Adaptive estimation of the density matrix in quantum homodyne tomography with noisy data

P Alquier\textsuperscript{1}, K Meziani\textsuperscript{2} and G Peyré\textsuperscript{2}

\textsuperscript{1} School of Mathematical Sciences, University College, Dublin, Ireland
\textsuperscript{2} CEREMADE, UMR CNRS 7534, Université Paris Dauphine, Place du Maréchal De Lattre De Tassigny, F-75775 PARIS Cedex 16, France

E-mail: meziani@ceremade.dauphine.fr, pierre.alquier@ucd.ie and gabriel.peyre@ceremade.dauphine.fr

Received 31 January 2013, in final form 30 April 2013
Published 24 June 2013
Online at stacks.iop.org/IP/29/075017

Abstract

In the framework of noisy quantum homodyne tomography with efficiency parameter $\eta = \frac{1}{2} < \eta \leq 1$, we propose a novel estimator of a quantum state whose density matrix elements $\rho_{m,n}$ decrease like $Ce^{-B(m+n)^{r}/2}$, for fixed $C \geq 1$, $B > 0$ and $0 < r \leq 2$. In contrast to previous works, we focus on the case where $r$, $C$ and $B$ are unknown. The procedure estimates the matrix coefficients by a projection method on the pattern functions, and then by soft-thresholding the estimated coefficients. We prove that under the $L_2$-loss our procedure is adaptive rate-optimal, in the sense that it achieves the same rate of convergence as the best possible procedure relying on the knowledge of $(r, B, C)$. Finite sample behaviour of our adaptive procedure is explored through numerical experiments.

(Some figures may appear in colour only in the online journal)

1. Introduction

This paper deals with a severely ill-posed inverse problem which comes from quantum optics. Quantum optics is a branch of quantum mechanics which studies physical systems at the atomic and subatomic scales. As the language used by physicists\textsuperscript{3} differs from the one that is used by statisticians, we start with general notions on quantum mechanics. The interested reader can get further acquainted with quantum concepts through textbooks or review articles [1–4].

1.1. Physical background

1.1.1. Quantum mechanics. In quantum mechanics, the quantum state of a system is a mathematical object which encompasses all the information about the system. The most

\textsuperscript{3} For example, they speak about ‘states’ or ‘observable’ instead of ‘laws’ or ‘random variables’...
common representation of a quantum state is an operator $\rho$ on a complex Hilbert space $\mathcal{H}$ (called the space of states) satisfying the following three conditions.

(i) Self-adjoint: $\rho = \rho^*$, where $\rho^*$ is the adjoint of $\rho$.
(ii) Positive: $\rho \geq 0$, or equivalently $\langle \psi, \rho \psi \rangle \geq 0$ for all $\psi \in \mathcal{H}$.
(iii) Trace one: $\text{Tr}(\rho) = 1$.

A quantum state $\rho$ encodes the probabilities of the measurable properties, or ‘observables’ (energy, position, ...) of the considered quantum system. Generally, in quantum mechanics the expected results of the measurements of an observable are not deterministic values but predictions about probability distributions, that is, the probability of obtaining each of the possible outcomes when measuring an observable.

An observable $X$ is described by a self-adjoint operator on the space of states $\mathcal{H}$ and

$$X = \sum_{\alpha} x_{\alpha} P_{\alpha},$$

where the eigenvalues $\{x_{\alpha}\}_\alpha$ of the observable $X$ are real and $P_{\alpha}$ is the projection onto the one-dimensional space generated by the eigenvector of $X$ corresponding to the eigenvalue $x_{\alpha}$. Then, when performing a measurement of the observable $X$ of a quantum state $\rho$, the result is a random variable $X$ with values in the set of the eigenvalues of the observable $X$. For a quantum system prepared in the state $\rho$, $X$ has the following probability distribution and expectation function:

$$P_\rho (X = x_\alpha) = \text{Tr}(P_{\alpha} \rho) \quad \text{and} \quad E_\rho (X) = \text{Tr}(X \rho).$$

An important element which affects the result of the measurement process is the purity of quantum states. A state is called pure if it cannot be represented as a mixture (convex combination) of other states, i.e., if it is an extreme point of the convex set of states. All other states are called mixed states. We give examples of states in section 3.

1.1.2. Quantum optics. In this paper, the quantum system we work with is a monochromatic light in a cavity described by a quantum harmonic oscillator. In the framework of quantum optics, the space of states is known to be the separable Hilbert space $\mathcal{H} = L_2(\mathbb{R})$, i.e. the space of square integrable complex-valued functions on the real line. A particular orthonormal basis $\{\psi_j\}_{j \in \mathbb{N}}$,—called the Fock basis—comes with this Hilbert space. This physically very meaningful basis is defined for all $j \in \mathbb{N}$ as follows:

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

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

We may give an equivalent representation for a quantum state $\rho$ in terms of the associated Wigner function $W_\rho$ (see [5]). The Wigner function $W_\rho$ is a real function of two variables and may be defined by its Fourier transform $\mathcal{F}_2$ with respect to both variables

$$\tilde{W}_\rho (u, v) := \mathcal{F}_2 [W_\rho] (u, v) = \text{Tr}(\rho \exp(iuQ + ivP)),$$

where $Q$ and $P$ are respectively the electric and magnetic fields. These two observables we are concerned by, do not commute. As non-commuting observables, they may not be simultaneously measurable. Therefore, by performing measurements on $(Q, P)$, we cannot obtain a probability density of the result $(Q, P)$. However, for $\phi \in [0, \pi]$, we can measure the quadrature observables $X_\phi := Q \cos \phi + P \sin \phi$, and then the above Wigner function plays the role of a quasi-probability density. It does not satisfy all the properties of a conventional
probability density but satisfies boundedness properties unavailable for classical densities. For instance, the Wigner function can and normally does go negative for states which have no classical model. The Wigner function is such that

- \( W_\rho : \mathbb{R} \rightarrow \mathbb{R} \)
- \( \int \int W_\rho(q, p) \, dq \, dp = 1 \)

Furthermore, its Radon transform is always a probability density

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

with respect to \( \frac{1}{\lambda} \), with \( \lambda \) being the Lebesgue measure on \( \mathbb{R} \times [0, \pi] \).

Now we can make explicit the links between the state \( \rho \) and the Radon transform \( p_\rho(x|\phi) \) of the Wigner function \( W_\rho \) associated with \( \rho \). In the Fock basis (1), the entries \( \rho_{j,k} \) of the finite density matrix \( \rho \) are given by

\[
\rho_{j,k} = \frac{1}{\pi} \int_0^{2\pi} p_\rho(x|\phi) f_{j,k}(x) e^{-i(k-j)\phi} \, d\phi \, dx
\]

for all \( j, k \in \mathbb{N} \). The functions \( f_{j,k} = f_{k,j} \), in expression (3), are bounded real functions commonly called pattern functions in quantum homodyne literature. A concrete expression for their Fourier transform \( \hat{f}_{k,j} \) using Laguerre polynomials \( L_n^\alpha(\cdot) \) is (cf [6]), for \( j \geq k \),

\[
\hat{f}_{k,j}(t) = \pi (-i)^{j-k} \left( \frac{2^{k-j}k!}{j!} \right) |t|^{j-k} e^{-\frac{t^2}{2}} L_{j-k}^{j-k} \left( \frac{t^2}{2} \right).
\]

We recall that the Laguerre polynomial of degree \( n \) and order \( \alpha \) is defined by

\[
L_n^\alpha(x) := (n!)^{-1} e^{x} x^{-\alpha} \frac{d^n}{dx^n} (e^{-x} x^n).\]

1.1.3. Quantum homodyne tomography. In this paper, we address the problem of reconstructing the density matrix \( \rho \) of a monochromatic light in a cavity. As the observables \( Q \) and \( P \) cannot be measured simultaneously, we measure the quadrature \( X_\phi := Q \cos \phi + P \sin \phi \), where \( \phi \in [0, \pi] \). Each of these quadratures could be measured on a laser beam by a technique put in practice for the first time in [7] and called quantum homodyne tomography (QHT). The theoretical foundation of quantum homodyne tomography was outlined in [8].

