

PAPER • OPEN ACCESS

Spectral thresholding quantum tomography for low rank states

To cite this article: Cristina Butucea *et al* 2015 *New J. Phys.* **17** 113050

View the [article online](#) for updates and enhancements.

Related content

- [Rank-based model selection for multiple ions quantum tomography](#)
Mdlin Gu, Theodore Kypraios and Ian Dryden
- [Statistically efficient tomography of low rank states with incomplete measurements](#)
Anirudh Acharya, Theodore Kypraios and Mdlin Gu
- [Statistical analysis of compressive low rank tomography with random measurements](#)
Anirudh Acharya and Mdlin Gu

Recent citations

- [Pseudo-Bayesian quantum tomography with rank-adaptation](#)
The Tien Mai and Pierre Alquier
- [Corrigendum: Spectral thresholding quantum tomography for low rank states \(2015 *New J. Phys.* **17** 113050\)](#)
Cristina Butucea *et al*

Spectral thresholding quantum tomography for low rank states

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 New J. Phys. 17 113050

(<http://iopscience.iop.org/1367-2630/17/11/113050>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

This content was downloaded by: kimgraves

IP Address: 89.202.245.164

This content was downloaded on 03/06/2016 at 15:02

Please note that [terms and conditions apply](#).



CORRIGENDUM

Corrigendum: Spectral thresholding quantum tomography for low rank states (2015 *New J. Phys.* **17** 113050)

OPEN ACCESS

PUBLISHED
3 June 2016Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Cristina Butucea¹, Mădălin Guță² and Theodore Kypraios²¹ Université Paris-Est Marne-la-Vallée, LAMA(UMR 8050), UPEMLV F-77454, Marne-la-Vallée, France² University of Nottingham, School of Mathematical Sciences, University Park, Nottingham NG7 2RD, UK

In this corrigendum to the paper Butucea *et al* (2015 *New J. Phys.* **17** 113050) we point out an error in one of the theoretical results describing the upper bound to the operator norm error of the least squares estimator. We provide a corrected version of the upper bound with a new convergence rate, and discuss the implications for other results which rely on the above upper bound.

Proposition 1 as stated in the paper is incorrect, in particular the dependence of the upper bound $\nu(\epsilon)^2$ on the number of atoms k is not valid. The error lies in the evaluation of the upper bound W of the variance term in the concentration bound. Below we provide a new version of proposition 1 with a corrected rate $\nu_c(\epsilon)^2$ replacing the rate $\nu(\epsilon)^2$ stated in the paper. Ignoring the logarithmic factors, the new upper bound scales as $3^k/N$ compared to erroneous rate $2^k/N$, where k is the number of atoms and $N = n3^k$ is the total number of measurements. We note that although the corrected bound is weaker than the one claimed in the paper, it is still an improvement compared to the previously known bound [2] which scaled as $4^k/N$.

We will now discuss the implication of the correction to subsequent results in the paper. Proposition 2, theorem 1, corollary 1, and theorem 2 establish error rates for estimators obtained by normalising, penalising or thresholding the least square estimator. The proofs of these results use the operator norm error rate $\nu(\epsilon)^2$ as a generic expression, and are therefore not affected by its concrete dependence on the number of atoms k . Therefore proposition 2, theorem 1, corollary 1, and theorem 2 hold true when the operator norm rate is taken to have expression $\nu_c(\epsilon)^2$ in proposition 1 below. In particular, the upper bounds on the Frobenius square norm error in corollary 1 and theorem 2, will scale as $r \cdot \nu_c(\epsilon)^2 = r3^k/N$ rather than $rd/N = r2^k/N$. The remaining results including the lower bound in theorem 3 and the simulation results are independent of proposition 1 and do not require any correction.

Proposition 1. Let $\hat{\rho}_n^{(ls)}$ be the linear estimator of ρ . Then, for any $\epsilon > 0$, the following operator norm inequality holds, for n large enough, with probability larger than $1 - \epsilon$ under \mathbb{P}_ρ

$$\|\hat{\rho}_n^{(ls)} - \rho\| \leq \nu_c(\epsilon),$$

where

$$\nu_c(\epsilon)^2 = \frac{4 \cdot 3^k}{N} \log\left(\frac{2^{k+1}}{\epsilon}\right)$$

with $N := n \cdot 3^k$ the total number of measurements. The same bound holds when $k = k(n)$ as long as $\nu(\epsilon) \rightarrow 0$.

Proof of proposition 1. Note that the empirical frequencies can be written as $f(\mathbf{o}|\mathbf{s}) = \frac{1}{n} \sum_i I(X_{s,i} = \mathbf{o})$, where the random variables $X_{s,i}$ are independent for all settings \mathbf{s} and all i from 1 to n . To estimate the risk of the linear estimator we write

$$\begin{aligned} \hat{\rho}_n^{(ls)} - \rho &= \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} (f(\mathbf{o}|\mathbf{s}) - p(\mathbf{o}|\mathbf{s})) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b})}} \sigma_{\mathbf{b}} \\ &= \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} \frac{1}{n} \sum_i (I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b})}} \sigma_{\mathbf{b}} \\ &:= \sum_{\mathbf{s}} \sum_i W_{\mathbf{s},i}. \end{aligned}$$

where $W_{s,i}$ are independent and centered Hermitian random matrices. We will apply the following extension of the Bernstein matrix inequality [1] due to Tropp, see also [4, 6].

Proposition 2 (Bernstein inequality, Tropp). *Let Y_1, \dots, Y_n be independent, centered, $m \times m$ Hermitian random matrices. Suppose that, for some constants $V, W > 0$ we have $\|Y_j\| \leq V$, for all j from 1 to n , and that $\|\sum_j \mathbb{E}(Y_j^2)\| \leq W$. Then, for all $t \geq 0$,*

$$\mathbb{P}(\|Y_1 + \dots + Y_n\| \geq t) \leq 2 m \exp\left(-\frac{t^2/2}{W + tV/3}\right).$$

In our setup, $W_{s,i}$ play the role of Y_j . We bound $\|W_{s,i}\| \leq V$ for all \mathbf{s} and i and $\|\sum_{s,i} \mathbb{E}(W_{s,i}^* W_{s,i})\| \leq W$, where V, W are evaluated below. We have

$$\begin{aligned} \|W_{s,i}\| &\leq \frac{1}{n} \sum_{\mathbf{b}} \sum_{\mathbf{o}} \left| \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b})}} \right| \cdot |I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})| \cdot \|\sigma_{\mathbf{b}}\| \\ &\leq \frac{1}{n} \sum_{\mathbf{b}} \frac{1}{2^k 3^{d(\mathbf{b})}} \prod_{j \notin E_{\mathbf{b}}} \delta_{b_j, s_j} \sum_{\mathbf{o}} |I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})| \\ &\leq \frac{2}{n 2^k} \sum_{\ell=0}^k \sum_{E:|E|=\ell} \frac{1}{3^\ell} = \frac{2}{n 2^k} \sum_{\ell=0}^k \binom{k}{\ell} \frac{1}{3^\ell} = \frac{2}{n 2^k} \left(1 + \frac{1}{3}\right)^k = 2 \frac{2^k}{n \cdot 3^k} := V. \end{aligned}$$

Now, denote by $B(\mathbf{o}|\mathbf{s}) = \sum_{\mathbf{b}} 2^{-k} 3^{-d(\mathbf{b})} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) \sigma_{\mathbf{b}}$ so that $W_{s,i} = \sum_{\mathbf{o}} B(\mathbf{o}|\mathbf{s}) (I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s}))$. Then

$$\begin{aligned} &\left\| \sum_{\mathbf{s}} \sum_i \mathbb{E}(W_{s,i}^* W_{s,i}) \right\| \\ &= \frac{1}{n^2} \left\| \sum_{\mathbf{s}} \sum_i \sum_{\mathbf{o}, \mathbf{o}'} B^*(\mathbf{o}|\mathbf{s}) \cdot \text{Cov}(I(X_{s,i} = \mathbf{o}), I(X_{s,i} = \mathbf{o}')) \cdot B(\mathbf{o}'|\mathbf{s}) \right\| \\ &= \frac{1}{n} \left\| \sum_{\mathbf{s}} \sum_{\mathbf{o}, \mathbf{o}'} B^*(\mathbf{o}|\mathbf{s}) \cdot \text{Cov}(I(X_{s,1} = \mathbf{o}), I(X_{s,1} = \mathbf{o}')) \cdot B(\mathbf{o}'|\mathbf{s}) \right\| \\ &\leq \frac{1}{n} \left\| \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) B^*(\mathbf{o}|\mathbf{s}) \cdot B(\mathbf{o}|\mathbf{s}) \right\| := \frac{1}{n} \|T\|. \end{aligned} \tag{1}$$

In the last inequality we used that

$$\text{Cov}(\mathbf{s})_{\mathbf{o}, \mathbf{o}'} := \text{Cov}(I(X_{s,1} = \mathbf{o}), I(X_{s,1} = \mathbf{o}')) = p(\mathbf{o}|\mathbf{s}) \delta_{\mathbf{o}, \mathbf{o}'} - p(\mathbf{o}|\mathbf{s}) \cdot p(\mathbf{o}'|\mathbf{s})$$

which implies the following inequality between $2^k \times 2^k$ matrices: $\text{Cov}(\mathbf{s}) \leq p(\mathbf{s})$ where $p(\mathbf{s})$ is the diagonal matrix with elements $p(\mathbf{s})_{\mathbf{o}, \mathbf{o}'} = p(\mathbf{o}|\mathbf{s}) \delta_{\mathbf{o}, \mathbf{o}'}$.

By expressing $B(\mathbf{o}|\mathbf{s})$ in terms of $\sigma_{\mathbf{b}}$ as above, we get

$$\frac{1}{n} \|T\| = \frac{1}{n} \left\| \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \sum_{\mathbf{b}, \mathbf{b}'} \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b})}} \frac{A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b}')}} \sigma_{\mathbf{b}} \sigma_{\mathbf{b}'} \right\|.$$

Before giving the upper bound we introduce some combinatorial notations which will be used below. Let $\mathbf{b} \in \{x, y, z, I\}^k$ and recall that $E_{\mathbf{b}} := \{i : b_i = I\} \subset \{1, \dots, k\}$. We say that \mathbf{b} agrees with a setting \mathbf{s} if $b_j = s_j$ for all $j \in E_{\mathbf{b}}^c$. In this case \mathbf{b} is completely determined by the set $E_{\mathbf{b}}$, for a fixed \mathbf{s} . This fact will be used to replace the sums over \mathbf{b} and \mathbf{b}' with those over $E_{\mathbf{b}}$ and $E_{\mathbf{b}'}$ in T . Indeed since $A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})$ is proportional to $\prod_{j \notin E_{\mathbf{b}}} \delta_{b_j, s_j}$, the sums over \mathbf{b} and \mathbf{b}' in T are restricted to sequences which agree with \mathbf{s} . We denote by $E_{\mathbf{b}} \cap E_{\mathbf{b}'} := (E_{\mathbf{b}} \setminus E_{\mathbf{b}'}) \cup (E_{\mathbf{b}'} \setminus E_{\mathbf{b}})$ and $E_{\mathbf{b}} \Delta E_{\mathbf{b}'}$ the symmetric difference and respectively the intersection of $E_{\mathbf{b}}$ and $E_{\mathbf{b}'}$. With these notations we have

$$A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s}) \sigma_{\mathbf{b}} \sigma_{\mathbf{b}'} = \prod_{j \in E_{\mathbf{b}} \Delta E_{\mathbf{b}'}} o_j \cdot \sigma_{\mathbf{g}}$$

where $\mathbf{g} = \mathbf{g}(\mathbf{s}, E_{\mathbf{b}} \Delta E_{\mathbf{b}'})$ is the sequence with $E_{\mathbf{g}} = (E_{\mathbf{b}} \Delta E_{\mathbf{b}'})^c$, and it agrees with \mathbf{s} . With these notations we have

$$\begin{aligned}
 T &= \frac{1}{2^{2k}} \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \sum_{E_b, E_{b'}} \frac{\prod_{j \in E_b \Delta E_{b'}} o_j}{3^{|E_b|+|E_{b'}|}} \sigma_{\mathbf{g}(\mathbf{s}, E_b \Delta E_{b'})} \\
 &= \frac{1}{2^{2k}} \sum_{E_b, E_{b'}} \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \frac{\prod_{j \in E_b \Delta E_{b'}} o_j}{3^{|E_b|+|E_{b'}|}} \sigma_{\mathbf{g}(\mathbf{s}, E_b \Delta E_{b'})} \\
 &= \frac{1}{2^k} \sum_{E_b, E_{b'}} \frac{3^{k-|E_b \Delta E_{b'}|}}{3^{|E_b|+|E_{b'}|}} \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \frac{\prod_{j \in E_b \Delta E_{b'}} o_j}{2^k \cdot 3^{k-|E_b \Delta E_{b'}|}} \sigma_{\mathbf{g}(\mathbf{s}, E_b \Delta E_{b'})} \tag{2}
 \end{aligned}$$

In the last expression we rewrite the sum over settings \mathbf{s} as a double sum over $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{s}}^c$ where $\tilde{\mathbf{s}}$ is the restriction of \mathbf{s} to $E_b \Delta E_{b'}$, and $\tilde{\mathbf{s}}^c$ is the restriction to $(E_b \Delta E_{b'})^c$. Note that $\mathbf{g}(\mathbf{s}, E_b \Delta E_{b'})$ depends on \mathbf{s} only through the component $\tilde{\mathbf{s}}$. Then

$$\begin{aligned}
 \sum_{\mathbf{s}} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \frac{\prod_{j \in E_b \Delta E_{b'}} o_j}{2^k \cdot 3^{k-|E_b \Delta E_{b'}|}} \sigma_{\mathbf{g}(\mathbf{s}, E_b \Delta E_{b'})} &= \sum_{\tilde{\mathbf{s}}} \left(\sum_{\tilde{\mathbf{s}}^c} \sum_{\mathbf{o}} p(\mathbf{o}|\mathbf{s}) \frac{\prod_{j \in E_b \Delta E_{b'}} o_j}{2^k \cdot 3^{k-|E_b \Delta E_{b'}|}} \right) \sigma_{\mathbf{g}(\tilde{\mathbf{s}}, E_b \Delta E_{b'})} \\
 &= \sum_{\tilde{\mathbf{s}}} \rho_{\mathbf{g}(\tilde{\mathbf{s}}, E_b \Delta E_{b'})} \sigma_{\mathbf{g}(\tilde{\mathbf{s}}, E_b \Delta E_{b'})} = \frac{1}{2^k} \sum_{\mathbf{g}: E_{\mathbf{g}} = (E_b \Delta E_{b'})^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}}
 \end{aligned}$$

In the second equality we have used formulas (2.3) and (2.6) in the paper [3], to evaluate the interior sum as a Fourier coefficient of ρ . In the third equality we replaced the sum over $\tilde{\mathbf{s}}$ with an equivalent sum over sequences \mathbf{g} such that $E_{\mathbf{g}} = (E_b \Delta E_{b'})^c$.

