Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography

Karim Lounici, Katia Meziani, Gabriel Peyré

Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography

Karim Lounici[*,??] **, Katia Meziani**[†,??] **and Gabriel Peyré** [‡,??]
EMAIL: ***klounici@math.gatech.edu*, *meziani@ceremade.dauphine.fr*, *gabriel.peyre@ceremade.dauphine.fr***

Abstract: In quantum optics, the quantum state of a light beam 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. In the framework of noisy quantum homodyne tomography with efficiency parameter $1/2 < \eta \leq 1$, we study the theoretical performance of a kernel estimator of the Wigner function. We prove that it is minimax efficient, up to a logarithmic factor in the sample size, for the $\mathbb{L}_\infty$-risk over a class of infinitely differentiable functions. We also compute the lower bound for the $\mathbb{L}_2$-risk. We construct an adaptive estimator, i.e. which does not depend on the smoothness parameters, and prove that it attains the minimax rates for the corresponding smoothness of the class of functions up to a logarithmic factor in the sample size. Finite sample behaviour of our adaptive procedure is explored through numerical experiments.

Keywords and phrases: Non-parametric minimax estimation, Adaptive estimation, Inverse problem, $\mathbb{L}_2$ and $\mathbb{L}_\infty$ Risks, Quantum homodyne tomography, Wigner function, Radon transform, Quantum state.

Quantum optics is a branch of quantum mechanics which studies physical systems at the atomic and subatomic scales. Unlike classical mechanics, the result of a physical measurement is generally random. Quantum mechanics does not predict a deterministic course of events, but rather the probabilities of various alternative possible events. It provides predictions on the outcome measures, therefore exploring measurements involves non-trivial statistical methods and inference on the result of a measurement should be done on identically prepared quantum systems.
In this paper, we study a *severely ill-posed inverse problem* that has arised in quantum optics. Let $(Z_1, \Phi_1), \ldots, (Z_n, \Phi_n)$ be n pairs of independent identically distributed random variables with values in $\mathbb{R} \times [0, \pi]$ satisfying

$$Z_\ell := X_\ell + \sqrt{2\gamma}\, \xi_\ell,$$

where X_l admits density $p(x, \phi)$ w.r.t. the Lebesgue measure on $\mathbb{R} \times [0, \pi]$, ξ_l is a standard normal variable independent of X_l and $\gamma \in (0, 1)$ is a known scalar. Due to the particular structure of this quantum optics problem, the density $p(x, \phi)$ satisfies

$$p(x, \phi) = \frac{1}{\pi} \mathcal{R}[W](x, \phi) \mathbb{1}_{[0,\pi]}(\phi),$$

where $W : \mathbb{R}^2 \to \mathbb{R}$ is the unknown function to be estimated based on indirect observations $(Z_1, \Phi_1), \ldots, (Z_n, \Phi_n)$ and $\mathcal{R}[W]$ is the Radon transform of W. The Radon transform will be properly defined in Section 1 below. The target W is called the Wigner function and is used to describe the quantum state of a physical system of interest.
For the interested reader, we provide in Section 1 a short introduction to the needed quantum notions. This section may be skipped at first reading. Section 2 introduces the statistical model by making the link with quantum theory. The interested reader can get further acquaintance with quantum concepts through the textbooks or the review articles of Helstrom (1976); Holevo (1982); Barndorff-Nielsen, Gill and Jupp (2003) and Leonhardt (1997).

[*]Supported in part by Simons Collaboration Grant 315477 and NSF CAREER Grant DMS-1454515.
[†]Supported in part by "Calibration" ANR-2011-BS01-010-01.
[‡]Supported by the European Research Council (ERC project SIGMA-Vision).

1. Physical background

In quantum mechanics, the measurable properties (ex: spin, energy, position, ...) of a quantum system are called "*observables*". The probability of obtaining each of the possible outcomes when measuring an observable is encoded in the *quantum state* of the considered physical system.

1.1. Quantum state and observable

The mathematical description of the quantum state of a system is given in the form of a density operator ρ on a complex Hilbert space $\mathcal{H}$ (called the space of states) satisfying the three following conditions:

1. Self adjoint: $\rho = \rho^*$, 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: $\mathrm{Tr}(\rho) = 1$.

Notice that $\mathcal{D}(\mathcal{H})$, the set of density operators ρ on $\mathcal{H}$, is a convex set. The extreme points of the convex set $\mathcal{D}(\mathcal{H})$ are called *pure states* and all other states are called *mixed states*.

In this paper, the quantum system we are interested in is a monochromatic light in a cavity. In this setting of quantum optics, the space of states $\mathcal{H}$ we are dealing with is the space of square integrable complex valued functions on the real line. A particular orthonormal basis for this Hilbert space is the Fock basis $\{\psi_j\}_{j\in\mathbb{N}}$:

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

where $H_j(x) := (-1)^j e^{x^2} \frac{d^j}{dx^j} e^{-x^2}$ denote the j-th Hermite polynomial. In this basis, a quantum state is described by an infinite density matrix $\rho = [\rho_{j,k}]_{j,k\in\mathbb{N}}$ whose entries are equal to

$$\rho_{j,k} = \langle \psi_j, \rho\psi_k \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product. The quantum states which can be created currently in laboratory are matrices whose entries are decreasing exponentially to 0, *i.e.*, these matrices belong to the natural class $\mathcal{R}(C, B, r)$ defined below, with $r = 2$. Let us define for $C \geq 1$, $B > 0$ and $0 < r \leq 2$, the class $\mathcal{R}(C, B, r)$ is as follows

$$\mathcal{R}(C, B, r) := \{\rho \text{ quantum state} : |\rho_{m,n}| \leq C \exp(-B(m+n)^{r/2})\}. \tag{2}$$

In order to describe mathematically a measurement performed on an observable of a quantum system prepared in state ρ, we give the mathematical description of an observable. An observable $\mathbf{X}$ is a self adjoint operator on the same space of states $\mathcal{H}$ and

$$\mathbf{X} = \sum_{a}^{dim\mathcal{H}} x_a \mathbf{P}_a,$$

where the eigenvalues $\{x_a\}_a$ of the observable $\mathbf{X}$ are real and $\mathbf{P}_a$ is the projection onto the one dimensional space generated by the eigenvector of $\mathbf{X}$ corresponding to the eigenvalue x_a.
As a quantum state ρ encompasses all the probabilities of the observables of the considered quantum system, when performing a measurement of the observable $\mathbf{X}$ of a quantum state ρ, the result is a random variable X with values in the set of the eigenvalues of the observable $\mathbf{X}$. For a quantum system prepared in state ρ, X has the following probability distribution and expectation function

$$\mathbb{P}_\rho(X = x_a) = \mathrm{Tr}(\mathbf{P}_a \rho) \quad \text{and} \quad \mathbb{E}_\rho(X) = \mathrm{Tr}(\mathbf{X}\rho).$$

Note that the conditions defining the density matrix ρ insure that $\mathbb{P}_\rho$ is a probability distribution. In particular, the characteristic function is given by

$$\mathbb{E}_\rho(e^{itX}) = \mathrm{Tr}(\rho e^{it\mathbf{X}}).$$

1.2. Quantum homodyne tomography and Wigner function

In quantum optics, a monochromatic light in a cavity is described by a quantum harmonic oscillator. In this setting, the observables of interest are usually $\mathbf{Q}$ and $\mathbf{P}$ (resp. the electric and magnetic fields). But according to Heisenberg's uncertainty principle, $\mathbf{Q}$ and $\mathbf{P}$ are non-commuting observables, they may not be simultaneously measurable. Therefore, by performing measurements on $(\mathbf{Q}, \mathbf{P})$, we cannot get a probability density of the result (Q, P). However, for all phase $\phi \in [0, \pi]$ we can measure the quadrature observables

$$\mathbf{X}_\phi := \mathbf{Q}\cos\phi + \mathbf{P}\sin\phi.$$

Each of these quadratures could be measured on a laser beam by a technique developed by *Smithey* and called **Quantum Homodyne Tomography** (QHT). The theoretical foundation of quantum homodyne tomography was outlined by Vogel and Risken (1989).
When performing a QHT measurement of the observable $\mathbf{X}_\phi$ of the quantum state ρ, the result is a random variable X_ϕ whose density conditionally to $\Phi = \phi$ is denoted by $p_\rho(\cdot|\phi)$. Its characteristic function is given by

$$\mathbb{E}_\rho(e^{itX_\phi}) = \mathrm{Tr}(\rho e^{it\mathbf{X}_\phi}) = \mathrm{Tr}(\rho e^{it(\mathbf{Q}\cos\phi + \mathbf{P}\sin\phi)}) = \mathcal{F}_1[p_\rho(\cdot|\phi)](t),$$

where $\mathcal{F}_1[p_\rho(\cdot|\phi)](t) = \int e^{itx} p_\rho(x|\phi)dx$ denotes the Fourier transform with respect to the first variable. Moreover if Φ is chosen uniformly on $[0, \pi]$, the joint density probability of (X_ϕ, Φ) with respect to the Lebesgue measure on $\mathbb{R} \times [0, \pi]$ is

$$p_\rho(x, \phi) = \frac{1}{\pi} p_\rho(x|\phi) \mathbb{1}_{[0,\pi]}(\phi).$$

An equivalent representation for a quantum state ρ is the function $W_\rho : \mathbb{R}^2 \to \mathbb{R}$ called the Wigner function, introduced for the first time by Wigner (1932). The Wigner function may be obtained from the momentum representation

$$\widetilde{W}_\rho(u, v) \ := \ \mathcal{F}_2[W_\rho](u, v) = \mathrm{Tr}\left(\rho e^{i(u\mathbf{Q} + v\mathbf{P})}\right), \tag{3}$$

where $\mathcal{F}_2$ is the Fourier transform with respect to both variables. By the change of variables (u, v) to polar coordinates $(t\cos\phi, t\sin\phi)$, we get

$$\widetilde{W}_\rho(t\cos\phi, t\sin\phi) = \mathcal{F}_1[p_\rho(\cdot|\phi)](t) = \mathrm{Tr}(\rho e^{it\mathbf{X}_\phi}). \tag{4}$$

The origin of the appellation quantum homodyne tomography comes from the fact that the procedure described above is similar to positron emission tomography (PET), where the density of the observations is the Radon transform of the underlying distribution

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

where $\mathcal{R}[W_\rho]$ denotes the Radon transform of W_ρ. The main difference with PET is that the role of the unknown distribution is played by the Wigner function which can be negative.
The physicists consider the Wigner function as a quasi-probability density of (Q, P) if one can measure simultaneously $(\mathbf{Q}, \mathbf{P})$. Indeed, the Wigner function satisfies

$$W_\rho : \mathbb{R}^2 \to \mathbb{R}, \quad \iint W_\rho(q, p)dqdp = 1, \tag{6}$$

and other boundedness properties unavailable for classical densities. However, the Wigner function can and normally does take negative values for states which are not associated to any classical model. This property of the Wigner function is used by Physicists as a criterion to discriminate nonclassical states of the field.

In the Fock basis, we can write W_ρ in terms of the density matrix $[\rho_{jk}]$ as follows (see Leonhardt (1997) for the details).

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

where for $j \geq k$,

$$W_{j,k}(q,p) \;=\; \frac{(-1)^j}{\pi}\left(\frac{k!}{j!}\right)^{\frac{1}{2}} \left(\sqrt{2}(ip-q)\right)^{j-k} e^{-\left(q^2+p^2\right)} L_k^{j-k}\left(2q^2+2p^2\right). \tag{7}$$

and $L_k^\alpha(x)$ the generalized Laguerre polynomial of degree k and order α.

1.3. Pattern functions

The ideal result of the QHT measurement provides (X_ϕ, Φ) of joint probability density with respect to the Lebesgue measure on $\mathbb{R} \times [0,\pi]$ equal to

$$p_\rho(x,\phi) = \frac{1}{\pi} p_\rho(x|\phi) \mathbb{I}_{[0,\pi]}(\phi) = \frac{1}{\pi} \mathcal{R}[W_\rho].(x,\phi) \mathbb{I}_{[0,\pi]}(\phi) \tag{8}$$

The density $p_\rho(\cdot,\cdot)$ can be written in terms of the entries of the density matrix ρ (see Leonhardt (1997))

$$p_\rho(x,\phi) = \sum_{j,k=0}^{\infty} \rho_{j,k} \psi_j(x) \psi_k(x) e^{-(j-k)\phi}, \tag{9}$$

where $\{\psi_j\}_{j\in\mathbb{N}}$ is the Fock basis defined in (1). Conversely (see D'Ariano, Macchiavello and Paris (1994); Leonhardt (1997) for details), we can write

$$\rho_{j,k} = \int_0^\pi\!\!\int p_\rho(x,\phi) f_{j,k}(x) e^{-(j-k)\phi} dx d\phi, \tag{10}$$

where the functions $f_{j,k} : \mathbb{R} \to \mathbb{R}$ introduced by Leonhardt, Paul and D'Ariano (1995) are called the *"pattern functions"*. An explicit form of the Fourier transform of $f_{j,k}(\cdot)$ is given by Richter (2000): for all $j \geq k$

$$\widetilde{f}_{j,k}(t) \;= \widetilde{f}_{k,j}(t) = \pi(-i)^{j-k} \sqrt{\tfrac{2^{k-j}k!}{j!}} |t| t^{j-k} e^{-\frac{t^2}{4}} L_k^{j-k}(\tfrac{t^2}{2}), \tag{11}$$

Note that by writing $t = ||w|| = ||(q,p)|| = \sqrt{q^2+p^2}$ in the equation (7), we can define for all $j \geq k$

$$l_{j,k}(t) := |W_{j,k}(q,p)| = \frac{2^{\frac{j-k}{2}}}{\pi}\left(\frac{k!}{j!}\right)^{\frac{1}{2}} t^{j-k} e^{-t^2} \left|L_k^{j-k}(2t^2)\right|. \tag{12}$$

Therefore, there exists a useful relation, for all $j \geq k$

$$\left|\widetilde{f}_{j,k}(t)\right| = \pi^2 |t| l_{j,k}(t/2). \tag{13}$$

Moreover Aubry, Butucea and Meziani (2009) have given the following Lemma which will be useful to prove our main results.

Lemma 1 (Aubry, Butucea and Meziani (2009))**.**
For all $j,k \in \mathbb{N}$ and $J := j+k+1$, for all $t \geq 0$,

$$l_{j,k}(t) \leq \frac{1}{\pi} \begin{cases} 1 & \text{if } 0 \leq t \leq \sqrt{J}, \\ e^{-(t-\sqrt{J})^2} & \text{if } t \geq \sqrt{J}. \end{cases} \tag{14}$$

2. Statistical model

In practice, when one performs a QHT measurement, a number of photons fails to be detected. These losses may be quantified by one single coefficient $\eta \in [0,1]$, such that $\eta = 0$ when there is no detection and $\eta = 1$ corresponds to the ideal case (no loss). The quantity $(1-\eta)$ represents the proportion of photons which are not detected due to various losses in the measurement process. The parameter η is supposed to be known, as physicists argue that their machines actually have high detection efficiency, *i.e.* $\eta \approx 0.9$. In this paper we consider the regime where more photons are detected than lost, that is $\eta \in (1/2, 1]$. Moreover, as the detection process is inefficient, an independent Gaussian noise interferes additively with the ideal data X_ϕ. Note that the Gaussian nature of the noise is imposed by the Gaussian nature of the vacuum state which interferes additively.
To sum up, for $\Phi = \phi$, the effective result of the QHT measurement is for a known efficiency $\eta \in (1/2, 1]$,

$$Y = \sqrt{\eta}\, X_\phi + \sqrt{(1-\eta)/2}\, \xi \tag{15}$$

where ξ is a standard Gaussian random variable, independent of the random variable X_ϕ having density, with respect to the Lebesgue measure on $\mathbb{R} \times [0, \pi]$, equal to $p_\rho(\cdot,\cdot)$ defined in equation (8). For the sake of simplicity, we re-parametrize (15) as follows

$$Z := Y/\sqrt{\eta} = X_\phi + \sqrt{(1-\eta)/(2\eta)}\, \xi := X_\phi + \sqrt{2\gamma}\, \xi, \tag{16}$$

where $\gamma = (1-\eta)/(4\eta)$ is known and $\gamma \in [0, 1/4)$ as $\eta \in (1/2, 1]$. Note that $\gamma = 0$ corresponds to the ideal case.
Let us denote by $p_\rho^\gamma(\cdot,\cdot)$ the density of (Z, Φ) which is the convolution of the density of X_ϕ with $N^\gamma(\cdot)$ the density of a centered Gaussian distribution having variance 2γ, that is

$$\begin{aligned} p_\rho^\gamma(z,\phi) &= \left[\frac{1}{\pi}\mathcal{R}[W_\rho](\cdot,\phi)\mathbb{I}_{[0,\pi]}(\phi)\right] * N^\gamma(z) = p_\rho(\cdot,\phi) * N^\gamma(z) \\ &= \int p_\rho(z-x,\phi)\, N^\gamma(x)dx. \end{aligned} \tag{17}$$

For $\Phi = \phi$, a useful equation in the Fourier domain, deduced by the previous relation (17) and equation (4) is

$$\mathcal{F}_1[p_\rho^\gamma(\cdot,\phi)](t) = \mathcal{F}_1[p_\rho(\cdot,\phi)](t)\widetilde{N}^\gamma(t) = \widetilde{W_\rho}(t\cos(\phi), t\sin(\phi))\widetilde{N}^\gamma(t), \tag{18}$$

where $\mathcal{F}_1$ denotes the Fourier transform with respect to the first variable and the Fourier transform of $N^\gamma(\cdot)$ is $\widetilde{N}^\gamma(t) = e^{-\gamma t^2}$.