The experimental set-up, described in figure 1, consists of mixing the cavity pulse prepared in the state \( \rho \) with an additional laser of high intensity \( |z| \gg 1 \) called the local oscillator. After the mixing, the beam is split again and each of the two emerging beams is measured by one of the two photodetectors which give integrated currents \( I_1 \) and \( I_2 \) proportional to the number of photons. The result of the measurement is produced by taking the difference of the two currents and rescaling it by the intensity \( |z| \). Just before the mixing, the experimentalist may choose the phase \( \Phi \) of the local oscillator, randomly, uniformly distributed on \( [0, \pi] \). In the case of noiseless measurement and for a phase \( \Phi = \phi \), the result \( X_\phi = \frac{H_1 - H_2}{\sqrt{2}} \) has a density \( p_{\phi}(x|\phi) \) corresponding to measuring \( X_\phi \).

In practice, 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 physicists argue that their machines actually have high detection efficiency, around 0.8/0.9. Thus, we suppose \( \eta \) known. As the detection process is inefficient, an independent Gaussian noise interferes additively with the ideal data \( X_\phi \). Thus, for \( \Phi = \phi \), the effective result of the QHT measurement (figure 1) is for a known efficiency \( \eta \in [0.5, 1] \),

\[
Y = \sqrt{\eta} X_\phi + \sqrt{(1 - \eta)/2} \xi,
\]

where \( \xi \) is a standard Gaussian random variable, independent of \( X_\phi \).
1.2. Statistical model

This paper aims at reconstructing the density matrix of a monochromatic light in a cavity prepared in the state $\rho$. 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 the state $\rho$ is done by QHT as described in section 1.1.3. In the ideal setting, the results of such experiments would be $n$ independent identically distributed random variables $(X_1, \Phi_1), \ldots, (X_n, \Phi_n)$ with values in $\mathbb{R} \times [0, \pi]$ and distribution $P_\rho$ having density with respect to $\lambda$ ($\lambda$ being the Lebesgue measure on $\mathbb{R} \times [0, \pi]$) equal to

$$p_\rho(x, \varphi) = \frac{1}{\pi} p_\rho(x|\varphi) = \frac{1}{\pi} \mathcal{R}[W_\rho](x, \varphi),$$

where $\mathcal{R}$ is the Radon transform defined in equation (2). As underlined in section 1.1.3, we do not observe $(X_\ell, \Phi_\ell)_{\ell=1,\ldots,n}$ but the noisy version $(Y_\ell, \Phi_\ell)_{\ell=1,\ldots,n}$ where

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

Here, $\xi_\ell$s are independent standard Gaussian random variables, independent of all $(X_\ell, \Phi_\ell)$, $\ell = 1, \ldots, n$. The detection efficiency $\eta \in [0, 1]$ is a known parameter and $1-\eta$ represents the proportion of photons which are not detected due to various losses in the measurement process.

Let us denote by $p^\eta_\rho(y, \varphi)$ the density of $(Y_\ell, \Phi_\ell)$. Then, for $\Phi = \varphi$, the conditional density $p^\eta_\rho(\cdot|\varphi)$ is the convolution of the density $\frac{1}{\sqrt{\eta}} p_\rho(\cdot|\varphi)\sqrt{\eta} X$, with $N^\eta$ being the density of a centred Gaussian distribution having variance $(1-\eta)/2$, that is,

$$p^\eta_\rho(y|\varphi) = \left( \frac{1}{\sqrt{\eta}} p_\rho \left( \frac{\cdot}{\sqrt{\eta}} | \varphi \right) \ast N^\eta \right)(y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\eta}} p_\rho \left( \frac{y-x}{\sqrt{\eta}} | \varphi \right) N^\eta(x) \, dx.$$

For $\Phi = \varphi$, a useful equation in the Fourier domain, deduced by the previous relation (7), is

$$\mathcal{F}_1[p^\eta_\rho(\cdot|\varphi)](t) = \mathcal{F}_1[p_\rho(\cdot|\varphi)](t) \tilde{N}^\eta(t),$$

where $\mathcal{F}_1$ denotes the Fourier transform with respect to the first variable and $\tilde{N}^\eta(t) = e^{-\frac{1}{4\eta} t^2}$ is the Fourier transform of $N^\eta(x)$, the density of a centered Gaussian density having variance $(1-\eta)/2\eta = \gamma$. 

Figure 1. QHT measurement scheme.
In order to estimate the elements of the density matrix defined in (3) from the data \((Y_\ell, \Phi_\ell)_{\ell=1,...,n}\), we define a realistic class of quantum states \(\mathcal{R}(C, B, r)\). For \(C \geq 1, B > 0\) and \(0 < r \leq 2\), the class \(\mathcal{R}(C, B, r)\) is defined as follows:

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

(9)

Note that the class \(\mathcal{R}(C, B, r)\) has been translated in terms of Wigner functions in [9], where it has been proved that the fast decay of the elements of the density matrix implies both rapid decay of the Wigner function and of its Fourier transform.

However, contrary to previous works, we do not assume here that the constants \(r, B\) and \(C\) are known. From now on, we denote by \(\langle \cdot, \cdot \rangle\) and \(\| \cdot \|\) the usual Euclidean scalar product and norm.

1.3. Outline of the results

This paper deals with the problem of adaptive estimation of the density matrix \(\rho\) in QHT when taking into account the detection losses occurring in the measurement, leading to an additional Gaussian noise in the measurement data. In order to compute the performance of our procedure in \(L_2\) risk, we defined in the previous section a realistic class of quantum states \(\mathcal{R}(C, B, r)\) in which the elements of the density matrix decrease rapidly. From the physical point of view, all the states which have been produced in the laboratory to date belong to such a class, and a more detailed argument can be found in [10] as to why this assumption is realistic and in [9] as how to translate this class in terms of associated Wigner functions.

The problem of reconstructing the quantum state of a light beam has been extensively studied in quantum statistics and physical literature. Methods for reconstructing a quantum state are based on the estimation of either the density matrix \(\rho\) or the Wigner function \(W_\rho\).

The estimation of the density matrix from averages of data has been considered in the framework of ideal detection (\(\eta = 1\)) in [11–14]. Max-likelihood methods have been studied in [14–17] and a procedure using adaptive tomographic kernels to minimize the variance has been proposed in [18]. In a more general case of an efficiency parameter \(\eta\) belonging to the interval \([1/2, 1]\), the estimation of the density matrix of a quantum state of light has been discussed in [19, 16, 20] and considered in [21] via the pattern functions for the diagonal elements. The problem of goodness-of-fit testing in quantum statistics has been considered in [22]. 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.

For the problem of pointwise estimation of the Wigner function, we mention the work [23] in the case of ideal detection that corresponds to \(\eta = 1\), where a kernel estimator is given and its sharp minimax optimality over a class of Wigner functions characterized by their smoothness is established. The same problem in the noisy setting \(\eta \in [1/2, 1]\) was treated in [10], where the minimax rates were obtained. The estimation of a quadratic functional of the Wigner function, as an estimator of the purity, was explored in [24].