Note that any pair (E, E') is uniquely determined by three disjoint subsets, $(E \setminus E', E' \setminus E, E \cap E')$, or equivalently by the symmetric difference $D := E \Delta E'$ together with $F := E \setminus E' \subset D$ and $M := E \cap E'$. The sum over E, E' in (2) is computed by summing over all triples D, F, M satisfying the above conditions:

$$\begin{aligned}
 T &= \frac{1}{2^{2k}} \sum_{E, E'} \frac{3^{k-|E \Delta E'|}}{3^{|E|+|E'|}} \sum_{\mathbf{g}: E_{\mathbf{g}} = (E \Delta E')^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}} \\
 &= \frac{1}{2^{2k}} \sum_D \sum_{F \subset D} \sum_{M: M \cap D = \emptyset} \frac{3^{k-|D|}}{3^{|D|+2|M|}} \sum_{\mathbf{g}: E_{\mathbf{g}} = D^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}} \\
 &= \frac{1}{2^{2k}} \sum_D 2^{|D|} \sum_{M: M \cap D = \emptyset} \frac{3^{k-|D|}}{3^{|D|+2|M|}} \sum_{\mathbf{g}: E_{\mathbf{g}} = D^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}}
 \end{aligned}$$

Indeed, the sum over $F \subset D$ gives a factor $2^{|D|}$ since the summands do not depend on F . Next, the sum over M is performed by summing over the size $m = |M|$ and the binomial coefficient represents the number of sets M of a given size.

$$\begin{aligned}
 T &= \frac{1}{2^{2k}} \sum_D 2^{|D|} \sum_{m=0}^{k-|D|} \binom{k-|D|}{m} \frac{3^{k-|D|}}{3^{|D|+2m}} \sum_{\mathbf{g}: E_{\mathbf{g}} = D^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}} \\
 &= \left(\frac{3}{4}\right)^k \sum_D \left(\frac{2}{9}\right)^{|D|} \left(\frac{10}{9}\right)^{k-|D|} \sum_{\mathbf{g}: E_{\mathbf{g}} = D^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}} \\
 &= \left(\frac{5}{6}\right)^k \sum_D \left(\frac{1}{5}\right)^{|D|} \sum_{\mathbf{g}: E_{\mathbf{g}} = D^c} \text{Tr}(\rho \sigma_{\mathbf{g}}) \sigma_{\mathbf{g}} \\
 &= \left(\frac{5}{6}\right)^k 2^k \mathcal{D}^{\otimes k}(\rho). \tag{3}
 \end{aligned}$$

The final sum goes over subsets D and over sequences \mathbf{g} such that $E_{\mathbf{g}} = D^c$, and is similar to the Fourier decomposition of ρ except that each terms is weighted by the factor $5^{-|D|}$. In fact, a closer inspection shows that the weighted sum is nothing but the output state of a product of depolarising channels acting in parallel in the state ρ , where an individual depolarising channel is defined by

$$\mathcal{D} : \frac{1}{2}(I + r_x \sigma_x + r_y \sigma_y + r_z \sigma_z) : \mapsto \frac{1}{2} \left(I + \frac{r_x}{5} \sigma_x + \frac{r_y}{5} \sigma_y + \frac{r_z}{5} \sigma_z \right).$$

For an arbitrary quantum channel \mathcal{T} , let $\nu_p(\mathcal{T}) := \sup_{\tau} \text{Tr}(T(\tau)^p)^{1/p}$ be its p -norm, where the supremum is taken over all input states τ ; in particular for $p \rightarrow \infty$ this becomes the ∞ -norm $\nu_{\infty}(\mathcal{T}) := \sup_{\tau} \|T(\tau)\|$. For the depolarising channel \mathcal{D} defined above, the ∞ -norm can be computed easily by applying the channel to an arbitrary pure state and is equal to $\nu_{\infty}(\mathcal{D}) = 3/5$. Moreover, it is known [5] that the depolarising channel has multiplicative p -norm, i.e. $\nu_p(\mathcal{D}^{\otimes k}) = \nu_p(\mathcal{D})^k$, which implies that $\|\mathcal{D}^{\otimes k}(\rho)\| \leq (3/5)^k$. Together with (3) this gives upper bound

$$\frac{1}{n} \|T\| \leq \frac{1}{n} \left(\frac{5}{6}\right)^k \cdot 2^k \cdot \left(\frac{3}{5}\right)^k = \frac{3^k}{n \cdot 3^k} = \frac{3^k}{N}. \quad (4)$$

Putting together (1) and (4) we obtain

$$\left\| \sum_{\mathbf{s}} \sum_i \mathbb{E}(W_{\mathbf{s},i}^* W_{\mathbf{s},i}) \right\| \leq \frac{3^k}{n \cdot 3^k} =: W.$$

We apply now the matrix Bernstein inequality in proposition 2 to get, for any $t > 0$:

$$\mathbb{P}_{\rho}(\|\hat{\rho}_n^{(k)} - \rho\| \geq t) \leq 2^{k+1} \exp\left(-\frac{n \cdot t^2/2}{1 + t \cdot (2/3)^{k+1}}\right).$$

We choose $t > 0$ such that

$$2^{k+1} \exp\left(-\frac{n \cdot t^2/2}{1 + t \cdot (2/3)^{k+1}}\right) \leq \varepsilon,$$

which leads, for n large enough, to $t = \nu_{\varepsilon}(\varepsilon)$ such that

$$\nu_{\varepsilon}(\varepsilon)^2 = \frac{4 \cdot 3^k}{N} \log\left(\frac{2^{k+1}}{\varepsilon}\right).$$

□

Acknowledgments

MG's work was supported by the EPSRC Grant No. EP/J009776/1.

References

- [1] Ahlswede R and Winter A 2002 Strong converse for identification via quantum channels *IEEE Trans. Inf. Theory* **48** 569–79
- [2] Alquier P, Butucea C, Hebiri M, Meziani K and Morimae T 2013 Rank penalized estimation of a quantum system *Phys. Rev. A* **88** 032113
- [3] Butucea C, Guta M and Kypraios T 2015 Spectral thresholding quantum tomography for low rank states *New J. Phys.* **17** 113050
- [4] Gross D 2011 Recovering low-rank matrices from few coefficients in any basis *IEEE Transactions on Information Theory* **57** 1548–66
- [5] King C 2002 A remark on low rank matrix recovery and noncommutative Bernstein type inequalities *J. Math. Phys.* **43** 4641
- [6] Koltchinskii V 2013 A remark on low rank matrix recovery and noncommutative Bernstein type inequalities *IMS Collections* **9** 213–26



PAPER

Spectral thresholding quantum tomography for low rank states

OPEN ACCESS

RECEIVED
17 July 2015REVISED
6 October 2015ACCEPTED FOR PUBLICATION
27 October 2015PUBLISHED
19 November 2015Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).Any further distribution of
this work must maintain
attribution to the
author(s) and the title of
the work, journal citation
and DOI.Cristina Butucea¹, Mădălin Guță^{2,3} and Theodore Kypraios²¹ Université Paris-Est Marne-la-Vallée, LAMA(UMR 8050), UPEMLV F-77454, Marne-la-Vallée, France² University of Nottingham, School of Mathematical Sciences, University Park, Nottingham NG7 2RD, UK³ Author to whom any correspondence should be addressed.E-mail: Madalin.guta@nottingham.ac.uk

Keywords: quantum tomography, low rank states, thresholding estimation, ion trap measurements, rank penalization

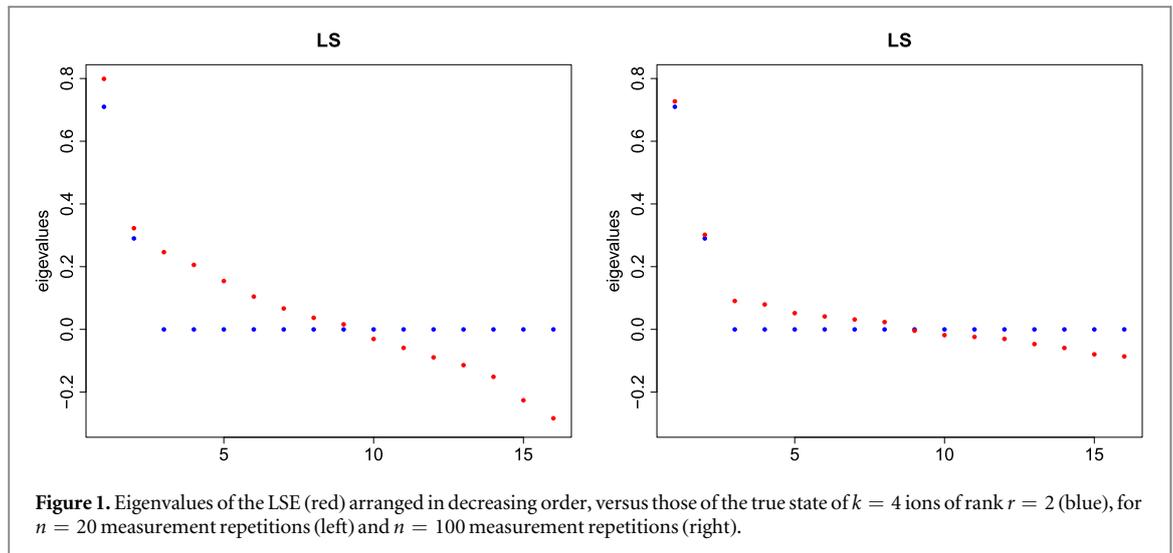
Abstract

The estimation of high dimensional quantum states is an important statistical problem arising in current quantum technology applications. A key example is the tomography of multiple ions states, employed in the validation of state preparation in ion trap experiments (Häffner *et al* 2005 *Nature* **438** 643). Since full tomography becomes unfeasible even for a small number of ions, there is a need to investigate lower dimensional statistical models which capture prior information about the state, and to devise estimation methods tailored to such models. In this paper we propose several new methods aimed at the efficient estimation of low rank states and analyse their performance for multiple ions tomography. All methods consist in first computing the least squares estimator, followed by its truncation to an appropriately chosen smaller rank. The latter is done by setting eigenvalues below a certain ‘noise level’ to zero, while keeping the rest unchanged, or normalizing them appropriately. We show that (up to logarithmic factors in the space dimension) the mean square error of the resulting estimators scales as $r \cdot d/N$ where r is the rank, $d = 2^k$ is the dimension of the Hilbert space, and N is the number of quantum samples. Furthermore we establish a lower bound for the asymptotic minimax risk which shows that the above scaling is optimal. The performance of the estimators is analysed in an extensive simulations study, with emphasis on the dependence on the state rank, and the number of measurement repetitions. We find that all estimators perform significantly better than the least squares, with the ‘physical estimator’ (which is a bona fide density matrix) slightly outperforming the other estimators.

1. Introduction

Recent years have witnessed significant developments at the overlap between quantum theory and statistics: from new state estimation (or tomography) methods [2–8], design of experiments [9–11], quantum process and detector tomography [12, 13] construction of confidence regions (error bars) [14–16], quantum tests [17, 18] entanglement estimation [19], asymptotic theory [20–23]. The importance of quantum state tomography, and the challenges raised by the estimation of high dimensional systems were highlighted by the landmark experiment [1] where entangled states of up to eight ions were created and fully characterized. However, as full quantum state tomography of large systems becomes unfeasible [24], there is significant interest in identifying physically relevant, lower dimensional models, and in devising efficient model selection and estimation methods in such setups [7, 8, 25–28]. In this paper we reconsider the multiple ions tomography (MIT) problem by proposing and analysing several new methods for estimating low rank states in a statistically efficient way. We emphasize that, while the theoretical and simulations results are specific to the ion tomography setup, the general methods based on combining linear (least squares) estimation with spectral thresholding (or eigenvalues truncation) can be applied to any informationally complete tomography scenario. Below, we briefly review the MIT setup, after which we proceed with presenting the key ideas and results of the paper.

In MIT [1], the goal is to statistically reconstruct the *joint state* of k ions (modelled as two-level systems), from *counts data* generated by performing a large number of measurements on identically prepared systems. The unknown state ρ is a $d \times d$ density matrix (complex, positive trace-one matrix) where $d = 2^k$ is the dimension of



the Hilbert space of k ions. The experimenter can measure an arbitrary Pauli observable σ_x , σ_y or σ_z of each ion, *simultaneously* on all k ions. Thus, each measurement setting is labelled by a sequence $\mathbf{s} = (s_1, \dots, s_k) \in \{x, y, z\}^k$ out of 3^k possible choices. The measurement produces an outcome $\mathbf{o} = (o_1, \dots, o_k) \in \{+1, -1\}^k$, whose probability is equal to the corresponding diagonal element of ρ with respect to the orthonormal basis (ONB) determined by the measurement setting \mathbf{s} . The measurement procedure and statistical model can be summarized as follows. For each setting \mathbf{s} the experimenter performs n repeated measurements and collects the counts of different outcomes $N(\mathbf{o}|\mathbf{s})$, so that the total number of quantum samples used is $N := n \times 3^k$. The resulting dataset is a $2^k \times 3^k$ table whose columns are independent and contain all the counts in a given setting. A commonly used [1] estimation method is *maximum likelihood* (ML) which selects the state for which the probability of the observed data is the highest among all states. However, while this method seems to perform well in practice, and has efficient numerical implementations [29], it does not provide confidence intervals (error bars) for finite samples, and it has been criticized for its tendency to produce rank-deficient states [2].

The goal of this paper is to find alternative estimators which can be efficiently computed, and work well for *low rank states*. The reason for focusing on low rank states is that they form a realistic model for physical states created in the lab, where experimentalists often aim at preparing a pure (rank-one) state. While this is generally difficult, the realized states tend to have rapidly decaying eigenvalues, so that they can be well approximated by low rank states. Our strategy is to combine an easy but ‘noisy’ estimation method—the least square estimator (LSE)—with an appropriate spectral truncation (tuned using available data only) which involves setting certain eigenvalues of the LSE to zero, while possibly adjusting the remaining ones. It turns out that this can lead to significant reduction of the mean square error (MSE) of the estimator. We discuss some of the key results on the LSE and the proposed truncation estimators below.

