the system or pure states as we will see below. In general, a state is described by a *density matrix*, which is a compact operator ρ on $\mathcal{H}$ having the following properties:

1. Selfadjoint: $\rho = \rho^*$, where ρ^* is the adjoint of ρ.
2. Positive: $\rho \geq 0$, or equivalently $\langle \psi, \rho\psi \rangle \geq 0$ for all $\psi \in \mathcal{H}$.
3. Trace one: $\mathrm{Tr}(\rho) = 1$.

The positivity property implies that all the eigenvalues of ρ are nonegative and by the trace property, they sum up to one. In the case of the finite dimensional Hilbert space $\mathbb{C}^d$ the density matrix is simply a positive semi-definite $d \times d$ matrix of trace one. The reader may have noticed that the above requirements are reminiscent of the properties of probability distributions. This connection will be strengthened in a moment when we discuss the distribution of measurement results.

Before that we will take a look at the structure of the space of states on a given Hilbert space $\mathcal{H}$. Clearly, the convex combination $\lambda\rho_1+(1-\lambda)\rho_2$ of two density matrices ρ_1 and ρ_2 is a density matrix again and it corresponds to the state obtained as result of randomly performing one of the two preparation procedures with probabilities λ and respectively $1-\lambda$. The extremals of the convex set of states are called pure states and are represented by one dimensional orthogonal projection operators. Indeed an arbitrary density matrix can be brought to the diagonal form

$$\rho = \sum_{i=1}^{\dim\mathcal{H}} \lambda_i \mathbf{P}_i,$$

where $\mathbf{P}_i$ is the projection onto the one dimensional space generated by the eigenvector $e_i \in \mathcal{H}$ of ρ and $\lambda_i \geq 0$ is the corresponding eigenvalue, i.e. $\rho e_i = \lambda_i e_i$.

The predictions made by quantum mechanics can be tested in the lab by performing measurements on quantum systems. We will now give the mathematical description of a measurement process. Let us consider a measurement with space of outcomes given by the measure space (Ω, Σ). The laws of quantum mechanics say that the result of the measurement performed on a system prepared in state ρ is *random* and has probability distribution P_ρ over (Ω, Σ) such that the map

$$\rho \mapsto P_\rho,$$

is affine, i.e. it maps convex combinations of states into the corresponding convex combination of probability distributions. This can be naturally interpreted as saying that for any mixed state $\lambda\rho_1 + (1-\lambda)\rho_2$, the distribution of the results will reflect the randomized preparation as well.

The most common measurement is that of an observable such as energy, position, spin, etc. An observable is described by a *selfadjoint operator* $\mathbf{X}$ on the Hilbert space $\mathcal{H}$ and we suppose here for simplicity that it has a discrete spectrum, that is, it can be written in the diagonal form

$$\mathbf{X} = \sum_{i=1}^{\dim\mathcal{H}} x_a \mathbf{P}_a. \tag{1}$$

with $x_a \in \mathbb{R}$ the eigenvalues of $\mathbf{X}$, and $\mathbf{P}_a$ one dimensional projections onto the eigenvectors of $\mathbf{X}$. The result of the measurement of the observable $\mathbf{X}$ will be denoted by X and is a random variable with values in the set $\Omega = \{x_1, x_2, \dots\}$ of eigenvalues of $\mathbf{X}$. When the system is prepared in the state ρ the result X has the distribution

$$P_\rho[X = x_a] = \mathrm{Tr}(\mathbf{P}_a \rho). \tag{2}$$

Notice that the conditions defining the density matrices insure that P_ρ is indeed a probability distribution. In particular the expectation on X in the state ρ is

$$\mathbb{E}_\rho[X] := \sum_x x P_\rho\,[X = x_a] = \mathrm{Tr}(\mathbf{X}\rho), \tag{3}$$

and the characteristic function

$$\mathbb{E}_\rho[\exp(itX)] = \mathrm{Tr}\,[\exp(it\mathbf{X})\rho]\,. \tag{4}$$

Measurements with continuous outcomes as wel as outcomes in an arbitrary measure space can be described in a similar way by using the spectral theory of selfadjoint operators (see Holevo [11]).

Suppose that a preparation procedure produces an unknown state ρ. It is clear that in general no individual measurement can completely determine the state but only give us statistical information about P_ρ and thus indirectly about ρ. The problem of state estimation should be then considered in the context of measurements on a big number of systems which are identically prepared in the state ρ. In this paper we consider the simplest situation when we perform identical and independent meausurements on all the n systems separately. The results are i.i.d. random variables with distribution P_ρ which we use to estimate ρ.

2.2 Quantum homodyne tomography and the Wigner function

The statistical problem analyzed in this paper is that of estimating a function W_ρ : $\mathbb{R}^2 \to \mathbb{R}$ depending on the state ρ of a quantum system, from i.i.d. data $(Y_1, \Phi_1), \ldots (Y_n, \Phi_n)$ with distribution P_ρ^η on $\mathbb{R} \times [0, \pi]$. In this subsection we will give an account of the physical origin of this problem.

An important example of a quantum system is monochromatic light in a cavity, whose state is described by density matrices on the Hilbert space of complex valued square integrable functions on the line $\mathbb{L}_2(\mathbb{R})$. The function W_ρ is called the *Wigner function* and depends in a one-to-one fashion on the state ρ of the light in the cavity. In quantum optics this alternative parametrization of the state is very appealing for many reasons, for example important physical features of the state can be easier recognized from the shape of the function W_ρ than from the expression of the density matrix ρ. Besides, a whole machinery exists for calculating probability distributions of observables directly in terms of the Wigner function rather than the density matrix.

Two important observables of this quantum system are the electric and magnetic fields whose corresponding selfadjoint operators on $\mathbb{L}_2(\mathbb{R})$ are given by

$$\mathbf{Q}\psi(x) = x\psi(x), \quad \text{and respectively} \quad \mathbf{P}\psi(x) = -i\frac{d\psi}{dx}.$$

The Wigner function $W_\rho : \mathbb{R}^2 \to \mathbb{R}$ is is much like a joint probability density for these quantities, for instance its marginals along any direction $\phi \in [0, \pi]$ in the plane which are given by the *Radon transform* of W_ρ

$$\mathcal{R}[W_\rho](x, \phi) = \int_{-\infty}^{\infty} W_\rho(x \cos\phi - t \sin\phi,\, x \sin\phi + t \cos\phi) dt, \tag{5}$$

are bona-fide probability densities and correspond to the measurement of the *quadrature* observables $\mathbf{X}_\phi := \mathbf{Q}\cos\phi + \mathbf{P}\sin\phi$. However in quantum mechanics non-commuting observables such as $\mathbf{Q}$ and $\mathbf{P}$ *cannot* be measured simultaneously, thus we cannot speak of their joint probability distribution. This fact is reflected at the level of the Wigner function which needs not be positive, indeed it might contain patches of "negative probability".

Thus for a given quantum system prepared in state ρ we can measure only one of the quadratures $\mathbf{X}_\phi$ for some phase ϕ and we obtain a result with probability density $p(x \mid \phi) = \mathcal{R}[W_\rho](x, \phi)$. Let us consider now that we have n quantum systems prepared in the same state ρ and we measure the quadrature $\mathbf{X}_{\Phi_i}$ on the i-th system with phases Φ_i chosen independently with uniform distribution on $[0, \pi]$. We obtain independent identically distributed results $(X_1, \Phi_1), \dots, (X_n, \Phi_n)$ with density $p_\rho(x, \phi) = p_\rho(x \mid \phi)$ with respect to the measure $\frac{1}{\pi}\lambda$, where λ is the Lebesgue measure on $\mathbb{R} \times [0, \pi]$. The Radon transform

$$\mathcal{R} : W_\rho \mapsto p_\rho(x, \phi),$$

is well known in statistics for its role in tomography problems such as Positron Emission Tomography (PET), Vardi *et al.* [17], and has a broad spectrum of other applications ranging from astronomy to geophysics, Deans [5]. In PET one estimates a probability density f on $\mathbb{R}^2$ related to the tissue distribution in a cross-section of the human body, from i.i.d. observations $(X_1, \Phi_1), \dots, (X_n, \Phi_n)$, with probability density equal to $\mathcal{R}[f]$. The observations are obtained 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 $\phi \in [0, \pi]$. The difference with our situation is that the role of the unknown distribution is played by the Wigner function which as we mentioned is not necessarily positive in the usual sense but carries an intrinsic positivity constraint in the sense that it corresponds to a density matrix. We will elaborate on this point and other properties of the Wigner functions in the next subsection.

The experimental method used for obtaining the data (X, Φ) is called *quantum homodyne tomography* (QHT) and was theoretically proposed in Vogel *et al.* [18] and put in practice for the first time by Smithey *et al.* [15]. The optical set-up sketched in

Figure 1 consists of an additional laser of high intensity $|z| \gg 1$ called local oscillator, a beam splitter through which the cavity pulse prepared in state ρ is mixed with the laser, and two photodetectors each measuring one of the two beams and producing currents $I_{1,2}$ proportional to the number of photons. An electronic device produces the result of the measurement by taking the difference of the two currents and rescaling it by the intensity $|z|$.

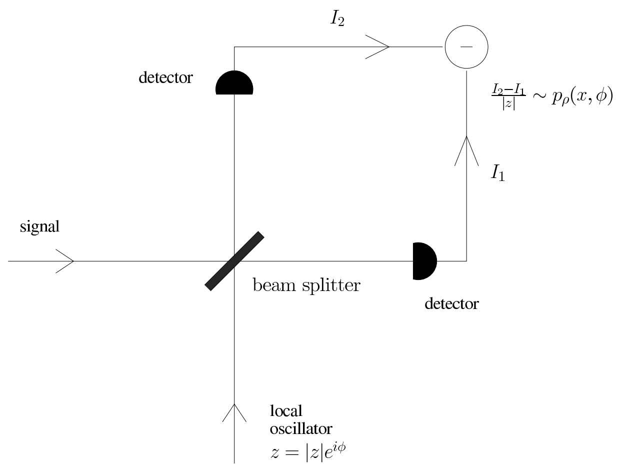

Figure 1: Quantum Homodyne Tomography measurement set-up

A simple quantum optics computation in Leonhardt [14] shows that if the relative phase between the laser and the cavity pulse is chosen to be ϕ then $(I_1 - I_2)/|z|$ has density $p_\rho(x \,|\, \phi)$ corresponding to measuring $\mathbf{X}_\phi$.