Recently, the more general case \(\eta \in [0, 1]\) was investigated in [9]. The authors provided the rates of convergence in \(L_2\) loss for both an estimator of the Wigner function and an estimator of the density matrix. Interestingly, the rates are polynomial in the case \(r = 2\), whereas they are intermediate for \(r \in [0, 2]\), where intermediate means that they are slower than any power of \(n\) but faster than any power of \(\log n\). However, the physicists argue that their machines actually have high detection efficiency, around 0.9. So we do not deal in this paper with values of \(\eta\) smaller than 1/2. It is to be noted that the estimator proposed in [9] depends on the knowledge of \(B\) and \(r\). This is a serious limitation since in practice, one will face situations where one...
wants to reconstruct a density matrix without assuming knowledge of \( B \) and \( r \). This is known in statistics as ‘adaptive estimation’. In this work, we tackle the problem of adaptive estimation over the classes of quantum states \( \mathcal{R}(C, B, r) \). Our estimator is actually a soft-thresholded version of the estimator in [9] which allows us to reach adaptation.

Coefficients thresholding is now a classical tool in statistics. It was introduced in a series of papers [25–27] in the context of function estimation via wavelet coefficients. We refer to [28] for a comprehensive introduction to thresholding and wavelets. These methods were extended to inverse problems [29–32]; see [33] for an introduction and a survey of the most recent results.

The remainder of the paper is organized as follows. In section 2, we present our adaptive thresholding procedure and state our main theoretical results. In particular, we establish upper bounds on the \( L_2 \) risk of our procedure and achieve the convergence rates over a broad family of set \( \mathcal{R}(C, B, r) \) which have been obtained in [9]. These bounds are nonasymptotic and hold true with large probability. The theoretical investigation is complemented by numerical experiments reported in section 3. The proofs of the main results are deferred to the appendix.

2. Density matrix estimation

We assume that now \( n \) independent identically distributed random pairs \((Y_i, \Phi_1)_i=1,...,n\) are observed, where \( \Phi_1 \) is uniformly distributed in \([0, \pi]\) and the conditional density of \( Y_1 \) given \( \Phi_1 = \eta \) is \( p(\eta, r, \Phi_1) \), cf (7). The goal is to estimate the density matrix \([\rho_{j,k}]_{j,k}\) defined by (3) and investigate the convergence rate of the proposed estimator. To achieve this goal, we follow the framework of [9] by assuming that the quantum state \( \rho \) is in some class \( \mathcal{R}(C, B, r) \) defined in (9). The notable difference of the present setting is that the precise knowledge of \( C, B \) and \( r \) is not required by our estimating procedure.

2.1. Adapted pattern functions

In order to reconstruct the entries of the density matrix from the noisy observations \((Y_\ell, \Phi_\ell)\) by a projection-type estimator on the pattern functions, we have to adapt the pattern functions as follows. From now on, we shall use the notation \( \gamma = \gamma(\eta) := \frac{1 - \eta}{4\eta} \). We denote by \( f_{\eta,j,k} \) the function which has the following Fourier transform:

\[
\tilde{f}_{\eta,j,k}(t) := \tilde{f}_{\eta,j,k}(t) e^{\gamma t^2},
\]

where \( \tilde{f}_{\eta,j,k} \) are the pattern functions defined in equation (4).

2.2. Estimation procedure

The estimation procedure we introduce in this section will depend on one tuning parameter \( N := N(n) \), the precise value of which will be given later. We define the set of indices \( J(N) \subset \mathbb{N}^2 \) by

\[
J(N) := \{(j, k) \in \mathbb{N}^2, 0 \leq j + k \leq N - 1\}.
\]

We first define an initial estimator \( \hat{\rho}_0 \) of \( \rho \) by setting

\[
\hat{\rho}_0^{\eta,j,k} := \frac{1}{n} \sum_{\ell=1}^n G_{j,k}(\frac{Y_\ell}{\sqrt{\eta}}, \Phi_\ell) \quad \forall (j, k) \in J(N),
\]

where \( (G_{j,k})_{j,k} \) are constructed using the pattern functions in (10) and

\[
G_{j,k}(x, \phi) := f_{j,k}^\eta(x) e^{-i(j-k)\phi}.
\]
Note that this procedure introduced by [9] estimates the matrix coefficients by replacing the theoretical by its empirical counterpart. To define our final procedure of estimation, let us introduce some notation. From now, we denote by \( \|\cdot\|_\infty \) the supremum norm for functions, i.e. for any \( f \),

\[
\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|.
\]

Let \( \varepsilon \in (0, 1) \) be a prescribed tolerance level. The final estimation procedure applies the soft-thresholding operator to the initial one:

\[
\hat{\rho}^{\eta}_{j,k} = \frac{\hat{\rho}^{\eta}_{j,k}}{|\hat{\rho}^{\eta}_{j,k}|} \left( |\hat{\rho}^{\eta}_{j,k}| - t_{j,k} \right)_+ ,
\]

with the convention \( 0/0 = 0 \), and where the thresholds are defined as

\[
t_{j,k} = 2 \frac{\| f^{\eta}_{j,k} \|_\infty}{\log(2N(N+1)\varepsilon^{-1})}. \tag{15}
\]

Thus, our estimator of the density matrix is given by

\[
\hat{\rho}^{\eta} = \left[ \hat{\rho}^{\eta}_{j,k} \right]_{j,k}.
\]

2.3. Main results

To characterize the behaviour of the estimator \( \hat{\rho}^{\eta} \), we measure the quality of estimation in an \( \ell_2 \)-norm. For any density matrix \( \nu = (\nu_{j,k})_{j,k \geq 0} \), we define the \( \ell_2 \)-norm of \( \nu \) as

\[
\|\nu\|_2 = \sqrt{\sum_{j,k \geq 0} |\nu_{j,k}|^2}.
\]

We first state a risk bound that holds with a large probability and will allow us to obtain the rates of convergence on the classes \( \mathcal{R}(C, B, r) \).

**Proposition 2.1.** With probability at least \( 1 - \varepsilon \), we have

\[
\|\hat{\rho}^{\eta} - \rho\|_2^2 \leq \inf_{I \subseteq \mathcal{J}(N)} \left\{ \frac{4}{n} \sum_{(j,k) \in I} t_{j,k}^2 + \sum_{(j,k) \notin I} |\rho_{j,k}|^2 \right\},
\]

where the set \( \mathcal{J}(N) \) is defined in (11).

The proof is given in appendix A. Note that this result holds true for any value of the tuning parameter \( N \). Choosing this parameter in a suitable manner leads to a rate of convergence that coincides with the one obtained in [9] for a nonadaptive procedure. This result is stated in the following theorem.

**Theorem 2.1.** Let us put \( r_0 \in (0, 2) \), \( B_0 > 0 \) and let us choose

\[
N = N(n) : = \left\lceil \frac{\log(n)}{2B_0} \right\rceil, \tag{16}
\]

where \( \lceil x \rceil \) denotes the integer part of \( x \) such that \( \lfloor x \rfloor \leq x < \lfloor x \rfloor + 1 \). Let us assume that \( \rho \in \mathcal{R}(C, B, r) \), for some unknown \( C \geq 1, B \geq B_0, r \in [r_0, 2] \). Then, there are constants \( C_1, C_2, C_3 > 0 \) such that with probability at least \( 1 - \varepsilon \), we have the following.