This paper aims at reconstructing the Wigner function W_ρ of a monochromatic light in a cavity prepared in state ρ from n observations. As we cannot measure precisely the quantum state in a single experiment, we perform measurements on n independent identically prepared quantum systems. The measurement carried out on each of the n systems in state ρ is done by QHT as described in Section 1. In practice, the results of such experiments would be n independent identically distributed random variables $(Z_1, \Phi_1), \ldots, (Z_n, \Phi_n)$ such that

$$Z_\ell := X_\ell + \sqrt{2\gamma}\, \xi_\ell. \tag{19}$$

with values in $\mathbb{R} \times [0, \pi]$ and distribution $\mathbb{P}_\rho^\gamma$ admitting density $p_\rho^\gamma(\cdot,\cdot)$ defined in (17) with respect to the Lebesgue measure on $\mathbb{R} \times [0, \pi]$. For all $\ell = 1, \ldots, n$, the ξ_ℓ's are independent standard Gaussian random variables, independent of all (X_ℓ, Φ_ℓ).

In order to study the theoretical performance of our different procedures, we use the fact that the unknown Wigner function belongs to the class of very smooth functions $\mathcal{A}(\beta, r, L)$ (similar to

those of Butucea, Guţă and Artiles (2007); Aubry, Butucea and Meziani (2009)) described via its Fourier transform:

$$\mathcal{A}(\beta, r, L) := \left\{ f : \mathbb{R}^2 \to \mathbb{R},\ \iint |\widetilde{f}(u,v)|^2 e^{2\beta\|(u,v)\|^r} du dv \leqslant (2\pi)^2 L \right\}, \tag{20}$$

where $\widetilde{f}(\cdot,\cdot)$ denotes the Fourier transform of f with respect to both variables and $\|(u,v)\| = \sqrt{u^2+v^2}$ denote the usual Euclidean norm. Note that this class is reasonable from a physical point of view. Indeed, it follows from Propositions 1 and 2 in Aubry, Butucea and Meziani (2009) that any Wigner functio whose density matrix belongs to the realistic class $\mathcal{R}(C, B, r)$ lies in a class $\mathcal{A}(\beta', r, L')$ where $\beta' > 0$ and $L' > 0$ depend only on B, C, r. To the best of our knowledge, there exists no converse result proving that the density matrix of any Wigner function in the class $\mathcal{A}(\beta', r, L')$ belongs to $\mathcal{R}(C, B, r)$.

Previous works and outline of the results

The problem of reconstructing the quantum state of a light beam has been extensively studied in physics literature and in quantum statistics. We only mention papers with a theoretical analysis of the performance of their estimation procedure. Additional references to physics papers can be found therein. Methods for reconstructing a quantum state are based on the estimation of either the density matrix ρ or the Wigner function W_ρ. In order to assess the performance of a procedure, a realistic class of quantum states $\mathcal{R}(C, B, r)$ has been defined in many papers such as in (2) where the elements of the density matrix decrease rapidly. From the physics point of view, all the states which have been produced in the laboratory up to now belong to such a class with $r = 2$, and a more detailed argument can be found in the paper of Butucea, Guţă and Artiles (2007).

The estimation of the density matrix from averages of data has been considered in the framework of ideal detection ($\eta = 1$ *i.e.* $\gamma = 0$) by Artiles, Gill and Guţă (2005) while the noisy setting has been investigated by Aubry, Butucea and Meziani (2009) for the Frobenius norm risk. More recently in the noisy setting, an adaptive estimation procedure over the classes of quantum states $\mathcal{R}(C, B, r)$, *i.e.* without assuming the knowledge of the regularity parameters, has been proposed by Alquier, Meziani and Peyré (2013) and an upper bound for Frobenius risk has been given. The problem of goodness-of-fit testing in quantum statistics has been considered in Meziani (2008). In this noisy setting, the latter paper derived a testing procedure from a projection-type estimator where the projection is done in L_2 distance on some suitably chosen pattern functions.

The Wigner function is an appealing tool to Physicists to determine particular features of the quantum state of a system. Therefore, this work is of practical interest. For instance, non classical quantum state correspond to negative parts of the Wigner function. This paper deals with the problem of reconstruction of the Wigner function W_ρ in the context of QHT when taking into account the detection losses occurring in the measurement, leading to an additional Gaussian noise in the measurement data ($\eta \in (1/2, 1]$). In the absence of noise ($\gamma = 0$), Guţă and Artiles (2007) obtained the sharp minimax rate of pointwise estimation over the class of Wigner functions $\mathcal{A}(\beta, 1, L)$ for a kernel based procedure. The same problem in the noisy setting was treated by Butucea, Guţă and Artiles (2007), they obtain minimax rates for the pointwise risk over the class $\mathcal{A}(\beta, r, L)$ for the procedure defined in (21). Moreover, a truncated version of their estimator is proposed by Aubry, Butucea and Meziani (2009) where an upper bound is computed for the $\mathbb{L}_2$-norm risk over the class $\mathcal{A}(\beta, r, L)$. The estimation of a quadratic functional of the Wigner function, as an estimator of the purity, was explored in Meziani (2007).

The reconstruction problem considered in this paper belongs to the class of linear inverse problems. It requires to solve simultaneously a tomography problem and a density deconvolution problem. We refer to Cavalier (2008) for a survey of the literature on general inverse problems in statistics.

Tomography problems, such as noisy integral equation of the form $y = \mathcal{R}[f](x,\phi) + \xi$ where $(x,\phi) \in \mathbb{R} \times [0,\pi]$, ξ is some random noise and f is the unknown function to be recovered, have been investigated in Korostelëv and Tsybakov (1991, 1993); Klemelä and Mammen (2010) and the references cited therein. For density type tomography problems closer to our setting, Johnstone and Silverman (1990) considered uncorrupted observations, corresponding to $\gamma = 0$ in (19), and established the minimax rate of the inverse Radon transform over Sobolev classes of density functions for the quadratic risk. Under a similar framework, Donoho and Low (1992) obtained the pointwise minimax rate of reconstruction.
The deconvolution problem has been studied extensively in the literature. We refer to Bissantz, Dümbgen, Holzmann and Munk (2007); Bissantz and Holzmann (2008); Butucea and Tsybakov (2008a,b); Carroll and Hall (1988); Delaigle and Gijbels (2004); Diggle and Hall (1993); Fan (1991, 1993); Goldenshluger (1999); Hesse and Meister (2004); Johnstone, Kerkyacharian, Picard and Raimondo (2004); Johnstone and Raimondo (2004); Meister (2008); Pensky and Sapatinas (2009); Pensky and Vidakovic (1999); Stefanski (1990); Stefanski and Carroll (1990). Most of these papers concern the quadratic risk or the pointwise risk. Lounici, K. and Nickl (2011) established the first minimax uniform risk estimation result for a wavelet deconvolution density estimator over Besov classes of density functions.

The remainder of the article is organized as follows. In Section 3, we establish in Theorem 1 the first $\mathbb{L}_\infty$-norm risk upper bound for the estimation procedure (21) of the Wigner function while in Theorem 2 we establish the first minimax lower bounds for the estimation of the Wigner function for the $\mathbb{L}_2$-norm and the $\mathbb{L}_\infty$-norm risks. As a consequence of our results, we determined the minimax $\mathbb{L}_\infty$-norm and L_2-norm rates of estimation for this noisy QHT problem up to a logarithmic factor in the sample size. We propose in Section 4 a Lepski-type procedure that adapts to the unknown smoothness parameters $\beta > 0$ and $r \in (0,2]$ of the Wigner function of interest. The only previous result on adaptation is due to Butucea, Guţă and Artiles (2007) but concerns the simplest case $r \in (0,1)$ where the estimation procedure (21) with a proper choice of the parameter h independent of β, r is naturally minimax adaptive up to a logarithmic factor in the sample size n. Theoretical investigations are complemented by numerical experiments reported in Section 5. The proofs of the main results are deferred to the Appendix.

3. Wigner function estimation and minimax risk

From now on, we work in the practical framework and we assume that n independent identically distributed random pairs $(Z_i, \Phi_i)_{i=1,\dots,n}$ are observed, where Φ_i is uniformly distributed in $[0,\pi]$ and the joint density of (Z_i,Φ_i) is $p_\rho^\gamma(\cdot,\cdot)$ (see (17)). As Butucea, Guţă and Artiles (2007), we use the modified usual tomography kernel in order to take into account the additive noise on the observations and construct a kernel K_h^γ which performs both deconvolution and inverse Radon transform on our data, asymptotically such that our estimation procedure is

$$\widehat{W}_h^\gamma(q,p) = \frac{1}{2\pi n} \sum_{\ell=1}^{n} K_h^\gamma \left([z,\Phi_\ell] - Z_\ell \right), \tag{21}$$

where $0 \le \gamma < 1/4$ is a fixed parameter and $h > 0$ tends to 0 when $n \to \infty$ in a proper way to be chosen later. The kernel is defined by

$$\widetilde{K}_h^\gamma(t) = |t| e^{\gamma t^2} \mathbb{I}_{|t| \le 1/h}, \tag{22}$$

where $z = (q,p)$ and $[z,\phi] = q\cos\phi + p\sin\phi$.
From now on, $\|\cdot\|_\infty$ and $\|\cdot\|_2$ and $\|\cdot\|_1$ will denote respectively the $\mathbb{L}_\infty$-norm, the $\mathbb{L}_2$-norm and the $\mathbb{L}_1$-norm. As the $\mathbb{L}_\infty$-norm risk can be trivially bounded as follows

$$\|\widehat{W}_h^\gamma - W_\rho\|_\infty \le \|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty + \left\| \mathbb{E}\left[\widehat{W}_h^\gamma\right] - W_\rho \right\|_\infty, \tag{23}$$

and in order to study the $\mathbb{L}_\infty$-norm risk of our procedure $\widehat{W}_h^\gamma$, we study in Propositions 1 and 2, respectively the bias term and the stochastic term.

Proposition 1. *Let $\widehat{W}_h^\gamma$ be the estimator of W_ρ defined in (21) and $h>0$ tends to 0 when $n\to\infty$. Then,*

$$\left\|\mathbb{E}\left[\widehat{W}_h^\gamma\right]-W_\rho\right\|_\infty \le \sqrt{\frac{L}{(2\pi)^2\beta r}}h^{(r-2)/2}e^{-\beta h^{-r}}(1+o(1)),$$

where $W_\rho\in\mathcal{A}(\beta,r,L)$ defined in (20) and $r\in(0,2]$.

The proof is deferred to Appendix A.1.

Proposition 2. *Let $\widehat{W}_h^\gamma$ be the estimator of W_ρ defined in (21) and $0<h<1$. Then, there exists a constant C_1, depending only on γ such that*

$$\mathbb{E}\left[\|\widehat{W}_h^\gamma-\mathbb{E}[\widehat{W}_h^\gamma]\|_\infty\right]\le C_1e^{\gamma h^{-2}}\left(\sqrt{\frac{\log n}{n}}+\frac{\log n}{n}\right). \tag{24}$$

Moreover, for any $x>0$, we have with probability at least $1-e^{-x}$ that

$$\|\widehat{W}_h^\gamma-\mathbb{E}[\widehat{W}_h^\gamma]\|_\infty\le C_2e^{\gamma h^{-2}}\max\left\{\sqrt{\frac{\log(n)+x}{n}},\frac{\log(n)+x}{n}\right\}, \tag{25}$$

where $C_2>0$ depends only on γ.

The proof is deferred to Appendix A.2. The following theorem establishes the upper bound of the $\mathbb{L}_\infty$-norm risk.

Theorem 1. *Assume that W_ρ belongs to the class $\mathcal{A}(\beta,r,L)$ defined in (20) for some $r\in]0,2]$ and $\beta,L>0$. Consider the estimator (21) with $h^*=h^*(r)$ such that*

$$\begin{cases} \frac{\gamma}{(h^*)^2}+\frac{\beta}{(h^*)^r}=\frac{1}{2}\log(n/\log n) & \text{if} \quad 0<r<2,\\ h^*=\left(\frac{2(\beta+\gamma)}{\log(n/\log n)}\right)^{1/2} & \text{if} \quad r=2. \end{cases} \tag{26}$$

Then we have

$$\mathbb{E}\left[\|\widehat{W}_{h^*}^\gamma-W_\rho\|_\infty\right]\le Cv_n(r),$$

where $C>0$ can depend only on γ,β,r,L and the rate of convergence v_n is such that

$$v_n(r)=\begin{cases} (h^*)^{(r-2)/2}e^{-\beta(h^*)^{-r}} & \text{if} \quad 0<r<2,\\ \left(\frac{\log n}{n}\right)^{\frac{\beta}{2(\beta+\gamma)}} & \text{if} \quad r=2. \end{cases} \tag{27}$$

Note that for $r\in(0,2)$ the rate of convergence v_n is faster than any logarithmic rate in the sample size but slower than any polynomial rate. For $r=2$, the rate of convergence is polynomial in the sample size.

Proof of Theorem 1: Taking the expectation in (23) and using Propositions 1 and 2, we get for all $0<h<1$

$$\begin{aligned}\mathbb{E}\left[\|[\widehat{W}_h^\gamma-W_\rho\|_\infty\right] &\le \mathbb{E}\left[\|\widehat{W}_h^\gamma-\mathbb{E}[\widehat{W}_h^\gamma]\|_\infty\right]+\|\mathbb{E}\left[\widehat{W}_h^\gamma\right]-W_\rho\|_\infty\\ &\le Ce^{\gamma h^{-2}}\sqrt{\frac{1}{n}}(1+o(1))+C_Bh^{(r-2)/2}e^{-\beta h^{-r}}(1+o(1))\end{aligned}$$

where $C_B=\sqrt{\frac{L}{(2\pi)^2\beta r}}$, $h\to0$ as $n\to\infty$ and $W_\rho\in\mathcal{A}(\beta,r,L)$. The optimal bandwidth parameter $h^*(r):=h^*$ is such that

$$h^*=\arg\inf_{h>0}\left\{C_Bh^{(r-2)/2}e^{-\beta h^{-r}}+C_1e^{\gamma h^{-2}}\sqrt{\frac{\log n}{n}}\right\}. \tag{28}$$

Therefore, by taking derivative, we get

$$\frac{\gamma}{(h)^2}+\frac{\beta}{(h)^r}=\frac{1}{2}\log(n/\log n)+C_1(1+o(1)).$$

For $0<r<2$, (26) provides an accurate approximation of the optimum h^* when the number of observations n is large. By plugging the result into (28), we get

$$(h^*)^{(r-2)/2}e^{-\beta(h^*)^{-r}}=(h^*)^{(r-2)/2}\sqrt{\frac{\log n}{n}}e^{\gamma(h^*)^{-2}}.$$

It follows that the bias term is much larger than the stochastic term for $0<r<2$. It is easy to see that for $r=2$, we have $h^*=\left(\frac{2(\beta+\gamma)}{\log(n/\log n)}\right)^{1/2}$ and that the bias term and the stochastic term are of the same order. □

We derive now a minimax lower bound. We consider specifically the case $r=2$ since it is relevant with quantum physic applications. The only known lower bound result for the estimation of a Wigner function is due to Butucea, Guţă and Artiles (2007) and concerns the pointwise risk. In Theorem 2 below, we obtain the first minimax lower bounds for the estimation of a Wigner function $W_\rho\in\mathcal{A}(\beta,2,L)$ with the $\mathbb{L}_2$-norm and $\mathbb{L}_\infty$-norm risks.

Theorem 2. *Assume that $(Z_1,\Phi_1),\cdots,(Z_n,\Phi_n)$ coming from the model (16) with $\gamma\in[0,1/4)$. Then, for any $\beta,L>0$ and $p\in\{2,\infty\}$, there exists a constant $c:=c(\beta,L,\gamma)>0$ such that for n large enough*

$$\inf_{\widehat{W}_n}\sup_{W_\rho\in\mathcal{A}(\beta,2,L)}\mathbb{E}\|\widehat{W}_n-W_\rho\|_p\ \geq\ cn^{-\frac{\beta}{2(\beta+\gamma)}},$$

where the infimum is taken over all possible estimators $\widehat{W}_n$ based on the i.i.d. sample $\{(Z_i,\Phi_i)\}_{i=1}^n$.

We believe similar arguments can be applied to the case $0<r<2$ up to several technical modifications. This is left for future work. The proof is deferred to Appendix B. This theorem guarantees that the $\mathbb{L}_\infty$-norm upper bound derived in Theorem 1 and also that the $\mathbb{L}_2$-norm risk upper bound of Aubry, Butucea and Meziani (2009) are minimax optimal up to a logarithmic factor in the sample size n.

4. Adaptation to the smoothness

As we see in (28), the optimal choice of the bandwidth h^* depends on unknown smoothness parameters β and $r\in(0,2]$. We propose here to implement a Lepskii type procedure to select an adaptive bandwidth h. The Lepski method was introduced in Lepskiĭ (1991, 1992) and has become since then a popular method to solve various adaptation problems. We will show that the estimator obtained with this bandwidth achieves the optimal minimax rate for the $\mathbb{L}_\infty$-norm risk. Our adaptive procedure is implemented in Section 5.