2.3 Properties of Wigner functions

In this subsection we collect some facts about Wigner functions which might help the reader appreciate to what extent the functions we want to estimate are different from the ones encountered in computerized tomography. The physics literature on Wigner functions and other types of "phase space functions" is vast but a starting point for the interested reader may be the monograph by Leonhardt [14]. Unfortunately, there exists no direct characterization of a Wigner function so we will begin with its definition in terms of the density matrix. Its Fourier transform $\mathcal{F}_2$ with respect to *both* variables has by definition the following expression

$$\widetilde{W}_\rho(u,v) := \mathcal{F}_2[W_\rho](u,v) = \mathrm{Tr}\big(\rho \exp(-iu\mathbf{Q} - iv\mathbf{P})\big). \tag{6}$$

By changing to the polar coordinates $(u, v) = (t\cos\phi, t\sin\phi)$ and using equation (4) together with the fact that $p_\rho(\cdot, \phi)$ is the density for measuring $\mathbf{X}_\phi$ we have

$$\widetilde{W}_\rho(u, v) = \mathrm{Tr}\big(\rho \exp(-it\mathbf{X}_\phi)\big) = \mathcal{F}_1[p_\rho(\cdot, \phi)](t), \tag{7}$$

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

The Wigner function plays an important role in quantum optics as an alternative representation of quantum states and a tool for calculating an observable's expectation similarly to the way it is done in classical probability: for any observable $\mathbf{X}$ there exists a function $W_{\mathbf{X}} : \mathbb{R}^2 \to \mathbb{R}$ such that when measuring $\mathbf{X}$ we obtain a random variable X with expectation

$$\mathbb{E}_\rho[X] := \mathrm{Tr}(\mathbf{X}\rho) = 2\pi \iint W_{\mathbf{X}}(q, p) W_\rho(q, p) dq dp.$$

An interesting consequence of this formula is deduced by taking $\mathbf{X}$ to be the projection onto the vector which represents the vacuum state, that is the state of the cavity when no photons are inside. The corresponding function is $W_{\mathbf{X}}(q, p) = \exp(-q^2 - p^2)$. Then the left side of the previous equation is positive because $\mathcal{X}$ is a positive operator which implies that the negative patches of W_ρ around the origin must be balanced by positive ones in such a way that the integral remains positive. A similar property holds for any point in the plane and any other choice of a positive operator $\mathcal{X}$ leads to a positivity constraint on W_ρ. Localized oscillations of the Wigner function are a signature of non-classical states such as states with a fixed number of photons or the so called Schrödinger cat states like the one shown in Figure 2.

On the other hand there exist probability densities which are not Wigner functions, for example the latter cannot be too "peaked", cf. Leonhardt [14]:

$$|W_\rho(q, p)| \leq \frac{1}{\pi}, \qquad \text{for all } (q, p) \in \mathbb{R}^2. \tag{8}$$

Another important property is the isometry (up to a constant) between the linear span of density matrices and that of Wigner functions with respect to the $\mathbb{L}_2$-distances, in particular

$$\|W_\rho - W_\tau\|_2^2 =: \iint |W_\rho(q, p) - W_\tau(q, p)|^2 dp dq = \frac{1}{2\pi}\|\rho - \tau\|_2^2 := \frac{1}{2\pi}\sum_{jk=0}^{\infty} |\rho_{jk} - \tau_{jk}|^2. \tag{9}$$

for any density matrices ρ, τ.

We will now explain the meaning of the coefficients ρ_{jk} appearing in the last sum. The space $\mathbb{L}_2(\mathbb{R})$ carries a distinguished orthonormal basis $\{\psi_j\}_{j\geq 0}$ whose vectors have

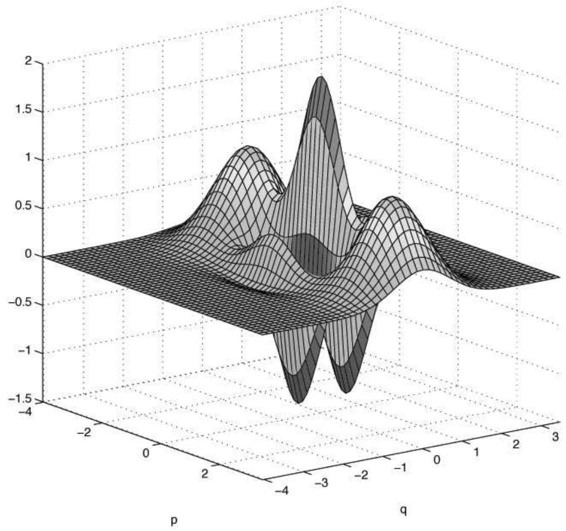

Figure 2: Wigner function of a Schrödinger cat state

the physical interpretation of pure states with precisely j photons

$$\psi_j(x) = \frac{1}{\sqrt{\sqrt{\pi}\, 2^j j!}} H_j(x) e^{-x^2/2}, \tag{10}$$

where $H_j(x)$ are the Hermite polynomials, see e.g. Erdelyi *et al.* [6]. A general density matrix ρ can be seen as an infinite dimensional matrix with coefficients $\rho_{jk} = \langle \psi_j, \rho\psi_k \rangle$ for $j, k \geq 0$ such that $\sum_{k\geq 0} \rho_{kk} = 1$ (trace one), and $[\rho_{jk}] \geq 0$ (positive definite matrix). In particular the diagonal elements $p_k = \rho_{kk}$ represent the probability of measuring k photons for a system in state ρ.

The density $p_\rho(x, \phi)$ is given in terms of the matrix elements of ρ by

$$p_\rho(x, \phi) = \sum_{j,k=0}^{\infty} \rho_{jk} p_{jk}(x, \phi) = \sum_{j,k=0}^{\infty} \rho_{jk} \psi_j(x) \psi_k(x) e^{-i(j-k)\phi}, \tag{11}$$

and a similar formula holds for the Wigner function

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

with W_{jk} such that $\mathcal{R}[W_{jk}] = p_{jk}$.

Some examples of quantum states which can be created at this moment in laboratory are given in Table 1 of Gill *et al.* [1]. Typically, the corresponding Wigner functions have a Gaussian tail but need not be positive. For example the state of one-photon in the cavity is described by the density matrix with $\rho_{1,1} = 1$ and all other elements zero which is equal to the orthogonal projection onto the vector ψ_1. The corresponding

Wigner function is

$$W(q,p) = \frac{1}{\pi}(2q^2 + 2p^2 - 1)\exp(-q^2 - p^2).$$

As a consequence of (8) not all two dimensional Gaussian distributions are Wigner functions but only those for which the determinant of the covariance matrix is bigger or equal than $\frac{1}{4}$. Equality is obtained for a remarkable set of states called *squeezed* states having Wigner functions

$$W(q,p) = \frac{1}{\pi}\exp(-e^{2\xi}(q-\alpha)^2 - e^{-2\xi}p^2).$$

This is a consequence of the celebrated *Heisenberg's uncertainty relations* which say that the non-commuting observables $\mathbf{P}$ and $\mathbf{Q}$ cannot have probability distributions such that the product of their variances is smaller than $\frac{1}{4}$.

2.4 Noisy observations

The quantum homodyne tomography measurement as presented in Subsection 2.2 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 [14]) shows that they can be quantified by a single *efficiency* coefficient $0 < \eta < 1$ and the change in the observations amounts to replacing X_ℓ by the noisy observations

$$Y_\ell := \sqrt{\eta}X_\ell + \sqrt{(1-\eta)/2}\;\xi_\ell,$$

with ξ_ℓ a sequence of i.i.d. standard Gaussians which are independent of all X_j and Φ_j. The problem is again to estimate W_ρ from the data $(Y_1, \Phi_1), \ldots (Y_n, \Phi_n)$. The efficiency-corrected probability density is then the convolution

$$p_\rho^\eta(y,\phi) = (\pi(1-\eta))^{-1/2}\int_{-\infty}^{\infty} p_\rho(x,\phi)\exp\left[-\frac{\eta}{1-\eta}(x - \eta^{-1/2}y)^2\right]\, dx. \tag{12}$$

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

$$\begin{tikzcd}
\rho \arrow[r, leftrightarrow] & W_\rho \arrow[rr, "\mathcal{R}"] \arrow[dr, "\mathcal{F}_2"'] & & p_\rho(x,\phi) \arrow[dl, "\mathcal{F}_1"] \arrow[r, "\star N^\eta"] & p_\rho^\eta(y,\phi) : (Y_\ell, \Phi_\ell),\ \ell = 1, \ldots, n. \\
& & \widetilde{W}_\rho & &
\end{tikzcd}$$

3 Statistical procedure and results

For convenience we summarize now the statistical problem tackled in this paper.

Consider $(X_1, \Phi_1), \ldots, (X_n, \Phi_n)$ independent identically distributed random variables with values in $\mathbb{R} \times [0, \pi]$ and distribution P_ρ having density $p_\rho(x, \phi)$ with respect to $\frac{1}{\pi}\lambda$, λ being the Lebesgue measure on $\mathbb{R} \times [0, \pi]$, given by

$$p_\rho(x, \phi) = \mathcal{R}[W_\rho](x, \phi),$$

where $\mathcal{R}$ is the Radon transform defined in equation (5) and $W_\rho : \mathbb{R}^2 \to \mathbb{R}$ is a so called Wigner function which we want to estimate. The space of all possible Wigner functions is parametrized by infinite dimensional matrices $\rho = [\rho_{jk}]_{j,k=0}^{\infty}$ such that $\mathrm{Tr}\rho = 1$ (trace one) and $\rho \geq 0$ (positive definite), in the way indicated by equation (6). Moreover the correspondence between ρ and W_ρ is one to one and isometric with respect to the $\mathbb{L}_2$ norms as in equation (9). The properties of Wigner functions have been discussed in subsection 2.3, in particular the fact that W_ρ may take negative values.

What we observe are not the variables (X_ℓ, Φ_ℓ) but the noisy ones $(Y_1, \Phi_1), \ldots, (Y_n, \Phi_n)$, where

$$Y_\ell := \sqrt{\eta} X_\ell + \sqrt{(1-\eta)/2}\ \xi_\ell, \tag{13}$$

with ξ_ℓ a sequence of independent identically distributed standard Gaussians which are independent of all (X_j, Φ_j). The parameter $0 < \eta < 1$ is known and we denote by p_ρ^η the density of (Y_ℓ, Φ_ℓ) given by the convolution (12). The aim is to recover the Wigner function W_ρ from the noisy observations.

In order to apply minimax estimation technology we will assume that the unknown Wigner function is infinitely differentiable and belongs to the following class described via its Fourier transform:

$$\mathcal{A}(\beta, r, L) = \left\{ W_\rho \text{ Wigner function } : \int \left| \widetilde{W}_\rho(w) \right|^2 e^{2\beta\|w\|^r} dw \leq (2\pi)^2 L \right\},$$

where $0 < r \leq 2$, $\beta, L > 0$ are real numbers. From now on we denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the usual Euclidian scalar product and norm, while $C(\cdot)$ will denote positive constants depending on parameters given in the parentheses. From the physical point of view the the choice of a class of very smooth Wigner functions seems to be quite reasonable considering that typical states ρ prepared in the laboratory do satisfy this type of condition.