- For \( \eta = 1 \) and \( r \in [r_0, 2] \),

\[
\|\hat{\rho}^{\eta} - \rho\|_2^2 \leq C_1 n^{-1} \left( \log(n) \right)^{C_2} \log(\log(n) \varepsilon^{-1}).
\]
For \( \eta \in (\frac{1}{2}, 1) \) and \( r = 2 \),
\[
\| \tilde{\rho}^N - \rho \|^2_2 \leq C_2 n^{-\frac{\eta}{2}} (\log(n) + (\log(n))^{1/3} \log(\log(n) \varepsilon^{-1})).
\]

For \( \eta \in (\frac{1}{2}, 1) \) and \( r \in (r_0, 2) \),
\[
\| \tilde{\rho}^N - \rho \|^2_2 \leq C_3 e^{-2BM(n)^{r/2}} (\log(n)^{2-r/2} + \log(n)^{1/3} \log(\log(n) \varepsilon^{-1})),
\]
where \( M(n) \) satisfies \( 8y M(n) + 2B M(n)^{r/2} = \log(n) \). In particular, note that
\[
M(n) = \frac{1}{8y} \log(n) - \frac{2B}{(8y)^{1+r/2}} \log(n)^{r/2} + o(\log(n)^{r/2}).
\]

The proof is given in appendix B. Let us give some comments on this result highlighting its relations to previous work. First of all, note that the convergence rate is polynomial in the cases \((\eta, r) \in [2] \times [r_0, 2] \) and \((\eta, r) \in (1/2, 1) \times [2] \). Furthermore, the rate is parametric, up to a logarithmic factor, in the first case. It is slower in the second case, but becomes closer to the parametric rate when \( B \) is very large. The benefits of the adaptation are particularly striking in this case. Indeed, if, for example, the only available information is that \( B \geq 1/3 \) and \( \eta = 3/4 \), then the estimator proposed in [9] will converge at the rate \( n^{1/2} \log n \), even if the true state \( \rho \) belongs to the class \( R(C, B, 2) \) with a very large constant \( B \geq 1/3 \). Contrarily to this, our estimator will converge at the rate \( n^{-m} \log n \) which can be very close to the parametric rate \( n^{-1} \) if \( B \) is large.

One can also note that when \((\eta, r) \in (1/2, 1) \times (r_0, 2) \), the rate we obtain is slower than any power of \( n^{-1} \), but faster than any power of \( (\log(n))^{-1} \). We will say that these rates are intermediate. They coincide, up to a \( \log((\log(n))/\varepsilon) \) factor, with the rates obtained in [9, 34]. Another interesting feature of the previous result is that it provides a risk bound with a high probability, whereas existing results are all concerned with bounding the expected risk.

Interestingly, the same procedure achieves the nearly parametric rate in the case of pure state as well.

**Theorem 2.2.** Under the same choice for \( N \) in theorem 2.1,
\[
N = N(n) := \left\lfloor \frac{(\log(n))^{\frac{1}{2}}}{2B_0} \right\rfloor,
\]
if \( \rho \) is a pure state, i.e., if \( \rho_{j_0,j_0} = 1 \) for some \( j_0 \) and all the other \( \rho_{j,k} \)'s are 0, then we have, as soon as \( N > \max(j_0, 2) \), with probability at least \( 1 - \varepsilon \),
\[
\| \tilde{\rho}^N - \rho \|^2_2 < \frac{64}{nr_0} \left\| f_{j_0,j_0} \right\|_\infty^2 \log \left( \frac{2 \log(n)}{B_0 \varepsilon} \right).
\]

The proof is given in appendix C.

3. Experimental evaluation

3.1. Examples considered in the experiments

We present in table 1 the examples of pure quantum states, which can be created at this moment in laboratory and belong to the class \( R(C, B, r) \) with \( r = 2 \). Table 1 gives also their density matrix coefficients \( \rho_{j,k} \) and probability densities \( p_x(x|\phi) \).

Among the pure states we consider the vacuum state, which is the pure state of zero photons. Note that the vacuum state would provide a random variable of the Gaussian probability density \( p_x(x|\phi) \) via the ideal measurement of QHT (see section 1.1.3). That explains the Gaussian nature of the noise in the effective result of the QHT measurement.
We consider also the single photon state which is the pure state of one photon and the coherent-q0 state, which characterizes the laser pulse with the number of photons Poisson distributed with an average of $M$ photons. Note that the well-known Schrödinger cat state is described by a linear superposition of two coherent vectors (see e.g. [35]).

### 3.2. Pattern functions $f_{j,k}^n$

We evaluate the pattern functions $f_{j,k}^n$ numerically on a 1D regular grid of $Q = 4096$ points. We use expressions (4) and (10) to evaluate $f_{j,k}$ and $\tilde{f}_{j,k}$ on the 1D frequency grid of $Q$ discretized $t$ points. The adapted pattern functions $f_{j,k}^n$ are computed on the 1D spatial grid of $Q$ discretized $x$ points by applying to $\tilde{f}_{j,k}$ the inverse fast Fourier transform (FFT) in $O(Q \log(Q))$ operations. Some pattern and adapted functions are depicted in Figure 2.

### 3.3. Implementation of our procedure

The deconvolved estimator $\hat{\rho}_{j,k}^n$ defined in (12) is computed by evaluating

$$G_{j,k}(x, \phi) = f_{j,k}^n(x) \ e^{-i(j-k)\phi}$$

at a point $x$ using a cubic spline interpolation of the values of $f_{j,k}^n$ on the discrete grid of $Q$ points.

In the following section, we assess the performance of the threshold estimator $\hat{\rho}_{j,k}^n$. We perform this evaluation by creating noisy samples $Y_t$ as defined in (6). The initial samples $X_t$ are drawn from the distribution $p_x(x|\phi)$ (see Table 1) using the rejection method. The value of $N = N(n)$ is set following (16). We use $r_0 = 2$ and $B_0 = 1/2$ for all the numerical experiments. A toolbox that implements this procedure and reproduces all the figures of this paper is available online\footnote{www.ceremade.dauphine.fr/~peyre/codes/}.

---

**Table 1. Examples of quantum states.**

| State Type          | Density | Description |
|---------------------|---------|-------------|
| Vacuum state        | $\rho_{0,0} = 1$ rest zero. |
| $p_{\phi}(x|\phi) = e^{-x^2}/\sqrt{\pi}$. |
| Single photon state | $\rho_{1,1} = 1$ rest zero. |
| $p_{\phi}(x|\phi) = 2xe^{-x^2}/\sqrt{\pi}$. |
| Coherent-q0 state   | $q_0 \in \mathbb{R}$ |
| $\rho_{j,k} = e^{-iq_0^2}(q_0/\sqrt{2})^{j+k}/\sqrt{j!k!}$. |
| $p_{\phi}(x|\phi) = \exp(-(x - q_0 \cos(\phi))^2)/\sqrt{\pi}$. |
| Thermal state       | $\beta > 0$ |
| $\rho_{j,k} = \delta_k(1 - e^{-\beta}) e^{-\beta^2}$. |
| $p_{\phi}(x|\phi) = \sqrt{\tanh(\beta/2)/\pi} \exp(-x^2 \tanh(\beta/2))$. |
| Schrödinger cat q0   | $q_0 > 0$ |
| $\rho_{j,k} = 2(q_0/\sqrt{2})^{j+k}/(\sqrt{\pi}(\exp(q_0^2/2) + \exp(-q_0^2/2)))$, for $j$ and $k$ even, rest zero, |
| $p_{\phi}(x|\phi) = (\exp(-(x - q_0 \cos(\phi))^2) + \exp(-(x + q_0 \cos(\phi))^2)) + 2 \cos(2q_0 x \sin(\phi)) \exp(-x^2 - q_0^2 \cos^2(\phi))/2\sqrt{\pi}(1 + \exp(-q_0^2))$. |
matrices \( \tilde{\rho}^\eta \) of some quantum state for several values of \( n \). In practice, the experimental number is typically of order \( 10^8 \).

### 3.4. Studies of the performance of our estimation procedure

We estimate numerically the (relative) root mean square error (RMSE)

\[
\text{RMSE}(n) = \frac{\| \tilde{\rho}^\eta - \rho \|_2}{\| \rho \|_2}
\]

of our soft-thresholding estimator. More precisely, figure 4 shows the evolution with \( n \) of the expected value of the RMSE. This expected value is evaluated by an empirical mean with Monte Carlo simulation using 50 replications for each value of \( n \). To evaluate the deviation with respect to this mean, we also display the confidence interval at \( \pm 3 \) times the standard deviation of the RMSE.