Let $M\geq 2$, and $0<h_M<\cdots<h_1<1$ a grid of $]0,1[$, we build estimators $\widehat{W}^\gamma_{h_m}$ associated to bandwidth h_m for any $1\leq m\leq M$. For any fixed $x>0$, let us define $r_n(x)=\max\left(\sqrt{\frac{\log(n)+x}{n}},\frac{\log(n)+x}{n}\right)$. We denote by $\mathcal{L}_\kappa(\cdot)$, the Lepski functional such that

$$\begin{aligned}\mathcal{L}_\kappa(m)\ =\ &\max_{j>m}\left\{\|\widehat{W}^\gamma_{h_m}-\widehat{W}^\gamma_{h_j}\|_\infty-2\kappa e^{\gamma h_j^{-2}}r_n(x+\log M)\right\}\\&+2\kappa e^{\gamma h_m^{-2}}r_n(x+\log M),\end{aligned}\tag{29}$$

where $\kappa>0$ is a fixed constant. Therefore, our final adaptive estimator denoted by $\widehat{W}^\gamma_{h_{\widehat{\mathbf{m}}}}$ will be the estimator defined in (21) for the bandwidth $h_{\widehat{\mathbf{m}}}$. The bandwidth $h_{\widehat{\mathbf{m}}}$ is such that

$$\widehat{\mathbf{m}}=\mathrm{argmin}_{1\leq m\leq M}\mathcal{L}_\kappa(m).\tag{30}$$

Note that the following result is valid for any $\beta > 0$ and $r \in (0, 2]$.

Theorem 3. *Assume that $W_\rho \in \mathcal{A}(\beta, r, L)$. Take $\kappa > 0$ sufficiently large and $M \geq 2$. Choose $0 < h_M < \cdots < h_1 < 1$. Then, for the bandwidth $h_{\widehat{\mathbf{m}}}$ with $\widehat{\mathbf{m}}$ defined in (30) and for any $x > 0$, we have with probability at least $1 - e^{-x}$*

$$\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \leq C \min_{1\leq m\leq M} \left\{ h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}} r_n(x + \log M) \right\}, \tag{31}$$

where $C > 0$ is a constant depending only on γ, β, r, L.
In addition, we have in expectation

$$\mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \right] \leq C' \min_{1\leq m\leq M} \left\{ h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}} r_n(\log M) \right\}, \tag{32}$$

where $C' > 0$ is a constant depending only on γ, r, β, L.

The proof is deferred to the Appendix C.

The idea is now to build a sufficiently fine grid $0 < h_M < \cdots < h_1 < 1$ to achieve the optimal rate of convergence over the range $\beta > 0$. Take $M = \lfloor \sqrt{\log n/(2\gamma)} \rfloor$. We consider the following grid for the bandwitdh parameter h:

$$h_1 = 1/2, \quad h_m = \frac{1}{2}\left(1 - (m-1)\sqrt{\frac{2\gamma}{\log n}}\right), \quad 1 \leq m \leq M. \tag{33}$$

We build the corresponding estimators $\widehat{W}^{\gamma}_{h_m}$ and we apply the Lepski procedure (29)-(30) to obtain the estimator $\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}}$. The next result guarantees that this estimator is minimax adaptive over the class

$$\Omega := \{(\beta, r, L),\ \beta > 0,\ 0 < r \leq 2,\ L > 0\}.$$

Corollary 1. *Let the conditions of Theorem 3 be satisfied. Then the estimator $\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}}$ for the bandwidth $h_{\widehat{\mathbf{m}}}$ with $\widehat{\mathbf{m}}$ defined in (30) and for any $(\beta, r, L) \in \Omega$ satisfies*

$$\text{limsup}_{n\to\infty} \sup_{W_\rho \in \mathcal{A}(\beta,r,L)} \mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \right] \leq C v_n(r),$$

where $v_n(r)$ is the rate defined in (27) and C is a positive constant depending only on r, L, β and γ.

Proof of Corollary 1 : First note that for all $m = 1, \cdots, M$ and as

$$h_m \in ((\gamma/(2\log n))^{1/2}, 1/2],$$

the bias term $h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}}$ is larger than the stochastic term $e^{\gamma h_m^{-2}} r_n(\log M)$ up to a numerical constant. Let define

$$\widetilde{m} := \text{argmax}_{1\leq m\leq M}\{|h_m - h^*| : h_m \leq h^*\},$$

where $\widetilde{m}$ is well defined. Indeed, we have

$$\begin{aligned} \frac{h_M}{h^*} &= \frac{(1/2)\left(1 - M(2\gamma/\log n)^{1/2} + (2\gamma/\log n)^{1/2}\right)}{(\log n/(2\gamma) - (\beta/\gamma)(h^*)^{-r})^{-1/2}} \\ &= \frac{1}{2}\left(1 - M + ((\log n)/(2\gamma))^{1/2}\right)\left(1 - (2\beta/(\log(n))(h^*)^{-r}\right)^{1/2}. \end{aligned}$$

Moreover, as $0 \leq ((\log n)/(2\gamma))^{1/2} - M \leq 1$, we get

$$\frac{h_M}{h^*} \leq \left(1 - (2\beta/(\log(n))(h^*)^{-r}\right)^{1/2} \leq 1.$$

Therefore, from (32),

$$\mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty\right] \;\le\; C h_{\widetilde{m}}^{r/2-1} e^{-\frac{\beta}{h_{\widetilde{m}}^r}} \le C h_{\widetilde{m}}^{r/2-1} e^{-\frac{\beta}{h_{\widetilde{m}}^r}} v_n(r) v_n(r)^{-1}$$
$$= \; C \left(\frac{h_{\widetilde{m}}}{h^*}\right)^{r/2-1} e^{-\beta(h_{\widetilde{m}}^{-r} - (h^*)^{-r})} v_n(r).$$

By the definition of $\widetilde{m}$, it follows that $h_{\widetilde{m}}^{-r} \ge (h^*)^{-r}$, then

$$\mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty\right] \;\le\; C \left(\frac{h_{\widetilde{m}}}{h^*}\right)^{r/2-1} v_n(r) = C \left(\frac{h_{\widetilde{m}} - h^*}{h^*} + 1\right)^{r/2-1} v_n(r).$$

By construction $|h_{\widetilde{m}} - h^*| \le (\gamma/(2\log n))^{1/2}$, then we have

$$\mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty\right] \;\le\; C \left(1 - \frac{(\gamma/(2\log n))^{1/2}}{h^*}\right)^{r/2-1} v_n(r).$$

As $(h^*)^{-1} \le (\log n/(2\gamma))^{1/2}$, it holds that $1 - \frac{(\gamma/(2\log n))^{1/2}}{h^*} \ge 1/2$. Therefore, there exists a numerical constant $C' > 0$ such that, for any $0 < r \le 2$, we have $\mathbb{E}\left[\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty\right] \le C' v_n(r)$. □

5. Experimental evaluation

We test our method on two examples of Wigner functions, corresponding to the single-photon and the Schrödinger's cat states, and that are respectively defined as

$$W_\rho(q,p) = -(1 - 2(q^2+p^2))e^{-q^2-p^2},$$
$$W_\rho(q,p) = \frac{1}{2}e^{-(q-q_0)^2-p^2} + \frac{1}{2}e^{-(q+q_0)^2-p^2} + \cos(2q_0p)e^{-q^2-p^2}.$$

We used $q_0 = 3$ in our numerical tests. The toolbox to reproduce the numerical results of this article is available online[1]. Following the paper of Butucea, Guţă and Artiles (2007) and in order to obtain a fast numerical procedure, we implemented the estimator $\widehat{W}^{\gamma}_h$ defined in (21) on a regular grid. More precisely, 2-D functions such as W_ρ are discretized on a fine 2-D grid of 256×256 points. We use the Fast Slant Stack Radon transform of Averbuch et al. (2008), which is both fast and faithful to the continuous Radon transform $\mathcal{R}$. It also implements a fast pseudo-inverse which accounts for the filtered back projection formula (21). The filtering against the 1-D kernel (22) is computed along the radial rays in the Radon domain using Fast Fourier transforms. We computed the Lepski functional (29) using the values $x = \log(M)$ and $\kappa = 1$.

[1] `https://github.com/gpeyre/2015-AOS-AdaptiveWigner`

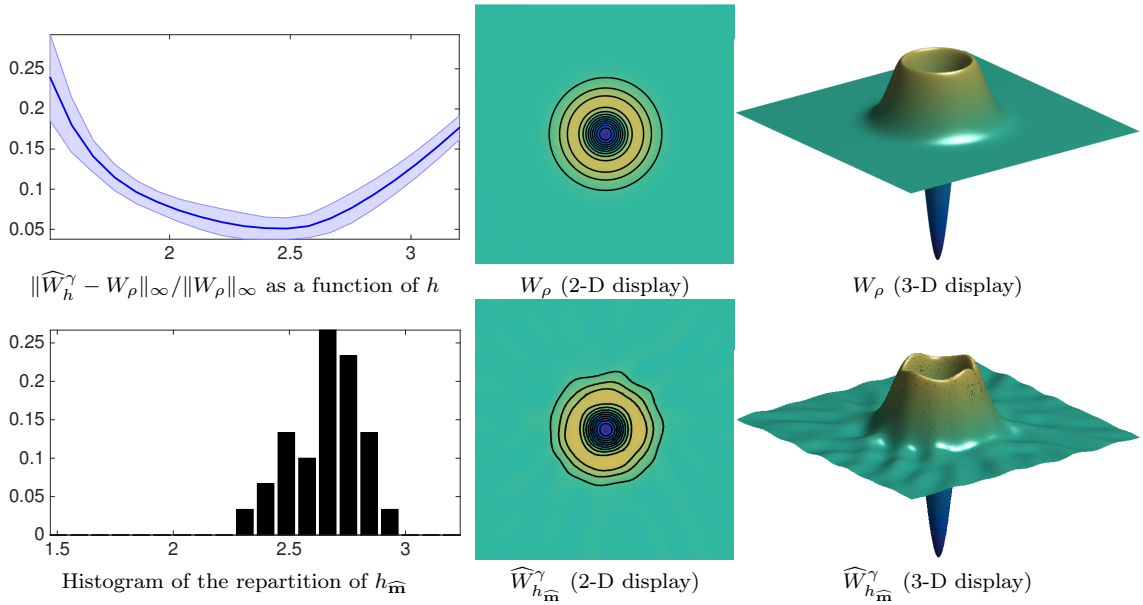

Figure 1: Single photon cat state estimation, with $\eta = 0.9$, $n = 100 \times 10^3$. *Left, top:* display of $\|\widehat{W}_h^\gamma - W_\rho\|_\infty / \|W_\rho\|_\infty$ as a function of $1/h$. The central curve is the mean of this quantity, while the shaded area displays the $\pm 2\times$ standard deviation of this quantity. *Left, right:* histogram of the empirical repartition of $\widehat{\mathbf{m}}$ computed by the Lepski procedure (30). *Center:* display as a 2-D image using level sets of W_ρ (top) and $\widehat{W}^\gamma_{h_{\widehat{\mathbf{m}}}}$ (bottom). *Right:* same, but displayed as an elevation surface.

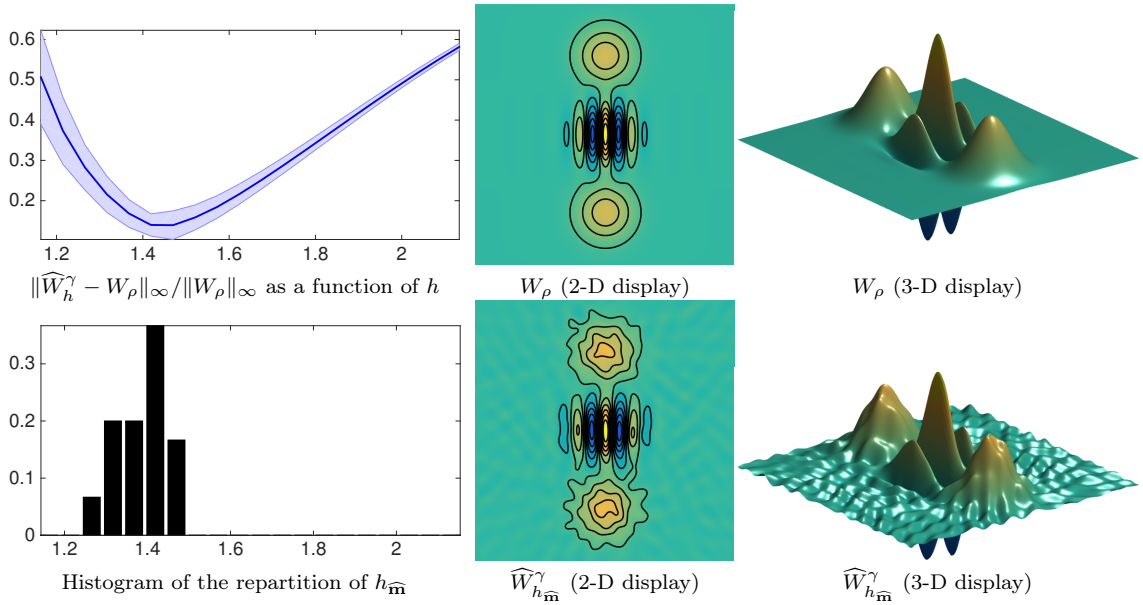

Figure 2: Schrödinger's cat state estimation, with $\eta = 0.9$, $n = 500 \times 10^3$. We refer to Figure 1 for the description of the plots.

Figures 1 and 2 reports the numerical results of our method on both test cases. The left part compares the error $\|\widehat{W}_h^\gamma - W_\rho\|_\infty$ (displayed as a function of h) to the parameters $h_{\widehat{\mathbf{m}}}$ selected by the Lepski procedure (30) . The error $\|\widehat{W}_h^\gamma - W_\rho\|_\infty$ (its empirical mean and its standard deviation) is computed in an "oracle" manner (since for these examples, the Wigner function to estimate W_ρ is known) using 20 realizations of the sampling for each tested value $(h_i)_{i=1}^M$. The histogram of values $h_{\widehat{\mathbf{m}}}$ is computed by solving (29) for 20 realizations of the sampling. This comparison shows, on both test cases, that the method is able to select a parameter value $h_{\widehat{\mathbf{m}}}$ which lies around the optimal parameter value (as indicated by the minimum of the $\mathbb{L}_\infty$-norm risk). The central and right parts show graphical displays of $\widehat{W}_{h_{\widehat{\mathbf{m}}}}^\gamma$, where $\widehat{\mathbf{m}}$ is selected using the Lepski procedure (30), for a given sampling realization.

Acknowledgement

We thank Laetitia Comminges for her careful reading of a preliminary version of the paper and her helpful comments that led to a logarithmic improvement in Theorem 2.

References

Abramowitz, M. and Stegun, I.A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover, New York, 1964.

Alquier, P., Meziani, K. and Peyré,G., *Adaptive Estimation of the Density Matrix in Quantum Homodyne Tomography with Noisy Data.* Inverse Problems, 29, 7, 075017, 2013.

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

Aubry, J.-M. and Butucea, C. and Meziani, K., *State estimation in quantum homodyne tomography with noisy data.* Inverse Problems, 25, 1, 2009.

Averbuch, A., Coifman, R.R., Donoho, D.L., Israeli, M., Shkolnisky, Y. and Sedelnikov, I. , *A framework for discrete integral transformations: II. The 2D discrete Radon transform.. SIAM J. Sci. Comput.*, 30(2), 785–803, 2008.

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

Bergh, J. and Löfström, J., *Interpolation spaces. An introduction.* Grundlehren der Mathematischen Wissenschaften, No. 223,Springer-Verlag, Berlin, 1976.

Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A., *Non-parametric confidence bands in deconvolution density estimation.* J. R. Stat. Soc. Ser. B Stat. Methodol., 69, 483–506, 2007.

Bissantz, N. and Holzmann, H., *Statistical inference for inverse problems.* Inverse Problems,24 034009, 2008.

Bourdaud, G. and Lanza de Cristoforis, M. and Sickel, W., *Superposition operators and functions of bounded p-variation. Rev. Math. Iberoam.*, 2, 455–487, 2006.

Bousquet, O., *A Bennett concentration inequality and its application to suprema of empirical processes.* C. R. Math. Acad. Sci. Paris, 334, 6, 495–500, 2002.

Butucea, C. and Guţă, M. and Artiles, L., *Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data.* Ann. Statist., 2, 35, 465–494, 2007.

Butucea C. and Tsybakov A. B., *Sharp Optimality in Density Deconvolution with Dominating Bias. I.* Theory of Probability & Its Applications, 52, 1, 24–39, 2008.

Butucea, C. and Tsybakov, A. B.. *Sharp optimality in density deconvolution with dominating bias. II.* Theory of Probability & Its Applications, 52, 237–249, 2008.

Carroll, R. J. and Hall, P., *Optimal rates of convergence for deconvolving a density.* J. Amer. Statist. Assoc., 83, 1184–1186, 1988.

Cavalier, L., *Nonparametric statistical inverse problems.* Inverse Problems, 24 034004, 2008.

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

Delaigle, A. and Gijbels, I., *Practical bandwidth selection in deconvolution kernel density estimation.* Comput. Statist. Data Anal, 45, 249–267, 2004.

Diggle, P. J. and Hall, P. (1993). *A Fourier approach to nonparametric deconvolution of a density estimate.* J. Roy. Statist. Soc. Ser. B, 55 523–531, 1993.

Donoho, D.L. and Low, M.G., *Renormalization exponents and optimal pointwise rates of convergence.* Ann. Statist., 20, 944–970, 1992.

Erdélyi A., Magnus W., Oberhettinger F., Tricomi F.G., *Higher transcendental functions.* McGraw-Hill Book Company, Inc., New York-Toronto- London , Vols. I, II, 1953.

Fan, J., *On the optimal rates of convergence for nonparametric deconvolution problems.* Ann. Statist., 19, 1257–1272, 1991.