We describe now the estimation method used in this paper. For the problem of estimating a probability density $f : \mathbb{R}^2 \to \mathbb{R}$ directly from data (X_ℓ, Φ_ℓ) with density $\mathcal{R}[f]$ we refer to the literature on X-ray tomography and PET, studied by Vardi *et* *al.* [17], Korostelev and Tsybakov [13], Johnstone and Silverman [12], Cavalier [4] and to many other references therein. In the context of tomography of bounded objects with noisy observations Goldenshluger and Spokoiny [7] solved the problem of estimating the borders of the object (the support). For the problem of Wigner function estimation when no noise is present, we mention the parallel work by Guţă and Artiles [9]. They use a kernel estimator and compute sharp minimax results over the class $\mathcal{A}(\beta, 1, L)$.

Let N^η denote the density of the rescaled noise $\sqrt{(1-\eta)/2}\xi$ and $\widetilde{N}^\eta$ its Fourier transform. Denote by $p_\rho^\eta(y,\phi)$ the probability density of (Y_ℓ, Φ_ℓ) in (12). Then

$$p_\rho^\eta(y,\phi) \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{\eta}} p_\rho\left(\frac{y-x}{\sqrt{\eta}}, \phi\right) N^\eta(x)dx \;=: \left(\frac{1}{\sqrt{\eta}} p_\rho\left(\frac{\cdot}{\sqrt{\eta}}, \phi\right) * N^\eta\right)(y),$$

where $p * q(y) = \int p(y-x)q(x)dx$ denotes the convolution of two arbitrary functions p and q. Via a change in variable we can write $p_\rho^\eta(y,\phi)$ as in (12). In the Fourier domain this relation becomes

$$\mathcal{F}_1[p_\rho^\eta(\cdot,\phi)](t) \;=\; \mathcal{F}_1[p_\rho(\cdot,\phi)](t\sqrt{\eta})\widetilde{N}^\eta(t),$$

where $\mathcal{F}_1$ denotes the Fourier transform with respect to the first variable.

In this paper, we modify the usual tomography kernel in order to take into account of the additive noise on the observations and construct a kernel K_h^η which performs both deconvolution and inverse Radon transform on our data, asymptotically. Let us define the estimator:

$$\widehat{W}_{h,n}^\eta(q,p) = \frac{1}{\pi n}\sum_{\ell=1}^{n} K_h^\eta\left(q\cos\Phi_\ell + p\sin\Phi_\ell - \frac{Y_\ell}{\sqrt{\eta}}\right), \tag{14}$$

where $0 < \eta < 1$ is a fixed parameter, and the kernel is defined by

$$K_h^\eta(u) = \frac{1}{4\pi}\int_{-1/h}^{1/h} \frac{\exp(-iut)|t|}{\widetilde{N}^\eta(t/\sqrt{\eta})}dt, \quad \widetilde{K}_h^\eta(t) = \frac{1}{2}\frac{|t|}{\widetilde{N}^\eta(t/\sqrt{\eta})} I(|t| \le 1/h), \tag{15}$$

and $h > 0$ tends to 0 when $n \to \infty$ in a proper way to be chosen later. For simplicity, let us denote $z = (q,p)$ and $[z,\phi] = q\cos\phi + p\sin\phi$, then the estimator can be written:

$$\widehat{W}_{h,n}^\eta(z) = \frac{1}{\pi n}\sum_{\ell=1}^{n} K_h^\eta\left([z,\Phi_\ell] - \frac{Y_\ell}{\sqrt{\eta}}\right).$$

This is a one-step procedure for treating two successive inverse problems. The main difference with the no-noise problem treated by Guţă and Artiles [9] is that the deconvolution is more 'difficult' than inverse Radon transform. In the literature on inverse problems, this problem would be qualified as severely ill-posed, meaning that the noise is dramatically (exponentially) smooth and makes the estimation problem much harder. Technically, the no-noise kernel-type estimator has dominating variance, while in the case of noisy observations the bias dominates the variance, as we will see later on.

In Subsection 3.1 we analyze the Mean Squared Error (MSE) at some fixed point. Our results concern minimax efficiency and adaptive optimality for this problem. We compute an upper bound for the convergence rate of the proposed estimator by minimizing the sum of upper bounds (uniform over the whole class) of the bias and of the variance. The optimality in rate of our estimator follows from the lower bounds which are proven in Section 5. The meaning of the lower bounds results is that no other estimation technique could outperform our method uniformly over all Wigner functions in the given class, asymptotically. Moreover, we prove the lower bounds including the asymptotic constant (sharp minimax).

Although our results on minimax sharp rates can be seen as an extension of those in Guţă and Artiles [9] to the case of noisy observations, the techniques for proving the optimality of the method (lower bound) are essentially different. We are in dominating bias setup, more similar to the deconvolution problem in Butucea and Tsybakov [3] to which we refer for the details of some of the computations. We concentrate instead on the main construction involved in the lower bound, that is the choice of two hypotheses belonging to the fixed class of Wigner functions such that their values in a fixed point are sufficiently different while their corresponding models have close to each other likelihoods as prescribed in Butucea and Tsybakov [3].

In Subsection 3.2, we look at a global estimation risk, the Mean Integrated Square (MISE) or $\mathbb{L}_2$-risk and in this case we need to use the asymptotic behaviour of Wigner functions. Under the additional assumption that the Wigner functions in the class have sub-Gaussian tails, we suggest a truncated estimator and compute upper bounds for the minimax $\mathbb{L}_2$ risk. They are very similar to the rates known to be optimal in the convolution problem in Butucea and Tsybakov [3] and therefore we expect them to be optimal as well for our problem, in the minimax sense. However, our construction for the pointwise lower bounds uses slowly decaying Wigner functions and does not satisfy the additional assumption of sub-gaussian tails. Thus, tt remains an open problem to find the minimax rates for the $\mathbb{L}_2$ risk. The solution to this problem is directly relevant to $\mathbb{L}_2$ estiamtion of the density matrix because of the isometry between the two representations of the state.

Despite the generality of a minimax sharp estimator, for practical purpose, it is not obvious how to choose the smoothness parameters r and β, though $r = 1$ seems a good choice for several examples we can provide. Therefore, an adaptive method (i.e. free of prior knowledge of parameters β, r and L provided that they are in some set) is designed for classes with $r \leq 1$ in Subsection 3.3. They behave as well as the previous estimators, provided that we know maximal values of parameters. In particular, this estimator is optimal adaptive and efficient for the pointwise risk. We note that in general such procedures do not always exist. We are fortunate in our case and this is mainly due to the dominating bias.

Let us mention that this paper provides a much more flexible estimator, in the sense that the reader may easily adapt the method to other noise distribution, e.g. s-supersmooth distributions with $r < s$, and still get the convergence rates from Butucea and Tsybakov [3]. Moreover, in the $\mathbb{L}_2$ problem, the subgaussianity may be replaced by any other assumption on the asymptotic behaviour; the reader may still compute the radius of truncation, the best bandwidth and the rates that this new estimator attain.

3.1 Pointwise estimation

In this section we give minimax and adaptive results for the pointwise risk (MSE) for the estimator $\widehat{W}_{h,n}^{\eta}$ in (14). Next proposition contains upper bounds for the two components of the risk, the bias and variance, as functions of the parameter h and the number n of samples. The bounds are uniform over all Wigner functions in the class $\mathcal{A}(\beta, r, L)$.

Proposition 1 *Let (Y_ℓ, Φ_ℓ), $\ell = 1, \ldots, n$ be i.i.d. data coming from the model (13) and let $\widehat{W}_{h,n}^{\eta}$ be an estimator (with $h \to 0$ as $n \to \infty$) of the underlying Wigner function W_ρ lying in the class $\mathcal{A}(\beta, r, L)$, with $0 < r \leq 2$. Then*

$$\begin{aligned}
\sup_{z\in\mathbb{R}^2} \sup_{W_\rho\in\mathcal{A}(\beta,r,L)} \left| E[\widehat{W}_{h,n}^{\eta}(z)] - W_\rho(z) \right|^2 &= \frac{Lh^{r-2}}{4\pi\beta r} \exp\left(-\frac{2\beta}{h^r}\right)(1+o(1)), \\
\sup_{W_\rho\in\mathcal{A}(\beta,r,L)} E\left[\left| \widehat{W}_{h,n}^{\eta}(z) - E[\widehat{W}_{h,n}^{\eta}(z)] \right|^2 \right] &\leq \frac{2\eta^2}{\pi^2(1-\eta)^2 n} \exp\left(\frac{1-\eta}{2\eta}\frac{1}{h^2}\right)(1+o(1)),
\end{aligned}$$

where $z \in \mathbb{R}^2$ and $o(1) \to 0$ as $h \to 0$ and $n \to \infty$.

Note that if we denote by $\gamma = (1-\eta)/(4\eta)$ then the upper bound for the variance term becomes $(8\pi^2\gamma^2 n)^{-1} \exp(2\gamma/h^2)$.

Given n, η and the class $\mathcal{A}(\beta, r, L)$, we now select the parameter h_{opt} for which the sum of the upper bounds derived above attains its minimum. The pointwise convergence rate of $\widehat{W}_{h,n}^{\eta}$ with $h = h_{opt}$ is then shown to be minimax by proving an additional lower bound.

Theorem 1 *Let $\beta > 0$, $L > 0$, $0 < r < 2$ and (Y_ℓ, Φ_ℓ), $\ell = 1, \dots, n$ be i.i.d. data coming from the model (13). Then $\widehat{W}_{h,n}^{\eta}$ defined in (14) with kernel K_h^{η} in (15) and bandwidth $h = h_{opt}$ solution of the equation*

$$\frac{2\beta}{h_{opt}^r} + \frac{1-\eta}{2\eta}\frac{1}{h_{opt}^2} = \log n, \tag{16}$$

satisfies the following upper bounds in pointwise distance

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left|\widehat{W}_{h,n}^{\eta}(z) - W_\rho(z)\right|^2\right] \varphi_n^{-2}(z) \le 1,$$

where the pointwise rate is

$$\varphi_n^2(z) = \frac{L}{4\pi\beta r}\left(\frac{2\eta}{1-\eta}\log n\right)^{1-r/2} \exp\left(-\frac{2\beta}{h_{opt}^r}\right).$$

Moreover, the previous rate is minimax efficient, i.e. the following lower bounds hold

$$\liminf_{n\to\infty} \inf_{\widehat{W}_n} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left|\widehat{W}_n(z) - W_\rho(z)\right|^2\right] \varphi_n^{-2}(z) \ge 1, \quad \forall z \in \mathbb{R},$$

where $\inf_{\widehat{W}_n}$ is taken over all possible estimators $\widehat{W}_n$ of the Wigner function W_ρ.