The threshold values \( t_{j,k} \) that are used in (14) to define our estimator are somewhat conservative. In practice, smaller values offer better decay of the RMSE. Figure 4 displays in dashed red (resp. dashed green) the decay of the RMSE obtained using thresholds \( 0.8t_{j,k} \) (resp. \( 0.5t_{j,k} \)). We found in these three examples and for \( \eta = 0.9 \) that using \( 0.5t_{j,k} \) gives consistently the lowest RMSE among other choices of thresholds proportional to the \( t_{j,k} \) values.

We found numerically that the decay of the RMSE with \( n \) almost perfectly fits a power-law, which (up to logarithmic factor) is in accordance with the upper bounds of corollary 2.1. Following this corollary in the setting \( \eta \in \left( \frac{1}{2}, 1 \right) \) and \( r = 2 \), we fit a power law of the form

\[
\mathbb{E}(\text{RMSE}(n)) \approx n^{-\frac{8}{3\eta r + 4\sigma}}.
\]

We perform a linear regression in a log–log domain to estimate \( \tilde{B} \). Table 2 reports the estimated value of \( \tilde{B} \) we found using this procedure.
Figure 3. First row: $\rho$. Following rows: estimated $\tilde{\rho}$ for $B_0 = 0.5$, $\eta = 0.9$, $\epsilon = 1$ and $n$ respectively equal to $n = 10^5$ (row 2), $n = 10^6$ (row 3), $n = 10^8$ (row 4). (a) Coherent $q_0 = 3$ (b) Schrödinger cat $q_0 = 3$ (c) thermal $\beta = 1/4$.

Table 2. Estimated values of $\tilde{B}$ when using $\eta = 0.9$, $\epsilon = 1$ and $N = 30$.

| Coherent $q_0 = 3$ | Schrödinger cat $q_0 = 3$ | Thermal $\beta = 1/10$ | Thermal $\beta = 1/4$ |
|--------------------|--------------------------|-------------------------|-----------------------|
| $\tilde{B} \approx 0.174$ | $\tilde{B} \approx 0.227$ | $\tilde{B} \approx 0.037$ | $\tilde{B} \approx 0.082$ |
Figure 4. Blue curve: evolution of $E(RMSE(n))$ as a function of $n$ for $\eta = 0.9$, $\varepsilon = 1$ and $N = 30$. The blue shaded area represent the confidence interval at ±3 times the standard deviation of $RMSE(n)$. Red (resp. green) curve: evolution of $E(RMSE(n))$ obtained when replacing the threshold $t_{jk}$ by $0.8t_{jk}$ (resp. $0.5t_{jk}$) in the estimator in (14). (a) Coherent $q_0 = 3$ (b) Schrödinger cat $q_0 = 3$ (c) thermal $\beta = 1/10$ (d) thermal $\beta = 1/4$.

Acknowledgments

We would like to thank Arnak Dalalyan for his relevant remarks and for his rereading of the present paper. The research of Katia Meziani is partly supported by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

Appendix A. Proof of proposition 2.1

The proofs follow the main lines of [36, 37]. First, we need a set of preliminary lemmas.

A.1. Some preliminary results

First, we recall Hoeffding’s inequality for bounded random variables.
Lemma 1. Let us assume that \( Z_1, \ldots, Z_n \) are independent real-valued random variables with \( a_i \leq |Z_i| \leq b_i \). Then, for any \( \lambda > 0 \),
\[
P\left( \left| \frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq \lambda \right) \leq 2 \exp \left( -\frac{2\lambda^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right).
\]

As a consequence, we have the following inequality for complex random variables.

Lemma 2. Let us assume that \( Z_1, \ldots, Z_n \) are independent complex-valued random variables with \( |Z_i| \leq c \). Then, for any \( t > 0 \),
\[
P\left( \left| \frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq t \right) \leq 4 \exp \left( -\frac{nt^2}{4c^2} \right).
\]

Proof. We have
\[
P\left( \left| \frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq \lambda \right) \leq P\left( \left| \text{Re} \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq \frac{\lambda}{\sqrt{2}} \right) + P\left( \left| \text{Im} \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq \frac{\lambda}{\sqrt{2}} \right)
\]
\[
\leq P\left( \left| \sum_{i=1}^{n} \text{Re}(Z_i) - E(\text{Re}(Z_i)) \right| \geq \frac{\lambda}{\sqrt{2}} \right) + P\left( \left| \sum_{i=1}^{n} \text{Im}(Z_i) - E(\text{Im}(Z_i)) \right| \geq \frac{\lambda}{\sqrt{2}} \right)
\]

Now, we apply Hoeffding’s inequality to the random variables \( \text{Re}(Z_i) \) which satisfy \(-c \leq \text{Re}(Z_i) \leq c\). So we have
\[
P\left( \left| \sum_{i=1}^{n} \text{Re}(Z_i) - E(\text{Re}(Z_i)) \right| \geq \frac{\lambda}{\sqrt{2}} \right) \leq 2 \exp \left( -\frac{2\left( \frac{\lambda}{\sqrt{2}} \right)^2}{\sum_{i=1}^{n} (2c)^2} \right) = 2 \exp \left( -\frac{\lambda^2}{4nc^2} \right).
\]

We have exactly the same result for \( \text{Im}(Z_i) \) so finally
\[
P\left( \left| \sum_{i=1}^{n} (Z_i - E(Z_i)) \right| \geq \lambda \right) \leq 4 \exp \left( -\frac{\lambda^2}{4nc^2} \right).
\]

Put \( t = \lambda/n \) to end the proof. \( \Box \)

Lemma 3. For some fixed \( \varepsilon \in (0, 1) \), let us define the set
\[
\Omega_{\varepsilon} := \left\{ \forall (j, k) \in J(N), |\hat{\rho}_{j,k} - \rho_{j,k}| \leq t_{j,k} \right\},
\]
where the \( (t_{j,k})_{j,k} \) are defined in (15) and the set \( J(N) \) is defined in (11). Then, we have
\[
P(\Omega_{\varepsilon}) \geq 1 - \varepsilon.
\]

Proof. Lemma 3 is proved by using Hoeffding’s inequality. To this aim, we have to first note that
\[
E_p[\hat{\rho}_{j,k}^\theta] = \rho_{j,k}.
\]
Indeed, by using (13), (8), (10) and (3), we have

\[
E_p[\tilde{\rho}_{j,k}^n] = E_p\left[G_{j,k}\left(\frac{Y}{\sqrt{n}}\right)\right] = E_p\left[f_{j,k}^n\left(\frac{Y}{\sqrt{n}}\right)e^{-i(j-k)\phi}\right]
\]

\[
= \frac{1}{\pi} \int_0^\pi e^{-i(j-k)\phi} \int f_{j,k}(y) \sqrt{n}p_y^n(y)\sqrt{n}\phi dy d\phi
\]

\[
= \frac{1}{\pi} \int_0^\pi e^{-i(j-k)\phi} \frac{1}{2\pi} \int \tilde{f}_{j,k}(t) \left[\sqrt{n}p_y^n(\sqrt{n}\phi)\right](t) dt d\phi
\]

\[
= \frac{1}{\pi} \int_0^\pi e^{-i(j-k)\phi} \frac{1}{2\pi} \int \tilde{f}_{j,k}(t) e^{\nu^2} F_1[p_y(\cdot|\phi)](t)N^n(t) dt d\phi
\]

Moreover, we easily obtain from the definition of \(G_{j,k}\) in (13) that for all \(\ell = 1, \ldots, N\) and \(\forall(j,k) \in J(N)\)

\[
\left|G_{j,k}\left(\frac{Y}{\sqrt{n}}, \Phi\right)\right| \leq \|f_{j,k}\|_{\infty}.
\]