The LSE $\hat{\rho}_n^{(\text{ls})}$ is obtained by inverting the linear map $A : \rho \mapsto \mathbb{P}_\rho$ between the state and the probability distribution of the data, where the unknown probabilities are replaced by the empirical (observed) frequencies of the measurement data. The resulting estimator is unbiased, and is ‘optimal’ in the sense that it minimizes the *prediction error*, i.e. the euclidian distance between the empirical frequencies and the predicted probabilities. However, one of the disadvantages of the LSE is that it does not take into account the physical properties of the state, i.e. its positivity and trace-one property. More importantly, as we explain below, the LSE has a relatively large *estimation error* for the class of low rank states, and performs well only on very mixed states. This is illustrated in figure 1 where the eigenvalues of $\hat{\rho}_n^{(\text{ls})}$ are plotted (in decreasing order) against those of the true state ρ , the latter being chosen to have rank $r = 2$. We see that while the non-zero eigenvalues of ρ are estimated reasonably well, the LSE is poor in estimating the zero-eigenvalues, and as consequence, it has a large estimation variance.

Our goal is to design more precise estimators, which have the LSE as a starting point, but take into account the ‘sparsity’ properties of the unknown state. Figure 1 suggests that the non-zero eigenvalues of the LSE which are below a certain ‘statistical threshold’, can be considered as statistical noise and may be set to zero in order to improve the estimation error. To find this noise level, we establish a concentration inequality (see proposition 1) which shows that the operator-norm error $\|\hat{\rho}_n^{(\text{ls})} - \rho\|^2$ is upper bounded by a rate ν^2 which (up to logarithmic factors in d) is proportional to d/N .

The first estimator we propose, is a *rank penalized* one obtained by diagonalizing the LSE, arranging its eigenvalues in decreasing order of their absolute values, and setting to zero all those eigenvalues whose absolute

values are below the threshold ν

$$\hat{\rho}_n^{(\text{ls})} = \sum_{i=1}^d \hat{\lambda}_i |\hat{\psi}_i\rangle \langle \hat{\psi}_i| \longrightarrow \hat{\rho}_n^{(\text{pen})} = \sum_{|\hat{\lambda}_i| \geq \nu} \hat{\lambda}_i |\hat{\psi}_i\rangle \langle \hat{\psi}_i|.$$

The same outcome can be obtained as solution of the following penalized estimation problem: among all selfadjoint matrices, choose the one that is close to the LSE but in the same time it has low rank, so that it minimizes over τ the norm-two (also known as Frobenius norm) distance squared, penalized by the rank

$$D(\tau) := \|\tau - \hat{\rho}_n^{(\text{ls})}\|_2^2 + \nu^2 \cdot \text{rank}(\tau).$$

In particular the estimator's rank is determined by the data. In theorem 1 we show that if ρ is of *unknown* rank $r \leq d$, then the MSE $\mathbb{E}\|\hat{\rho}_n^{(\text{pen})} - \rho\|_2^2$ is upper-bounded (up to logarithmic factors) by the rate $(r \cdot d)/N$. This captures the expected optimal dependence on the number of parameters for a state of rank r . Indeed, in section 5 we show that no estimator can improve the above rate for all states of rank r , see theorem 3 for the *asymptotic minimax* lower bound.

The penalized estimator has however the drawback that it may not represent a physical state. To remedy this, and further improve its statistical accuracy, we propose a *physical estimator* which is the solution of the following optimization problem. We seek the *density matrix* which is closest to the LSE $\hat{\rho}_n^{(\text{ls})}$, and whose non-zero eigenvalues are larger than the threshold 4ν . It turns out that the solution can be found via a simple iterative algorithm whereby at each step the eigenvalues of $\hat{\rho}_n^{(\text{ls})}$ below the threshold are set to zero, and the remaining eigenvalues are normalized by shifting with a common constant, while the eigenvectors are not changed throughout the process. In theorem 2 we show that the physical estimator satisfies a similar upper bound to the penalized one.

In section 6 we present results of extensive numerical investigations of the two proposed estimators. In addition we consider the *oracle* 'estimator' and the *cross-validated* estimator $\hat{\rho}_n^{(\text{cv})}$ (see next paragraph for a brief explanation of the cross-validation method). The oracle estimator is simply the spectral truncation of the LSE that is closest to the true state ρ , and is obtained by setting to zero a number of eigenvalues with small magnitude. It is not a proper estimator since it uses the true unknown density matrix ρ , but it is very useful as a benchmark. The cross-validated estimator is also a spectral truncation of the LSE which aims to find the optimal truncation rank (i.e. rank of the truncated estimator) by minimizing an estimate of the norm-two square error over all ranks.

Cross-validation is a widely used collection of procedures for model validation and selection [30] which involve splitting the data into different (independent) batches, some of which are used for estimation while the others are used for testing the model. We found that cross-validation can help in better tuning the constant factor of the threshold rate of the penalized and physical estimators. As expected from the theoretical results, we find that all estimators perform significantly better than the LSE on low rank states; moreover the physical estimator has slightly smaller estimation error than the others, including the oracle estimator. We also find that all methods converge to the correct rank in the limit of large number of repetitions but through different routes: the penalized estimator tends to underestimate, while the physical one tends to overestimate the rank, for small number of samples.

Having discussed the upper bounds on the estimators' MSE, we would like to know how they compare with the best possible estimation procedure. For each estimation procedure $\hat{\rho}_n$ one can consider its maximum MSE over the class $\mathcal{S}_{d,r}$ of rank r states

$$R_{\max}(\hat{\rho}_n; r) = \sup_{\rho \in \mathcal{S}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2.$$

Its minimum over all estimators is called the minimax risk $R_{\min\max}(r, n)$. Asymptotic statistics theory [31] shows that the limiting value of the minimax risk rescaled by the number of samples $N = n \cdot 3^k$, satisfies the following lower bound

$$\liminf_{n \rightarrow \infty} NR_{\min\max}(r, n) \geq \sup_{\rho \in \mathcal{S}_{d,r}} \text{Tr}(I(\rho)^{-1}G(\rho)).$$

The left side is the *asymptotic minimax* risk while on the right side $I(\rho)$ is the Fisher information corresponding to all measurement settings taken together, and $G(\rho)$ is a positive matrix describing the quadratic approximation of the norm-two (Frobenius) distance squared, around ρ . In theorem 3 we show that the above lower bound is further bounded from below by $2r(d - r)$ which shows that (up to logarithmic factors) the upper bounds of the penalized and physical estimators have the same scaling as the asymptotic minimax risk.

Recently, a number of papers discussed related aspects of quantum tomography problems. The idea of the penalized estimator has been proposed in [8], which provided a weaker upper bound for its MSE. Here we improve on the MSE upper bounds, provide minimax lower bounds, and propose and analyse new classes of estimators, e.g. a 'physical' estimator with improved estimation performance. Reference [7] analyses model selection methods for finite rank models and ML estimation. Reference [32] proposes a different estimator and establishes a comparable upper bound for its MSE. The class of low rank states is also employed in compressed

sensing quantum tomography [25, 28, 33], but their statistical model is based on expectations of Pauli observables rather than measurement counts.

The paper is organized as follows. In section 2 we describe the measurement procedure and introduce the statistical model of MIT. In section 3 we define the linear (least squares) estimator and derive an upper bound on its operator norm error which improves on a previous bound of [8]. In section 4 we define the penalized and threshold estimators and derive upper bounds for their MSEs with respect to the norm-two (Frobenius) distance squared. The performance of the different methods is analysed in section 6. An asymptotic lower bound for the minimax risk is derived in section 5, based on the Fisher information of the measurement data. The upper and lower bounds match in the scaling with the number of parameters and number of total measurements, up to a logarithmic factor. We give a detailed description of the numerical implementation of the algorithms, including the cross-validation routines used for tuning the pre-factor of the penalty and threshold constants. We illustrate the simulation results with box plots of the MSEs for the least squares, oracle, cross-validation, penalization and threshold estimator, for states of ranks 1, 2, 6 and 10, and for different choices of measurement repetitions $n = 20, 100$. Additionally, we plot the empirical distribution of the chosen rank for different estimators, showing the concentration on the true rank as the number of repetitions increases.

2. Multiple ions tomography

This paper deals with the problem of estimating the joint quantum state of k two-dimensional systems (qubits), as encountered in ion trap quantum tomography [1]. The two-dimensional system is determined by two energy levels of an ion, while the remaining levels can be ignored as they remain unpopulated during the experiment. The joint Hilbert space of the ions is therefore the tensor product $(\mathbb{C}^2)^{\otimes k} \cong \mathbb{C}^d$ where $d = 2^k$, and the state is a density matrix ρ on this space, i.e. a positive $d \times d$ matrix of trace one.

Our statistical model is derived from standard ion trap measurement procedures, and takes into account the specific statistical uncertainty due to finite number of measurement repetitions. We consider that for each individual qubit, the experimenter can measure one of the three Pauli observables $\sigma_x, \sigma_y, \sigma_z$. A measurement set-up is then defined by a *settings* $\mathbf{s} = (s_1, \dots, s_k) \in \mathcal{S}_k := \{x, y, z\}^k$ which specifies which of the three Pauli observables is measured for each ion. For each fixed setting, the measurement produces random outcomes $\mathbf{o} \in \mathcal{O}_k := \{+1, -1\}^k$ with probability

$$p_\rho(\mathbf{o}|\mathbf{s}) := \text{Tr}(\rho P_{\mathbf{o}}^{\mathbf{s}}) = \langle e_{\mathbf{o}}^{\mathbf{s}} | \rho | e_{\mathbf{o}}^{\mathbf{s}} \rangle, \quad (2.1)$$

where $P_{\mathbf{o}}^{\mathbf{s}}$ are the one-dimensional projections

$$P_{\mathbf{o}}^{\mathbf{s}} = |e_{o_1}^{s_1}\rangle \langle e_{o_1}^{s_1}| \otimes \dots \otimes |e_{o_k}^{s_k}\rangle \langle e_{o_k}^{s_k}| \quad (2.2)$$

and $|e_o^s\rangle$ are the eigenvectors of the Pauli matrices, i.e. $\sigma_s |e_{\pm}^s\rangle = \pm |e_{\pm}^s\rangle$.

The measurement procedure consists of choosing a setting \mathbf{s} , and performing n repeated measurements in that setting, on identically prepared systems in state ρ . This provides information about the diagonal elements of ρ with respect to the chosen measurement basis, i.e. the probabilities $p_\rho(\mathbf{o}|\mathbf{s})$. In order to identify the other elements, the procedure is then repeated for all 3^k possible settings.

Before describing the statistical model of the measurement counts data, we start by discussing in more detail the relation between the unknown parameter ρ and the probabilities $p(\mathbf{o}|\mathbf{s})$. Consider the ‘extended’ set of Pauli operators $\{\sigma_x, \sigma_y, \sigma_z, \sigma_I := \mathbf{1}\}$ which form a basis in $M(\mathbb{C}^2)$. We construct the tensor product basis in $M(\mathbb{C}^d)$ with elements $\sigma_{\mathbf{b}} = \sigma_{b_1} \otimes \dots \otimes \sigma_{b_k}$ where $\mathbf{b} \in \{x, y, z, I\}^k$ and note that the following orthogonality relations hold $\text{Tr}(\sigma_{\mathbf{b}} \sigma_{\mathbf{c}}) = d \delta_{\mathbf{b}, \mathbf{c}}$. The state ρ can be expanded in this basis as

$$\rho = \sum_{\mathbf{b} \in \{I, x, y, z\}^k} \rho_{\mathbf{b}} \sigma_{\mathbf{b}}, \quad \text{where } \rho_{\mathbf{b}} = \text{Tr}(\rho \sigma_{\mathbf{b}}) / d. \quad (2.3)$$

Equation (2.1) can then be written as

$$p_\rho(\mathbf{o}|\mathbf{s}) = \sum_{\mathbf{b} \in \{I, x, y, z\}^k} \rho_{\mathbf{b}} \text{Tr}(\sigma_{\mathbf{b}} P_{\mathbf{o}}^{\mathbf{s}}) = \sum_{\mathbf{b} \in \{I, x, y, z\}^k} \rho_{\mathbf{b}} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}).$$

The coefficients $A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})$ can be computed explicitly as

$$A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) = \text{Tr}(\sigma_{\mathbf{b}} P_{\mathbf{o}}^{\mathbf{s}}) = \prod_{j \in E_{\mathbf{b}}} o_j \cdot \delta_{b_j, s_j}, \quad (2.4)$$

where $E_{\mathbf{b}} := \{i : b_i = I\}$. Let $\tilde{\rho} \in \mathbb{C}^{4^k}$ be the representation of ρ as the vector of coefficients $\rho_{\mathbf{b}}$, and let \mathbf{p}_ρ be the corresponding vector of probabilities for all settings ($p_\rho(\mathbf{o}|\mathbf{s}) : (\mathbf{o}, \mathbf{s}) \in \mathcal{O}_k \times \mathcal{S}_k$), with settings, and outcomes within settings ordered in lexicographical order. The measurement is then described by the linear map $\mathbf{A} : \mathbb{C}^{4^k} \rightarrow \mathbb{C}^{3^k} \otimes \mathbb{C}^{2^k}$ with matrix elements $\mathbf{A}_{\mathbf{b}, (\mathbf{o}, \mathbf{s})} := A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})$ defined in (2.4), such that

$$\mathbf{p}_\rho = \mathbf{A}\tilde{\rho}. \quad (2.5)$$

The linear map \mathbf{A} is injective and we solve the previous equation via the optimization problem:

$$\tilde{\rho} = \arg \inf_{\tilde{\tau} \in \mathbb{C}^{d^k}} \|\mathbf{p} - \mathbf{A}\tilde{\tau}\|^2,$$

which comes down to multiplying equation (2.5) by the Moore–Penrose pseudoinverse of \mathbf{A} and gives

$\tilde{\rho} = (\mathbf{A}^* \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^* \cdot \mathbf{p}_\rho$. The following lemma has appeared in [8] but for completeness we include its proof in appendix A.1.

Lemma 1. *Let \mathbf{A} be the linear map defined in equation (2.5). Then $\mathbf{A}^* \cdot \mathbf{A}$ is diagonal and its elements are*

$$\left[\mathbf{A}^* \cdot \mathbf{A} \right]_{\mathbf{b}, \mathbf{b}'} = 2^{k3^d(\mathbf{b})} \delta_{\mathbf{b}, \mathbf{b}'}, \quad \text{for all } \mathbf{b} \in \{I, x, y, z\}^k,$$

where $d(\mathbf{b}) := |E_{\mathbf{b}}|$ is the number of I 's in the sequence \mathbf{b} .