Proof. The proof of the lower bounds is given in Section 5.
Sketch of proof of the upper bounds. By Proposition 1 we write

$$\sup_{z\in\mathbb{R}^2} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left|\widehat{W}_{h,n}^{\eta}(z) - W_\rho(z)\right|^2\right] \le C_B h^{r-2} \exp\left(-\frac{2\beta}{h^r}\right) + \frac{C_V}{n}\exp\left(\frac{1-\eta}{2\eta}\frac{1}{h^2}\right),$$

where C_B and C_V denote the constant terms, depending on β, r, L and η. We select the best bandwidth as $h_{opt} = \arg\inf_{h>0}\{C_B h^{r-2}\exp(-2\beta/h^r) + C_V/n \exp((1-\eta)/(2\eta h^2))\}$. By taking derivatives, h_{opt} is a positive real number satisfying

$$\frac{2\beta}{h^r} + \frac{1-\eta}{2\eta h^2} = \log n + C(1 + o(1)), \text{ as } n \to \infty,$$

where $C > 0$ depends on β, r, L and η. This allows us to check easily that

$$h_{opt}^{r-2}\exp\left(-\frac{2\beta}{h_{opt}^r}\right) = h_{opt}^{r-2}\frac{C(1+o(1))}{n}\exp\left(\frac{1-\eta}{2\eta h_{opt}^2}\right),$$

i.e. the bias is asymptotically larger than the variance, for all $0 < r < 2$. If we replace h_{opt} by the solution of equation (16), the upper bounds will remain asymptotically the same (see Butucea and Tsybakov [3]). ■

3.2 $\mathbb{L}_2$ risk estimation

We establish next the properties of the same estimator when the quality of estimation is measured in $\mathbb{L}_2$ distance. In the literature $\mathbb{L}_2$ tomography is usually performed for bounded supported functions, see Korostelev and Tsybakov [13] and Johnstone and Silverman [12]. However no Wigner function can have a bounded support! Instead, we assume that it is fast decreasing. Thus, we modify the estimator by truncating it over a disc with increasing radius, as $n \to \infty$. Let us denote

$$D(s_n) = \{z = (q,p) \in \mathbb{R}_2 : \|z\| \leq s_n\},$$

where $s_n \to \infty$ as $n \to \infty$ will be defined in Theorem 2. Let now

$$\widehat{W}_{h,n}^{\eta,*}(z) = \widehat{W}_{h,n}^{\eta}(z) I_{D(s_n)}(z). \tag{17}$$

From now on, we will denote for any function f,

$$\|f\|_{D(s_n)}^2 = \int_{D(s_n)} f^2(z)dz,$$

and by $\overline{D}(s_n)$ the complementary set of $D(s_n)$ in $\mathbb{R}^2$. Then,

$$\begin{aligned}
E\left[\left\|\widehat{W}_{h,n}^{\eta,*} - W_\rho\right\|_2^2\right] &= E\left[\left\|\widehat{W}_{h,n}^{\eta} - W_\rho\right\|_{D(s_n)}^2\right] + \|W_\rho\|_{\overline{D}(s_n)}^2 \\
&= E\left[\left\|\widehat{W}_{h,n}^{\eta} - E\left[\widehat{W}_{h,n}^{\eta}\right]\right\|_{D(s_n)}^2\right] + \left\|E\left[\widehat{W}_{h,n}^{\eta}\right] - W_\rho\right\|_{D(s_n)}^2 \\
&\quad + \|W_\rho\|_{\overline{D}(s_n)}^2.
\end{aligned}$$

When replacing the $\mathbb{L}_2$ norm with the above restricted integral, the upper bound of the bias of the estimator is unchanged, whereas the variance part is infinitely larger than the deconvolution variance in Butucea and Tsybakov [3]. As the bias is dominating over the variance in this setup, we can still choose a suitable sequence s_n so that the same bandwidth is optimal associated to the same optimal rate, provided that W_ρ decreases fast enough asymptotically. We suppose, additionnally, that W_ρ is A-subgaussian, i.e. for some fixed real z_0 and some constant $A > 0$

$$|W_\rho(z)| \leq c \exp(-A\|z\|^2), \quad \forall \|z\|^2 \geq z_0.$$

Let us denote $\mathcal{A}(\beta, r, L, A)$ the class of $A-$subgaussian Wigner functions belonging to $\mathcal{A}(\beta, r, L)$. The following proposition gives upper bounds for the three components of the $\mathbb{L}_2$ risk uniformly over the class $\mathcal{A}(\beta, r, L, A)$.

Proposition 2 *Let (Y_ℓ, Φ_ℓ), $\ell = 1, \ldots, n$ be i.i.d. data coming from the model (13) and let $\widehat{W}_{h,n}^{\eta}$ be an estimator (with $h \to 0$ as $n \to \infty$) of the underlying Wigner function*

W_ρ. We suppose W_ρ lies in the class $\mathcal{A}(\beta, r, L, A)$, with $0 < r < 2$. Then, for $s_n \to \infty$ as $n \to \infty$ and n large enough,

$$\begin{aligned}
\sup_{W_\rho \in \mathcal{A}(\beta,r,L,A)} \left\| E[\widehat{W}^\eta_{h,n}] - W_\rho \right\|^2_{D(s_n)} &\leq L \exp\left(-\frac{2\beta}{h^r}\right)(1+o(1)), \\
\sup_{W_\rho \in \mathcal{A}(\beta,r,L,A)} E\left[\left\| \widehat{W}^\eta_{h,n} - E\left[\widehat{W}^\eta_{h,n}\right]\right\|^2_{D(s_n)}\right] &\leq \frac{M(\eta) s_n^2}{16\pi^2 \gamma n h} \exp\left(\frac{2\gamma}{h^2}\right)(1+o(1)), \\
\sup_{W_\rho \in \mathcal{A}(\beta,r,L,A)} \|W^\rho\|^2_{\bar{D}(s_n)} &\leq C e^{-2As_n^2},
\end{aligned}$$

where $\gamma = (1-\eta)/(4\eta) > 0$, $M(\eta) > 0$ is the constant appearing in Lemma 5 depending only on η and $o(1) \to 0$ as $h \to 0$ and $n \to \infty$.

In the following Theorem we use the phenomenon which was noticed already: deconvolution with Gaussian type noise is a much harder problem than inverse Radon transform (the tomography part). We expect then to attain the deconvolution rates with $\mathbb{L}_2$ risk. We define the kernel estimator with bandwidth h_{opt} in (18) known to be optimal for $\mathbb{L}_2$ deconvolution with Gaussian noise (see Butucea and Tsybakov [3]). The additional terms due to tomography will be smaller than the dominant term, due to the choice of the radius s_n of the disc $D(s_n)$, and we attain indeed the rate.

Tomography induces the drawback that we cannot prove the optimality of this estimator in this context. Indeed, the Wigner functions that we construct for the pointwise lower bounds in Section 5 do not fit into the class of A-subgaussian functions we considered here, due to their polynomial asymptotic decay.

Theorem 2 *Let $\beta > 0$, $L > 0$, $0 < r < 2$, $A > 0$ and (Y_ℓ, Φ_ℓ), $\ell = 1, \ldots, n$ be i.i.d. data coming from the model (13). Then $\widehat{W}^{\eta,*}_{h,n}$ defined in (17) with kernel K^η_h in (15), $s_n = \sqrt{\log n}$ and bandwidth $h = h_{opt}$ solution of the equation*

$$\frac{2\beta}{h^r_{opt}} + \frac{1-\eta}{2\eta h^2_{opt}} = \log n - (\log\log n)^2, \tag{18}$$

satisfies the following upper bounds in $\mathbb{L}_2$ distance

$$\limsup_{n\to\infty} \sup_{W_\rho \in \mathcal{A}(\beta,r,L,A)} E\left[\left\| \widehat{W}^{\eta,*}_{h,n} - W_\rho \right\|^2\right] \varphi_n^{-2}(\mathbb{L}_2) \leq 1,$$

where the $\mathbb{L}_2$ rate is

$$\varphi_n^2(\mathbb{L}_2) = L \exp\left(-\frac{2\beta}{h^r_{opt}}\right).$$

Sketch of proof of the upper bounds. By Proposition 2, we get

$$\sup_{W_\rho \in \mathcal{A}(\beta,r,L,A)} E\left[\left\| \widehat{W}^\eta_{h,n} - W_\rho \right\|^2\right] \leq L \exp\left(-\frac{2\beta}{h^r}\right) + \frac{C_V s_n^2}{n} \exp\left(\frac{1-\eta}{2\eta}\frac{1}{h^2}\right) + Ce^{-2As_n^2},$$

where C_V and C denote the constant terms, depending on β, r, L, A and η. Now let us see that for $h = h_{opt}$ in (18)

$$\frac{C_V s_n^2}{n} \exp\left(\frac{1-\eta}{2\eta}\frac{1}{h_{opt}^2}\right) = C_V \frac{\log n}{n} \exp\left(\log n - (\log\log n)^2 - \frac{2\beta}{h_{opt}^r}\right)$$
$$= o(1) \exp\left(-\frac{2\beta}{h_{opt}^r}\right).$$

Moreover, $\exp(-2As_n^2) = n^{-2A} = o(1)\exp\left(-2\beta/h_{opt}^r\right)$. Indeed, the last term is slower than any polynomial, but faster than any logarithm. So, the dominant term is of the order of $\varphi_n^2(\mathbb{L}_2)$. ■

3.3 Adaptive estimators

In the previous theorems the kernel estimator $\widehat{W}_{h,n}^{\eta}$ has a bandwidth $h = h_{opt}$ which is the solution of the equations (16) and (18) respectively, depending on the parameters β and r of the class. In the next theorem we will show that there exists an adaptive estimator, i.e. not depending on parameters, performing as well as the former estimators, provided that these parameters lie in a certain set. Indeed let us define

$$\mathcal{B}_1 = \{(\beta, r, L) : \beta > 0,\ 0 < r < 1,\ L > 0\},$$

respectively,

$$\mathcal{B}_2 = \{(\beta, r, L) : 0 < \beta \le B,\ r = 1,\ L > 0\},$$

where $B > 0$.