Then, for

\[
t_{j,k} = 2\left\|f_{j,k}\right\|_{\infty} \log \left(\frac{2^{N+1}}{\epsilon}\right)
\]

and according to lemma 2

\[
P(\left|\tilde{\rho}_{j,k}^n - \rho_{j,k}\right| \geq t_{j,k}) \leq 4 \exp \left[-\frac{n^2_{j,k}^2}{4\|f_{j,k}\|_{\infty}^2}\right] = \frac{2\epsilon}{N(N+1)}.
\]

By the classical union bound argument:

\[
P(\Omega_{\epsilon}^c) \leq \sum_{(j,k) \in J(N)} P(\left|\tilde{\rho}_{j,k}^n - \rho_{j,k}\right| \geq t_{j,k}) \leq \sum_{(j,k) \in J(N)} \frac{2\epsilon}{N(N+1)} \leq \epsilon.
\]

**Lemma 4.** For some fixed \(\epsilon \in (0, 1)\) and \(\forall(j,k) \in J(N)\), with \(J(N)\) defined in (11), we define the set

\[
R_{j,k}^\epsilon := \{\nu \text{ density matrix}, |\nu_{j,k} - \tilde{\rho}_{j,k}^n| \leq t_{j,k}\},
\]

where the \((t_{j,k})_{j,k}\) are defined in (15). Then, on the event \(\Omega_{\epsilon}\) defined in lemma 3 and \(\forall(j,k) \in J(N)\)

\((i)\) \(\rho \in R_{j,k}^\epsilon\).

\((ii)\) \(R_{j,k}^\epsilon\) is a closed and convex set.

\((iii)\) For \(\Pi_{j,k}^\epsilon\) the orthogonal projection onto \(R_{j,k}^\epsilon\) and for any density matrix \(\nu\),

\[
\|\rho - \Pi_{j,k}^\epsilon(\nu)\|_2^2 \leq \|\rho - \nu\|_2^2.
\]

**Proof.** The first point is just a consequence of lemma 3. The second point comes from the definition of \(R_{j,k}^\epsilon\).

Moreover, it is well known that for any closed and convex set \(C\), if \(\Pi_C\) is the orthogonal projection on \(C\), then the following property holds:

\[
\forall x \in C, \forall y, \quad \|\Pi_C(y) - x\|_2 \leq \|y - x\|_2.
\]

This concludes the proof of the third point.
Lemma 5. For \( \varepsilon \in (0, 1) \), any fixed \((j, k) \in J(N)\), with \(J(N)\) defined in (11), and any density matrix \(\nu\), we denote by \(\nu'\) the projection of \(\nu\) into \(R^\varepsilon_{j,k}\)

\[
\nu' := \Pi^\varepsilon_{j,k}(\nu) = [\nu'_{\ell,m}]_{\ell,m},
\]

with \(R^\varepsilon_{j,k}\) defined in lemma 4. Then, the entries \(\nu'_{\ell,m}\) of \(\nu'\) are equal to

\[
\nu'_{\ell,m} = \begin{cases} 
    v_{j,k} + \frac{\hat{\rho}^\eta_{j,k} - v_{j,k}}{\|\hat{\rho}^\eta_{j,k} - v_{j,k}\|_1} \left((\hat{\rho}^\eta_{j,k} - v_{j,k}) - t_{j,k}\right)_+ & \text{if } (\ell, m) = (j, k), \\
    v_{\ell,m} & \text{otherwise},
\end{cases}
\]

with the convention \(0/0 = 0\).

Proof. The projection \(\nu'\) of \(\nu\) into \(R^\varepsilon_{j,k}\) satisfies

\[
\nu' = \arg \min_{x \in R^\varepsilon_{j,k}} \|\nu - x\|_2^2 = \arg \min_{x \in R^\varepsilon_{j,k}} \sum_{\ell,m=0}^{\infty} |x_{\ell,m} - \nu_{\ell,m}|^2.
\]

As the constraint \(x \in R^\varepsilon_{j,k}\) is only a constraint on \(x_{j,k}\), it is clear that for \((\ell, m) \neq (j, k)\) the minimum is reached for \(x_{j,k} = v_{j,k}\). Finally,

\[
v'_{j,k} = \arg \min_{x_{j,k}:|\hat{\rho}^\eta_{j,k} - v_{j,k}| \leq |t_{j,k}|} |v_{j,k} - x_{j,k}|^2.
\]

The solution \(v'_{j,k}\) is obvious:

\[
v'_{j,k} = \begin{cases} 
    v_{j,k} + \frac{\hat{\rho}^\eta_{j,k} - v_{j,k}}{\|\hat{\rho}^\eta_{j,k} - v_{j,k}\|_1} \left((\hat{\rho}^\eta_{j,k} - v_{j,k}) - t_{j,k}\right)_+ & \text{if } |v_{j,k} - \hat{\rho}^\eta_{j,k}| \leq t_{j,k}, \\
    v_{j,k} + \hat{\rho}^\eta_{j,k} - v_{j,k} \left(\|\hat{\rho}^\eta_{j,k} - v_{j,k}\|_1 - t_{j,k}\right)_+ & \text{otherwise}
\end{cases}
\]

This ends the proof. \(\square\)
**Definition 1.** For \( m > 0 \) being an integer, let
\[
I := \{(j_1, k_1), \ldots, (j_m, k_m)\} \subseteq J(N)
\]
be a set of indices, where \( J(N) \) is the set defined in (11). That \( \forall \ell \neq i, (j_\ell, k_\ell) \neq (j_i, k_i) \). For \( \varepsilon \in (0, 1) \) and for any density matrix \( \nu \), we denote by \( \Pi_\varepsilon(\nu) \) the successive projections of \( \nu \) into spaces \( (R^e_{j_i, k_i})_{j_i, k_i} \), i.e.
\[
\Pi_\varepsilon(\nu) := \Pi_{j_m, k_m}^e \Pi_{j_{m-1}, k_{m-1}}^e \cdots \Pi_{j_1, k_1}^e \Pi_{j_0, k_0}(\nu).
\]

Note that for any set of indices \( I \) and from lemma 5, the application of the successive projections \( \Pi_\varepsilon \) to a density matrix \( \nu \) does not depend on the order of the successive projections.

**Lemma 6.** For \( \varepsilon \in (0, 1) \), for \( J(N) \) defined in (11) and for \( \tilde{\nu}^0 \) defined in (14), we have
\[
\tilde{\nu}^0 = \Pi_\varepsilon(J(N))(0),
\]
where \( 0 \) is the zero-infinite matrix.