Fan, J., *Adaptively local one-dimensional subproblems with application to a deconvolution problem.* Ann. Statist., 21, 600–610, 1993.

Feingold, D. G. and Varga, R. S., *Block Diagonally Dominant Matrices and Generalizations of the Gerschgorin Circle Theorem.* Pacific J. Math., 12, 1241–1250, 1962.

Giné, E. and Nickl, R., *Uniform limit Theorems for wavelet density estimators.* Ann. Probab., 37, 4,1605–1646, 2009.

Goldenshluger, A., *On pointwise adaptive nonparametric deconvolution.* Bernoulli, 5, 907–925, 1999.

Guţă, M. and Artiles, L., *Minimax estimation of the Wigner in quantum homodyne tomography with ideal detectors.* Math. Methods Statist., 16, 1,1–15, 2007.

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

Hesse, C. H. and Meister, A., *Optimal iterative density deconvolution.* J. Nonparametr. Statist., 16, 879–900, 2004.

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

Johnstone, I. M., Kerkyacharian, G., Picard, D. and Raimondo, M., *Wavelet deconvolution in a periodic setting.* J. R. Stat. Soc. Ser. B Stat. Methodol., 66, 547–573, 2004.

Johnstone, I. M. and Raimondo, M., *Periodic boxcar deconvolution and Diophantine approximation.* Ann. Statist., 32, 1781–1804, 2004.

Johnstone, I. M. and Silverman, B.W., *Speed of estimation in positron emission tomography and related inverse problems.* Ann. Statist., 18, 1, 251–280, 1990.

Klemelä, J. and Mammen, E., *Empirical risk minimization in inverse problems.* Ann. Statist., 38, 1, 482–511, 2010.

Korostelëv, A. P. and Tsybakov, A. B., *Optimal rates of convergence of estimates in a probabilistic formulation of the tomography problem.* Problemy Peredachi Informatsii, 27, 92–103, 1991.

Korostelëv, A. P. and Tsybakov, A. B., *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics 82. Springer, Berlin, 1993.

Krasikov, I., *Inequalities for orthonormal Laguerre polynomials.* J. Approx. Theory, 144, 1, 1–26, 2007.

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

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

Lepskiĭ, O. V., *Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates.* Teor. Veroyatnost. i Primenen., 36, 645–659, 1991.

Lepskiĭ, O. V., *Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates.* Teor. Veroyatnost. i Primenen., 37, 468–481, 1992.

Lounici, K. and Nickl, R., *Global uniform risk bounds for wavelet deconvolution estimators.* Ann. Statist., 39, 1, 201–231, 2011.

Meister, A., Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. Inverse Problems, 24 015003, 2008.

Meziani, K., *Nonparametric goodness-of fit testing in quantum homodyne tomography with noisy data.* Electron. J. Stat., 2, 1195–1223, 2008.

Meziani, K. *Nonparametric Estimation of the Purity of a Quantum State in Quantum Homodyne Tomography with Noisy Data.* Math. Meth. of Stat., 4, 16, 1–15, 2007.

Muckenhoupt, B., *Asymptotic forms for Laguerre polynomials.* Proc. Amer. Math. Soc., 24, 288–292, 1970.

Pensky, M. and Sapatinas, T., *Functional deconvolution in a periodic setting: Uniform case.* Ann. Statist., 37, 73–104, 2009.

Pensky, M. and Vidakovic, B., *Adaptive wavelet estimator for nonparametric density deconvolution.* Ann. Statist., 27, 2033–2053, 1999.

Richter, T., *Realistic pattern functions for optical homodyne tomography and determination of specific expectation values.* Phys. Rev. A, 61, 2000.

Stefanski, L. A., *Rates of convergence of some estimators in a class of deconvolution problems.* Statist.

Probab. Lett., 9, 229–235, 1990.
Stefanski, L. A. and Carroll, R. J., *Deconvoluting kernel density estimators.* Statistics, 21, 169–184, 1990.
Tsybakov, A.B., *Introduction to Nonparametric Estimation.* Springer Series in Statistics, New York, 2009.
Vogel, K. and Risken, H. , *Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase.* Phys. Rev. A, 40, 2847–2849, 1989.
Watson, G. N., *A treatise on the theory of Bessel functions.* Cambridge University Press, Cambridge, 1995.
Zorich, V.A., *Mathematical analysis. II.* Springer, Heidelberg, 2016.
Wigner, E., *On the quantum correction for thermodynamic equations.* Phys. Rev., 40, 749–759, 1932.

Appendix A: Proof of Propositions

A.1. Proof of Proposition 1

First remark that by the Fourier transform formula for $w = (q, p) \in \mathbb{R}^2$ and $x = (x_1, x_2)$

$$W_\rho(w) \;=\; \frac{1}{(2\pi)^2} \iint \widetilde{W}_\rho(x) e^{-i(qx_1+px_2)} dx. \tag{34}$$

Let $\widehat{W}_h^\gamma$ be the estimator of W_ρ defined in (21), then

$$\begin{aligned}\mathbb{E}\left[\widehat{W}_h^\gamma(w)\right] &= \frac{1}{2\pi}\mathbb{E}\left[K_h^\gamma([w,\Phi_1]-Z_1)\right] = \frac{1}{2\pi}\int_0^\pi\!\!\int K_h^\gamma([w,\phi]-z)p_\rho^\gamma(z,\phi)dzd\phi \\ &= \frac{1}{2\pi}\int_0^\pi K_h^\gamma * p_\rho^\gamma(\cdot,\phi)([w,\phi])d\phi.\end{aligned}$$

In the Fourier domain, the convolution becomes a product, combining with (18), we obtain

$$\mathbb{E}\left[\widehat{W}_h^\gamma(w)\right] \;=\; \int_0^\pi \frac{1}{(2\pi)^2}\int \widetilde{K}_h^\gamma(t)\mathcal{F}_1[p_\rho^\gamma(\cdot,\phi)](t)e^{-it[w,\phi]}dtd\phi.$$

As $\widetilde{N}^\gamma(t) = e^{-\gamma t^2}$, definition (22) of the kernel combined with (18) gives

$$\begin{aligned}\mathbb{E}\left[\widehat{W}_h^\gamma(w)\right] &= \int_0^\pi \frac{1}{(2\pi)^2}\int \widetilde{K}_h^\gamma(t)\widetilde{W}_\rho(t\cos(\phi),t\sin(\phi))\widetilde{N}^\gamma(t)e^{-it[w,\phi]}dtd\phi \\ &= \int_0^\pi \frac{1}{(2\pi)^2}\int_{|t|\le 1/h}|t|\widetilde{W}_\rho(t\cos(\phi),t\sin(\phi))e^{-it[w,\phi]}dtd\phi.\end{aligned}$$

Therefore, by the change of variable $x = (t\cos(\phi), t\sin(\phi))$, it follows

$$\mathbb{E}\left[\widehat{W}_h^\gamma(w)\right] \;=\; \frac{1}{(2\pi)^2}\int_{||x||\le 1/h}\widetilde{W}_\rho(x)e^{-i(qx_1+px_2)}dx. \tag{35}$$

From equations (34) and (35), we have

$$\begin{aligned}\left|\mathbb{E}\left[\widehat{W}_h^\gamma(w)\right] - W_\rho(w)\right| &\le \frac{1}{(2\pi)^2}\int_{||x||>1/h}\left|\widetilde{W}_\rho(x)\right|dx \\ &\le \frac{1}{(2\pi)^2}\left[\iint\left|\widetilde{W}_\rho(x)\right|^2 e^{2\beta||x||^r}dx\right]^{1/2}\left[\int_{||x||>1/h} e^{-2\beta||x||^r}dx\right]^{1/2} \\ &\le \sqrt{\frac{L}{(2\pi)^2\beta r}}h^{(r-2)/2}e^{-\beta h^{-r}}(1+o(1)), \quad h\to 0,\end{aligned}$$

by applying Lemma 7 (see Section D.5 below) and as $W_\rho \in \mathcal{A}(\beta, r, L)$ the class defined in (20).

A.2. Proof of Proposition 2

We recall first the notion of covering numbers for a functional class. For any probability distribution Q, we denote by $L^2(Q)$ the set of real-valued functions on $\mathbb{R}$ embedded with the $L^2(Q)$-norm $\|\cdot\|_{L^2(Q)} = \left(\int_{\mathbb{R}} |\cdot|^2 dQ\right)^{1/2}$. For any functional class $\mathcal{H}$ in $L^2(Q)$, the covering number $N(\epsilon, \mathcal{H}, L^2(Q))$ denotes the minimal number of $L^2(Q)$-balls of radius less than or equal to ϵ, that cover $\mathcal{H}$.
The following lemma is needed to prove the Proposition 2.

Lemma 2. *Let $\delta_h := h^{-1}e^{\frac{\gamma}{h^2}} > 0$ for any $0 < h \leq 1$, then the class*

$$\mathcal{H}_h \;=\; \{\delta_h^{-1} K_h^\gamma(\cdot - t), t \in \mathbb{R}\}, \quad h > 0 \tag{36}$$

is uniformly bounded by $U := \frac{h}{2\gamma\pi}$. Moreover, for every $0 < \epsilon < A$ and for finite positive constants A, v depending only on γ,

$$\sup_Q N(\epsilon, \mathcal{H}_h, L^2(Q)) \;\leq\; (A/\epsilon)^v, \tag{37}$$

where the supremum extends over all probability measures Q on $\mathbb{R}$.

The proof of this Lemma can be found in D.1. To prove (24), we have to bound the following quantity :

$$\begin{aligned}
\mathbb{E}[|K_h^\gamma([z,\Phi_\ell] - Z_\ell)|^2] \;&\leq\; \|K_h^\gamma\|_\infty^2 \leq \|\widetilde{K}_h^\gamma\|_1^2 = \left[\int_{|t|\leq h^{-1}} |t| e^{\gamma t^2} dt\right]^2 \\
&= \left[2\int_0^{h^{-1}} t e^{\gamma t^2} dt\right]^2 = \left(\gamma^{-1} e^{\gamma h^{-2}} - \frac{1}{\gamma}\right)^2 \leq \frac{1}{\gamma^2} e^{2\gamma h^{-2}}.
\end{aligned} \tag{38}$$

Moreover for $\delta_h = h^{-1}e^{\gamma h^{-2}}$, we have

$$\delta_h^{-2}\mathbb{E}[|K_h^\gamma([z,\Phi_\ell] - Z_\ell)|^2] \;\leq\; \frac{h^2}{\gamma^2}. \tag{39}$$

By Lemma 2, it follows that the class $\mathcal{H}_h$ is VC. Next, we note that the supremum over $\mathbb{R}$ is the same as a countable supremum since K_h^γ is continuous. Hence, we can apply (85) to get

$$\begin{aligned}
\mathbb{E}\left[\|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty\right] \;&=\; \mathbb{E}\sup_{z\in\mathbb{R}^2}\left|\frac{1}{2\pi n}\sum_{l=1}^n \left(K_h^\gamma([z,\Phi_\ell] - Z_\ell) - \mathbb{E}\left[K_h^\gamma([z,\Phi_\ell] - Z_\ell)\right]\right)\right| \\
&=\; \frac{\delta_h}{2\pi n}\mathbb{E}\sup_{z\in\mathbb{R}^2}\left|\sum_{l=1}^n \left(\delta_h^{-1}K_h^\gamma([z,\Phi_\ell] - Z_\ell) - \mathbb{E}[\delta_h^{-1}K_h^\gamma([z,\Phi_\ell] - Z_\ell)]\right)\right| \\
&\leq\; \frac{C(\gamma)\delta_h}{2\pi n}\left(\sigma\sqrt{n\log\frac{AU}{\sigma}} + U\log\frac{AU}{\sigma}\right),
\end{aligned} \tag{40}$$

where $U = \frac{h}{2\gamma\pi}$ is the envelope of the class $\mathcal{H}_h$ defined in Lemma 2. By choosing

$$\sigma^2 := \frac{h^2}{\gamma^2} \geq \sup_{z\in\mathbb{R}^2}\mathbb{E}\left[\left(\delta_h^{-1}K_h^\eta([z,\Phi_\ell] - Z_\ell)\right)^2\right]$$

in (40) we get the result in expectation (24).

Now, prove the result in probability (25).

In view of the previous display (38), we have

$$\begin{aligned}\mathrm{Var}\left(\gamma(h\delta_h)^{-1}\left|K_h^\gamma([\cdot,\Phi_1]-Z_1)-\mathbb{E}\left[K_h^\gamma([\cdot,\Phi_1]-Z_1)\right]\right|\right) &\leq \gamma^2(h\delta_h)^{-2}\mathbb{E}\left[\left|K_h^\gamma([\cdot,\Phi_1]-Z_1)\right|^2\right]\\ &\leq \gamma^2(h\delta_h)^{-2}\frac{1}{\gamma^2}e^{2\gamma h^{-2}}=1.\end{aligned}$$

As $U=\frac{1}{2\gamma\pi}$ and by (72), it follows

$$\gamma(h\delta_h)^{-1}\left\|K_h^\gamma([\cdot,\Phi_1]-Z_1)-\mathbb{E}[K_h^\gamma([\cdot,\Phi_1]-Z_1)]\right\|_\infty \leq \gamma(h\delta_h)^{-1}\|K_h^\gamma\|_\infty \leq \gamma h^{-1}U \leq 1.$$

We use Talagrand's inequality as in Theorem 2.3 of Bousquet (2002). Let us define

$$\mathcal{Z} := \frac{n\gamma}{h\delta_h}\|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty.$$

Then, for any $x>0$ and with probability at least $1-e^{-x}$, we obtain

$$\begin{aligned}\mathcal{Z} &\leq \mathbb{E}\left[\mathcal{Z}\right]+\sqrt{2xn+4x\mathbb{E}[\mathcal{Z}]}+\frac{x}{3} \leq \mathbb{E}\left[\mathcal{Z}\right]+\sqrt{2xn}+2\sqrt{x\mathbb{E}[\mathcal{Z}]}+\frac{x}{3}\\ &\leq 2\mathbb{E}\left[\mathcal{Z}\right]+\sqrt{2xn}+\frac{4x}{3},\end{aligned}$$

where we have used the decoupling inequality $2ab\leq a^2+b^2$ with $a=\sqrt{x}$ and $b=\sqrt{\mathbb{E}[\mathcal{Z}]}$. Thus, with probability at least $1-e^{-x}$, we get

$$\|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty = \frac{h\delta_h}{n\gamma}Z \quad \leq 2\mathbb{E}\left[\|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty\right]+\tfrac{e^{\gamma h^{-2}}}{\gamma}\left(\sqrt{2\tfrac{x}{n}}+\tfrac{4x}{3n}\right).$$

Plugging our control (24) on $\mathbb{E}[\|\widehat{W}_h^\gamma - \mathbb{E}[\widehat{W}_h^\gamma]\|_\infty]$, the result in probability follows.

Appendix B: Proof of Theorem 2 - Lower bounds

B.1. Proof of Theorem 2 - Lower bounds for the $\mathbb{L}_2$-norm

The proof for the minimax lower bounds follows a standard scheme for deconvolution problem as in the paper of Butucea, Guţă and Artiles (2007); Lounici, K. and Nickl (2011). However, additional technicalities arise to build a proper set of Wigner functions and then to derive a lower bound. From now on, for the sake of brevity, we will denote $\mathcal{A}(\beta,2,L)$ by $\mathcal{A}(\beta,L)$ as we consider only the practical case $r=2$. Let $W_0\in\mathcal{A}(\beta,L)$ be a Wigner function. Its associated density function will be denoted by $p_0(x,\phi)=\frac{1}{\pi}\mathcal{R}[W_0](x,\phi)\mathbb{1}_{[0,\pi]}(\phi)$.

We suggest the construction of a family of two Wigner functions W_0 and W_1 such that for all $w\in\mathbb{R}^2$:

$$W_1(w)=W_0(w)+V_h(w),$$

where the construction of W_0 and V_h are given in Appendices B.1.1 and B.1.2 and the parameter $h=h(n)\to 0$ as $n\to\infty$. We denote by

$$p_m(x,\phi)=\frac{1}{\pi}\mathcal{R}[W_m](x,\phi)\mathbb{1}_{[0,\pi]}(\phi),\quad m=0,1$$

the density function associated to the Wigner functions W_0 and W_1. As we consider the noisy framework (16) and in view of (17), we set for $m=0,1$

$$p_m^\gamma(z,\phi)=\left[p_m(\cdot,\phi)*N^\gamma\right](z)$$

If the following conditions (**C1**) to (**C3**) are satisfied, then Theorem 2.6 in the book of Tsybakov (2009) gives the lower bound.

(**C1**) $W_0, W_1 \in \mathcal{A}(\beta, L)$.
(**C2**) We have $||W_1 - W_0||_2^2 \geq 4\varphi_n^2$, with $\varphi_n^2 = \mathcal{O}\left(n^{-\frac{\beta}{\beta+\gamma}}\right)$.
(**C3**) We have

$$n\mathcal{X}^2(p_1^\gamma, p_0^\gamma) := n\int_0^\pi\int \frac{(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi))^2}{p_0^\gamma(z,\phi)} dz d\phi \leq \frac{1}{4}.$$

Proofs of these three conditions are provided in Appendices B.1.3 to B.1.5.