Theorem 3 *Let (Y_ℓ, Φ_ℓ), $\ell = 1, \ldots, n$ be i.i.d. data coming from the model (13). Then $\widehat{W}_{h,n}^{\eta}$ with $h = h_{ad}^i$, $i = 1, 2$*

$$h_{ad}^1 = \left(\frac{2\eta\log n}{1-\eta} - \sqrt{\frac{2\eta\log n}{1-\eta}}\right)^{-1/2} \text{ and } h_{ad}^2 = \left(\frac{2\eta\log n}{1-\eta} - \frac{4B\eta}{1-\eta}\sqrt{\frac{2\eta\log n}{1-\eta}}\right)^{-1/2}$$

is an optimal adaptive estimator over the set of parameters $\mathcal{B}_i$, $i = 1, 2$, respectively. That is, the estimator attains the same upper bounds, for all $(\beta, r, L) \in \mathcal{B}_i$

$$\limsup_{n\to\infty} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left|\widehat{W}_{h_{ad}^i,n}^{\eta}(z) - W_\rho(z)\right|^2\right] \varphi_n^{-2}(z) \le C_i,\ \forall z \in \mathbb{R}^2,$$
$$\limsup_{n\to\infty} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left\|\widehat{W}_{h_{ad}^i,n}^{\eta} - W_\rho\right\|^2\right] \varphi_n^{-2}(\mathbb{L}_2) \le C_i,$$

where the rates $\varphi_n^{-2}(z)$ and $\varphi_n^{-2}(\mathbb{L}_2)$ are given in Theorems 1 and 2, respectively, and the constants are, respectively,

$$C_1 = 1 \text{ and } C_2 = \exp\left(\frac{4\beta B\eta}{1-\eta} - \frac{4\beta^2\eta}{1-\eta}\right).$$

An important consequence is that in conjunction with the lower bounds in Theorem 1, the estimator $\widehat{W}^{\eta}_{h^1_{ad},n}$ is optimal adaptive and efficient over the set $\mathcal{B}_1$ for the pointwise risk. This means it attains the minimax rate and the constant $C_1 = 1$ for an estimator free of β, r and L provided that these parameters are in the class $\mathcal{B}_1$.

4 Proofs of upper bounds

Proof of Proposition 1. Since our data are i.i.d., we write

$$\begin{aligned} E[\widehat{W}^{\eta}_{h,n}(z)] &= \frac{1}{\pi}\int_0^{\pi}\int K_h^{\eta}([z,\phi]-y/\sqrt{\eta})p_{\rho}^{\eta}(y,\phi)dyd\phi \\ &= \frac{1}{\pi}\int_0^{\pi} K_h^{\eta} * (\sqrt{\eta}\, p_{\rho}^{\eta}(\cdot\sqrt{\eta},\phi))([z,\phi])d\phi. \end{aligned}$$

Now, write the convolution in the integral as an inverse Fourier transform. Indeed, it has Fourier transform (see (15)):

$$\begin{aligned} \mathcal{F}\left[K_h^{\eta} * \left(\sqrt{\eta}\, p_{\rho}^{\eta}(\cdot\sqrt{\eta},\phi)\right)\right](t) &= \widetilde{K}_h^{\eta}(t)\mathcal{F}_1[p_{\rho}^{\eta}(\cdot,\phi)](t/\sqrt{\eta}) \\ &= \frac{1}{2}|t|\mathcal{F}_1[p_{\rho}(\cdot,\phi)](t)I(|t|\le 1/h). \end{aligned}$$

Replace this into the expected value of our estimator and use (7)

$$\begin{aligned} E[\widehat{W}^{\eta}_{h,n}(z)] &= \frac{1}{4\pi^2}\int_0^{\pi}\int_{-1/h}^{1/h} e^{-it[z,\phi]}|t|\widetilde{W}_{\rho}(t\cos\phi, t\sin\phi)\, dt\, d\phi \\ &= \frac{1}{4\pi^2}\int\int e^{-i(qu+pv)}\widetilde{W}_{\rho}(u,v)I(\sqrt{u^2+v^2}\le 1/h)\, du\, dv \\ &= \frac{1}{4\pi^2}\int e^{-i\langle z,w\rangle}\widetilde{W}_{\rho}(w)I(\|w\|\le 1/h)dw, \end{aligned} \tag{19}$$

where we denote by $w = (u, v)$.

Recall that we also have

$$W_{\rho}(z) = \frac{1}{4\pi^2}\int e^{-i\langle z,w\rangle}\widetilde{W}_{\rho}(w)\, dw,$$

and then we write for the pointwise bias of our estimator:

$$\begin{aligned} \left|E[\widehat{W}^{\eta}_{h,n}](z) - W_{\rho}(z)\right|^2 &= \frac{1}{(4\pi^2)^2}\left|\int e^{-i\langle z,w\rangle}\left\{\mathcal{F}\left[E[\widehat{W}^{\eta}_{h,n}]\right](w) - \widetilde{W}_{\rho}(w)\right\}\, dw\right|^2 \\ &\le \frac{1}{(4\pi^2)^2}\int\left|\widetilde{W}_{\rho}(w)\right|^2 e^{2\beta\|w\|^r}\, dw\int_{\|w\|>1/h} e^{-2\beta\|w\|^r}\, dw \\ &\le \frac{Lh^{r-2}}{4\pi\beta r}e^{-2\beta/h^r}(1+o(1)),\ \text{as } h\to 0, \end{aligned}$$

by the assumption on our class.

As for the variance of our estimator:

$$\begin{aligned} V\left[\widehat{W}^{\eta}_{h,n}(z)\right] &= E\left[\left|\widehat{W}^{\eta}_{h,n}(z) - E\left[\widehat{W}^{\eta}_{h,n}(z)\right]\right|^2\right] \\ &= \frac{1}{\pi^2 n}\left\{E\left[\left|K_h^{\eta}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right|^2\right] - \left|E\left[K_h^{n}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right]\right|^2\right\} \end{aligned} \tag{20}$$

On the one hand, by using the Fourier transform computed in (19) and Cauchy-Schwarz we get

$$\begin{aligned} \left|E\left[K_h^{n}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right]\right| &\le \frac{1}{4\pi^2}\int\left|\widetilde{W}_\rho(w)\right| I\left(\|w\|\le 1/h\right)dw \\ &\le \frac{\sqrt{L}}{2\pi}\left(\int_{\|w\|\le 1/h} e^{-2\beta\|w\|^r}dw\right)^{1/2} \le M, \quad \forall\, z\in\mathbb{R}^2 \end{aligned} \tag{21}$$

where $M = M(\beta, r, L)$ is a constant depending only on the parameters of the class of functions.

On the other hand, the dominant term in the variance will be

$$E\left[\left|K_h^{\eta}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right|^2\right] = \int_0^{\pi}\int\left(K_h^{\eta}([z,\phi]-y/\sqrt{\eta})\right)^2 p_\rho^{\eta}(y,\phi)dyd\phi. \tag{22}$$

At this point, let us denote

$$G(t) := \mathcal{F}\left[K_h^{\eta}([z,\phi]-\cdot/\sqrt{\eta})\right](t) = \sqrt{\eta}e^{it[z,\phi]\sqrt{\eta}}\widetilde{K}_h^{\eta}(-t\sqrt{\eta}).$$

Replace in (22) by taking into account that for a probability density $p_\rho^{\eta}(\cdot,\phi)$ we have $|\mathcal{F}_1[p_\rho^{\eta}(\cdot,\phi)]| \le 1$,

$$\begin{aligned} E\left[\left|K_h^{\eta}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right|^2\right] &= \int_0^{\pi}\frac{1}{2\pi}\left|\int G * G(t)\mathcal{F}_1[p_\rho^{\eta}(\cdot,\phi)](t)dt\right|d\phi \\ &\le \frac{1}{2}\left(\int|G(t)|dt\right)^2 \le \frac{1}{2}\left(\frac{\eta}{2}\int_{|t|\le 1/(h\sqrt{\eta})}\frac{|t|}{\widetilde{N}^{\eta}(t)}dt\right)^2. \end{aligned}$$

Finally we obtain,

$$E\left[\left|K_h^{\eta}\left([z,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right|^2\right] \le \frac{1}{2}\left(2\eta\int_0^{1/(h\sqrt{\eta})}\frac{t}{2}\exp\left(t^2\frac{1-\eta}{4}\right)dt\right)^2. \tag{23}$$

Let us note here that, more generally, for any positive a, s and for any $A\in\mathbb{R}$, we have

$$\int_0^x t^A\exp(at^s)dt = \frac{1}{as}x^{A+1-s}\exp(ax^s)(1+o(1)), \quad \text{as } x\to\infty. \tag{24}$$

This asymptotic evaluation of the integral is based on integration by parts.

We use formula (24) for the integral in (23) as $1/h \to \infty$ and together with (20) and (21) we obtain

$$V\left[\widehat{W}_{h,n}^{\eta}(z)\right] \leq \frac{2\eta^2}{\pi^2(1-\eta)^2 n} \exp\left(\frac{1-\eta}{2\eta}\frac{1}{h^2}\right)(1+o(1)),\ n \to \infty.$$

∎

Proof of Proposition 2. We have proven in (19) that

$$\mathcal{F}\left[E[\widehat{W}_{h,n}^{\eta}]\right](w) = \widetilde{W}_{\rho}(w) I(\|w\| \leq 1/h).$$

Now we write for the $\mathbb{L}_2$ bias of our estimator:

$$\begin{aligned}
\|E[\widehat{W}_{h,n}^{\eta}] - W_{\rho}\|_{D(s_n)}^2 &\leq \|E[\widehat{W}_{h,n}^{\eta}] - W_{\rho}\|_2^2 = \frac{1}{4\pi^2}\|\mathcal{F}\left[E[\widehat{W}_{h,n}^{\eta}]\right] - \widetilde{W}_{\rho}\|_2^2 \\
&= \frac{1}{4\pi^2}\int \left|\widetilde{W}_{\rho}(w)\right|^2 I(\|w\| > 1/h)\, dw \\
&\leq e^{-2\beta/h^r}\frac{1}{4\pi^2}\int \left|\widetilde{W}_{\rho}(w)\right|^2 e^{2\beta\|w\|^r}\, dw \leq L e^{-2\beta/h^r},
\end{aligned}$$

by the assumption on our class.
As for the variance of our estimator:

$$\begin{aligned}
&V\left[\widehat{W}_{h,n}^{\eta}\right] = E\left[\left\|\widehat{W}_{h,n}^{\eta} - E\left[\widehat{W}_{h,n}^{\eta}\right]\right\|_{D(s_n)}^2\right] \\
&= \frac{1}{\pi^2 n}\left\{E\left[\left\|K_h^{\eta}\left([\cdot,\Phi] - \frac{Y}{\sqrt{\eta}}\right)\right\|_{D(s_n)}^2\right] - \left\|E\left[K_h^{n}\left([\cdot,\Phi] - \frac{Y}{\sqrt{\eta}}\right)\right]\right\|_{D(s_n)}^2\right\}. \quad (25)
\end{aligned}$$

By using two-dimensional Plancherel formula and the Fourier transform computed in (19), we get:

$$\left\|E\left[K_h^{n}\left([\cdot,\Phi] - \frac{Y}{\sqrt{\eta}}\right)\right]\right\|_{D(s_n)}^2 \leq \pi^2\int |W(w)|^2 dw \leq \pi/2. \quad (26)$$