**Proof.** This is obvious from the definition of \( \tilde{\nu}^0 \) and from lemma 5 applied to \( \nu = 0 \). \( \square \)

**A.2. Proof of proposition 2.1**

**Proof.** For \( J(N) \) the set of indices defined in (11), let \( I \) be a subset of \( J(N) \), \( I \subseteq J(N) \). For a fixed \( \varepsilon \in (0, 1) \), we have by lemma 6 and by successive applications of the inequality (A.1) to all pair of indices \( (j, k) \neq I \)
\[
\|\tilde{\nu}^0 - \rho\|^2 \leq \|\Pi_\varepsilon(\nu) - \rho\|^2 \leq \|\Pi_\varepsilon(\nu) - \rho\|^2_2.
\]
Moreover, from lemma 5 applied to \( \nu = 0 \), we obtain
\[
(\Pi_\varepsilon(\nu))_{j,k} = \begin{cases} \rho^{0}_{j,k} - t_{j,k}, & \text{if } (j, k) \in I, \\ 0, & \text{otherwise}. \end{cases}
\]
Therefore, from (A.2) we obtain
\[
\|\tilde{\nu}^0 - \rho\|^2 \leq \sum_{j,k=0}^{\infty} |\rho_{j,k} - (\Pi_\varepsilon(\nu))_{j,k}|^2
\]
\[
= \sum_{(j,k) \neq I} \left| \rho_{j,k} - \frac{\tilde{\nu}^0_{j,k}}{|\tilde{\nu}^0_{j,k}|}(|\tilde{\nu}^0_{j,k}| - t_{j,k}) \right|^2 + \sum_{(j,k) \neq I} |t_{j,k}|^2
\]
\[
:= \sum_{(j,k) \neq I} |A_{j,k}|^2, \quad \sum_{(j,k) \neq I} |A_{j,k}|^2, \quad (A.3)
\]
where
\[
A_{j,k} = \rho_{j,k} - \frac{\tilde{\nu}^0_{j,k}}{|\tilde{\nu}^0_{j,k}|}(|\tilde{\nu}^0_{j,k}| - t_{j,k})
\]
\[
= \begin{cases} \rho_{j,k}, & \text{if } |\tilde{\nu}^0_{j,k}| \leq t_{j,k}, \\ \rho_{j,k} - \frac{\tilde{\nu}^0_{j,k}}{|\tilde{\nu}^0_{j,k}|}(|\tilde{\nu}^0_{j,k}| - t_{j,k}), & \text{otherwise}. \end{cases}
\]
Moreover
\[
|A_{j,k}| \leq \begin{cases} \left| \tilde{\nu}^0_{j,k} - \rho_{j,k} \right| + \left| \tilde{\nu}^0_{j,k} \right|, & \text{if } |\tilde{\nu}^0_{j,k}| \leq t_{j,k}, \\ \left| \tilde{\nu}^0_{j,k} - \rho_{j,k} \right| + |t_{j,k}|, & \text{otherwise.} \end{cases}
\]
For any \((j, k) \in I\) and on the event \(\Omega_e\) defined in lemma 3, it holds that
\[
|\tilde{\rho}_{j,k}^\eta - \rho_{j,k}| \leq t_{j,k}.
\]
Therefore, from (A.3)
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq \sum_{(j,k) \in I} (2t_{j,k})^2 + \sum_{(j,k) \notin I} |\rho_{j,k}|^2.
\]
We conclude the proof by taking the infimum over the set \(I \subseteq J(N)\).

\(\square\)

Appendix B. Proof of theorem 2.1

\textbf{Proof.} For \(r_0 \in (0, 2), B_0 > 0\) and \(N\) as in (16), let \(M\) be an integer s.t. \(M < N\). We define the set
\[
J(M) := \{(j, k) \in \mathbb{N}^2, 0 \leq j + k \leq M\}.
\]
Then, for \(\varepsilon \in (0, 1)\) and by applying proposition 2.1 to \(I = J(M)\), with probability larger than \(1 - \varepsilon\), we obtain
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq \inf_{0 \leq M \leq N - 1} \left\{ \frac{4}{n} \sum_{0 \leq j + k \leq M} t_{j,k}^2 + \sum_{j+k>M} |\rho_{j,k}|^2 \right\}.
\]  
(B.1)

(a) For \(\eta = 1\) and \(r \in [r_0, 2]\), as \(f_{j,k}^\eta = f_{j,k}\) for \(\eta = 1\), we have by pluging (D.1) and (D.3) into (B.1)
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq \inf_{0 \leq M \leq N - 1} \left\{ \frac{16}{n} \sum_{0 \leq j+k \leq M} \| f_{j,k} \|_\infty^2 \log (2N(N+1)/\varepsilon) + \sum_{j+k>M} |\rho_{j,k}|^2 \right\}.
\]  
(B.2)

for some constant \(c_1 > 0\).

For \(N\) such in (16) and by taking \(M = (\log(n)/2B)^{2/r} < N\), it leads to
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq C_1 \log((\log(n)/\varepsilon))(\log(n))^{\frac{M}{n}} n^{-1}
\]  
for some constant \(C_1 > 0\).

(b) For \(\eta \in (1/2, 1)\) and \(r = 2\), next, we deal with the case \(1 > \eta > 1/2\). We plug (D.1) and (D.2) into (B.1) to obtain in the case \(r = 2\)
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq \inf_{0 \leq M < N} \left\{ \frac{C_2}{n} \log((\log(n)/\varepsilon)) M^4 \varepsilon^{4/3} + CM \varepsilon^{-2BM} \right\},
\]  
for some constant \(c_2 > 0\).

By taking \(M = M(n)\) s.t.
\[
M = \frac{\log(n)}{2(4\gamma + B)},
\]
we obtain
\[
\| \tilde{\rho}^\eta - \rho \|_2^2 \leq C_2 n^{-\frac{\eta}{1+\eta}} \log((\log(n))/\varepsilon)(\log(n))^{1/2} + \log(n)
\]  
for some constant \(C_2 > 0\).
(c) For \( \eta \in (1/2, 1) \) and \( r \in (r_0, 2) \).

Finally, in the case \( \eta \in (1/2, 1) \) and \( r \in (r_0, 2) \) and by plugging (D.1) and (D.2) into (B.1) we obtain

\[
\| \hat{\rho}^\eta - \rho \|_2^2 \leq \inf_{0 \leq M < N} \left\{ \frac{C_3}{n} \log(N/\varepsilon)M \hat{\gamma}M + CM^2 \gamma e^{-2BM^2} \right\},
\]

for some constant \( C_3 > 0 \).

For \( M \) a solution of the equation \( 8\gamma M + 2BM^2 = \log(n) \) and for

\[
M(n) = \frac{1}{8\gamma} \log(n) - \frac{2B}{(8\gamma)^{1+r/2}} \log(n)^{r/2} + o(\log(n)^{r/2})
\]

in particular, we obtain

\[
\| \hat{\rho}^\eta - \rho \|_2^2 \leq C_3 \exp^{-2BM^2} \left( \log(n)^{2-r/2} + \log(n)^{1/3} \log(N/\varepsilon) \right),
\]

for some constant \( C_3 > 0 \).

\[\square\]

Appendix C. Proof of theorem 2.2

Proof. We apply theorem 2.1 for \( I = \{ (j_0, j_0) \} \). We obtain, with probability larger than \( 1 - \varepsilon \),

\[
\| \hat{\rho}^\eta - \rho \|_2^2 \leq \frac{16}{n} \sum_{(j,k) = (j_0,j_0)} \| f_{j,k} \|_\infty^2 \log(2N(N + 1)/\varepsilon) + \sum_{(j,k) \neq (j_0,j_0)} \rho_{j,k}^2
\]

\[
= \frac{16}{n} \| f_{j_0,j_0} \|_\infty^2 \log(2N(N + 1)/\varepsilon) + 0.
\]

For \( n \) large enough, \( N = N(n) \geq 2 \). Then, \( (N + 1) < 2N \) and

\[
\| \hat{\rho}^\eta - \rho \|_2^2 \leq \frac{16}{n} \| f_{j_0,j_0} \|_\infty^2 \log(4N^2/\varepsilon)
\]

\[
= \frac{16}{n} \| f_{j_0,j_0} \|_\infty^2 [2 \log(N) + \log(4/\varepsilon)]
\]

\[
\leq \frac{16}{n} \| f_{j_0,j_0} \|_\infty^2 \left[ \frac{4}{r_0} \log \left( \frac{\log(n)}{2B_0} \right) + \log(4/\varepsilon) \right]
\]

where we replaced \( N \) by its definition. As \( r_0 < 2, 4/r_0 > 1 \) and we have the following rough upper bound:

\[
\| \hat{\rho}^\eta - \rho \|_2^2 \leq \frac{16}{n} \| f_{j_0,j_0} \|_\infty^2 \left[ \frac{4}{r_0} \log \left( \frac{\log(n)}{2B_0} \right) + \frac{4}{r_0} \log(4/\varepsilon) \right]
\]

\[
= \frac{64}{nr_0} \| f_{j_0,j_0} \|_\infty^2 \log \left( \frac{2 \log(n)}{B_0 \varepsilon} \right).
\]

\[\square\]

Appendix D. Technical lemmas

Useful lemmas are the following.