From the decomposition (2.3) and lemma 1 we find

$$\rho = \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} p_\rho(\mathbf{o}|\mathbf{s}) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^{k3^d(\mathbf{b})}} \sigma_{\mathbf{b}} = \sum_{\mathbf{b}} \tilde{\rho}_{\mathbf{b}} \sigma_{\mathbf{b}}. \quad (2.6)$$

The above formula allows to reconstruct the matrix elements from the measurement probabilities. However, since the experiment only provides random counts from these probabilities, we need to construct a statistical model for the measurement data. After n repetitions of the measurement with setting \mathbf{s} , we collect independent, identically distributed observations $X_{i|\mathbf{s}} \in \mathcal{O}_k$, for i from 1 to n . The data can be summarized by the set of counts $\{N(\mathbf{o}|\mathbf{s}) : \mathbf{o} \in \mathcal{O}_k\}$, where $N(\mathbf{o}|\mathbf{s}) = \sum_i \delta_{X_{i|\mathbf{s}}, \mathbf{o}}$ is the number of times that outcome \mathbf{o} has occurred. After repeating this for each setting $\mathbf{s} \in \mathcal{S}_k$, we collect all the data in the counts data-set $D := \{N(\mathbf{o}|\mathbf{s}) : (\mathbf{o}, \mathbf{s}) \in \mathcal{O}_k \times \mathcal{S}_k\}$.

Since successive preparation-measurement cycles are independent of each other, the probability of a given set of counts $N(\mathbf{o}|\mathbf{s})$ (obtained by repeating n times the measurement in setting \mathbf{s}) is a multinomial distribution, and the probability of a certain dataset D is the product of such multinomials over the different settings:

$$\mathbb{P}_\rho(D) = \mathbb{P}_\rho\left(\{N(\mathbf{o}|\mathbf{s}) : (\mathbf{o}, \mathbf{s}) \in \mathcal{O}_k \times \mathcal{S}_k\}\right) = \prod_{\mathbf{s}} \frac{n!}{\prod_{\mathbf{o}} N(\mathbf{o}|\mathbf{s})!} \prod_{\mathbf{o}} p_\rho(\mathbf{o}|\mathbf{s})^{N(\mathbf{o}|\mathbf{s})}. \quad (2.7)$$

The statistical problem is to estimate the state ρ from the measurement data summarized by the counts dataset D . The most commonly used estimation method is ML. The ML estimator is defined by

$$\hat{\rho}_n^{(\text{ml})}(D) := \arg \max_{\sigma} \mathbb{P}_\sigma(D),$$

where the maximum is taken over all *density matrices* τ on \mathbb{C}^d , and can be computed by using standard maximization routines, or the iterative algorithms proposed in [29, 34]. However, ML becomes impractical for about $k = 10$ ions, and the iterative algorithm has the drawback that it cannot be adapted to models where prior information about the state is encoded in a lower dimensional parametrization of the relevant density matrices, e.g. when the states are low rank. In the next section we discuss an alternative method, the LSE, and derive an upper bound on its MSE. After this, we will show that by ‘post-processing’ the LSE using penalization and thresholding methods, its performance can be considerably improved when the unknown state has low rank.

3. The LSE

Recall that the vectorized version $\tilde{\rho}$ of the state ρ satisfies (2.5), it is therefore the solution of the optimization

$$\tilde{\rho} = \arg \inf_{\tilde{\tau} \in \mathbb{C}^{d^k}} \|\mathbf{p} - \mathbf{A}\tilde{\tau}\|^2,$$

giving $\rho = \sum_{\mathbf{b}} \tilde{\rho}_{\mathbf{b}} \cdot \sigma_{\mathbf{b}}$ in (2.6). If the number of repetitions n is large compared with the dimension d , then the outcomes’ empirical frequencies are good approximations of the corresponding probabilities, i.e.

$f(\mathbf{o}|\mathbf{s}) := N(\mathbf{o}|\mathbf{s})/n \rightarrow p(\mathbf{o}|\mathbf{s})$ by the law of large numbers. Therefore, by replacing \mathbf{p} by the vector of frequencies \mathbf{f} in the previous display, we can define the linear estimator also known as the LSE of ρ

$$\tilde{\rho}_n^{(\text{ls})} := \arg \inf_{\tilde{\tau} \in \mathbb{C}^{d^k}} \|\mathbf{f} - \mathbf{A}\tilde{\tau}\|^2,$$

which has the explicit expression

$$\hat{\rho}_n^{(\text{ls})} = \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} f(\mathbf{o}|\mathbf{s}) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^{k3^d(\mathbf{b})}} \sigma_{\mathbf{b}}. \quad (3.1)$$

Note that in this case it comes down to replacing the unknown probability \mathbf{p} in equation (2.6) with the empirical frequencies \mathbf{f} (also known as the plug-in method).

In spite of this ‘optimality’ property and its computational efficiency, the LSE has the disadvantage that in general it is not a state, i.e. it is not trace-one and may have negative eigenvalues. A more serious disadvantage is that its risk—measured for instance by the MSE $\mathbb{E}(\|\hat{\rho}_n^{(\text{ls})} - \rho\|_2^2)$ —is large compared with other estimators such as the ML estimator. This is due to the fact that the LSE does not use the physical properties of the unknown parameter ρ , that is positivity and trace-one. As we will see below, the modified estimators proposed in section 4 outperform the LSE while adding only a small amount of computational complexity. Moreover, the second estimator will be a density matrix.

In the remainder of this section we provide concentration bounds on the square error of the LSE, which will later be used in obtaining the upper bounds of the improved estimators. The following proposition improves the rate $k(4/3)^k/n$ obtained in [8] to $k(2/3)^k/n$.

Proposition 1. *Let $\hat{\rho}_n^{(\text{ls})}$ be the LSE of ρ . Then, for any $\varepsilon > 0$ small enough the following operator norm inequality holds with probability larger than $1 - \varepsilon$ under \mathbb{P}_ρ*

$$\|\hat{\rho}_n^{(\text{ls})} - \rho\| \leq \nu(\varepsilon),$$

where

$$\nu(\varepsilon)^2 = \frac{2}{n} \left(\frac{2}{3}\right)^k \log\left(\frac{2^{k+1}}{\varepsilon}\right) = \frac{2d}{N} \log\left(\frac{2d}{\varepsilon}\right)$$

with $N := n \cdot 3^k$ the total number of measurements. The same bound holds when $k = k(n)$ as long as $\nu(\varepsilon) \rightarrow 0$.

Proof. See appendix A.2.

As a side remark we note that projecting the LSE onto the space of Hermitian matrices with trace 1, does not change the rate of convergence from proposition 1. The following proposition allows us to assume that, without loss of generality, the LSE has also trace 1.

Proposition 2. *Under the notation and assumptions in proposition 1, let*

$$\hat{\rho}_n^{(\text{ls},n)} = \arg \min_{\tau: \text{tr}(\tau)=1} \|\tau - \hat{\rho}_n^{(\text{ls})}\|_2^2. \quad (3.2)$$

Then with probability larger than $1 - \varepsilon$ we have $\|\hat{\rho}_n^{(\text{ls},n)} - \rho\| \leq 2\nu(\varepsilon)$.

Proof. See appendix A.3.

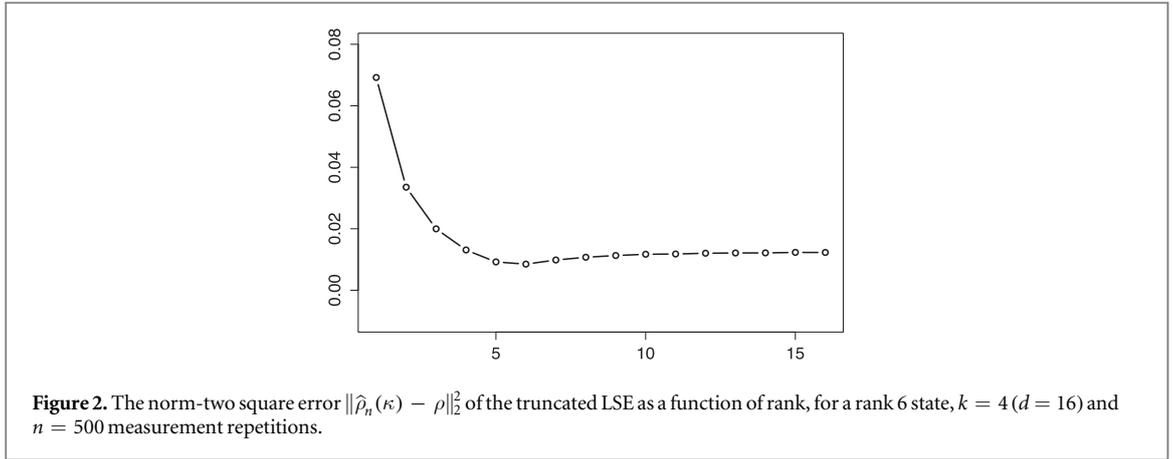
4. Rank-penalized and threshold projection estimator

In this section we investigate two ways to improve the LSEs. The first method is to project the LSE onto the space of finite rank Hermitian matrices of an appropriate rank. We prove upper bounds for its risk with respect to the norm-two (Frobenius) distance squared. Building on the knowledge about the rank-penalized estimator, we define the second estimator which is the projection of the LSE on the space of physical states whose eigenvalues are larger than a certain positive noise threshold. We give a simple and fast algorithm producing a proper density matrix from the data, which also inherits the good theoretical properties of the rank-penalized estimator.

4.1. Rank-penalized estimator

We introduce here the rank-penalized nonlinear estimator, which can be computed from the LSE by truncation to an appropriately chosen rank.

As noted earlier, while the LSE is unbiased, it has a large variance due to the fact that it does not take into account the physical constraints encoded in the unknown parameter ρ . A possible remedy is to ‘project’ the LSE onto the space of physical states, i.e. positive, trace-one matrices. This method will be discussed in the following subsection. Another improvement can be obtained by taking into account the ‘sparsity’ properties of the unknown state. For instance, in many experimental situations the goal is to create a particular low rank, or even pure state. The fact that such states can be characterized with a smaller number of parameters than a general density matrix, has two important consequences. Firstly, they can be estimated by measuring an ‘informationally incomplete’ set of observables, as demonstrated in [25, 26]. Secondly, the prior information can be used to design estimators with reduced estimation error compared with generic methods which do not take into account the structure of the state. Roughly speaking, this is because each unknown parameter brings its own contribution to the overall error of the estimator.



However, the downside of working with a lower dimensional model is that it contains built-in assumptions which may not be satisfied by the true (unknown) physical state. Preparing a pure state is strictly speaking rarely achievable due to various experimental imperfections, so using a pure state statistical model is in fact an oversimplification and can lead to erroneous conclusions about the true state. On the other hand, one can argue that when the (small) experimental noises are taken into account, the actual state is ‘effectively’ low rank, i.e. it has a small number of significant eigenvalues and a large number of eigenvalues which are so close to zero that they cannot be distinguished from it. Then, the interesting question is how to decide on where to make the cut-off between statistically relevant eigenvalues and pure statistical noise. This is a common problem in statistics which is closely related to that of model selection [30]. Below we describe the rank-penalized estimator addressing this problem, and show that its theoretical and practical performance is superior to the LSE, and is close to what one would expect from an optimal estimator. In addition, its computation requires only the diagonalization of the LSE.

Before presenting a simple algorithm for computing the estimator, we briefly discuss the idea behind its definition. Let

$$\hat{\rho}_n^{(\text{ls})} = \sum_{i=1}^d \hat{\lambda}_i |\hat{\psi}_i\rangle \langle \hat{\psi}_i|. \quad (4.1)$$

Be the spectral decomposition of the LSE, with eigenvalues ordered such that $|\hat{\lambda}_1| \geq \dots \geq |\hat{\lambda}_d|$. For each given rank $\kappa \in \{1, \dots, d\}$ we can project $\hat{\rho}_n^{(\text{ls})}$ onto the space of matrices of rank κ by computing the matrix which is the closest to $\hat{\rho}_n^{(\text{ls})}$ with respect to the norm-two distance

$$\hat{\rho}_n(\kappa) := \arg \min_{\tau: \text{rank}(\tau) = \kappa} \|\tau - \hat{\rho}_n^{(\text{ls})}\|_2^2.$$

Although the projection is not a linear operator, $\hat{\rho}_n(\kappa)$ is easy to compute, and is obtained by truncating the spectral decomposition (4.1) to the most significant κ eigenvalues

$$\hat{\rho}_n(\kappa) = \sum_{i=1}^{\kappa} \hat{\lambda}_i |\hat{\psi}_i\rangle \langle \hat{\psi}_i|.$$

The question is now how to choose the rank κ in order to obtain a good estimator. In figure 2 we illustrate the dependence of the norm-two square error $e(\kappa) := \|\hat{\rho}_n(\kappa) - \rho\|_2^2$ on the rank, for a particular dataset generated with a rank 6, 4-ions state. As the rank is increased starting with $\kappa = 1$ (pure states), the error decreases steeply as $\hat{\rho}_n(\kappa)$ becomes less biased, it reaches a minimum close to the true rank, and increases slowly as added parameters increase the variance of the estimator. However, since the state ρ is unknown, the norm-two error and optimal rank for which the minimum is achieved, are unknown. To go around this, we can estimate the error $e(\kappa)$ from the data by means of e.g. cross-validation, as it will be described in section 6. However, in this section we follow a different path, and we define the rank-penalized estimator [8, 35] as the minimizer over κ of the following expression:

$$\|\hat{\rho}_n(\kappa) - \hat{\rho}_n^{(\text{ls})}\|_2^2 + \nu^2 \cdot \kappa = \sum_{i=\kappa+1}^d \hat{\lambda}_i^2 + \nu^2 \cdot \kappa,$$

where ν is a constant which will be tuned appropriately. The first term quantifies the fit of the truncated estimator with respect to the LSE, while the second term is a penalty which increases with the complexity of the model, i.e. the rank. The rank penalized estimator $\hat{\rho}_n^{(\text{pen})}$ is thus the solution of the simple optimization problem

$$\hat{\rho}_n^{(\text{pen})} := \hat{\rho}_n(\hat{\kappa}), \quad \text{where} \quad \hat{\kappa} := \arg \min_{\kappa=1, \dots, d} \left\{ \sum_{i=\kappa+1}^d \hat{\lambda}_i^2 + \nu^2 \cdot \kappa \right\} = \max \left\{ \kappa : \hat{\lambda}_\kappa^2 \geq \nu^2 \right\}. \quad (4.2)$$

This means that the eigenvalues below a certain noise threshold are set to zero while those above the threshold remain unchanged. The following theorem is our first main result, and shows that the appropriate threshold is given by the upper bound on the operator norm error of the LSE, as established in proposition 1.

Theorem 1. *Let $\theta > 0$ be an arbitrary constant, let $c(\theta) := 1 + 2/\theta$, and let $\varepsilon > 0$ be a small parameter. Then with probability larger than $1 - \varepsilon$, we have*

$$\|\hat{\rho}_n^{(\text{pen})} - \rho\|_2^2 \leq \min_{\kappa=1, \dots, d} \left\{ c^2(\theta) \sum_{j>\kappa} \lambda_j^2(\rho) + 2c(\theta)\nu(\varepsilon)^2\kappa \right\}, \quad (4.3)$$

where $\hat{\rho}_n^{(\text{pen})}$ is the penalized estimator defined in (4.2) with threshold $\nu(\varepsilon)^2$ given by

$$\nu(\varepsilon)^2 := \frac{2}{n} \left(\frac{2}{3} \right)^k \log \left(\frac{2^{k+1}}{\varepsilon} \right) = \frac{2d}{N} \log \frac{2d}{\varepsilon},$$

which is assumed to be $o(1)$ with increasing n and k .