In the last inequality we have used the fact that $\|W\|_2^2 = \frac{1}{2\pi}\mathrm{Tr}(\rho^2) \leq \frac{1}{2\pi}$ where ρ is the density matrix corresponding to the Wigner function W (see equation (9)).
On the other hand, the dominant term in the variance will be given by

$$\begin{aligned}
E\left[\left\|K_h^{\eta}\left([\cdot,\Phi] - \frac{Y}{\sqrt{\eta}}\right)\right\|_{D(s_n)}^2\right] &= \int_0^{\pi}\int\int_{D(s_n)} \left(K_h^{\eta}([z,\phi] - y/\sqrt{\eta})\right)^2 dz\, p_{\rho}^{\eta}(y,\phi) dy d\phi \\
&= \int_0^{\pi}\int_{D(s_n)}\int \left(K_h^{\eta}(u)\right)^2 \sqrt{\eta}\, p_{\rho}^{\eta}(([z,\phi] - u)\sqrt{\eta},\phi) du dz d\phi \\
&= \int \left(K_h^{\eta}(u)\right)^2 \int_{D(s_n)}\int_0^{\pi} p_{\rho}(\cdot,\phi) * NN^{\eta}([z,\phi] - u) d\phi dz du \\
&\leq M(\eta)\pi s_n^2 \int (K_h^{\eta}(u))^2 du,
\end{aligned}$$

using Lemma 5 and the constant $M(\eta) > 0$ depending only on η, defined therein. Indeed, let us note that $\sqrt{\eta}p_\rho^\eta(\cdot\sqrt{\eta},\phi)$ is the density of $Y/\sqrt{\eta} = X + \sqrt{(1-\eta)/(2\eta)}\varepsilon$ and let us call NN^η the Gaussian density of the noise as normalized in this last equation. Let us first compute $\|K_h^\eta\|_2^2$ by applying Plancherel formula and (24):

$$\begin{aligned}
\|K_h^\eta\|_2^2 &= \frac{1}{2\pi}\int |\widetilde{K}_h^\eta(t)|^2 dt = \frac{1}{2\pi}\int_{|t|\le 1/h} \frac{t^2}{4\widetilde{N}^2(t\sqrt{(1-\eta)/(2\eta)})}dt \\
&= \frac{1}{4\pi}\int_0^{1/h} t^2 \exp\left(t^2\frac{1-\eta}{2\eta}\right)dt \\
&= \frac{1}{4\pi h}\frac{\eta}{1-\eta}\exp\left(\frac{1-\eta}{2\eta h^2}\right)(1+o(1)),\ \text{ as } h\to 0.
\end{aligned}$$

We replace in the second order moment, then as $h \to 0$

$$E\left[\left\|K_h^\eta\left([\cdot,\Phi]-\frac{Y}{\sqrt{\eta}}\right)\right\|^2_{D(s_n)}\right] \le \frac{M(\eta)s_n^2}{16\gamma h}\exp\left(\frac{2\gamma}{h^2}\right)(1+o(1)). \qquad (27)$$

The result about the variance of the estimator is obtained from (25)-(27).

At last, for n large enough $s_n \ge z_0$ and thus

$$\begin{aligned}
\|W_\rho\|^2_{D(s_n)} &\le \int_{\|z\|>s_n} \exp(-2A\|z\|_2^2)dz \\
&\le \int_0^{2\pi}\int_{s_n}^\infty t\exp(-2At^2)dtd\phi \le \frac{\pi}{2A}e^{-2As_n^2}(1+o(1)).
\end{aligned}$$

■

Proof of Theorem 3. Let us discuss the pointwise risk problem briefly. Over $\mathcal{B}_1$,

$$E\left[\left|\widehat{W}^\eta_{h^1_{ad},n}(z) - W_\rho(z)\right|^2\right] \le \frac{L}{4\pi\beta r}(h^1_{ad})^{r-2}\exp\left(-\frac{2\beta}{(h^1_{ad})^r}\right) + \frac{2\eta^2}{\pi^2(1-\eta^2)n}\exp\left(\frac{1-\eta}{2\eta(h^1_{ad})^2}\right),$$

and it is easy to check that, for $(\beta,r,L)\in\mathcal{B}_1$

$$\begin{aligned}
\exp\left(-\frac{2\beta}{(h^1_{ad})^r}\right) &\le \exp\left(-\frac{2\beta}{h^r_{opt}}\right)(1+o(1)), \\
\frac{1}{n}\exp\left(\frac{1-\eta}{2\eta(h^1_{ad})^2}\right) &= \exp\left(-\sqrt{\frac{\eta-1}{2\eta}}\log n\right) = o(1)\exp\left(-\frac{2\beta}{h^r_{opt}}\right).
\end{aligned}$$

Thus, $\widehat{W}^\eta_{h^1_{ad},n}$ attains precisely the rate φ_n^2 ($C_1 = 1$).

Over the set $\mathcal{B}_2$ we have $r = 1$ and simple calculation show that

$$h_{opt} = \left(\frac{2\eta\log n}{1-\eta} - \frac{4\beta\eta}{1-\eta}\sqrt{\frac{1-\eta}{2\eta\log n}}\right)^{-1/2},$$

is a correct approximation in this case, giving a variance infinitely smaller than the bias which is of order

$$\varphi_n^2 = \frac{L}{4\pi\beta r}\sqrt{\frac{1-\eta}{2\eta\log n}}\exp\left(-2\beta\sqrt{\frac{2\eta\log n}{1-\eta}+\frac{4\beta^2\eta}{1-\eta}}\right)(1+o(1)).$$

As for the estimator with bandwidth h_{ad}^2:

$$E\left[\left|\widehat{W}^{\eta}_{h^2_{ad},n}(z) - W_\rho(z)\right|^2\right] \leq \frac{L}{4\pi\beta r}\sqrt{\frac{1-\eta}{2\eta\log n}}\exp\left(-2\beta\sqrt{\frac{2\eta\log n}{1-\eta}} + \frac{4\beta B\eta}{1-\eta}\right),$$

hence the results. The same reasoning holds for the $\mathbb{L}_2$ risk. ∎

5 Proof of the lower bound for the pointwise risk

In this section we will construct a pair of Wigner functions W_1 and W_2 depending on a parameter $\tilde{h}$ such that $\tilde{h} \to 0$ as $n \to \infty$. The choice of $\tilde{h}$ (see equation (37)) is such that it insures the existence of the lower bound in Theorem 1, and it should not be confused with the window h appearing in the expression of the estimator which is optimal with respect to the upper bounds. We choose W_1 and W_2 of the form

$$W_1(z) = W_0(z) + V_{\tilde{h}}(z) \quad \text{and} \quad W_2(z) = W_0(z) - V_{\tilde{h}}(z),$$

where W_0 is a fixed Wigner function corresponding to the density matrix ρ_0. The function V_z is not a Wigner function of a density matrix but belongs to the linear span of the space of Wigner functions and thus has a corresponding matrix $\tau^{\tilde{h}}$ in the linear span of density matrices. The choice of $W_0, V_{\tilde{h}}$ is such that

$$\rho_1 = \rho_0 + \tau^{\tilde{h}} \quad \text{and} \quad \rho_2 = \rho_0 - \tau^{\tilde{h}},$$

are density matrices (positive and trace equal to one) with Radon transforms p_1 and p_2. Suppose that the following conditions are fulfilled:

$$W_1 \text{ and } W_2 \qquad \text{belong to the class } \mathcal{A}(\beta, r, L), \tag{28}$$

$$|W_2(z) - W_1(z)| \geq 2\varphi_n(z)(1+o(1)), \text{ as } n \to \infty, \tag{29}$$

$$n\chi^2 := n\int_0^\pi \int \frac{(p_2^\eta(y,\phi) - p_1^\eta(y,\phi))^2}{p_1^\eta(y,\phi)} dy d\phi = o(1), \text{ as } n \to \infty. \tag{30}$$

Then we reduce the minimax risk to these two functions, W_1 and W_2 and bound the max from below by the mean of the two risks, to get for some $0 < \tau < 1$

$$\begin{aligned}
&\inf_{\widehat{W}_n} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} E\left[\left|\widehat{W}_n(z) - W_\rho(z)\right|^2\right] \\
\geq\ & \left(\inf_{\widehat{W}_n} \frac{1}{2}\left(E_{\rho_1}\left[\left|\widehat{W}_n(z) - W_1(z)\right|\right] + (1-\tau)E_{\rho_1}\left[I\left[\frac{dP^\eta_{\rho_2}}{dP^\eta_{\rho_1}} \geq 1-\tau\right]\left|\widehat{W}_n(z) - W_2(z)\right|\right]\right)\right)^2 \\
\geq\ & \frac{(1-\tau)^2}{4} \cdot (2\varphi_n)^2 P^2_{\rho_1}\left[\frac{dP^\eta_{\rho_2}}{dP^\eta_{\rho_1}} \geq \tau\right](1+o(1)).
\end{aligned}$$

We used the triangular inequality to get rid of the estimator and (29). Following Lemma 4 in Butucea and Tsybakov [3], we know that the last probability in the display above is bounded from below by $1-\tau^2$ provided that $n\chi^2 \le \tau^4$. It is therefore sufficient to check (30), in order to find $\tau_n \to 0$, as $n \to \infty$ and give a lower bound of the minimax risk of order $\varphi_n^2(1+o(1))$, for any estimator $\widehat{W}_n$.

We construct first the functions $W_{1,2}$ and then prove (28)-(30) in Subsection 5.3.

5.1 Construction of the density matrix ρ_0

In this subsection we will construct a family of density matrices ρ^α from which we will later select $\rho_0 = \rho^{\alpha_0}$ used in the lower bound. We derive their asymptotic behavior in Lemmas 1 and 2, and we show that W_α belongs to the class $\mathcal{A}(\beta, r, L)$ for $\alpha > 0$ small enough.

Let us consider the Mehler formula, (see Erdelyi *et al.* [6], 10.13.22)

$$\sum_{n=0}^{\infty} z^n \frac{1}{\sqrt{\pi} n! 2^n} H_n(x)^2 e^{-x^2} = \frac{1}{\sqrt{\pi(1-z^2)}} \exp\left(-x^2 \frac{1-z}{1+z}\right), \tag{31}$$

where H_n are the Hermite polynomials. Integrating both terms with $f_\alpha(z) = \alpha(1-z)^\alpha$, for some $0 < \alpha \le 1$, we get

$$p_\alpha(x,\phi) := \sum_{n=0}^{\infty} \psi_n(x)^2 \int_0^1 f_\alpha(z) z^n \, dz = \int_0^1 \frac{f_\alpha(z)}{\sqrt{\pi(1-z^2)}} \exp\left(-x^2 \frac{1-z}{1+z}\right) dz, \tag{32}$$

where ψ_j are the orthonormal vectors defined in (10). The Fourier transform of p_α is

$$\widetilde{W}_\alpha(w) = \mathcal{F}_1[p_\alpha](\|w\|, \phi) = \int_0^1 \frac{f_\alpha(z)}{1-z} \exp\left(-\|w\|^2 \frac{1+z}{4(1-z)}\right) dz \tag{33}$$

Notice that the normalization condition $\int p_\alpha = 1$ is equivalent to $\widetilde{W}_\alpha(0) = 1$ which is satisfied for the chosen functions f_α, thus p_α is a probability density. From the first equality in (32) we deduce that p_α is the probability density corresponding to a diagonal density matrix ρ^α with elements

$$\rho^\alpha_{k,k} = \int_0^1 z^k f_\alpha(z) \, dz. \tag{34}$$

We look now at the behavior of $p_\alpha(x,\phi)$ with respect to x.