B.1.1. ***Construction of*** $\mathbf{W_0}$

The Wigner function W_0 is the same as in the paper of Butucea, Guţă and Artiles (2007). For the sake of completeness, we recall its construction here. The probability density function associated to any density matrix ρ in the ideal noiseless setting is given by equation (9). In particular, for diagonal density matrix ρ, the associated probability density function is

$$p_\rho(x,\phi) = \sum_{k=0}^{\infty} \rho_{kk}\psi_k^2(x).$$

For all $0 < \alpha, \lambda < 1$, we introduce a family of diagonal density matrices $\rho^{\alpha,\lambda}$ such that for all $k \in \mathbb{N}$

$$\rho_{kk}^{\alpha,\lambda} = \int_0^1 z^k \alpha \frac{(1-z)^\alpha}{(1-\lambda)^\alpha} \mathbb{1}_{\lambda \leq z \leq 1} dz. \tag{41}$$

Therefore the probability density associated to this diagonal density matrix $\rho^{\alpha,\lambda}$ can be written as follows

$$p_{\alpha,\lambda}(x,\phi) = \sum_{k=0}^{\infty} \rho_{kk}^{\alpha,\lambda}\psi_k^2(x) = \sum_{k=0}^{\infty} \psi_k^2(x) \int_0^1 z^k \alpha \frac{(1-z)^\alpha}{(1-\lambda)^\alpha} \mathbb{1}_{\lambda \leq z \leq 1} dz. \tag{42}$$

Moreover by the well known Mehler formula (see Erdélyi , Magnus, Oberhettinger and Tricomi (1953)), we have

$$\sum_{k=0}^{\infty} z^k \psi_k^2(x) = \frac{1}{\sqrt{\pi(1-z^2)}} \exp\left(-x^2 \frac{1-z}{1+z}\right).$$

Then, it follows

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

The following Lemma, proved in the paper of Butucea, Guţă and Artiles (2007), gives a control on the tails of the associated density $p_{\alpha,\lambda}(x,\phi) = p_{\alpha,\lambda}(x)$ as it does not depend on ϕ.

Lemma 3 (Butucea, Guta and Artiles (2007))**.** *For all $\phi \in [0,1]$ and all $0 < \alpha, \lambda < 1$ and $|x| > 1$ there exist constants c, C depending on α and λ such that*

$$c|x|^{-(1+2\alpha)} \leq p_{\alpha,\lambda}(x) \leq C|x|^{-(1+2\alpha)}.$$

In view of Lemma 3 of Butucea, Guţă and Artiles (2007), the Wigner function W_0 will be chosen in the set

$$\mathcal{W}^{\alpha,\lambda} = \left\{W^{\alpha,\lambda} = W_{\rho_{\alpha,\lambda}} : \text{Wigner function associated to } \rho_{\alpha,\lambda} : 0 < \alpha, \lambda < 1\right\},$$

where λ is such that W_0 is a Wigner function belonging to $\mathcal{A}(\beta, L)$ (see Section 6.1 in Butucea, Guţă and Artiles (2007) or the proof of Theorem 2 in Guţă and Artiles (2007)).

B.1.2. ***Construction of*** V_h ***for the*** $\mathbb{L}_2$***-norm***

Let

$$\delta := \log^{-1}(n). \tag{43}$$

We define two infinitely differentiable function g and g_1 such that:

- $g_1 : \mathbb{R} \to [0,1]$.
- The support of g_1 is $\mathrm{Supp}(g_1) = (\delta, 2\delta)$.
- And $\forall t \in \left[\frac{\delta}{3}, \frac{2\delta}{3}\right]$, $g_1(t) = 1$.
- $g : \mathbb{R} \to [-1,1]$ is on odd function, such that for some fixed $\epsilon > 0$, $g(x) = 1$ for any $x \geq \epsilon$.

Define also the following parameters:

$$a_1 := (h^{-2} + \delta)^{1/2}, \quad b_1 := (h^{-2} + 2\delta)^{1/2}. \tag{44}$$

$$\widetilde{a}_1 := (h^{-2} + (4/3)\delta)^{1/2}, \quad \widetilde{b}_1 := (h^{-2} + (5/3)\delta)^{1/2}. \tag{45}$$

$$C_0 := \sqrt{\pi L(\beta + \gamma)}. \tag{46}$$

We also introduce an infinitely differentiable function V_h such that:

- $V_h : \mathbb{R}^2 \to \mathbb{R}$ is an odd real-valued function.
- Set $t = \sqrt{w_1^2 + w_2^2}$, then the function V_h admits the following Fourier transform with respect to both variables

$$\widetilde{V}_h(w) := \mathcal{F}_2[V_h](w) := iaC_0 h^{-1} e^{\beta h^{-2}} e^{-2\beta|t|^2} g_1(|t|^2 - h^{-2}) g(w_2), \tag{47}$$

  where $a > 0$ is a numerical constant chosen sufficiently small. The bandwidth is such that

$$h \;=\; \left(\frac{\log n}{2(\beta+\gamma)}\right)^{-1/2}. \tag{48}$$

Note that $\widetilde{V}_h(w)$ is infinitely differentiable and compactly supported, thus it belongs to the Schwartz class $\mathcal{S}(\mathbb{R}^2)$ of fast decreasing functions on $\mathbb{R}^2$. The Fourier transform being a continuous mapping of the Schwartz class onto itself, this implies that V_h is also in the Schwartz class $\mathcal{S}(\mathbb{R}^2)$. Moreover, $\widetilde{V}_h(w)$ is an odd function with purely imaginary values. Consequently, V_h is an odd real-valued function. Thus, we get

$$\iint V_h(p,q)dpdq = \int \mathcal{R}[V_h](x,\phi)dx = 0, \tag{49}$$

for all $\phi \in [0,\pi]$ and $\mathcal{R}[V_h]$ the Radon transform of V_h.

Now, we can define the function W_1 as follows:

$$W_1(z) = W_0(z) + V_h(z), \tag{50}$$

where W_0 is the Wigner function associated to the density p_0 defined in (41).

As in (8), we also define

$$p_1(x,\phi) = \frac{1}{\pi}\mathcal{R}[W_1](x,\phi)\mathbb{1}_{(0,\pi}(\phi),$$

$$\text{and} \quad \rho^{(1)}_{j,k} = \int_0^\pi\!\!\int p_m(x,\phi) f_{j,k}(x) e^{(j-k)\phi} dx d\phi. \tag{51}$$

By Lemma 6 in Appendix D.4, the matrix $\rho^{(1)}$ is proved to be a density matrix. Therefore, in view of (9) and (49), the function W_1 is indeed a Wigner function.

B.1.3. ***Condition*** **(C1)**

By the triangle inequality, we have

$$\|\widetilde{W}_1 e^{\beta\|\cdot\|^2}\|_2 \leq \|\widetilde{W}_0 e^{\beta\|\cdot\|^2}\|_2 + \|\widetilde{V}_h e^{\beta\|\cdot\|^2}\|_2.$$

The first term in the above sum has be bounded in Lemma 3 of Butucea, Guţă and Artiles (2007) as follows

$$\|\widetilde{W}_0 e^{\beta\|\cdot\|^2}\|_2^2 \leq \pi^2 L. \tag{52}$$

To study the second term in the sum above, we consider the change of variables $w = (t\cos\phi, t\sin\phi)$ and as g is bounded by 1, we get, using (43), (44) and (46) that

$$\begin{aligned}
\|\widetilde{V}_h e^{\beta\|\cdot\|^2}\|_2^2 &\leq \iint \left[aC_0 h^{-1} e^{\beta h^{-2}}\right]^2 e^{-2\beta\|w\|^2} g_1^2(\|w\|^2 - h^{-2})dw \\
&\leq a^2 C_0^2 h^{-2} e^{2\beta h^{-2}} \int_0^{\pi}\int_{a_1}^{b_1} |t| e^{-2\beta|t|^2} dt \\
&\leq \pi a^2 C_0^2 h^{-2} e^{2\beta h^{-2}} e^{-2\beta a_1^2} \int_{a_1}^{b_1} t dt \leq \frac{\pi}{2} a^2 C_0^2 h^{-2} e^{-2\beta\delta}\left[b_1^2 - a_1^2\right] \\
&\leq \frac{\pi}{3} a^2 C_0^2 h^{-2} \delta e^{-2\beta\delta} \leq \pi^2 L,
\end{aligned} \tag{53}$$

for a small enough. It follows from (52) and (53) that $W_1 \in \mathcal{A}(\beta, L)$.

B.1.4. ***Condition*** **(C2)**

By applying Plancherel Theorem and the change of variables $w = (t\cos\phi, t\sin\phi)$, we have that

$$\begin{aligned}
\|W_1 - W_0\|_2^2 = \|V_h\|_2^2 &= \|V_h\|_2^2 = \frac{1}{4\pi^2}\int_0^{\pi}\int |t| \left|\widetilde{V}_h(t,\phi)\right|^2 dt d\phi \\
&= \frac{a^2 C_0^2}{4\pi^2} h^{-2} e^{2\beta h^{-2}} \int_0^{\pi}\int |t| e^{-4\beta t^2} g^2(t\sin\phi) g_1^2(t^2 - h^{-2}) dt d\phi.
\end{aligned} \tag{54}$$

Note that for a fixed $\mu \in (0, \pi/4)$, there exists a numerical constant $c > 0$ such that $\sin(\phi) > c$ on $(\mu, \pi - \mu)$. From now on, we denote by $\widetilde{A}_1$ the set

$$\widetilde{A}_1 := \left\{w \in \mathbb{R}^2 : \widetilde{a}_1 \leq \|w\|^2 \leq \widetilde{b}_1\right\}, \tag{55}$$

where $\widetilde{a}_1$ and $\widetilde{b}_1$ are defined in (45). By definition of g and for a large enough n, we have for any $(t, \phi) \in \widetilde{A}_1 \times (\mu, \pi - \mu)$ that $g^2(t\sin(\phi)) = 1$ with $t^2 = \|w\|^2$. Therefore, (54) can be lower bounded as follows

$$\begin{aligned}
\|W_1 - W_0\|_2^2 &\geq \frac{a^2 C_0^2}{4\pi^2} h^{-2} e^{2\beta h^{-2}} \int_{\mu}^{\pi-\mu}\int_{\widetilde{A}_1} |t| e^{-4\beta t^2} g_1^2(t^2 - h^{-2}) dt d\phi \\
&= \frac{\pi - 2\mu}{4\pi^2} a^2 C_0^2 h^{-2} e^{2\beta h^{-2}} \int_{\widetilde{A}_1} |t| e^{-4\beta t^2} g_1^2(t^2 - h^{-2}) dt.
\end{aligned} \tag{56}$$

On $\widetilde{A}_1$ and by construction of the function g_1, we have $g_1^2(t^2 - h^{-2}) = 1$. Hence, it results

$$\begin{aligned}
I &:= \int_{\widetilde{A}_1} |t| e^{-4\beta t^2} g_1^2(t^2 - h^{-2}) dt \\
&\geq e^{-4\beta\widetilde{b}_1^2} \int_{\widetilde{A}_1} |t| g_1^2(t^2 - h^{-2}) dt \\
&\geq e^{-4\beta\widetilde{b}_1^2} \int_{\widetilde{a}_1}^{\widetilde{b}_1} t dt \geq \frac{1}{2} e^{-4\beta\widetilde{b}_1^2}(\widetilde{b}_1^2 - \widetilde{a}_1^2) \geq \frac{1}{6}\delta e^{-4\beta\widetilde{b}_1^2}.
\end{aligned} \tag{57}$$

Combining (56) and (57), we get, since $C_0^2 h^{-2}\delta = \pi L/2$ that

$$\begin{aligned}\|W_1 - W_0\|_2^2 &\geq \frac{\pi - 2\mu}{24\pi^2} a^2 C_0^2 h^{-2} e^{2\beta h^{-2}} \delta e^{-4\beta \widetilde{b}_1^2} = \frac{\pi - 2\mu}{48\pi} a^2 L e^{2\beta h^{-2}} e^{-4\beta \widetilde{b}_1^2} \\ &= \frac{\pi - 2\mu}{48\pi} a^2 L e^{-2\beta h^{-2}} e^{-\frac{20}{3}\beta\delta}.\end{aligned}$$

It follows from (48) that

$$\|W_1 - W_0\|_2^2 \geq \frac{\pi - 2\mu}{48\pi} a^2 L n^{-\frac{\beta}{\beta+\gamma}} e^{-\frac{40}{3}\beta} \geq 4cn^{-\frac{\beta}{\beta+\gamma}} =: 4\varphi_n^2,$$

where $c > 0$ is a numerical constant possibly depending only on β.

B.1.5. ***Condition*** **(C3)**

Denote by $\widetilde{C} > 0$ a constant whose value may change from line to line and recall that N^γ is the density of the Gaussian distribution with zero mean and variance 2γ. Note that p_0 and N^γ do not depend on ϕ. Consequently, in the framework of noisy data defined in (16), $p_0^\gamma(z,\phi) = p_0^\gamma(z)\frac{1}{\pi}\mathbb{1}_{(0,\pi)}(\phi)$.

Lemma 4. *There exists numerical constants $c' > 0$ and $c'' > 0$ such that*

$$p_0^\gamma(z) \geq c' z^{-2}, \quad \forall |z| \geq 1 + \sqrt{2\gamma}, \tag{58}$$

and

$$p_0^\gamma(z) \geq c'', \quad \forall |z| \leq 1 + \sqrt{2\gamma}. \tag{59}$$

The proof of this Lemma is done in Appendix D.3. Using Lemma 4, the χ^2-divergence can be upper bounded as follows

$$\begin{aligned}n\mathcal{X}^2(p_1^\gamma, p_0^\gamma) &= n\int_0^\pi\int \frac{\left(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi)\right)^2}{p_0^\gamma(z,\phi)} dz d\phi \\ &\leq \frac{n}{c''}\int_0^\pi\int_{-(1+\sqrt{2\gamma})}^{1+\sqrt{2\gamma}} \left(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi)\right)^2 dz d\phi \\ &\quad + \frac{n}{c'}\int_0^\pi\int_{\mathbb{R}\setminus(1+\sqrt{2\gamma}, 1+\sqrt{2\gamma})} z^2 \left(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi)\right)^2 dz d\phi \\ &=: \frac{n}{c''} I_1 + \frac{n}{c'} I_2. \end{aligned} \tag{60}$$

Note that, as in (18) the Fourier transforms of p_1^γ and p_0^γ with respect to the first variable are respectively equal to

$$\begin{aligned}\mathcal{F}_1[p_1^\gamma(\cdot,\phi)](t) &= \widetilde{W}_1(t\cos\phi, t\sin\phi)\widetilde{N}^\gamma(t) \\ &= \left(\widetilde{V}_h(t\cos\phi, t\sin\phi) + \widetilde{W}_0(t\cos\phi, t\sin\phi)\right) e^{-\gamma t^2}, \end{aligned} \tag{61}$$

$$\mathcal{F}_1[p_0^\gamma(\cdot,\phi)](t) = \widetilde{W}_0(t\cos\phi, t\sin\phi) e^{-\gamma t^2}, \tag{62}$$

since $\widetilde{N}^\gamma(t) = e^{-\gamma t^2}$. Using Plancherel Theorem, equations (47), (61) and (62), the first integral I_1 in the sum (60) is bounded by

$$\begin{aligned}I_1 &\leq \int_0^\pi\int \left(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi)\right)^2 dz d\phi = \frac{1}{4\pi^2}\int_0^\pi\int \left|\mathcal{F}_1[p_1^\gamma(\cdot,\phi)](t) - \mathcal{F}_1[p_0^\gamma(\cdot,\phi)](t)\right|^2 dt d\phi \\ &= \frac{1}{4\pi^2}\int_0^\pi\int \left|\widetilde{V}_h(t\cos\phi, t\sin\phi)\right|^2 e^{-2\gamma t^2} dt d\phi \\ &= \frac{a^2 C_0^2}{4\pi^2} h^{-2} e^{2\beta h^2} \int_0^\pi\int e^{-4\beta t^2 - 2\gamma t^2} g_1^2\left(t^2 - h^{-2}\right) g^2(t\sin\phi) dt d\phi. \end{aligned}$$

By construction, the function g is bounded by 1 and the function g_1 admits as support $\mathrm{Supp}(g_1) = (\delta, 2\delta)$. Thus,

$$\begin{aligned} I_1 &\leq \frac{a^2C_0^2}{4\pi} e^{2\beta h^2} \int e^{-4\beta t^2 - 2\gamma t^2} g_1^2\left(t^2 - h^{-2}\right) dt \leq \frac{a^2C_0^2}{4\pi} h^{-2} e^{2\beta h^{-2}} \int_{a_1}^{b_1} e^{-4\beta t^2 - 2\gamma t^2} dt \\ &\leq \frac{a^2C_0^2}{4\pi}(b_1 - a_1) h^{-2} e^{2\beta h^{-2}} e^{-4\beta a_1^2 - 2\gamma a_1^2} \leq \frac{a^2C_0^2}{4\pi} \frac{b_1^2 - a_1^2}{2a_1} h^{-2} e^{2\beta h^{-2} - 4\beta a_1^2 - 2\gamma a_1^2}. \end{aligned}$$