Proof. The upper bound follows directly from proposition 1 combined with the following oracle inequality established in [8],

$$\|\hat{\rho}_n^{(\text{pen})} - \rho\|_2^2 \leq \min_{\kappa=1, \dots, d} \left\{ c^2(\theta) \min_{R: \text{rank}(R)=\kappa} \|R - \rho\|_2^2 + 2c(\theta) \cdot \kappa \cdot \nu(\varepsilon)^2 \right\}$$

which holds true provided that $(1 + \theta) \|\hat{\rho}_n^{(\text{ds})} - \rho\|^2 \leq \nu(\varepsilon)^2$. This event occurs with probability larger than $1 - \varepsilon$. □

Let us make some explanatory remarks on the above result. Firstly, the bound (4.3) applies to *all* states ρ , not only ‘small’ rank ones. Recall that $\mathcal{S}_{d,r}$ denotes the set of states of rank- r states on \mathbb{C}^d . In the special case when $\text{rank}(\rho) = r \leq d$, the theorem implies that, with probability larger than $1 - \varepsilon$,

$$\|\hat{\rho}_n^{(\text{pen})} - \rho\|_2^2 \leq 2c(\theta) \cdot r \cdot \nu(\varepsilon)^2.$$

If the rank r is much smaller than d , this bound is a significant improvement to the corresponding upper bound $d \cdot \nu(\varepsilon)^2$ for the LSE, which can be derived by combining the operator norm bound of proposition 1 and the matrix norms inequality $\|\tau\|_2^2 \leq d\|\tau\|^2$, for $\tau \in M(\mathbb{C}^d)$. Moreover, up to a constant factor the rate $\nu(\varepsilon)^2$ is equal to $d \cdot r \log(d)/N$ which is essentially the ratio of the number of parameters and total number of measurements. In section 5 we will show that apart from the log factor this rate is also optimal, and cannot be improved even if the rank of the state is *known*, which indicates that the estimator adapts to the complexity of the true parameter. Furthermore, we stress the fact that the bound (4.3) holds true for growing dimension $d = 2^k$ as well as the number of measurements n ; the bound remains meaningful as far as $d \log d/N \rightarrow 0$.

The second observation is that our procedure selects the true rank consistently. Denote by $\hat{\kappa}$ the rank of the resulting estimator $\hat{\rho}_n^{(\text{pen})}$. Following [8] we can prove that, if there exists some κ such that $\lambda_\kappa(\rho) > (1 + \delta)\sqrt{\nu(\varepsilon)}$ and $\lambda_{\kappa+1}(\rho) < (1 - \delta)\sqrt{\nu(\varepsilon)}$ for some $\delta \in (0, 1)$, then

$$\mathbb{P}(\hat{\kappa} = \kappa) = 1 - \mathbb{P}(\|\hat{\rho}_n^{(\text{ds})} - \rho\| \geq \delta\sqrt{\nu(\varepsilon)}).$$

This stresses the fact that the procedure detects the eigenvalues above a threshold related to the error of the LSE. If the true rank of ρ is r and if k and n are such that ν tends to 0 (which always occurs for fixed number of ions k), then $\lambda_r > \nu$ asymptotically and the probability that $\hat{\kappa} = r$ tends to 1.

We can also project $\hat{\rho}_n^{(\text{pen})}$ on the matrices with trace 1, to get

$$\hat{\rho}_n^{(\text{pen},n)} = \arg \min_{R \in \mathcal{S}} \|R - \hat{\rho}_n^{(\text{pen})}\|_2^2, \quad (4.4)$$

where \mathcal{S} is the set of all density matrices on \mathbb{C}^d . The following corollary shows that the key properties of the estimator are preserved if we additionally normalize it to trace-one after thresholding.

Corollary 1. *Under the notation and assumptions of theorem 1 if ρ is an arbitrary state in $\mathcal{S}_{d,r}$ and if k and n are such that $\lambda_r > \nu(\varepsilon)$ for some $\varepsilon \in (0, 1)$, then, with probability larger than $1 - \varepsilon$,*

$$\|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 \leq 8c(\theta) \cdot r \cdot \nu(\varepsilon)^2.$$

Moreover, there exists an absolute constant $C > 0$ such that

$$\sup_{\rho \in \mathcal{S}_{d,r}} E_{\rho} \|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 \leq C \frac{rd}{N} \log \left(\frac{2d}{\varepsilon} \right).$$

Proof. See appendix A.4. □

4.2. Physical threshold estimator

Although the rank-penalized estimator performs well in terms of its risk, it is not necessarily positive and trace-one and therefore it may not represent a physical state. In this section we propose and analyse the following ‘physical estimator’

$$\hat{\rho}_n^{(\text{phys})} = \arg \min_{\sigma \in \mathcal{S}(\nu)} \|\sigma - \tilde{\rho}_n^{(\text{ls})}\|_2^2, \quad (4.5)$$

where $\tilde{\rho}_n^{(\text{ls})}$ is the ‘normalized LSE’ defined in (3.2), and $\mathcal{S}(\nu)$ denotes the set of states at noise level ν

$$\mathcal{S}(\nu) = \left\{ \sigma \text{ density matrix with eigenvalues } \lambda_j \in \{0\} \cup (4\nu, 1], j = 1, \dots, d \right\}.$$

In particular, the space of all density matrices correspond to $\nu = 0$ and is denoted \mathcal{S} . The estimator $\hat{\rho}_n^{(\text{phys})}$ is therefore the physical state which is closest to the (normalized) LSE, and whose non-zero eigenvalues are above the threshold 4ν .

Before analysing the performance of the estimator, we describe its numerical implementation through the following simple iterative algorithm.

Let $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_d$ denote the eigenvalues of $\tilde{\rho}_n^{(\text{ls})}$. Let $\ell = 0$, and define $\hat{\lambda}_j^{(0)} = \tilde{\lambda}_j$ for $j = 1, \dots, d$.

For $\ell = 1, \dots, d$, do

if $\hat{\lambda}_{d-\ell+1}^{(\ell-1)} > 4\nu$, STOP;

else, put $\hat{\lambda}_{d-\ell+1}^{(\ell)} = 0$ and

$$\hat{\lambda}_j^{(\ell)} = \hat{\lambda}_j^{(\ell-1)} + \frac{1}{d-\ell} \left(1 - \sum_{k=1}^{d-\ell} \hat{\lambda}_k^{(\ell-1)} \right), \text{ for } j = 1, \dots, d-\ell;$$

$\ell = \ell + 1$.

The algorithm checks whether the smallest eigenvalue is larger than the noise level 4ν and if it is not, then sets its value to 0 and distributes the mass of the erased eigenvalue in such a way that they sum to 1. This algorithm is similar to that proposed by Smolin *et al* [4], with the important difference that we do not keep all positive eigenvalues but only *significantly* positive eigenvalues. Here, significant means larger than the noise threshold of the order of the operator-norm error of the LSE. The reason for keeping only eigenvalues above the threshold is similar to the case of the penalized estimator. By proposition 1, eigenvalues below the threshold are of the order of the estimation error $\nu(\varepsilon)$. Therefore, if the state is low rank, such eigenvalues are likely to correspond to zero eigenvalues of the true state and keeping them unchanged would induce significant statistical noise.

If the total number of iterations is $\hat{\ell} = d - \hat{r}$ then the estimator $\hat{\rho}_n^{(\text{phys})}$ has rank \hat{r} . Its eigenvalues are equal to 0 for $j > \hat{r}$, while, for $j \leq \hat{r}$ they are given by

$$\hat{\lambda}_j^{(\text{phys})} = \tilde{\lambda}_j + \frac{L}{2}, \quad \text{where} \quad \frac{L}{2} \hat{r} = \sum_{k>\hat{r}} \tilde{\lambda}_k = 1 - \sum_{k \leq \hat{r}} \tilde{\lambda}_k.$$

This implies that $\hat{\rho}_n^{(\text{phys})}$ has decreasing eigenvalues and $\hat{\lambda}_{\hat{r}}^{(\text{phys})} \geq 4\nu$. The following theorem shows that $\hat{\rho}_n^{(\text{phys})}$ is *rank-consistent* and its MSE has the same scaling as that of penalized estimator $\hat{\rho}_n^{(\text{pen})}$.

Theorem 2. Assume that the state ρ has rank r , i.e. belongs to $\mathcal{S}_{d,r}$. For small $\varepsilon > 0$, let $\nu = \nu(\varepsilon)$ be defined as in theorem 1, and assume that $\lambda_r > 8\nu(\varepsilon)$. Then, with \mathbb{P}_{ρ} probability larger than $1 - \varepsilon$ we have $\hat{r} = r$ and

$$\|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 \leq 48 \cdot r \cdot \nu(\varepsilon)^2.$$

Moreover, there exists an absolute constant $C > 0$ such that

$$\sup_{\rho \in \mathcal{S}_{d,r}} E_{\rho} \|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 \leq C \frac{rd}{N} \log \left(\frac{2d}{\varepsilon} \right).$$

Proof. See appendix A.5.

5. Lower bounds for rank-constrained estimation

The goal of this section is to investigate how the convergence rates of our estimators compare with that of an ‘optimal estimator’ for the statistical model consisting of all states of rank up to r . For this we will derive a lower bound for the maximum risk of any estimator.

In this section $\hat{\rho}_n$ will be an arbitrary estimator and the true state ρ is assumed to belong to the set $\mathcal{S}_{d,r}$ of rank- r states. To quantify the overall performance of $\hat{\rho}_n$, we define the maximum risk

$$R_{\max}(\hat{\rho}_n; r) = \sup_{\rho \in \mathcal{S}_{d,r}} \mathbb{E}_{\rho} \|\hat{\rho}_n - \rho\|_2^2.$$

In view of the previous upper bounds, we expect its asymptotic behaviour (in terms of the total number of measurements, for a large number of repetitions n) to be

$$R_{\max}(\hat{\rho}_n; r) = \frac{rd}{N} \cdot O(1), \quad N = n \cdot 3^k.$$

Taking this into account we define the (appropriately rescaled) minimax risk as

$$R_{\min\max}(r, k; n) := \inf_{\hat{\rho}_n} NR_{\max}(\hat{\rho}_n; r), \tag{5.1}$$

which describes the behaviour of the best estimator at the hardest to estimate state. The next theorem provides an asymptotic lower bound for the minimax risk. It shows that the maximum MSE of any estimator is at least of the order of $r(d - r)/N$, which for low rank states scales as #parameters/#samples, which up to logarithmic factors is the same as the upper bounds derived in theorems 1, corollary 1, and theorem 2.

Theorem 3. *The following lower bound holds for the asymptotic minimax risk holds*

$$\liminf_{n \rightarrow \infty} R_{\min\max}(r, k; n) \geq 2r(d - r).$$

Proof. The minimax risk captures the worst asymptotic behaviour of the rescaled risk, over *all* states of rank r . In order to bound the risk from *below*, we construct a (lower dimensional) subfamily of states $\mathcal{R}_{d,r} \subset \mathcal{S}_{d,r}$ such that the maximum risk for this subfamily provides the lower bound. Let

$$\rho_0 := \text{Diag}\left(\frac{1}{r}, \dots, \frac{1}{r}, 0, \dots, 0\right) \tag{5.2}$$

be a diagonal state with respect to the standard basis \mathbf{B} (consisting of tensor products of eigenvectors of the Pauli σ_z operator), and define $\mathcal{R}_{d,r}$ to be the set of matrices obtained by rotating ρ_0 with an arbitrary unitary U , i.e. $\mathcal{R}_{d,r} := \{\rho := U\rho_0 U^* \mid U \text{ unitary}\}$. This is a smooth, compact manifold of dimension $2r(d - r)$ known as a (complex) Grassmannian [36]. For each point $\rho = U\rho_0 U^*$ we consider the ONB \mathbf{B}_U of eigenvectors of ρ , which is obtained by rotating \mathbf{B} by the unitary U . With respect to this basis, we consider first the parametrization of an *arbitrary* density matrix ρ' by its matrix elements, more precisely by the diagonal, real and imaginary parts of the off-diagonal matrix elements, such that $\rho' \equiv \rho_{\theta}$ with

$$\begin{aligned} \theta &= (\theta^{(d)}, \theta^{(r)}, \theta^{(i)}) \\ &:= (\rho_{11}, \dots, \rho_{dd}; \text{Re } \rho_{1,2}, \dots, \text{Re } \rho_{d-1,d}; \text{Im } \rho_{1,2}, \dots, \text{Im } \rho_{d-1,d}). \end{aligned} \tag{5.3}$$

The norm-two distance is given by

$$\|\rho_{\theta_1} - \rho_{\theta_2}\|_2^2 = \|\theta_1^d - \theta_2^d\|^2 + 2\|\theta_1^r - \theta_2^r\|^2 + 2\|\theta_1^i - \theta_2^i\|^2 = (\theta_1 - \theta_2)^T G (\theta_1 - \theta_2),$$

where G is the constant diagonal weight matrix $G = \text{Diag}(1_d, 2 \cdot 1_{d(d-1)/2}, 2 \cdot 1_{d(d-1)/2})$. However, this parametrization does not take into account the prior information about the rank of the true state, and moreover, our key argument involves the even smaller family $\mathcal{R}_{d,r}$ of states. We will now focus on providing a *local* parametrization of $\mathcal{R}_{d,r}$ around $\rho = U\rho_0 U^*$. With respect to the basis \mathbf{B}_U , a state $\rho' \in \mathcal{R}_{d,r}$ in the neighbourhood of ρ has the form

$$\rho' = \rho + \Delta_{\text{off}} + \delta = \begin{pmatrix} \frac{1}{r} I_r & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \Delta \\ \Delta^\dagger & 0 \end{pmatrix} + \begin{pmatrix} O(\|\Delta\|^2) & 0 \\ 0 & O(\|\Delta\|^2) \end{pmatrix}, \tag{5.4}$$

where Δ is a matrix of free (complex) parameters, and the two $O(\|\Delta\|^2)$ blocks are $r \times r$ and respectively $(d - r) \times (d - r)$ matrices whose elements scale quadratically in Δ near $\Delta = 0$. The intuition behind this decomposition is that a small rotation of ρ produces off-diagonal blocks of the size of the ‘rotation angles’ while the change in the diagonal blocks are only quadratic in those angles. Since we are interested in the asymptotic

behaviour of estimators, the local approach is justified, and the leading contribution to the norm-two distance comes from the off-diagonal blocks. More precisely, if $\rho_1, \rho_2 \in \mathcal{R}_{d,r}$ are in the neighbourhood of ρ then

$$\|\rho_1 - \rho_2\|^2 = 2\|\tilde{\theta}_1^r - \tilde{\theta}_2^r\|^2 + 2\|\tilde{\theta}_1^i - \tilde{\theta}_2^i\|^2 + O(\|\tilde{\theta}_1\|^4, \|\tilde{\theta}_2\|^4),$$

where $\tilde{\theta}^r, \tilde{\theta}^i$ are the real and imaginary parts of the off-diagonal elements contained in the block Δ , i.e. for $i \leq r < j$. Locally, the manifold $\mathcal{R}_{d,r}$ can be parametrized by $\tilde{\theta} := (\tilde{\theta}^r, \tilde{\theta}^i)$.