Lemma 1 *For all $0 < \alpha \le 1$ and $|x| > 1$ there exist constants c, C depending on α, such that*

$$c|x|^{-(1+2\alpha)} \le p_\alpha(x,\phi) \le C|x|^{-(1+2\alpha)}. \tag{35}$$

Proof. We have

$$p_\alpha(x,\phi) = \frac{\alpha}{\sqrt{\pi}} \int_0^1 \frac{(1-z)^{\alpha-1/2}}{(1+z)^{1/2}} \exp\left(-x^2 \frac{1-z}{1+z}\right) dz,$$

which by the change of variables $u = x\sqrt{\frac{1-z}{1+z}}$ becomes

$$p_\alpha(x,\phi) = \frac{\alpha 2^{\alpha+1}|x|}{\sqrt{\pi}} \int_0^x \frac{u^{2\alpha}}{(u^2+x^2)^{\alpha+1}} \exp(-u^2)\, du.$$

The last integral is bounded for $|x| \geq 1$ as follows

$$\frac{\alpha 2^{\alpha+1}}{\sqrt{\pi}|x|^{2\alpha+1}} \int_0^1 u^{2\alpha} \exp(-u^2)\, du \leq p_\alpha(x,\phi) \leq \frac{\alpha 2^{\alpha+1}}{\sqrt{\pi}|x|^{2\alpha+1}} \int_1^\infty u^{2\alpha} \exp(-u^2)\, du.$$

■

A similar analysis can be done for the matrix elements of ρ^α. Let us consider some particular cases first. The matrix ρ^1, for $\alpha = 1$, has elements $\rho^1_{n,n} = \frac{1}{(n+1)(n+2)}$. For $\alpha = 1/2$, the coefficients of the corresponding matrix $\rho^{1/2}$ are

$$\rho^{1/2}_{nn} = \frac{1}{2} \int_0^1 z^n \sqrt{1-z} = \frac{2^{(2n+2)} n!}{(2n+1)!(2n+3)}.$$

By using Stirling's formula

$$n! = \sqrt{2\pi} n^{n+1/2} e^{-n} (1+o(1)), \text{ as } n \to \infty,$$

we get that $\rho^{1/2}_{n,n}$ decreases as $\sqrt{\pi} n^{-3/2}$ for $n \to \infty$.

Lemma 2 *For all* $0 < \alpha \leq 1$ *we have*

$$\rho^\alpha_{n,n} = \alpha \Gamma(\alpha+1) n^{-(1+\alpha)} (1+o(1)), \text{ as } n \to \infty.$$

Proof. We have

$$\rho^\alpha_{n,n} = \alpha \int_0^1 z^n (1-z)^\alpha = \alpha \frac{\Gamma(1+\alpha)\Gamma(1+n)}{\Gamma(2+\alpha+n)}.$$

We have $\Gamma(1+n) = n!$ and from Stirling formula, for large n

$$\Gamma(2+\alpha+n) = \sqrt{2\pi}(2+\alpha+n)^{\alpha+n+3/2} e^{-2+\alpha+n} (1+o(1)),$$

thus as $n \to \infty$, we obtain

$$\rho^\alpha_{n,n} = \alpha \Gamma(\alpha+1) n^{-(1+\alpha)} (1+o(1)), \text{ as } n \to \infty.$$

■

Lemma 3 *For any (β, r, L) such that $0 < r < 2$, there exists an $\alpha > 0$ such that W_α belongs to the class $\mathcal{A}(\beta, r, L)$.*

Proof. Using (33) and applying the generalized Minkowski inequality we get

$$\begin{aligned}\int e^{2\beta\|w\|^r}\left|\widetilde{W}_\alpha(w)\right|^2 dw &= 2\pi\int_0^\infty\left|\int_0^1 \sqrt{t}\frac{f_\alpha(z)}{1-z}\exp\left(-t^2\frac{1+z}{4(1-z)}+\beta t^r\right)dz\right|^2 dt\\ &\leq 2\pi\alpha^2\left[\int_0^1\left(\int_0^\infty t\exp\left(-t^2\frac{1+z}{2(1-z)}+2\beta t^r\right)dt\right)^{1/2}(1-z)^{\alpha-1}\,dz\right]^2.\end{aligned}$$

Now

$$\begin{aligned}\int_0^\infty t\exp\left(-t^2\frac{1+z}{2(1-z)}+2\beta t^r\right)dt &= \frac{1-z}{1+z}\int_0^\infty\exp\left(-u^2+2\beta(2u)^{r/2}\left(\frac{1-z}{1+z}\right)^{r/2}\right)du\\ &\leq (1-z)\int_0^\infty\exp(-u^2+2\beta(2u)^{r/2})du.\end{aligned}$$

Thus

$$\int e^{2\beta\|w\|^r}\left|\widetilde{W}_\alpha(w)\right|^2 dw \leq C(\beta,r)\alpha^2 \to 0,$$

as $\alpha \to 0$, where $C(\beta, r) > 0$. ■

5.2 Construction of $V_{\tilde{h}}$ and asymptotic properties of $\rho^{\tilde{h}}$

.

Let $V_{\tilde{h}}$ be the function defined on $\mathbb{R}^2$ whose Fourier transform is

$$\mathcal{F}_2[V_{\tilde{h}}](w) = \widetilde{V}_{\tilde{h}}(w) := J_{\tilde{h}}(t) = 2\sqrt{\pi\beta r L}\,\tilde{h}^{1-r/2}e^{\beta/\tilde{h}^r}e^{-2\beta|t|^r}J\left(|t|^r - \frac{1}{\tilde{h}^r}\right), \tag{36}$$

where $t = \|w\|$, and J is 3-times continuously differentiable function with bounded derivatives and such that

$$I_{[2\delta,D-2\delta]}(u) \leq J(u) \leq I_{[\delta,D-\delta]}(u),$$

for some $\delta > 0$ and $D > 4\delta$. The choice of the function $V_{\tilde{h}}$ is motivated by the results on lower bounds for deconvolution obtained in Butucea and Tsybakov [3]. The parameter $\tilde{h} \to 0$ as $n \to \infty$ is solution of the equation

$$\frac{2\beta}{\tilde{h}^r} + \frac{1-\eta}{2\tilde{h}^2} = \log n + (\log\log n)^2. \tag{37}$$

We think of $V_{\tilde{h}}$ as a function belonging to the linear span of the Wigner functions. Indeed, as shown in equation (9) the convex map sending a density matrix ρ to its corresponding Wigner function W_ρ can be extended by linearity to an isometry (up to a constant) with respect to the $\|\cdot\|_2$ on the two spaces. We can thus construct a matrix $\tau^{\tilde{h}}$ belonging to the linear span of the space of density matrices and whose corresponding Wigner is $V_{\tilde{h}}$. Because the function $V_{\tilde{h}}$ is invariant under rotations in the plane, the corresponding matrix has all off-diagonal elements equal to 0 and for the diagonal ones we can use the following formula from Leonhardt [14]

$$\tau_{nn}^{\tilde{h}} = 4\pi^2 \int_0^\infty L_n(t^2/2)e^{-t^2/4} t J_{\tilde{h}}(t)\, dt. \tag{38}$$

Lemma 4 *The matrix $\tau^{\tilde{h}}$ has the following asymptotic behavior*

$$\tau_{nn}^{\tilde{h}} = O(n^{-5/4})\, o_{\tilde{h}}(1). \tag{39}$$

Proof. We use the differential equation of the Laguerre polynomials, Gradshteyn and Ryzhik [8] 8.979:

$$L_n(x) = \frac{1}{n}\left((x-1)L_n'(x) - xL_n''(x)\right).$$

Thus

$$\begin{aligned} \frac{d}{dt}L_n(t^2/2) &= tL_n'(t^2/2), \\ \frac{d^2}{dt^2}L_n(t^2/2) &= L_n'(t^2/2) + t^2 L_n''(t^2/2) \end{aligned}$$

which implies

$$\frac{t^2}{2}L_n''(t^2/2) = \frac{1}{2}\frac{d^2}{dt^2}L_n(t^2/2) - \frac{1}{2}t^{-1}\frac{d}{dt}L_n(t^2/2),$$

and

$$L_n(t^2/2) = \frac{1}{2n}\left((t^2-1)t^{-1}\frac{d}{dt}L_n(t^2/2) - \frac{d^2}{dt^2}L_n(t^2/2)\right).$$

Using integration by parts we obtain

$$4\pi^2 \int_0^\infty L_n(t^2/2)e^{-t^2/4} t J_{\tilde{h}}(t)\, dt = \frac{1}{n}\int_0^\infty L_n(t^2/2)e^{-t^2/4}\left[P_1(t)J_{\tilde{h}}(t) + P_2(t)J_{\tilde{h}}'(t) + P_3(t)J_{\tilde{h}}''(t)\right] dt$$

with $P_i(t)$ polynomials with degree at most three, whose coefficients do not depend on $\tilde{h}$ or k. As the support of the function under the integral is contained in the interval $[1/\tilde{h}, \infty)$ we can use the following bound for the behavior of Laguerre polynomials (see Szegö [16] Theorem 8.9.12):

$$\sup_{x\in[1,\infty)} e^{-x/2}|L_n(x)| = O(n^{-1/4}).$$

The matrix $\tau^{\tilde{h}}$ has thus the following asymptotic behavior

$$\tau_{nn}^{\tilde{h}} \le Cn^{-5/4}\int_{1/\tilde{h}}^\infty |P_1(t)J_{\tilde{h}}(t) + P_2(t)J_{\tilde{h}}'(t) + P_3(t)J_{\tilde{h}}''(t)| = O(n^{-5/4})\, o_{\tilde{h}}(1). \tag{40}$$

$\blacksquare$

5.3 Lower bounds

Lemma 3 implies that for any α small enough the Wigner function W_α belongs to the class $\mathcal{A}(\beta, r, a^2 L)$. On the other hand, combining the results of Lemma 2 and Lemma 4 we get that for any $\alpha < 1/4$ the diagonal matrices $\rho_1 = \rho^\alpha + \tau^{\tilde{h}}$ and $\rho_2 = \rho^\alpha - \tau^{\tilde{h}}$ are positive and have trace one for $\tilde{h}$ sufficiently small. Thus, there exists an α_0 such that the corresponding ρ_1 and ρ_2 are density matrices and $W_0 = W_{\alpha_0} \in \mathcal{A}(\beta, r, a^2 L)$.