Lemma 7. For \( \rho \in \mathcal{R}(C, B, r) \), the set defined in (9), there exists a \( M_0 \) s.t. \( \forall M \geq M_0 \) implies

\[
\sum_{|j+k|M} |\rho_{j,k}|^2 \leq CM^2 \gamma e^{-2BM^2},
\]

(D.1)

where \( C = \frac{2C}{B} \).
Proof. For $\rho \in \mathcal{R}(C, B, r)$, we have by the definition of the class $\mathcal{R}(C, B, r)$ and by lemma 3 in [9]

$$\sum_{j+k>M} |\rho_{j,k}|^2 \leq C^2 \sum_{j+k>M} \exp(-2B(j+k)/r) \leq \frac{2C^2}{Br} M^{2r^2} e^{-2BM^2}.$$ 

□

Lemma 8. For $\eta \in (1/2, 1)$, there exists a positive constant $C_{\infty}^\eta > 0$ s.t.

$$\sum_{0 \leq j+k \leq M} \|f_{j,k}^\eta\|_\infty^2 \leq C_{\infty}^\eta M^{2\gamma} e^{8\gamma M},$$

where $\gamma = (1 - \eta)/(4\eta)$ and the $(f_{j,k}^\eta)_{j,k}$ are the adapted pattern functions defined in expression (10).

There exists a positive constant $C_{\infty} > 0$ s.t.

$$\sum_{0 \leq j+k \leq M} \|f_{j,k}\|_\infty^2 \leq C_{\infty} M^4,$$

where the $(f_{j,k})_{j,k}$ are the pattern functions defined in expression (4).

Proof. For the proof of this lemma, we refer to lemmas 4 and 5 in [9]. □

References
[1] Helstrom C W 1976 Quantum Detection and Estimation Theory (New York: Academic)
[2] Holevo A S 1982 Probabilistic and Statistical Aspects of Quantum Theory (Amsterdam: North-Holland) ISBN: 04444863338
[3] Barndorff-Nielsen O E, Gill R and Jupp P E 2003 On quantum statistical inference (with discussion) J. R. Stat. Soc. B 65 775–816
[4] Leonhardt U 1997 Measuring the Quantum State of Light (Cambridge: Cambridge University Press) ISBN 10: 0521497302
[5] Wigner E 1932 On the quantum correction for thermodynamic equations Phys. Rev. 40 749–59
[6] Richter T 2000 Realistic pattern functions for optical homodyne tomography and determination of specific expectation values Phys. Rev. A 61 063819
[7] Smulley D T, Beck M, Raymer M G and Faridani A 1993 Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum Phys. Rev. Lett. 70 1244–7
[8] Vogel K and Risken H 1989 Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase Phys. Rev. A 40 2847–9
[9] Aubry J-M, Butucea C and Meziani K 2009 State estimation in quantum homodyne tomography with noisy data Inverse Problems 25 015003
[10] Butucea C, Guţă M and Artiles L 2007 Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data Ann. Stat. 35 465–94
[11] D’Ariano G M, Macchiavello C and Paris M G A 1994 Detection of the density matrix through optical homodyne tomography without filtered back projection Phys. Rev. A 50 4298–302
[12] D’Ariano G M, Leonhardt U and Paul H 1995 Homodyne detection of the density matrix of the radiation field Phys. Rev. A 52 1801–4
[13] Leonhardt U, Paul H and D’Ariano G M 1995 Tomographic reconstruction of the density matrix via pattern functions Phys. Rev. A 52 4899–907
[14] Artiles L, Gill R and Guţă M 2005 An invitation to quantum tomography J. R. Stat. Soc. B 67 109–34
[15] Banaszek K, D’Ariano G M, Paris M G A and Sacchi M F 2000 Maximum-likelihood estimation of the density matrix Phys. Rev. A 61 R10304
[16] D’Ariano G M, Maccone L and Sacchi M F 2007 Homodyne tomography and the reconstruction of quantum states of light Quantum Information with Continuous Variables of Atoms and Light ed N J Cerf, G Leuchs and E S Polzik (London: Imperial College Press) p 141
[17] Guţă M I 2007 Maximum likelihood estimation of the density matrix through quantum tomography, unpublished
[18] D’Ariano G M and Paris M G A 1999 Adaptive quantum tomography Phys. Rev. A 60 518
[19] D’Ariano G M 1995 Tomographic measurement of the density matrix of the radiation field Quantum Semiclass. Opt. 7 693–704
[20] D’Ariano G M 2002 Tomographic methods for universal estimation in quantum optics International School of Physics Enrico Fermi vol 148 (IOS Press) p 385
[21] Richter T 1996 Pattern functions used in tomographic reconstruction of photon statistics revisited Phys. Lett. A 211 327–30
[22] Meziani K 2008 Nonparametric goodness-of-fit testing in quantum homodyne tomography with noisy data Electron. J. Stat. 2 1195–223
[23] Gută M and Artiles I 2007 Minimax estimation of the Wigner in quantum homodyne tomography with ideal detectors Math. Methods Stat. 16 1–15
[24] Meziani K 2007 Nonparametric estimation of the purity of a quantum state in quantum homodyne tomography with noisy data Math. Methods Stat. 16 1–15
[25] Donoho D and Johnstone I 1994 Ideal spatial adaptation via wavelet shrinkage Biometrika 81 425–55
[26] Donoho D and Johnstone I 1995 Adapting to unknown smoothness via wavelet shrinkage J. Am. Stat. Assoc. 90 1200–24
[27] Donoho D, Johnstone I, Kerkyacharian G and Picard D 1995 Wavelet shrinkage: asymptotia? J. R. Stat. Soc. B 57 301–69
[28] Härdle W, Kerkyacharian G, Picard D and Tsybakov A 1998 Wavelets, Approximation, and Statistical Applications (Lecture Notes in Statistics vol 129) (Berlin: Springer) (Math SciNet: MR1618204)
[29] Donoho D 1995 Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition Appl. Comput. Harmon. Anal. 2 101–26
[30] Johnstone I 1999 Wavelet shrinkage for correlated data and inverse problems: adaptivity results Stat. Sin. 9 51–93
[31] Kolaczyk E 1996 A wavelet shrinkage approach to tomographic image reconstruction J. Am. Stat. Assoc. 91 1079–90
[32] Cavalier L, Golubev G, Lepski O and Tsybakov A 2003 Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems Theory Probab. Appl. 48 426–46
[33] Cavalier L 2011 Inverse problems in statistics Inverse Problems and High-Dimensional Estimation, Stats in the Château Summer School, 31 August–4 September 2009 (Lecture Notes in Statistics vol 203) ed P Alquier, E Gautier and G Stoltz (Berlin: Springer) pp 3–96
[34] Meziani K 2008 Estimation non-paramétrique en tomographie quantique homodyne PhD Thesis Université Paris 7 (in English)
[35] Ourjoumtsev A, Tualle-Brouri R, Laurat J and Grangier P 2006 Generating optical Schrödinger kittens for quantum information processing Science 312 83–6
[36] Alquier P 2008 Iterative feature selection in least square regression estimation Ann. Inst. H. Poincaré B 44 47–88
[37] Alquier P 2008 Lasso, iterative feature selection and the correlation selector: Oracle inequalities and numerical performances Electron. J. Stat. 2 1129–52