Since $\mathcal{R}_{d,r} \subset \mathcal{S}_{d,r}$, the maximum risk for the model consisting of rank- r states is bounded from below by that of the (smaller) rotation model $\mathcal{R}_{d,r}$

$$\inf_{\hat{\rho}_n} \sup_{\rho \in \mathcal{S}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2 \geq \inf_{\hat{\rho}_n} \sup_{\rho \in \mathcal{R}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2. \quad (5.5)$$

Let π be the ‘uniform’ distribution over $\mathcal{R}_{d,r}$. To draw a sample from this distribution, one can choose a random unitary U from the Haar measure over unitaries, and defines $\rho := U\rho_0 U^*$. Then the maximum risk is bounded from below by the Bayes risk

$$\sup_{\rho \in \mathcal{R}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2 \geq \int_{\mathcal{R}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2 \pi(d\rho) \quad (5.6)$$

By applying the van Trees inequality in [37] (see also [38]) we get that

$$\int_{\mathcal{R}_{d,r}} \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2 \pi(d\rho) \geq \frac{1}{n} \int_{\mathcal{R}_{d,r}} \text{Tr}(\tilde{G}(\rho)\tilde{I}^{-1}(\rho))\pi(d\rho) - \frac{\alpha}{n^2}, \quad (5.7)$$

where $\alpha > 0$ is a constant which does not depend on n . Here, $\tilde{I}(\rho)$ is the (classical) Fisher information matrix of the the data obtained by performing one measurement for each setting, and $\tilde{G}(\rho)$ is the weight matrix corresponding to the quadratic approximation of the norm-two distance squared, around ρ . Both matrices are of dimensions $2r(d-r) = \dim(\mathcal{R}_{d,r})$, and depend on the chosen parametrization, but the trace is independent of it. Inserting (5.6) and (5.7) into (5.5), we get

$$\inf_{\hat{\rho}_n} \sup_{\rho \in \mathcal{S}_{d,r}} N \mathbb{E}_\rho \|\hat{\rho}_n - \rho\|_2^2 \geq \frac{N}{n} \int_{\mathcal{R}_{d,r}} \text{Tr}(\tilde{G}(\rho)^{1/2}\tilde{I}(\rho)^{-1}\tilde{G}(\rho)^{1/2})\pi(d\rho) - \frac{\alpha N}{n^2}.$$

Since $t \mapsto t^{-1}$ is an operator convex function we have

$$\int \tilde{G}(\rho)^{1/2}\tilde{I}(\rho)^{-1}\tilde{G}(\rho)^{1/2}\pi(d\rho) \geq \left(\int \tilde{G}(\rho)^{-1/2}\tilde{I}(\rho)\tilde{G}(\rho)^{-1/2}\pi(d\rho) \right)^{-1}$$

and by taking the limit $n \rightarrow \infty$ we obtain the asymptotic minimax lower bound

$$\liminf_{n \rightarrow \infty} R_{\min\max}(r, k) \geq 3^k \text{Tr} \left(\left(\int \tilde{G}(\rho)^{-1/2}\tilde{I}(\rho)\tilde{G}(\rho)^{-1/2}\pi(d\rho) \right)^{-1} \right), \quad (5.8)$$

where $R_{\min\max}(r, k; n)$ is the minimax risk defined in equation (5.1).

At this point we choose a convenient *local* parametrization around an arbitrary state $\rho \in \mathcal{R}_{d,r}$. As discussed in the beginning of the proof we showed that for this we can use the real and imaginary parts $\tilde{\theta} = (\tilde{\theta}^r, \tilde{\theta}^i)$ of the off-diagonal block Δ , and that the corresponding weight matrix is $\tilde{G}(\rho) = 2\mathbf{1}_{2r(d-r)}$. The lower bound (5.8) becomes

$$R_{\min\max}(r, k) \geq 3^k \cdot 2 \cdot \text{Tr} \left(\left(\int_{\mathcal{R}_{d,r}} \tilde{I}(\rho)\pi(d\rho) \right)^{-1} \right).$$

Another consequence of (5.4) is that the Fisher information matrix $\tilde{I}(\rho)$ is equal to the corresponding block of the Fisher information matrix I of the full (d^2 -dimensional) unconstrained model with parametrization θ defined in (5.3). We will now compute the average over states of the Fisher information with respect to the *Pauli bases measurements*, by showing that it is equal to the average Fisher information at ρ_0 , for the *random basis measurement*. As the different settings are measured independently, the Fisher information $\tilde{I}(\rho)$ is (and similarly for I)

$$\tilde{I}(\rho) = \sum_{\mathbf{s} \in \mathcal{S}_k} \tilde{I}(\rho|\mathbf{s}),$$

where $\tilde{I}(\rho|\mathbf{s})$ is the Fisher information corresponding to the von Neumann measurement with respect to the ONB defined by setting \mathbf{s} . More generally, with \mathbf{B}_U as defined above, we denote by $\tilde{I}(\rho|\mathbf{B}_U)$ the Fisher information corresponding to this basis. Due to the rotation symmetry, we have

$$\tilde{I}(U\rho U^*|\mathbf{B}_U) = \tilde{I}(\rho|\mathbf{B})$$

so

$$\begin{aligned}\int \pi(d\rho)\bar{I}(\rho) &= \sum_{\mathbf{s}} \int \pi(d\rho)\bar{I}(\rho|\mathbf{s}) = \sum_{\mathbf{s}} \int \mu^d(dU)\bar{I}(U\rho_0U^*|\mathbf{s}) \\ &= 3^k \int \mu(dU)\bar{I}(\rho_0|\mathbf{B}_U) = 3^k\bar{I},\end{aligned}$$

where $\mu^d(dU)$ is the unique Haar measure on the unitary group on \mathbb{C}^d .

The average Fisher information matrix \bar{I} (of the full, unconstrained model) is computed in section A.6 where we show that the block corresponding to $\tilde{\theta}$ parameters has average $\bar{I} = 2\mathbf{1}_{2r(d-r)}$ such that the lower bound is

$$R_{\min\max}(r, k) \geq 2r(d-r).$$

6. Numerical results

In this section we present the results of a simulation study which analyses the performance of the proposed estimation methods. The penalized and physical estimators discussed in the previous sections use a theoretical penalization and respective threshold rates proportional to ν^2 . However in practice we found that the performance of the estimators can be further improved when the rates are adjusted by multiplying with an appropriate constant c —whose choice is informed by the data—from a grid over a small interval which was chosen to be $[0, 3]$. The last two estimators are such versions of the theoretical ones with constant c chosen by using cross-validation methods which are explained in detail in section 6.1. We will compare the following 5 estimators described below.

- (1) The LSE $\hat{\rho}_n^{(\text{ls})}$ defined in (3.1).
- (2) The oracle ‘estimator’ $\hat{\rho}_n^{(\text{oracle})}$ defined below. This is strictly speaking *not* an estimator since it requires the knowledge of the state ρ itself, and can be computed only in simulation studies. However, the oracle is a useful benchmark for evaluating the performance of the other estimators.
- (3) The cross-validated projection estimator $\hat{\rho}_n^{(\text{cv})}$. Here we try to find the optimal truncation rank of the LSE, by using the cross-validation method.
- (4) The cross-validated penalized estimator $\hat{\rho}_n^{(\text{pen-cv})}$. This is a modification of the penalized estimator $\hat{\rho}_n^{(\text{pen})}$ defined in (4.2), where the value of the penalization constant is adjusted by cross-validation.
- (5) The cross-validated physical threshold estimator $\hat{\rho}_n^{(\text{phys-cv})}$. This is a modification of the physical estimator $\hat{\rho}_n^{(\text{phys})}$ defined in (4.5), where the value of the threshold constant is adjusted by cross-validation. We explore the estimators’ behaviour by simulating datasets from states with different ranks, and with different number of measurement repetitions per setting. The methodology is described in detail below.

6.1. Generation of random states and simulation of datasets

In order to generate a density matrix of rank r , we first create a rank r upper triangular matrix T in which

- (i) the off-diagonal elements of the first r rows are random complex numbers,
- (ii) the diagonal elements T_{22}, \dots, T_{rr} are real, positive random numbers,
- (iii) all elements of the rows $r+1, \dots, d$ are zero.

The matrix T is completed by setting T_{11} such that $T_{11} \geq 0$, and $\|T\|_2^2 = 1$. If these conditions cannot be satisfied we repeat the procedure by generating a new set of matrix elements for T . When successful, we set $\rho := T^*T$ which by construction is a density matrix of rank r . We note that it is not our purpose to generate matrices from a particular ‘uniform’ ensemble, but merely to have a state with reasonably random eigenvectors, and whose r eigenvalues are not significantly smaller than $1/r$. Following this procedure we have generated 4 states of 4 ions ($d = 2^4$) with ranks 1, 2, 6, 10. The rank 6 state for instance, has non-zero eigenvalues (0.47, 0.19, 0.12, 0.11, 0.07, 0.04).

For each state, we have then simulated a number of 100 independent datasets with a given number of repetitions chosen from the range 20, 100, 500, and 2500. In this way we can study the dependence of the MSE of each estimator on state (or rank) and number of repetitions.

6.2. Computation of estimators

We conducted the following simulation study for all the possible combinations between the states and the total number of cycles (i.e. $4 \times 4 = 16$ different scenarios). Below, we denote by r the rank of the ‘true’ state ρ , from which the data has been generated. The procedure has the following steps:

- (1) For a given number of repetitions n , we simulate 5 independent datasets D_1, \dots, D_5 , each with $n/5$ repetitions. By simply adding the number of counts for each setting and outcome, we obtain a dataset D of n repetitions. However, as we will see below, having 5 separate ‘smaller’ datasets is important for the purpose of applying cross-validation. Note that such a procedure can be easily implemented in an experimental setting.
- (2) We compute the LSE $\hat{\rho}_n^{(\text{ls})}$ based on the full dataset D with total number of cycles n .
- (3) We compute the *oracle estimator* as follows:

- (a) We compute the spectral decomposition (4.1) of $\hat{\rho}_n^{(\text{ls})}$, with the eigenvalues $\hat{\lambda}_i$ arranged in decreasing order of their absolute values. For each rank $1 \leq \kappa \leq d$ we define the truncated (least squares) matrix

$$\hat{\rho}_n(\kappa) = \sum_{i=1}^{\kappa} \hat{\lambda}_i |\hat{\psi}_i\rangle \langle \hat{\psi}_i|.$$

- (b) We then evaluate the norm-two distance squared $e(\kappa) := \|\rho - \hat{\rho}_n(\kappa)\|_2^2$ and define the oracle estimator as the truncated estimator with minimal norm two error

$$\hat{\rho}_n^{(\text{oracle})} = \hat{\rho}_n(\kappa_0), \quad \kappa_0 = \arg \min_{\kappa} e(\kappa).$$

Note that the oracle estimator relies on the knowledge of the true state ρ which is not available in a real data set-up. It is nevertheless useful as a benchmark for judging the performance of other estimators in simulation studies. At the next point we define the cross-validation estimator which tries to find the ‘optimal’ rank κ_0 by replacing the unknown state ρ with the LSE computed on a separate batch of data.

- (4) We compute the *cross validation estimator* as follows.

- (a) For each $j \in \{1, \dots, 5\}$ we compute the following estimators. While holding the batch D_j out, we compute the LSE $\hat{\rho}_{n,-j}^{(\text{ls})}$ for the dataset consisting of joining the remaining four batches together. Similarly to the point above, we define the rank κ truncation of this estimator by $\hat{\rho}_{n,-j}(\kappa)$. We also compute the LSE for the remaining batch j , denoted by $\hat{\rho}_{n,j}^{(\text{ls})}$.
- (b) For each rank κ we evaluate the ‘empirical discrepancy’

$$\text{CV}(\kappa) = \frac{1}{5} \sum_{i=1}^5 \left\| \hat{\rho}_{n,-i}(\kappa) - \hat{\rho}_{n,i}^{(\text{ls})} \right\|_2^2.$$

Since $\hat{\rho}_{n,-j}(\kappa)$ and $\hat{\rho}_{n,j}^{(\text{ls})}$ are independent, and the LSE is unbiased $\mathbb{E}[\hat{\rho}_{n,j}^{(\text{ls})}] = \rho$, the expected value of $\text{CV}(k)$ is

$$\begin{aligned} \mathbb{E}[\text{CV}(\kappa)] &= \mathbb{E}_{-1} \left[\text{Tr} \left(\hat{\rho}_{n,-1}(\kappa)^2 \right) \right] - 2 \text{Tr} \left(\mathbb{E}_{-1} \left[\hat{\rho}_{n,-1}(\kappa) \right] \cdot \mathbb{E}_1 \left[\hat{\rho}_{n,1}^{(\text{ls})} \right] \right) + \mathbb{E}_1 \left[\text{Tr} \left(\left(\hat{\rho}_{n,1}^{(\text{ls})} \right)^2 \right) \right] \\ &= \mathbb{E}_{-1} \left[\text{Tr} \left(\hat{\rho}_{n,-1}(\kappa)^2 \right) \right] - 2 \text{Tr} \left(\mathbb{E}_{-1} \left[\hat{\rho}_{n,-1}(\kappa) \right] \rho \right) + \text{Tr}(\rho^2) + C \\ &= \mathbb{E}_{-1} \left[\left\| \hat{\rho}_{n,-1}(\kappa) - \rho \right\|^2 \right] + C, \end{aligned}$$

where we denoted \mathbb{E}_{-1} and \mathbb{E}_1 the expectation over all batches except the first, and respectively over the first batch. Therefore, the average of $\text{CV}(\kappa)$ is equal to the MSE of the truncated estimator $\hat{\rho}_{n,-1}(\kappa)$, up to a constant C which is independent of κ .

- (c) Based on the above observation we use the the cross-validation method as a proxy for the oracle estimator. Concretely, we minimize $\text{CV}(\kappa)$ with respect to κ

$$\hat{\kappa}_{\text{cv}} = \arg \min_{\kappa} \text{CV}(\kappa),$$

and define the cross-validation estimator as the truncation to rank $\hat{\kappa}_{\text{cv}}$ of the *full data* LSE $\hat{\rho}_n^{(\text{cv})} := \hat{\rho}_n(\hat{\kappa}_{\text{cv}})$.