Proof of (28). By the triangle inequality

$$\left\|\mathcal{F}_2[W_{1,2}]e^{\beta\|\cdot\|^r}\right\|_2 \leq \left\|\mathcal{F}_2[W_0]e^{\beta\|\cdot\|^r}\right\|_2 + \left\|\mathcal{F}_2[V_{\tilde{h}}]e^{\beta\|\cdot\|^r}\right\|_2.$$

The first term in the sum above is less than $2\pi\sqrt{L}a$. For the second one we have

$$\begin{aligned}
\int \left|\mathcal{F}_2[V_{\tilde{h}}](w)\right|^2 e^{2\beta\|w\|^r} dw &= \int_0^\pi \int |t| \left|\mathcal{F}_2[V_{\tilde{h}}](t\cos\phi, t\sin\phi)\right|^2 e^{2\beta|t|^r} dt d\phi \\
&= \pi \int |t| \left|J_{\tilde{h}}(t)\right|^2 e^{2\beta|t|^r} dt \\
&\leq 4\pi^2 \beta r L \tilde{h}^{2-r} e^{2\beta/\tilde{h}^r} \int_{\delta \leq |t|^r - 1/\tilde{h}^r \leq D-\delta} |t| e^{-2\beta|t|^r} dt \\
&\leq 4\pi^2 \beta r L \tilde{h}^{2-r} e^{2\beta/\tilde{h}^r} 2 \int_{(1+\delta\tilde{h}^r)^{1/r}/\tilde{h}^r}^{\infty} t e^{-2\beta t^r} dt \\
&\leq 4\pi^2 L e^{-2\beta\delta}.
\end{aligned}$$

Thus, if we take $a = 1 - e^{-\beta\delta/2}$, we get $W_{1,2}$ in the class $\mathcal{A}(\beta, r, L(1 - e^{-\beta\delta/2} + e^{-\beta\delta}))$ included in $\mathcal{A}(\beta, r, L)$. ■

Proof of (29). Notice that

$$\begin{aligned}
|W_2(z) - W_1(z)|^2 &= \left|\frac{1}{4\pi^2} \int_{\mathbb{R}^2} e^{-i\langle z,w\rangle} (\widetilde{W}_2(w) - \widetilde{W}_1(w)) dw\right|^2 \\
&= \left|\frac{1}{4\pi^2} \int_0^{2\pi} \int_0^\infty e^{-it[z,\phi]} |t| (\widetilde{W}_2(t\cos\phi, t\sin\phi) - \widetilde{W}_1(t\cos\phi, t\sin\phi)) dt d\phi\right|^2 \\
&= \left|\frac{1}{2\pi^2} \int_0^{2\pi} \int_0^\infty e^{-it[z,\phi]} |t| J_{\tilde{h}}(t) dt d\phi\right|^2.
\end{aligned}$$

Take $z = 0$ without loss of generality:

$$\begin{aligned}
|W_2(z) - W_1(z)|^2 &= \left|\frac{1}{2\pi} \int_0^\pi \int |t| J_{\tilde{h}}(t) dt\right|^2 \\
&\geq 4\pi\beta r L \tilde{h}^{2-r} e^{2\beta/\tilde{h}^r} \left|\frac{1}{2\pi} \int_{2\delta \leq |t|^r - 1/\tilde{h}^r \leq D-2\delta} |t| e^{-2\beta|t|^r} dt\right|^2 \\
&\geq \frac{1}{\pi} L\beta r \tilde{h}^{2-r} e^{2\beta/\tilde{h}^r} \left(2 \int_{(1+2\delta\tilde{h}^r)^{1/r}/\tilde{h}}^{(1+(D-2\delta)\tilde{h}^r)^{1/r}/\tilde{h}} t e^{-2\beta t^r} dt\right)^2 \\
&\geq 4 \frac{L}{4\pi\beta r} \tilde{h}^{r-2} e^{-2\beta/\tilde{h}^r} [e^{-4\beta\delta}(1 + o(1) - e^{-2\beta(D-2\delta)}(1 + o(1))]^2,
\end{aligned}$$

which is larger than $4\varphi_n^2(\tilde{h})[e^{-4\beta\delta} - e^{-2\beta(D-2\delta)}]^2(1+o(1))$ for n large enough. ■

Proof of (30). We want to bound from above $n\chi^2 \leq \pi n \int (p_2^\eta(y) - p_1^\eta(y))^2/p_1^\eta(y)dy$. We have proven that $p_1(x) \geq Cx^{-2}$ for all $|x| \geq 1$. It is easy to prove, that after convolution with the gaussian density of the noise the asymptotic decay can not be faster

$$p_1^\eta(y) \geq \frac{c_1}{y^2}, \, \forall\, |y| \geq M,$$

for some fixed $M > 0$. Then we split the integration domain into $|y| \leq M$ and $|y| > M$ and get

$$n\chi^2 \leq Cn \left(C(M)\|p_2^\eta - p_1^\eta\|^2 + \int_{|y|>M} y^2(p_2^\eta(y) - p_1^\eta(y))^2 dy \right). \tag{41}$$

Let us see first that

$$\begin{aligned}
\|p_2^\eta - p_1^\eta\|^2 &= C \int |J_{\tilde{h}}(t)|^2 e^{-(1-\eta)t^2/2} dt \\
&\leq C\tilde{h}^{1-r} \exp\left(\frac{2\beta}{\tilde{h}^r}\right) \int_{(1+\delta\tilde{h}^r)^{1/r}/\tilde{h}}^{\infty} e^{-4\beta t^r - (1-\eta)t^2/2} dt \\
&\leq C\tilde{h}^{2-r} \exp\left(-\frac{2\beta}{\tilde{h}^r} - \frac{1-\eta}{2\tilde{h}^2}\right). \qquad (42)
\end{aligned}$$

Then

$$\begin{aligned}
\int_{|y|>M} y^2(p_2^\eta(y) - p_1^\eta(y))^2 dy &\leq \int \left(\frac{\partial}{\partial t}(J_{\tilde{h}}(t)e^{-(1-\eta)t^2/4})\right)^2 dt \\
&\leq C\tilde{h}^{1-r} \exp\left(\frac{2\beta}{\tilde{h}^r}\right) \int_{(1+\delta\tilde{h}^r)^{1/r}/\tilde{h}}^{\infty} t^2 e^{-4\beta t^r - (1-\eta)t^2/2} dt \\
&\leq C\tilde{h}^{-r} \exp\left(-\frac{2\beta}{\tilde{h}^r} - \frac{1-\eta}{2\tilde{h}^2}\right). \qquad (43)
\end{aligned}$$

It is enough to choose $\tilde{h}$ as solution of the equation (37) to get the expressions in (42) and (43) tend to 0 and together with (41) conclude. ■

6 Auxiliary results

Lemma 5 *For every $W_\rho \in \mathcal{A}(\beta, r, L)$ and $0 < \eta < 1$, we have that the corresponding probability density p_ρ satisfies*

$$0 \leq \int_0^\pi p_\rho(\cdot, \phi) * NN^\eta(x) d\phi \leq M(\eta),$$

for all $x \in \mathbb{R}$ eventually depending on ϕ, where $M(\eta) > 0$ is a constant depending only on fixed η.

Proof. Indeed, using inverse Fourier transform and the fact that $\left|\widetilde{W}_\rho(w)\right| \le 1$ we get:

$$\begin{aligned}\left|\int_0^\pi p_\rho(\cdot,\phi) * NN^\eta(x)d\phi\right| &\le \left|\int_0^\pi \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_1[p_\rho(\cdot,\phi)](t)\cdot \widetilde{NN}^\eta(t)dtd\phi\right| \\ &\le c(\eta)\int_0^\pi\int \left|\widetilde{W}_\rho(t\cos\phi, t\sin\phi)\right| \exp\left(-\frac{t^2(1-\eta)}{4\eta}\right) dtd\phi \\ &\le c(\eta)\int \frac{1}{\|w\|}\left|\widetilde{W}_\rho(w)\right| \exp\left(-\frac{\|w\|^2(1-\eta)}{4\eta}\right) dw \le M(\eta),\end{aligned}$$

where $c(\eta)$, $M(\eta)$ are positive constants depending only on $\eta \in (0,1)$. ■

Acknowledgments: Mădălin Guţă acknowledges the financial support received from the European IST Programme "Resources for quantum information" (RESQ), contract number IST-2001-37559.

References

[1] Artiles, L, Gill, R., Guţă, M., An invitation to quantum tomography, *J. Royal Statist. Soc. B (Methodological)*, **67**, (2005), 109–134.

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

[3] Butucea C., Tsybakov A.B., Sharp optimality for density deconvolution with dominating bias, arXiv:math.ST/0409471.

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

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

[6] Erdélyi A., Magnus W., Oberhettinger F., Tricomi F.G., *Higher transcendental functions. Vols. I, II*, McGraw-Hill Book Company, Inc., New York-Toronto-London (1953), based, in part, on notes left by Harry Bateman.

[7] Goldenshluger A., Spokoiny V., On the shape-from-moments problem and recovering edges from noisy Radon data, *Probab. Theory Related Fields*, **128**, no. 1, (2004), 123–140.

[8] Gradshteyn I.S., Ryzhik I.M., *Table of integrals, series, and products*, Academic Press Inc., San Diego, CA, sixth edition (2000), translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.

[9] Guţă M., Artiles. L, Minimax estimation of the Wigner in quantum homodyne tomography with ideal detectors (2005), in preparation.

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

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

[12] Johnstone I.M., Silverman B.W., Speed of estimation in positron emission tomography and related inverse problems, *Ann. Statist.*, **18**, no. 1, (1990), 251–280.

[13] Korostelëv A.P., Tsybakov A.B., *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*, Springer-Verlag, New York (1993).

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

[15] Smithey D.T., Beck, M., Raymer, M. G., Faridani, A., 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**, (1993), 1244–1247.

[16] Szegö G., *Orthogonal Polynomials*, Americam Mathematical Society, Providence (1975).

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

[18] Vogel K., Risken, H., Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, *Phys. Rev. A*, **40**, (1989), 2847–2849.

[1]Laboratoire de Probabilités et Modèles
Aléatoires (UMR CNRS 7599),
Université Paris VI
4, pl.Jussieu, Boîte courrier 188,
75252 Paris, France
email: butucea@ccr.jussieu.fr
and
Modal'X, Université Paris X
200, avenue de la République
92001 Nanterre Cedex, France

[2]Mathematisch Instituut
Universiteit Utrecht
Budapestlaan 6
3584 CD Utrecht, The Netherlands
email: guta@math.uu.nl

[3]Eurandom
Technische Universiteit Eindhoven
P.O. Box 513, 5600 MB Eindhoven
The Netehrlands
email: artiles@eurandom.tue.nl