Some basic algebra, (43), (44), (46) and (48) yield

$$\frac{n}{c''} I_1 \leq \frac{a^2 \widetilde{C}}{\sqrt{\log n}}, \tag{63}$$

for some constant $\widetilde{C} > 0$ which may depend on β, γ, L and c''. For the second term I_2 in the sum (60), with the same tools we obtain using in addition the spectral representation of the differential operator, that

$$\begin{aligned} I_2 &\leq \int_0^\pi\!\!\int z^2 \left(p_1^\gamma(z,\phi) - p_0^\gamma(z,\phi)\right)^2 dz d\phi \\ &= \int_0^\pi\!\!\int \left| \frac{\partial}{\partial t} \left(\mathcal{F}_1[p_1^\gamma(\cdot,\phi)] - \mathcal{F}_1[p_0^\gamma(\cdot,\phi)]\right)(t) \right|^2 dt d\phi \\ &= \int_0^\pi\!\!\int \left| \frac{\partial}{\partial t} \left(\widetilde{V}_h(t\cos\phi, t\sin\phi) e^{-\gamma t^2} \right) \right|^2 dt d\phi \\ &= \int_0^\pi\!\!\int \left| e^{-\gamma t^2} \frac{\partial}{\partial t}(\widetilde{V}_h)(t\cos\phi, t\sin\phi) - 2\gamma t e^{-\gamma t^2} \widetilde{V}_h(t\cos\phi, t\sin\phi) \right|^2 dt d\phi \\ &\leq 2\int_0^\pi\!\!\int e^{-2\gamma t^2} \left| I_{2,1} \right|^2 dt d\phi + 16\gamma^2 \int_0^\pi\!\!\int t^2 e^{-2\gamma t^2} \left| I_{2,2} \right|^2 dt d\phi, \end{aligned} \tag{64}$$

where $I_{2,2} = \widetilde{V}_h(t\cos\phi, t\sin\phi)$ and $I_{2,1}$, the partial derivative $\frac{\partial}{\partial t}(\widetilde{V}_h)(t\cos\phi, t\sin\phi)$, is equal to

$$iaC_0 h^{-1} e^{\beta h^{-2} - 2\beta t^2} \left[g_1(t^2 - h^{-2}) \left(-4\beta t g(t\sin\phi) + g'(t\sin\phi)\sin\phi \right) + 2t g_1'(t^2 - h^{-2}) g(t\sin\phi) \right].$$

Since g_1 and g belong to the Schwartz class, there exists a numerical constant $c_S > 0$ such that $\max\{\|g_1\|_\infty, \|g_1'\|_\infty, \|g\|_\infty, \|g'\|_\infty\} \leq c_S$. Furthermore, the support of the function g_1 is $\mathrm{Supp}(g_1) = (\delta, 2\delta)$, then

$$|I_{2,1}|^2 \leq a^2 c_S^4 C_0^2 h^{-2} e^{2\beta h^{-2} - 4\beta t^2} \left((4\beta + 2)|t| + 1\right)^2 \mathbb{1}_{(a_1, b_1)}(t), \tag{65}$$

with a_1 and b_1 defined in (44). Proceeding similarly, we have

$$\begin{aligned} |I_{2,2}|^2 &= \left| aC_0 h^{-1} e^{\beta h^{-2}} e^{-2\beta t^2} g_1(t^2 - h^{-2}) g(t\sin\phi) \right|^2 \\ &\leq a^2 c_S^4 C_0^2 h^{-2} e^{2\beta h^{-2} - 4\beta t^2} \mathbb{1}_{(a_1, b_1)}(t). \end{aligned} \tag{66}$$

Combining (65) and (66) with (64), as $0 \leq \delta \leq 1$

$$\begin{aligned} I_2 &\leq 2a^2 c_S^4 C_0^2 h^{-2} e^{2\beta h^{-2}} \int_0^\pi\!\!\int_{a_1}^{b_1} e^{-2\gamma t^2} e^{-4\beta t^2} \left[\left((4\beta + 2)|t| + 1\right)^2 + 8\gamma^2 t^2 \right] dt d\phi \\ &\leq 2\pi a^2 c_S^4 C_0^2 h^{-2} e^{2\beta h^{-2}} \int_{a_1}^{b_1} e^{-2(\gamma + 2\beta)t^2} \left[(1 + 4(2\beta + 1)t + (2 + 4\beta + 8\gamma^2)t^2 \right] dt. \end{aligned}$$

An integration by part gives

$$\begin{aligned} I_2 &\leq 2\pi a^2 c_S^4 C_0^2 h^{-2} e^{2\beta h^{-2}} e^{-(4\beta+2\gamma)a_1^2} \left[((4\beta+2)b_1+1)^2 + 8\gamma^2 b_1^2 \right] \int_{a_1}^{b_1} dt \\ &\leq 2\pi a^2 c_S^4 C_0^2 \tilde{c} h^{-2} \delta e^{-2(\beta+\gamma)h^{-2}} e^{-2(2\beta+\gamma)\delta}, \end{aligned}$$

where $\tilde{c} > 0$ depends only on γ, β.
Some basic algebra, (43), (44), (46) and (48) yield

$$\frac{n}{c'} I_2 \leq a^2 \widetilde{C}, \tag{67}$$

for some constant $\widetilde{C} > 0$ possibly depending on β, γ, L, c_S and c'. Combining (67) and (63) with (60), we get for n large enough

$$n\mathcal{X}^2(p_{k,h}^\gamma, p_0^\gamma) := n \int_0^\pi \int_{\mathbb{R}} \frac{(p_{k,h}^\gamma(z,\phi) - p_0^\gamma(z,\phi))^2}{p_0^\gamma(z,\phi)} dz d\phi \leq a^2 \widetilde{C},$$

where $\widetilde{C} > 0$ is a constant which may depend on β, γ, L c_S, c' and c''. Taking the numerical constant $a > 0$ small enough, we deduce from the previous display that

$$n\mathcal{X}^2(p_{k,h}^\gamma, p_0^\gamma) \leq \frac{1}{4}.$$

B.2. Proof of Theorem 2 - Lower bounds for the sup-norm

To prove the lower bound for the sup-norm, we need to slightly modify the construction of the Wigner function W_1 defined in (50). In our new construction, the Wigner function W_0, associated to the density p_0 defined in (41), stays unchanged as compared to the $\mathbb{L}_2$ case. However, the function V_h given in (47) is modified as follows. We replaced the functions g_1 and g respectively into $g_{1,\epsilon}$ and g_ϵ for some $0 < \epsilon < 1$.

We introduce an infinitely differentiable function $g_{1,\epsilon}$ such that:

- $g_{1,\epsilon} : \mathbb{R} \to [0,1]$.
- The support of $g_{1,\epsilon}$ is $\mathrm{Supp}(g_{1,\epsilon}) = (\delta, 2\delta)$.
- Using a similar construction as for function g_1, we can also assume that

$$g_{1,\epsilon}(t) = 1, \quad \forall t \in A_{1,\epsilon} := [(1+\epsilon)\delta, (2-\epsilon)\delta], \tag{68}$$

  and

$$\|g'_{1,\epsilon}\|_\infty \leq \frac{c}{\epsilon\delta}, \tag{69}$$

  for some numerical constant $c > 0$.
- An odd function $g_\epsilon : \mathbb{R} \to [-1,1]$ satisfies the same conditions as g above but we assume in addition that

$$\|g'_\epsilon\|_\infty \leq \frac{c}{\epsilon}, \tag{70}$$

  for some numerical constant $c > 0$.

The condition (70) will be needed to check Condition **(C3)**. Such a function can be easily constructed. Consider for instance a function g_ϵ such that its derivative satisfies

$$g'_\epsilon(t) = \left[\psi * \frac{1}{\epsilon} \mathbb{1}_{(0,\epsilon)} \right](t),$$

for any $t \in (0, \epsilon)$ where ψ is a mollifier. Integrate this function and renormalize it properly so that $g_\epsilon(t) = 1$ for any $t \geq \epsilon$. Complete the function by symmetry to obtain an odd function defined on the whole real line. Such a construction satisfies condition (70).

It is easy to see that Condition **(C1)** is always satisfied by the new test functions $W_{0,\epsilon}$ and $W_{1,\epsilon}$. We now check Condition **(C2)**. Set $C_h = iaC_0h^{-1}e^{\beta h^{-2}}$. Then, we have

$$\begin{aligned} W_{1,\epsilon}(z) - W_{0,\epsilon}(z) &= \frac{1}{4\pi^2}\iint e^{-i\langle z,w\rangle}\left(\widetilde{W}_{1,\epsilon}(w) - \widetilde{W}_{0,\epsilon}(w)\right)dw \\ &= \frac{1}{4\pi^2}\int_0^\pi\!\!\int e^{-it[z,\phi]}|t|C_he^{-2\beta t^2}g_{1,\epsilon}\left(t^2-h^{-2}\right)g_\epsilon(t)dtd\phi. \end{aligned}$$

Note that $A_1 = \lim_{\epsilon\to 0} A_{1,\epsilon}$ where A_1 is defined in (68). For all $z \in \mathbb{R}^2$, we define the following quantity

$$I(z) := \int_0^\pi\!\!\int e^{-it[z,\phi]}|t|C_he^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)\left[\mathbb{1}_{(0,\infty)}(t) - \mathbb{1}_{(-\infty,0)}(t)\right]dtd\phi.$$

Lebesgue dominated convergence Theorem guarantees that

$$\lim_{\epsilon\to 0}\left(\int_0^\pi\!\!\int e^{-it[z,\phi]}|t|C_he^{-2\beta t^2}g_{1,\epsilon}\left(t^2-h^{-2}\right)g_\epsilon(t)dtd\phi\right) = I(z).$$

Therefore, there exists an $\epsilon > 0$ (possibly depending on n, z) such that

$$|W_{1,\epsilon}(z) - W_{0,\epsilon}(z)| \geq \frac{1}{8\pi^2}\left|I(z)\right|.$$

Taking $z = (0, 2h)$, Fubini's Theorem gives

$$\begin{aligned} I(z) &= \frac{1}{4\pi^2}\int_0^\pi\!\!\int e^{-it2h\sin\phi}|t|C_he^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)\left[\mathbb{1}_{(0,\infty)}(t) - \mathbb{1}_{(-\infty,0)}(t)\right]dtd\phi \\ &= \frac{1}{4\pi^2}\int\left(\int_0^\pi e^{-it2h\sin\phi}d\phi\right)|t|C_he^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)\left[\mathbb{1}_{(0,\infty)}(t) - \mathbb{1}_{(-\infty,0)}(t)\right]dt. \end{aligned}$$

Note that

$$\int_0^\pi e^{-it2h\sin\phi}d\phi = \pi(iH_0(2ht) + J_0(2ht)),$$

where H_0 and J_0 denote respectively the Struve and Bessel functions of order 0. By definition, H_0 is an odd function while J_0 and $t \to |t|C_he^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)$ are even functions. Consequently, we get

$$\begin{aligned} I(z) &= \frac{1}{4\pi}iC_h\int|t|H_0(2ht)e^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)\left[\mathbb{1}_{(0,\infty)}(2ht) - \mathbb{1}_{(-\infty,0)}(t)\right]dt \\ &= \frac{1}{2\pi}iC_h\int_0^\infty tH_0(2ht)e^{-2\beta t^2}\mathbb{1}_{A_1}\left(t^2-h^{-2}\right)dt \\ &= \frac{iC_h}{2\pi}\int_{a_1}^{b_1} tH_0(2ht)e^{-2\beta t^2}dt, \end{aligned}$$

with a_1 and b_1 defined in (44). Note that $\forall t \in [a_1, b_1]$ and for a large enough n, it follows that $2ht \in [2, 3]$. Therefore, on $[a_1, b_1]$, the function $t \to H_0(2ht)$ is decreasing. In addition, (see Erdélyi , Magnus, Oberhettinger and Tricomi (1953)), we have

$$\min_{t\in[a_1,b_1]}\{H_0(2ht)\} > 1/2.$$

We easily deduce from the previous observations that

$$|I(z)| \;\geq\; \frac{|C_h|}{4\pi}\int_{a_1}^{b_1} te^{-2\beta t^2}dt \geq \frac{|C_h|}{16\pi\beta}\left(e^{-2\beta a_1^2}-e^{-2\beta b_1^2}\right).$$

Therefore, some simple algebra gives, for n large enough. that

$$|I(z)| \geq c2\beta a_1\delta(1-\beta a_1\delta)\delta|C_h|n^{-\frac{\beta}{\beta+\gamma}} \geq ac'n^{-\frac{\beta}{2(\beta+\gamma)}},$$

for some numerical constants $c, c' > 0$ depending only β. Taking the numerical constant $a > 0$ small enough independently of n, β, γ, we get that Condition **(C2)** is satisfied with $\varphi_n = cn^{-\frac{\beta}{2(\beta+\gamma)}}$. Concerning Condition **(C3)**, we proceed similarly as above for the $\mathbb{L}_2$-norm risk. The only modification appears in (65)-(66) where we now use (68)-(69) combined with the fact that

$$|\mathrm{Supp}(g'_\epsilon)| \leq 2\epsilon \text{ and } |\mathrm{Supp}(g'_{1,\epsilon})| \leq 2\delta\epsilon,$$

by construction of these functions. Therefore, the details will be omitted here.

Appendix C: Proof of Theorem 3 - Adaptation

The following Lemma is needed to prove Theorem 3.

Lemma 5. *For $\kappa > 0$, a constant, let $\mathcal{E}_\kappa$ be the event defined such that*

$$\mathcal{E}_\kappa = \bigcap_{m=1}^{M}\left\{\|\widehat{W}^\gamma_{h_m} - \mathbb{E}[\widehat{W}^\gamma_{h_m}]\|_\infty \leq \kappa e^{\gamma h_m^{-2}} r_n(x+\log M)\right\}. \tag{71}$$

Therefore, on the event $\mathcal{E}_\kappa$

$$\|\widehat{W}^\gamma_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \leq C \min_{1\leq m\leq M}\left\{h_m^{r/2-1}e^{-\beta h_m^{-r}} + e^{\gamma h_m^{-2}} r_n(x+\log M)\right\},$$

where $C > 0$ is a constant depending only on $\gamma, \beta, L, r, \kappa$ and $\widehat{W}^\gamma_{h_{\widehat{\mathbf{m}}}}$ is the adaptive estimator with the bandwidth $h_{\widehat{\mathbf{m}}}$ defined in (30).

The proof of the previous Lemma is done in D.2. For any fixed $m \in \{1, \cdots, M\}$, we have in view of Proposition 2 that

$$\mathbb{P}\left(\|\widehat{W}^\gamma_{h_m} - \mathbb{E}[\widehat{W}^\gamma_{h_m}]\|_\infty \leq Ce^{\gamma h_m^{-2}} r_n(x)\right) \geq 1 - e^{-x},$$

where $r_n(x) = \max\left(\sqrt{\frac{1+x}{n}}, \frac{1+x}{n}\right)$. By a simple union bound, we get

$$\mathbb{P}\left(\bigcap_{1\leq m\leq M}\left\{\|\widehat{W}^\gamma_{h_m} - \mathbb{E}[\widehat{W}^\gamma_{h_m}]\|_\infty \leq C_2 e^{\gamma {h_m}^{-2}} r_n(x)\right\}\right) \geq 1 - Me^{-x}.$$

Replacing x by $(x + \log M)$, implies

$$\mathbb{P}\left(\bigcap_{1\leq m\leq M}\left\{\|\widehat{W}^\gamma_{h_m} - \mathbb{E}[\widehat{W}^\gamma_{h_m}]\|_\infty \leq C_2 e^{\gamma {h_m}^{-2}} r_n(x+\log M)\right\}\right) \geq 1 - e^{-x}.$$

For $\kappa > C_2$, we immediately get that $\mathbb{P}(\mathcal{E}_\kappa) \geq 1 - e^{-x}$ and the result in probability (31) follows by Lemma 5. To prove the result in expectation (32), we use the property $\mathbb{E}[Z] = \int_0^\infty \mathbb{P}(Z \geq t)dt$, where Z is any positive random variable. We have indeed for any $1 \leq m \leq M$ that

$$\mathbb{P}\left(\|\widehat{W}^\gamma_{h_{\widehat{\imath}}} - W_\rho\|_\infty \geq C\left(h_m^{r/2-1}e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}} r_n(x+\log M)\right)\right) \leq e^{-x}, \quad \forall x > 0.$$

Note that

$$\begin{aligned} r_n(x+\log M) &= \max\left\{\sqrt{\frac{x+\log(eM)}{n}}, \frac{x+\log(eM)}{n}\right\} \\ &\leq \max\left\{\sqrt{\frac{\log eM}{n}}, \frac{\log eM}{n}\right\} + \max\left\{\sqrt{\frac{x}{n}} \vee \frac{x}{n}\right\} \\ &\leq r_n(\log M) + r_n(x-1). \end{aligned}$$

Combining the two previous displays, we get $\forall x > 0$

$$\mathbb{P}\left(\|\widehat{W}^{\gamma}_{h_{\hat{\imath}}} - W_\rho\|_\infty \geq C\left(h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}}\left[r_n(\log M) + r_n(x-1)\right]\right)\right) \leq e^{-x}.$$

Set $Y = \|\widehat{W}^{\gamma}_{h_{\hat{\imath}}} - W_\rho\|_\infty / C$, $a = h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}} r_n(\log M)$ and $b = e^{\gamma h_m^{-2}}$. We have

$$\mathbb{E}[Y] = a + \mathbb{E}[Y-a] = a + \int_0^\infty \mathbb{P}\left(Y - a \geq u\right) du = a + b\int_0^\infty \mathbb{P}\left(Y - a \geq bt\right) dt.$$

Set now $t = r_n(x-1)$. If $0 < t < 1$, then we have $t = \sqrt{\frac{x}{n}}$. If $t \geq 1$ then we have $t = \frac{x}{n}$. Thus we get by the change of variable $t = \sqrt{\frac{x}{n}}$ that

$$\int_0^1 \mathbb{P}\left(Y - a \geq bt\right) dt = \int_0^n \mathbb{P}\left(Y - a \geq b\sqrt{\frac{x}{n}}\right) \frac{1}{2\sqrt{xn}} dx \leq \frac{1}{2\sqrt{n}} \int_0^n \frac{e^{-x}}{\sqrt{x}} dx \leq \frac{c}{\sqrt{n}},$$

where $c > 0$ is a numerical constant. Similarly, we get by change of variable $t = \frac{x}{n}$

$$\int_1^\infty \mathbb{P}\left(Y - a \geq bt\right) dt = \int_n^\infty \mathbb{P}\left(Y - a \geq b\frac{x}{n}\right) \frac{1}{n} dx \leq \frac{1}{n} \int_n^\infty e^{-x} dx \leq \frac{c'}{n},$$

where $c' > 0$ is a numerical constant. Combining the last three displays, we obtain the result in expectation.