- (5) We compute the *cross-validated rank-penalized estimator* as follows.

- (a) Let c be a penalization constant chosen from a suitable set of discrete values in the interval $[0, 3]$. Similarly to the cross-validation procedure, we hold out batch j , and we compute the rank-penalized estimator (4.2), with penalty constant $c\nu^2$ for $j = 1, \dots, 5$. We denote these estimators by $\hat{\rho}_{n;-j}^{(\text{pen})}(c)$. We will also need the LSE $\rho_{n;j}^{(\text{ls})}$ for batch j computed above.

- (b) For each value of c we evaluate the empirical discrepancy

$$\text{CV}(c) = \frac{1}{5} \sum_{i=1}^5 \left\| \hat{\rho}_{n;-j}^{(\text{pen})}(c) - \rho_{n;j}^{(\text{ls})} \right\|_2^2$$

and minimize $\text{CV}(c)$ with respect to the constant c

$$\hat{c} = \arg \min_c \text{CV}(c).$$

Finally we compute the cross-validated rank penalized estimator $\hat{\rho}_n^{(\text{rk-cv})}$ which is defined as in (4.2), with constant $\hat{c}\nu^2$, on the whole dataset D .

- (6) We compute the *physical estimator* as follows.

- (a) As above we choose a constant c from a grid over the interval $[0, 3]$. We hold out batch j , and we compute the physical threshold estimator (4.5), using the algorithm below this equation, with threshold $c \cdot 4\nu$. We denote the resulting estimators by $\hat{\rho}_{n;-j}^{(\text{phys})}(c)$, for $j = 1, \dots, 5$.

- (b) For each value of c we evaluate the empirical discrepancy

$$\text{CV}(c) = \frac{1}{5} \sum_{i=1}^5 \left\| \hat{\rho}_{n;-j}^{(\text{phys})}(c) - \hat{\rho}_{n;j}^{(\text{ls})} \right\|_2^2.$$

We then minimize $\text{CV}(c)$ with respect to the constant c .

$$\hat{c} = \arg \min_c \text{CV}(c)$$

Finally we compute the cross-validated physical estimator $\hat{\rho}_n^{(\text{phys-cv})}$ which is defined as in (4.5), with constant $\hat{c} \cdot 4\nu$.

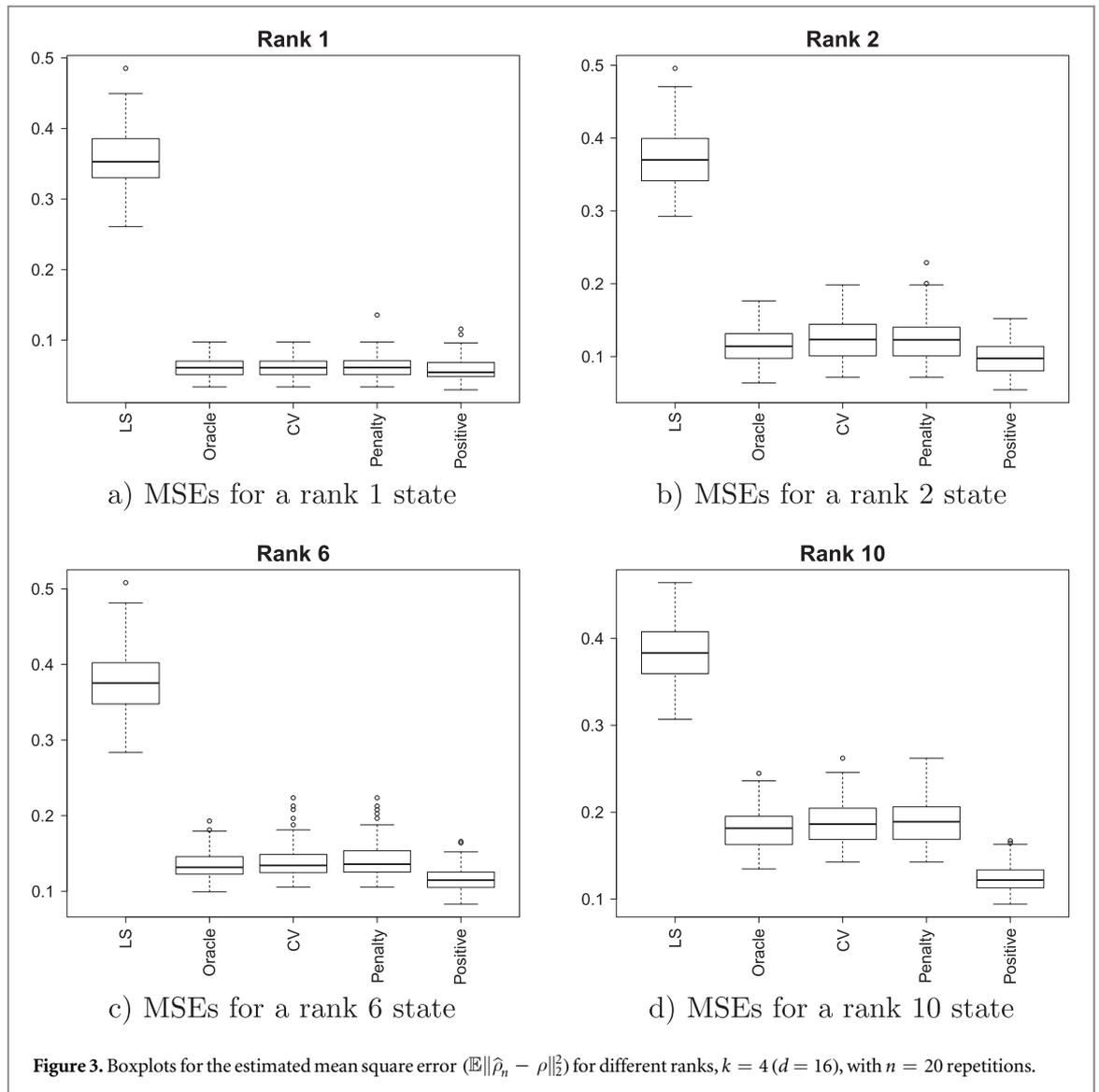
6.3. Simulation results

We collect here the results of the simulation study described in the previous section. As a figure of merit we focus on the MSE $\mathbb{E}(\|\hat{\rho}_n - \rho\|_2^2)$ of each estimator, which is estimated by averaging the square errors over the 100 independent repetitions of the procedure. We are also interested in how the different methods perform relative to each other, and whether the selected rank is consistent, i.e. it concentrates on the rank of the true state for large number of repetitions.

The four panels in figure 3 represent the boxplots of the square errors $\|\hat{\rho}_n - \rho\|_2^2$ for the different estimators, and different states, when the number of repetitions is $n = 20$. Similarly, figure 4 shows the same boxplots at $n = 100$. As expected, in both cases the least squares performs significantly worse than the other estimators, and the discrepancy is larger for small rank states. The remaining 4 estimators have similar MSE's with the physical one performing slightly better than the rest, followed by the oracle. Note also that the estimators' variances (indicated by the size of the boxes) are larger for the least squares than the other estimators. A similar behaviour has been observed for $n = 500, 2500$ repetitions.

Figure 5 illustrates the dependence of the MSE of a given estimator, as a function of n , for the four different states which have been analysed. Since the MSE decreases as n^{-1} we have chosen to plot the 'renormalized' MSE given by $n \cdot \mathbb{E}\|\hat{\rho}_n - \rho\|_2^2$, which converges to a constant value for large n . As expected the limiting value increases with the rank of the state, as a proxy for the number of parameters to be estimated.

The histograms in figure 6 show the probability distributions for the chosen rank of each given estimator, as a function of the number of measurement repetitions n , for the state of rank 6. We note that in all cases the proportion of times that the chosen rank is equal to the true rank of the state increases as with the number of repetitions. However, this convergence towards a 'rank-consistent' estimator is rather slow, as the proportion surpasses 80% only when $n = 2500$. Another observation is that the penalty and threshold estimators appear to have different behaviours: the former tends to underestimate the true rank, while the latter tends to overestimate it. As expected, the oracle estimator is more likely to choose the correct rank for large number of repetitions. Perhaps slightly more surprising, for small number of repetitions ($n = 20$), the oracle choose a pure state in most cases.

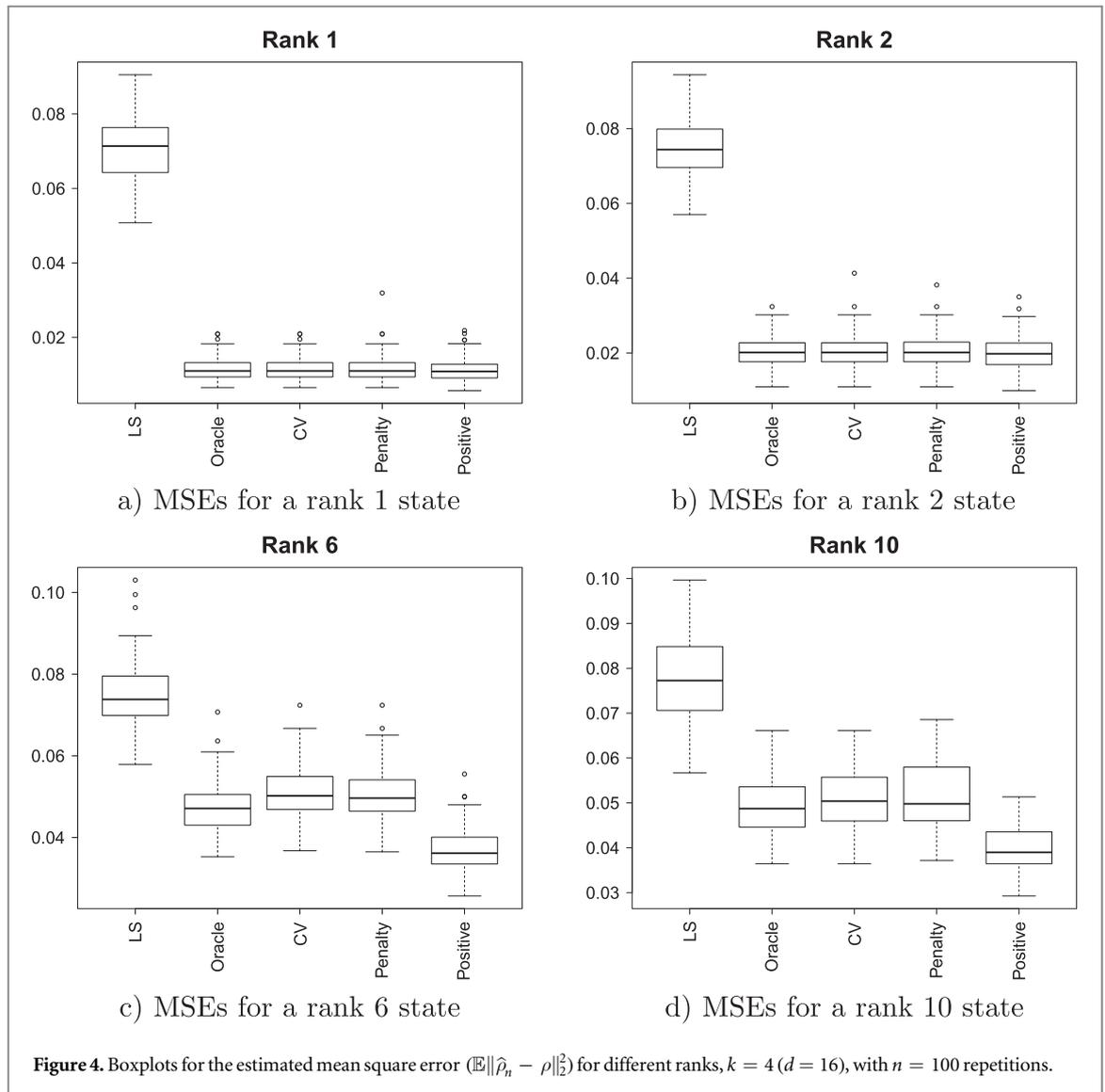


7. Conclusions and outlook

Quantum state tomography, and in particular MIT is an important enabling component of quantum engineering experiments. Since full quantum tomography becomes unfeasible for large dimensional systems, it is useful to identify lower dimensional models with good approximation properties for physically relevant states, and to develop estimation methods tailored for such models. In particular, quantum states created in the lab are often very well approximated by low rank density matrices, which are characterized by a number of parameter which is linear rather than quadratic in the space dimension.

In this work we analysed several estimation algorithms targeted at estimating low rank states in MIT. The procedure consists in computing the LSE, which is then diagonalized, truncated to an appropriate smaller rank by setting eigenvalues below a ‘noise threshold’ to zero, and normalized. Among the several truncation methods proposed, the best performing one is the ‘physical estimator’; this chooses the density matrix whose non-zero eigenvalues are above a certain threshold and is the closest to the LSE. We proved concentration bounds and upper bounds for the MSE of the penalized and physical estimators, as well as a lower bound for the asymptotic minimax rate for MIT. The results show that the proposed methods have an optimal dependence on rank and dimension, up to a logarithmic factor in dimension. In addition, the algorithms are easy to implement numerically and their computational complexity is determined by that of the LSE.

An interesting future direction is to extend the spectral thresholding methodology to a measurement setup where a smaller number of settings is measured, which is however sufficient to identify the unknown low rank state. Another direction involves the construction of confidence intervals / regions for such estimators, beyond the concentration bounds established here. Since the main ingredients are the LSE and the spectral thresholding algorithms, the proposed estimation methods can be applied to any quantum tomography scenario with



informationally complete measurements. It would be interesting to see how these methods perform in other settings.

Acknowledgments

MG's work was supported by the EPSRC Grant No. EP/J009776/1.

Appendix A.