Appendix D: Proof of Auxiliary Lemmas

D.1. Proof of Lemma 2

To prove the uniform bound of (36), we define

$$\delta_h = \max_{|t| \leq h^{-1}} \left\{|t| e^{\gamma t^2}\right\}.$$

Then, by definition of K_h^η and using the inverse Fourier transform formula, we have

$$\begin{aligned} \delta_h^{-1} \|K_h^\gamma\|_\infty &= \frac{1}{2\pi} \delta_h^{-1} \sup_{x \in \mathbb{R}} \left| \int e^{-itx} \widetilde{K}_h^\gamma(t) dt \right| \leq \frac{1}{2\pi} \delta_h^{-1} \int_{h^{-1}}^{h^{-1}} |t| e^{\gamma t^2} dt \\ &\leq \frac{1}{\pi} \delta_h^{-1} \int_0^{h^{-1}} t e^{\gamma t^2} dt \leq \frac{1}{2\gamma\pi} \delta_h^{-1} \int_0^{h^{-1}} 2\gamma t e^{\gamma t^2} dt \\ &\leq \frac{1}{2\gamma\pi} \delta_h^{-1} (e^{\gamma h^{-2}} - 1) \leq \frac{1}{2\gamma\pi} \delta_h^{-1} (e^{\gamma h^{-2}} - 1) \leq \frac{h}{2\gamma\pi} := U. \end{aligned} \tag{72}$$

For the entropy bound (37), we need to prove that $K_h^\gamma \in V_2(\mathbb{R})$, where $V_2(\mathbb{R})$ is the set of functions with finite quadratic variation (see Theorem 5 of Bourdaud, Lanza de Cristoforis and Sickel (2006)). To do this, it is enough to verify that $K_h^\gamma \in B_{2,1}^{1/2}(\mathbb{R})$ and the result is a consequence of

the embedding $B^{1/2}_{2,1}(\mathbb{R}) \subset V_2(\mathbb{R})$.

Let us define the Littlewood-Paley characterization of the seminorm $\|\cdot\|^{\bullet}_{1/2,2,1}$ as follows

$$\|g\|^{\bullet}_{1/2,2,1} := \sum_{\ell\in\mathbb{Z}} 2^{\ell/2}\|\mathcal{F}_1^{-1}[\alpha_\ell \mathcal{F}_1[g]]\|_2,$$

where $\alpha_\ell(\cdot)$ is a dyadic partition of unity with α_ℓ symmetric w.r.t to 0, supported in

$$[-2^{\ell+1}, -2^{\ell-1}] \cup [2^{\ell-1}, 2^{\ell+1}]$$

and $0 \le \alpha_\ell \le 1$ (see e.g. Theorem 6.3.1 and Lemma 6.1.7 in the paper of Bergh and Löfström (1976)). Then, $K_h^\gamma \in B^{1/2}_{2,1}(\mathbb{R})$, if and only if $\|K_h^\gamma\|^{\bullet}_{1/2,2,1}$ is bounded by a fixed constant. By isometry of the Fourier transform combined with definition of α_ℓ and K_h^γ, we get that

$$\begin{aligned}\|\mathcal{F}_1^{-1}[\alpha_\ell \mathcal{F}_1[K_h^\gamma]]\|_2 &= \|\alpha_\ell \mathcal{F}_1[K_h^\gamma]\|_2 = \|\alpha_\ell \widetilde{K}_h^\gamma\|_2 \\ &= \sqrt{\int_{[0,h^{-1}]\cap[2^{\ell-1},2^{\ell+1}]} \alpha_\ell(t)^2 2|t|^2 e^{2\gamma t^2} dt} \\ &\le \sqrt{\int_{[0,h^{-1}]\cap[2^{\ell-1},2^{\ell+1}]} 2t^2 e^{2\gamma t^2} dt}.\end{aligned}$$

A primitive of $t \to 2t^2 e^{2\gamma t^2}$ is $\frac{1}{2\gamma} t e^{2\gamma t^2} - \frac{1}{2\gamma}\int_0^t e^{2\gamma u^2} du$. Thus, we get that

$$\|\mathcal{F}_1^{-1}[\alpha_\ell \mathcal{F}_1[K_h^\gamma]]\|_2 \le \sqrt{\tfrac{1}{\gamma}} h^{-1/2} e^{\gamma h^{-2}}, \quad \forall \ell \in \mathbb{Z},$$

$$\text{and} \quad \|K_h^\gamma\|^{\bullet}_{1/2,2,1} \le \sqrt{\frac{1}{\gamma}} h^{-1/2} e^{\gamma h^{-2}} \sum_{\ell=-\infty}^{L_h} 2^{\ell/2},$$

where $L_h = \lfloor \log_2(h^{-1}) + 1 \rfloor$. A simple computation gives that

$$\sum_{\ell=-\infty}^{L_h} 2^{\ell/2} \le \frac{\sqrt{2}}{\sqrt{2}-1} + \frac{2^{(L_h+1)/2}-1}{\sqrt{2}-1} \le \frac{\sqrt{2}}{\sqrt{2}-1} + \frac{2}{\sqrt{2}-1} h^{-1/2}.$$

Combining the last two displays and since $h^{-1} \ge 1$, we get

$$\|K_h^\gamma\|^{\bullet}_{1/2,2,1} \le c\sqrt{\frac{1}{\gamma}} h^{-1} e^{\gamma h^{-2}},$$

where $c > 0$ is a numerical constant. This shows that $\delta_h^{-1}\|K_h^\gamma\|^{\bullet}_{1/2,2,1}$ is bounded by a fixed constant depending only on γ. Therefore $K_h^\gamma \in V_2(\mathbb{R})$ and the entropy bound (37) is obtained by applying Lemma 1 of Giné and Nickl (2009).

D.2. Proof of Lemma 5

We recall that the bandwidth $h_{\widehat{\mathbf{m}}}$ with $\widehat{\mathbf{m}}$ is defined in (30). Let $r_n(x) = \max\left(\sqrt{\frac{\log(n)+x}{n}}, \frac{\log(n)+x}{n}\right)$ and define

$$m^* := \text{argmin}_{1\le m\le M}\left\{h_m^{r/2-1} e^{-\frac{\beta}{h_m^r}} + e^{\gamma h_m^{-2}} r_n(x+\log M)\right\}, \tag{73}$$

and

$$B(m) = \max_{j:j>m} \left\{ \|\widehat{W}^{\gamma}_{h_m} - \widehat{W}^{\gamma}_{h_j}\|_\infty - 2\kappa e^{\gamma h_j^{-2}} r_n(x+\log M) \right\}.$$

In one hand, we have

$$\begin{aligned}
\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - \widehat{W}^{\gamma}_{h_{m^*}}\|_\infty \mathbb{1}_{\widehat{\mathbf{m}}>m^*} &= \left(\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - \widehat{W}^{\gamma}_{h_{m^*}}\|_\infty - 2\kappa e^{\gamma h_{\widehat{\mathbf{m}}}^{-2}} r_n(x+\log M) \right) \mathbb{1}_{\widehat{\mathbf{m}}>m^*} \\
&\quad + 2\kappa e^{\gamma h_{\widehat{\mathbf{m}}}^{-2}} r_n(x+\log M) \mathbb{1}_{\widehat{\mathbf{m}}>m^*} \\
&\leq \left(B(m^*) + 2\kappa e^{\gamma h_{\widehat{\mathbf{m}}}^{-2}} r_n(x+\log M) \right) \mathbb{1}_{\widehat{\mathbf{m}}>m^*}.
\end{aligned}$$

In the other hand, similarly, we have

$$\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - \widehat{W}^{\gamma}_{h_{m^*}}\|_\infty \mathbb{1}_{\widehat{\mathbf{m}}\leq m^*} \leq \left(B(\widehat{\mathbf{m}}) + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M) \right) \mathbb{1}_{\widehat{\mathbf{m}}\leq m^*}.$$

Combining the last two displays, and by definition of $\mathcal{L}_\kappa(\cdot)$ in (29), we get

$$\begin{aligned}
\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - \widehat{W}^{\gamma}_{h_{m^*}}\|_\infty &\leq \left(B(m^*) + 2\kappa e^{\gamma h_{\widehat{\mathbf{m}}}^{-2}} r_n(x+\log M) \right) \mathbb{1}_{\widehat{\mathbf{m}}>m^*} \\
&\quad + \left(B(\widehat{\mathbf{m}}) + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M) \right) \mathbb{1}_{\widehat{\mathbf{m}}\leq m^*} \\
&\leq B(m^*) + B(\widehat{\mathbf{m}}) + 2\kappa r_n(x+\log M)(e^{\gamma h_{\widehat{\mathbf{m}}}^{-2}} + e^{\gamma h_m^{-2}}) \\
&= \mathcal{L}(m^*) + \mathcal{L}(\widehat{\mathbf{m}}) \leq 2\mathcal{L}(m^*), \qquad (74)
\end{aligned}$$

where the last inequality follows from the definition of $\widehat{\mathbf{m}}$ in (30). By the definition of $B(\cdot)$, it follows

$$\begin{aligned}
\mathcal{L}(m^*) &= B(m^*) + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M) \\
&= \max_{j:j>m^*} \left\{ \|\widehat{W}^{\gamma}_{h_{m^*}} - \widehat{W}^{\gamma}_{h_j}\|_\infty - 2\kappa e^{\gamma h_j^{-2}} r_n(x+\log M) \right\} \\
&\leq \max_{j:j>m^*} \Big\{ \|\widehat{W}^{\gamma}_{h_{m^*}} - \mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}]\|_\infty + \|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + \|W_\rho - \mathbb{E}[\widehat{W}^{\gamma}_{h_j}]\|_\infty \\
&\quad + \|\mathbb{E}[\widehat{W}^{\gamma}_{h_j}] - \widehat{W}^{\gamma}_{h_j}\|_\infty] - 2\kappa e^{\gamma h_j^{-2}} r_n(x+\log M) \Big\} + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M).
\end{aligned}$$

On the event $\mathcal{E}_\kappa$, it follows that

$$\begin{aligned}
\mathcal{L}(m^*) &\leq \max_{j:j>m^*} \Big\{ \|\widehat{W}^{\gamma}_{h_{m^*}} - \mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}]\|_\infty + \|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + \|W_\rho - \mathbb{E}[\widehat{W}^{\gamma}_{h_j}]\|_\infty \\
&\quad - \kappa e^{\gamma h_j^{-2}} r_n(x+\log M) \Big\} + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M).
\end{aligned}$$

As $h_{m^*} > h_j$ for all $j > m^*$, we have $-e^{\gamma h_j^{-2}} < -e^{\gamma h_{m^*}^{-2}}$. Therefore, on the event $\mathcal{E}_\kappa$, we get

$$\mathcal{L}(m^*) \leq \|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + \max_{j:j>m^*} \left\{ \|\mathbb{E}[\widehat{W}^{\gamma}_{h_j}] - W_\rho\|_\infty \right\} + 2\kappa e^{\gamma h_{m^*}^{-2}} r_n(x+\log M) \qquad (75)$$

From (74) and on the event $\mathcal{E}_\kappa$, we have

$$\begin{aligned}
\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty &\leq \|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - \widehat{W}^{\gamma}_{h_{m^*}}\|_\infty + \|\widehat{W}^{\gamma}_{h_{m^*}} - W_\rho\|_\infty \leq |\widehat{W}^{\gamma}_{h_{m^*}} - W_\rho\|_\infty + 2\mathcal{L}(m^*) \\
&\leq \|\widehat{W}^{\gamma}_{h_{m^*}} - \mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}]\|_\infty + \|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + 2\mathcal{L}(m^*) \\
&\leq \kappa e^{\gamma h_m^{-2}} r_n(x+\log M) + \|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + 2\mathcal{L}(m^*).
\end{aligned}$$

Combining the last inequality with (75)

$$\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \leq 5\kappa e^{\gamma h_m^{-2}} r_n(x+\log M) + 3|\mathbb{E}[\widehat{W}^{\gamma}_{h_{m^*}}] - W_\rho\|_\infty + 2 \max_{j:j>m^*} \left\{ \|\mathbb{E}[\widehat{W}^{\gamma}_{h_j}] - W_\rho\|_\infty \right\}.$$

From Proposition 1, the bias is bounded by $t \to t^{r/2-1}e^{-\beta t^{-r}}$ an increasing function for sufficiently small $t > 0$, and as s $h_{m^*} > h_j$ for all $j > m^*$, we can write

$$\|\widehat{W}^{\gamma}_{h_{\widehat{\mathbf{m}}}} - W_\rho\|_\infty \;\le\; C\left(\kappa e^{\gamma h_m^{-2}} r_n(x + \log M) + h_{m^*}^{r/2-1} e^{-\beta h_{m^*}^{-r}}\right).$$

The result follows from (73), the definition of m^*.

D.3. Proof of Lemma 4

In view of Fatou's Lemma, we have

$$\begin{aligned}\operatorname{liminf}_{|z|\to\infty} z^2 p_0^{\gamma}(z) \;&\ge\; \int \operatorname{liminf}_{|z|\to\infty} z^2 p_0(z-x) N^{\gamma}(x)dx \\ &\ge\; \int_{-\sqrt{2\gamma}}^{\sqrt{2\gamma}} \operatorname{liminf}_{|z|\to\infty} z^2 p_0(z-x) N^{\gamma}(x)dx.\end{aligned}$$

Recall that $\gamma = \frac{1-\eta}{4\eta} \le 1/4$, then for $|z| \ge \sqrt{2\gamma}+1$ and any $x \in (-\sqrt{2\gamma}, \sqrt{2\gamma})$, it follows by Lemma 3 that $p_0(z-x) \ge c(z-x)^{-2}$. Thus,

$$\operatorname{liminf}_{|z|\to\infty} z^2 p_0^{\gamma}(z) \ge c\int_{-\sqrt{2\gamma}}^{\sqrt{2\gamma}} N^{\gamma}(x)dx = c\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \ge c' > 0,$$

where $c' > 0$ is a numerical constant . Choose now a numerical constant $\widetilde{c} \ge 0$ such that $\int_{-\widetilde{c}}^{\widetilde{c}} p_0(x)dx \ge 1/2$, therefore, for any $|z| \le 1 + \sqrt{2\gamma}$ and some numerical constant $c'' > 0$ we get

$$\begin{aligned}p_0^{\gamma}(z) \;&\ge\; \int_{-\widetilde{c}}^{\widetilde{c}} p_0(x) N^{\gamma}(z-x)dx \ge \min_{|y|\le M+1+\sqrt{2\gamma}} \{N^{\gamma}(y)\} \int_{-\widetilde{c}}^{\widetilde{c}} p_0(x)dx \\ &\ge\; \frac{1}{2} \min_{|y|\le M+1+\sqrt{2\gamma}} \{N^{\gamma}(y)\} \ge c'' > 0.\end{aligned}$$

D.4. Lemma 6

Lemma 6. *The density matrix $\rho^{(1)}$ defined in (51) satisfies the following conditions :*

(i) Self adjoint: $\rho^{(1)} = (\rho^{(1)})^*$.
(ii) Positive semi-definite: $\rho^{(1)} \ge 0$.
(iii) Trace one: $\mathrm{Tr}(\rho^{(1)}) = 1$.

Proof:
• Note first that V_h is not a Wigner function, however it belongs to the linear span of Wigner functions. Consequently, it admits the following representation

$$\frac{1}{\pi}\mathcal{R}[V_h](x,\phi)\mathbb{1}_{(0,\pi}(\phi) = \sum_{j,k=0}^{\infty} \tau_{j,k}^{(h)} \psi_j(x)\psi_k(x) e^{-i(j-k)\phi},$$

where

$$\tau_{j,k}^{(1)} = \int_0^{\pi}\!\!\int \frac{1}{\pi}\mathcal{R}[V_h](x,\phi) f_{j,k}(x) e^{-i(j-k)\phi} dx d\phi. \tag{76}$$

For the sake of brevity, we set from now on $\rho = \rho^{(1)}$ and $\tau = \tau^{(1)}$. Note that the matrix ρ satisfies $\rho_{j,k} = \rho_{j,k}^{(0)} + \tau_{j,k}$. Exploiting the above representation of τ, it is easy to see that $\tau_{j,k} = \overline{\tau}_{k,j}$ for

any $j, k \geq 0$. On the other hand, $\rho^{(0)}$ is a diagonal matrix with real-valued entries. This gives (i) immediately.

• We consider now (iii). First, note that $\mathcal{R}[V_h](\cdot, \phi)$ is an odd function for any fixed ϕ. Indeed, its Fourier transform with respect to the first variable

$$\mathcal{F}_1 \left[\mathcal{R}[V_h](\cdot, \phi)\right](t) = \widetilde{V}_h(t\cos\phi, t\sin\phi),$$

is an odd function of t for any fixed ϕ. Thus, it is easy to see that $\tau_{j,j} = 0$, for any $j \geq 0$. Since $\rho^{(0)}$ is already known to be a density matrix, this implies that

$$\mathrm{Tr}(\rho^{(1)}) = \mathrm{Tr}(\rho^{(0)}) + \mathrm{Tr}(\tau) = 1.$$

• Let now prove (ii). Define $r_k := \sum_{j=1\,:\,j\neq k}^{\infty} |\tau_{j,k}|, \quad \forall k \geq 1$. We will prove that $\rho_{k,k} \geq 2r_k$ for all $k \geq 1$ and combine this fact with Gershgorin's disk theorem to get the result.

We omit the numerical constants in our analysis since we are only interested in bounding the coefficients $\tau_{j,k}$. More specifically, we will prove there exists a numerical constant $c > 0$ such that we have for any $k \geq 2$ that

$$|\tau_{k,1}| \leq c\frac{a}{\sqrt{k!}}, \tag{77}$$

and also for any $k \geq 2$ and $l \geq 1$,

$$\tau_{k+2l,k} = 0, \quad |\tau_{k+2l+1,k}| \leq c\frac{a}{k^{l+\frac{1}{2}}}, \quad |\tau_{k+1,k}| \leq c\frac{a}{k^{\frac{5}{4}}}. \tag{78}$$

We combine now the previous display with Gershgorin's disk theorem to get the result. More precisely, since ρ is a Hermitian matrix (*iii*), it admits real eigenvalues. For any eigenvalue μ of ρ, in view of Theorem 5 in Appendix D.5, there exists an integer $k \geq 1$ such that

$$\left|\mu - \rho_{kk}^{(0)}\right| \leq \sum_{j=1\,:\,j\neq k}^{\infty} |\tau_{j,k}| =: r_k. \tag{79}$$

If $k = 1$, we get $r_1 \leq ca\sum_{j=2}^{\infty} \frac{1}{\sqrt{j!}} \leq c'a$ for some numerical constant $c' > 0$. If $k \geq 2$, we get in view of (11) that

$$\begin{aligned} r_k &= \sum_{j=1}^{k-1} |\tau_{j,k}| + \sum_{j=k+1}^{\infty} |\tau_{j,k}| \\ &\leq ca\left(\frac{1}{\sqrt{k!}} + \sum_{l=2}^{k-2} \frac{1}{(k-l)^{l+1/2}} + (k-1)^{-5/4} + k^{-5/4} + \sum_{l=2}^{\infty} \frac{1}{k^{l+\frac{1}{2}}}\right) \\ &\leq c'a\frac{1}{k^{5/4}}, \end{aligned}$$

for some numerical constant $c' > 0$.

Recall that $\rho^{(0)} = \rho^{\alpha,\lambda}$ for some $0 < \alpha, \lambda < 1$ where $\rho^{\alpha,\lambda}$ is defined in (41). Lemma 2 in the paper of Butucea, Guţă and Artiles (2007) guarantees that

$$\rho_{kk}^{\alpha,\lambda} = \frac{\alpha}{(1-\lambda)^{\alpha}}\Gamma(\alpha+1)k^{-(1+\alpha)}(1+o(1)),$$

as $n \to \infty$. We note that $\rho_{kk}^{(0)} > 0$ decreases polynomially at the rate $k^{-(1+\alpha)}$ as $k \to \infty$ whereas r_k decreases polynomially at the rate $k^{-\frac{5}{4}}$. Fix the numerical constant $\alpha \in (0, 1/4)$. Then, taking the numerical constant $a > 0$ small enough in (47) independently of k, we get $\rho_{kk} \geq 2r_k$ for any $k \geq 1$. Thus, any eigenvalue μ of ρ is non-negative. Consequently, ρ is positive semi-definite.

We now prove (77)-(78). In (80), we fix the variable ϕ and we apply Plancherel's theorem to the integral w.r.t. the variable t to get that

$$\tau_{j,k} = \int_0^\pi \frac{1}{\pi} \widetilde{V}_h(t\cos\phi, t\sin\phi)\widetilde{f}_{j,k}(t)e^{-i(j-k)\phi}dtd\phi. \tag{80}$$

Plugging the definition of the pattern functions (11) and (47) into the previous display, we get for any $j > k$ that

$$\tau_{j,k} = (-i)^{j-k}\sqrt{\frac{2^{k-j}k!}{j!}} \int_0^\pi \widetilde{V}_h(t\cos\phi, t\sin\phi)|t|t^{j-k}e^{-\frac{t^2}{4}}L_k^{j-k}(\frac{t^2}{2})e^{-i(j-k)\phi}dtd\phi. \tag{81}$$

Set $\mathcal{C}_{a,h} = aC_0h^{-1}e^{\beta h^{-2}}$. Plugging now (47) into the previous display, we get for any $j > k$ that

$$\tau_{j,k} = \mathcal{C}_{a,h}(-1)^{j-k}i^{j-k+1}\sqrt{\frac{2^{k-j}k!}{j!}} \int_0^\pi e^{-2\beta t^2}g_1(t^2-h^{-2})g(t\sin\phi)|t|t^{j-k}e^{-\frac{t^2}{4}}L_k^{j-k}(\frac{t^2}{2})e^{-i(j-k)\phi}dtd\phi.$$

By construction, for any fixed $\phi \in (0,\pi)$, we note that $t \to e^{-2\beta t^2}g_1(t^2-h^{-2})g(t\sin\phi)|t|e^{-\frac{t^2}{4}}L_k^{j-k}(\frac{t^2}{2})$ is an odd function. Hence, if $j-k$ is even, then we have $\tau_{j,k} = 0$.

Set $\alpha = j-k$. We will study separately the four different settings: (a) k and α are bounded; (b) k is bounded and α is large; (c) k is large and α is also large; (d) k is large and α is bounded.
Case (a). For any pair (j,k) such that $k \leq 34$ and $\alpha = j-k \leq 23$. In view of Theorem 1 in Krasikov (2007), we get that

$$|\tau_{j,k}| \leq \sqrt{12}\mathcal{C}_{a,h}\left(\int_{a_1}^{b_1} e^{-\beta t^2}dt\right)\sqrt{\frac{k!}{j!}}k^{\frac{1}{12}}(j+1)^{1/4} \leq C\frac{a}{j^{3/2}},$$

provided n is taken large enough.
Case (b). Assume that $k \leq 34$ and $\alpha = j-k \geq 24$. Again, in view of Theorem 1 in Krasikov (2007), we have

$$\sup_{x>0}\left\{x^{\alpha+1}e^{-x}\left(L_k^\alpha(x)\right)^2\right\} \leq 6k^{1/6}\sqrt{k+\alpha+1}.$$

Plugging the above display into the definition of $\tau_{j,k}$, we obtain that

$$\begin{aligned}|\tau_{j,k}| &\leq \sqrt{6}\mathcal{C}_{a,h}\left(\int_{a_1}^{b_1} e^{-\beta t^2}dt\right)\sqrt{\frac{2^{k-j}k!}{j!}}k^{\frac{1}{12}}(j+1)^{1/4}\sqrt{2^{j-k+1}} \\ &\leq \sqrt{12}\mathcal{C}_{a,h}\left(\int_{a_1}^{b_1} e^{-\beta t^2}dt\right)\sqrt{\frac{k!}{j!}}k^{\frac{1}{12}}(j+1)^{1/4} \leq C\frac{a}{j^{3/2}},\end{aligned}$$

where $C > 0$ is an absolute constant.
Case (c). Assume that $k \geq 35$ and $\alpha = j-k \geq 24$. Theorem 2 in Krasikov (2007) guarantees the existence of a numerical constant $C > 0$ such that:

$$\sup_{x>0}\left\{x^{\alpha+1}e^{-x}\left(L_k^\alpha(x)\right)^2\right\} \leq Ck^{-1/6}\sqrt{k+\alpha+1}.$$

Similarly as in the previous case, we obtain

$$|\tau_{j,k}| \leq C\mathcal{C}_{a,h}\left(\int_{a_1}^{b_1} e^{-\beta t^2}dt\right)\sqrt{\frac{k!}{j!}}k^{-\frac{1}{12}}(j+1)^{1/4} \leq C'\frac{a}{j^{3/2}},$$

for some numerical constant $C' > 0$.

Case (d). Assume that $k \geq 35$ and $\alpha = j - k \leq 24$. In view of Theorem 1 in Krasikov (2007), we have

$$M_k^\alpha(x) := \left[L_k^\alpha(x)\right]^2 e^{-x} x^{\alpha+1} \leq \frac{x(s^2-q^2)}{r(x)}, \quad \forall x \in (q^2, s^2),$$

where $s = \sqrt{k+\alpha+1} + \sqrt{k}$, $q = \sqrt{k+\alpha+1} - \sqrt{k}$ and $r(x) = (x-q^2)(s^2-x)$.

Note that $s \to \infty$ as $k \to \infty$ whereas $q \leq \sqrt{\alpha+1}$. Thus, for any n large enough such that $a_1^2/2 \geq \alpha + 1$, there exists k_0 large enough such that for any $k \geq k_0$, we have $\left(\frac{a_1^2}{2}, \frac{b_1^2}{2}\right) \subset (q^2, s^2)$. Thus, the above display gives

$$M_k^\alpha\left(\frac{t^2}{2}\right) \leq \frac{t^2}{2} \frac{s^2-q^2}{(t^2/2-q^2)(s^2-t^2/2)} \leq C$$

for some absolute constant $C > 0$.

Combining the above display with (81), we get that

$$|\tau_{j,k}| \leq C\mathcal{C}_{a,h} \left(\int_{a_1}^{b_1} e^{-2\beta t^2} dt\right) \sqrt{2\frac{k!}{j!}} \leq C' a \sqrt{\frac{k!}{j!}},$$

for some numerical constant $C' > 0$.

We consider now the case $3 \leq \alpha \leq 24$ (we recall that $\tau_{k+2,k} = 0$). We have in view of the previous display that

$$|\tau_{j,k}| = |\tau_{k+\alpha,k}| \leq C' a \sqrt{\frac{1}{(k+1)\cdots(k+\alpha)}} \leq C \frac{a}{k^{\alpha/2}}, \quad \forall k \geq 1.$$

for some numerical constant $C > 0$.

Note that the previous bound is also valid for $\alpha = 1$ but it is not sufficient for our needs. The case $\alpha = 1$ is actually the most difficult as the previous bounds on the Laguerre polynomials are not sufficient to yield the desired control. In that case, we should rather exploit the oscillatory properties of the Laguerre polynomials. For large k, the Laguerre polynomial function L_k^1 behaves essentially like a trigonometric function on any fixed compact interval. Combining this fact with the stationary phase principle, we can obtain a small enough bound.

In view of (1.1) in Muckenhoupt (1970), we have, as $k \to \infty$, for any $x \in (a_1^2/2, b_1^2/2)$ that

$$L_k^1(x) = \sqrt{2}(k+1)\frac{e^{x/2}}{\nu_k^{1/4}\sqrt{x}} \psi\left(\frac{x}{\nu_k}\right) \left[J_1\left(\sqrt{\nu_k x} + \epsilon(x,\nu_k)\right) + O\left(\frac{x^{1/4}}{\nu_k^{7/4}}\right)\right],$$

where $\nu_k = 4k+6$, $\epsilon(x,\nu_k) = O\left(\nu_k^{-1/2} x^{\frac{3}{2}}\right)$, $\psi(t) = \left(\frac{\tilde{g}(t)}{\tilde{g}'(t)}\right)^{1/2}$ with $\tilde{g}(t) = \arcsin(t)$ and J_1 is the Bessel function of order 1.

Note that $\psi(t) = \sqrt{t}(1+o(1))$ as $t \to 0$. The Bessel function J_1 admits the following asymptotic expansion as $t \to \infty$:

$$J_1(t) = \sqrt{\frac{2}{\pi t}} \left[\cos\left(t - \frac{3\pi}{4}\right) + \frac{3}{8t} \sin\left(t - \frac{3\pi}{4}\right) + O\left(\frac{1}{t^2}\right)\right].$$

See for instance Watson (1995) Section 7.1 page 195.

Combining the last two displays with (80), we get

$$\begin{aligned}
\tau_{k+1,k} &= \int_0^\pi \int A_k(t,\phi) \cos\left(\sqrt{\frac{\nu_k}{2}} t - \epsilon(t^2/2,\nu_k) - \frac{3\pi}{4}\right) dt d\phi \\
&\quad + \int_0^\pi \int B_k(t,\phi) \sin\left(\sqrt{\frac{\nu_k}{2}} t - \epsilon(t^2/2,\nu_k) - \frac{3\pi}{4}\right) dt d\phi + O\left(\int_0^\pi \int C_k(t,\phi) dt d\phi\right) \\
&= I_1 + I_2 + I_3, \qquad (82)
\end{aligned}$$

where

$$A_k(t,\phi) = -i\sqrt{\frac{k+1}{\pi}}\frac{1}{\nu_k}\widetilde{V}_h(t\cos\phi, t\sin\phi)\sqrt{|t|}t(1+o(1)),$$

and

$$B_k(t,\phi) = \frac{3}{8t}A_k(t,\phi), \quad C_k(t,\phi) = \frac{1}{t^2}|A_k(t,\phi)|.$$

Applying Theorem 4 with $m = 1$ gives

$$|I_1| \leq c\frac{a}{k^{3/2}}, \quad \text{and} \quad |I_2| \leq c\frac{a}{k^{3/2}}, \quad \forall k \geq 1,$$

for some numerical constant $c > 0$.

For the last integral in (82), a quick inspection gives

$$|I_3| \leq c\frac{a}{k^{5/4}}, \quad \forall k \geq 1,$$

for some numerical constant $c > 0$.

Combining the last two displays with (82) gives the desired bound.

D.5. Auxiliary results

For the sake of completeness, we collect here a few results used in our proofs.

The following lemma, due to Butucea and Tsybakov (2008a), describes the asymptotic behaviour of integrals of exponentially decreasing functions.

Lemma 7. *For any positive α, β, r, s and for any $A \in \mathbb{R}$ and $B \in \mathbb{R}$, we have*

$$\int_v^\infty u^A \exp\left(-\alpha u^r\right) du = \frac{1}{\alpha r} v^{A+1-r} \exp(-\alpha v^r)(1+o(1)), v \to \infty, \tag{83}$$

and

$$\int_0^v u^B \exp\left(\beta u^s\right) du = \frac{1}{\beta s} v^{B+1-s} \exp(\beta v^s)(1+o(1)), v \to \infty. \tag{84}$$

The following classical result describes the asymptotic behaviour of integrals with non-stationary phase functions. See for instance page 631 in Zorich (2016).

Theorem 4 (Localisation principle)**.** *Fix a compact domain $\Omega \subset \mathbb{R}$ and let $f \in \mathcal{C}_c^\infty(\Omega)$. Let $S \in \mathcal{C}^\infty(\Omega)$ be a function such that $S'(x) \neq 0$ for any $x \in \text{support}(f)$. Then, as $\nu \to \infty$, we have*

$$F(\nu) = \int_\Omega f(x)\cos(\nu S(x))dx = O\left(\frac{1}{\nu^m}\right), \quad \forall m \geq 1.$$

The same conclusion holds valid for the cosine function replaced by the sine function.

The following theorem provides a localization bound on the eigenvalues of square matrices. See for instance Feingold and Varga (1962).

Theorem 5 (Gershgorin Disk Theorem)**.** *Let A be an infinite square matrix and let μ be any eigenvalue of A. Then, for some $j \geq 1$, we have*

$$|\mu - A_{j,j}| \leq r_j(A),$$

where $r_j(A) = \sum_{k \geq 1 : k \neq j} |A_{j,k}|$.

We present some well-known results about the theory of empirical processes that are used in our proof. We refer the interested reader to Giné and Nickl (2009) for more details about this theory.

Let $Z_1, ..., Z_n$ be i.i.d. with law P on $\mathbb{R}$, and let $\mathcal{F}$ be a P-centered (i.e., $Pf = \int f dP = 0$ for all $f \in \mathcal{F}$) countable class of real-valued functions on $\mathbb{R}$, uniformly bounded by the constant U, called the envelope of the class. We say that $\mathcal{F}$ is a VC-type class for the envelope U and with VC-characteristics A, v if its $\mathcal{L}^2(Q)$ covering numbers satisfy that, for all probability measures Q and $\varepsilon > 0$, $N(\mathcal{F}, \mathcal{L}^2(Q), \varepsilon) \leq (AU/\varepsilon)^v$.

For such classes, assuming $Pf = 0$ for $f \in \mathcal{F}$, there exists a universal constant L such that

$$E := \mathbb{E} \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(Z_i) \right| \leq L \left(\sqrt{v} \sqrt{n\sigma^2} \sqrt{\log \frac{AU}{\sigma}} + vU \log \frac{AU}{\sigma} \right), \tag{85}$$

where σ is any positive number such that

$$\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathbb{E}(f^2(Z)).$$

See, e.g., Giné and Guillou (2001).

Talagrand's inequality bounds the deviation of the suprema of empirical processes. The following result is a version of this inequality is due to Bousquet (2002).

Theorem 6. *Assume that Z_i are identically distributed according to P. Let $\mathcal{F}$ be a countable class set of functions from a set $\mathcal{X}$ to $\mathbb{R}$ and assume that all functions f in $\mathcal{F}$ are $\mathbb{P}$-measurable, square-integrables and satisfy $\mathbb{E}[f(Z_1)] = 0$ with envelope equal to 1. Let $\sigma^2 \geq \sup_{f \in F} \mathrm{Var}(f(X_1))$ almost surely, then for all $x \geq 0$, we have*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(Z_i) \right| \geq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(Z_i) \right| \right] + \sqrt{2xnv} + \frac{x}{3} \right) \leq e^{-x},$$

with $v = n\sigma^2 + 2\mathbb{E}\left[\sup_{f \in \mathcal{F}} \left|\sum_{i=1}^{n} f(Z_i)\right|\right]$.