A.1. Proof of lemma 1

Proof. Since the matrix elements of \mathbf{A} are $A_{\mathbf{b},(\mathbf{o},\mathbf{s})} = A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})$, the product $\mathbf{A}^* \cdot \mathbf{A}$ has elements

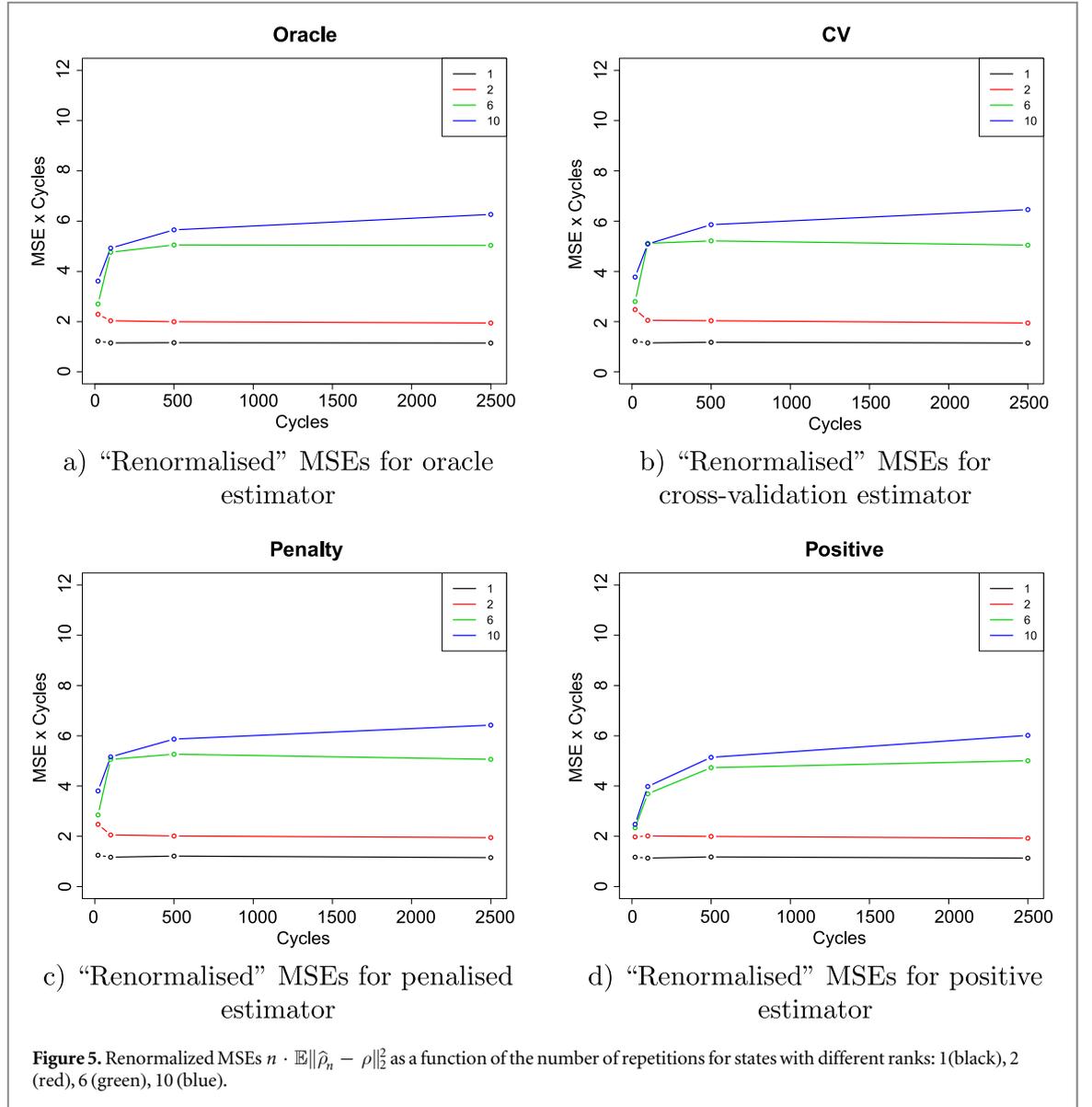
$$[\mathbf{A}^* \cdot \mathbf{A}]_{\mathbf{b},\mathbf{b}'} = \sum_{\mathbf{s}} \sum_{\mathbf{o}} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s}).$$

If $\mathbf{b} = \mathbf{b}'$ it is easy to see that

$$[\mathbf{A}^* \cdot \mathbf{A}]_{\mathbf{b},\mathbf{b}} = \sum_{\mathbf{s}} \sum_{\mathbf{o}} \prod_{j \notin E_{\mathbf{b}}} \delta_{s_j, b_j} = 2^k 3^{d(\mathbf{b})}.$$

If $\mathbf{b} \neq \mathbf{b}'$, we have either $E_{\mathbf{b}} = E_{\mathbf{b}'}$ or $E_{\mathbf{b}} \neq E_{\mathbf{b}'}$. On the one hand, in case the sets $E_{\mathbf{b}}$ and $E_{\mathbf{b}'}$ are equal, we have

$$A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s}) = \prod_{j \notin E_{\mathbf{b}}} \delta_{s_j, b_j} \cdot \delta_{s_j, b'_j}.$$



For each \mathbf{s} , the previous product is 0. Indeed, if different from 0 then $b_j = b'_j$ for all j not in E_b . As $b_j = b'_j = I$ for j in E_b , it implies that $\mathbf{b} = \mathbf{b}'$ which contradicts the assumption here.

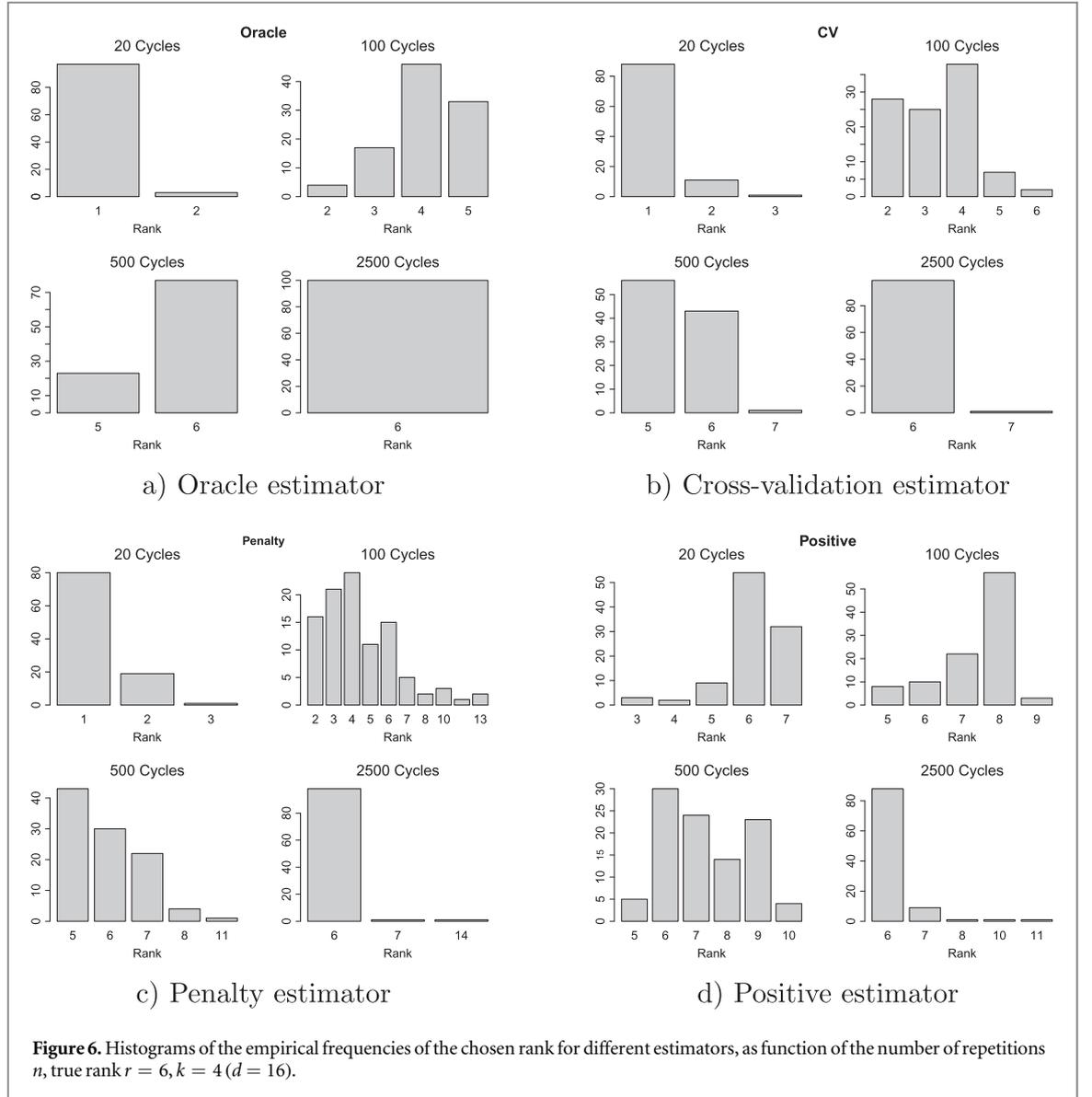
On the other hand, if the sets E_b and $E_{b'}$ are different, there exists at least one coordinate j_0 in the symmetric difference $E_b \Delta E_{b'}$ and the sum over outcomes \mathbf{o} will split over values of \mathbf{o} where $o_{j_0} = 1$ and values where $o_{j_0} = -1$:

$$\begin{aligned} & \sum_{\mathbf{o}} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s}) \\ &= \sum_{\mathbf{o}: o_{j_0}=1} \delta_{s_{j_0}, b_{j_0}} \prod_{j \notin E_b} o_j \delta_{s_j, b_j} \cdot \prod_{l \notin E_{b'}} o_l \delta_{s_l, b'_l} \\ & - \sum_{\mathbf{o}: o_{j_0}=-1} \delta_{s_{j_0}, b_{j_0}} \prod_{j \notin E_b} o_j \delta_{s_j, b_j} \cdot \prod_{l \notin E_{b'}} o_l \delta_{s_l, b'_l} = 0. \end{aligned}$$

We assumed here that j_0 belongs to $E_b \setminus E_{b'}$ and the same holds for j_0 in $E_{b'} \setminus E_b$. □

A.2. Proof of proposition 1

Proof of proposition 1. For more clarity, in this proof we will use the notation $I(a = b) = \delta_{a,b}$. Note that $f(\mathbf{o}|\mathbf{s}) = \frac{1}{n} \sum_i I(X_{s,i} = \mathbf{o})$, where the random variables $X_{s,i}$ are independent for all settings \mathbf{s} and all i from 1 to n . To estimate the risk of the linear estimator we write



$$\begin{aligned}
 \hat{\rho}_n^{(ls)} - \rho &= \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} (f(\mathbf{o}|\mathbf{s}) - p(\mathbf{o}|\mathbf{s})) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^{k3^d(\mathbf{b})}} \sigma_{\mathbf{b}} \\
 &= \sum_{\mathbf{b}} \sum_{\mathbf{o}} \sum_{\mathbf{s}} \frac{1}{n} \sum_i (I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})) \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^{k3^d(\mathbf{b})}} \sigma_{\mathbf{b}} \\
 &:= \sum_{\mathbf{s}} \sum_i W_{\mathbf{s},i},
 \end{aligned}$$

where $W_{\mathbf{s},i}$ are independent and centred Hermitian random matrices. We will apply the following extension of the Bernstein matrix inequality [39] due to [40], see also [32, 33].

Proposition 3. (Bernstein inequality, [40].) Let Y_1, \dots, Y_n be independent, centred, $m \times m$ Hermitian random matrices. Suppose that, for some constants $V, W > 0$ we have $\|Y_j\| \leq V$, for all j from 1 to n , and that $\|\sum_j \mathbb{E}(Y_j^2)\| \leq W$. Then, for all $t \geq 0$,

$$\mathbb{P}(\|Y_1 + \dots + Y_n\| \geq t) \leq 2m \exp\left(-\frac{t^2/2}{W + tV/3}\right).$$

In our setup we bound $\|W_{\mathbf{s},i}\| \leq V$ for all \mathbf{s} and i and $\|\sum_{\mathbf{s}} \mathbb{E}(W_{\mathbf{s},i}^* W_{\mathbf{s},i})\| \leq W$, where V, W are evaluated below. We have

$$\begin{aligned} \|W_{s,i}\| &\leq \frac{1}{n} \sum_{\mathbf{b}} \sum_{\mathbf{o}} \left| \frac{A_{\mathbf{b}}(\mathbf{o}|\mathbf{s})}{2^k 3^{d(\mathbf{b})}} \right| \cdot |I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})| \cdot \|\sigma_{\mathbf{b}}\| \\ &\leq \frac{1}{n} \sum_{\mathbf{b}} \frac{1}{2^k 3^{d(\mathbf{b})}} \prod_{j \notin E_{\mathbf{b}}} I(b_j = s_j) \sum_{\mathbf{o}} |I(X_{s,i} = \mathbf{o}) - p(\mathbf{o}|\mathbf{s})| \\ &\leq \frac{2}{n 2^k} \sum_{\ell=0}^k \sum_{\mathbf{b}: d(\mathbf{b})=\ell} \frac{1}{3^\ell} = \frac{2}{n 2^k} \sum_{\ell=0}^k \binom{k}{\ell} \frac{1}{3^\ell} = \frac{2}{n 2^k} \left(1 + \frac{1}{3}\right)^k = \frac{2}{n} \left(\frac{2}{3}\right)^k := V. \end{aligned}$$

Let us denote $B(\mathbf{o}|\mathbf{s}) := \sum_{\mathbf{b}} 2^{-k} 3^{-d(\mathbf{b})} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) \sigma_{\mathbf{b}}$. Then

$$\begin{aligned} &\left\| \sum_{\mathbf{s}} \sum_i \mathbb{E}(W_{s,i}^* W_{s,i}) \right\| \\ &= \frac{1}{n^2} \left\| \sum_{\mathbf{s}} \sum_i \sum_{\mathbf{o}, \mathbf{o}'} B^*(\mathbf{o}|\mathbf{s}) \cdot \text{Cov}(I(X_{s,i} = \mathbf{o}), I(X_{s,i} = \mathbf{o}')) \cdot B(\mathbf{o}'|\mathbf{s}) \right\| \\ &\leq \frac{1}{n^2} \left\| \sum_{\mathbf{s}} \sum_i \sum_{\mathbf{o}} B^*(\mathbf{o}|\mathbf{s}) \cdot B(\mathbf{o}|\mathbf{s}) \right\|. \end{aligned} \tag{A.1}$$

The last inequality follows from the fact that the covariance matrix can be written as the difference of two positive matrices $\text{Cov}_{\mathbf{o}, \mathbf{o}'} := \text{Cov}(I(X_{s,i} = \mathbf{o}), I(X_{s,i} = \mathbf{o}')) = p(\mathbf{o}|\mathbf{s})\delta_{\mathbf{o}, \mathbf{o}'} - p(\mathbf{o}|\mathbf{s})p(\mathbf{o}'|\mathbf{s})$ and since $p(\mathbf{o}|\mathbf{s}) \leq 1$, we get $\text{Cov} \leq \mathbf{1}$.

We replace $B(\mathbf{o}|\mathbf{s})$ in (A.1) and use lemma 1 to get

$$\begin{aligned} &\left\| \sum_{\mathbf{s}} \sum_i \mathbb{E}(W_{s,i}^* W_{s,i}) \right\| \\ &\leq \frac{1}{n} \left\| \sum_{\mathbf{s}} \sum_{\mathbf{o}} \sum_{\mathbf{b}, \mathbf{b}'} \frac{1}{2^k 3^{d(\mathbf{b})+d(\mathbf{b}')}} A_{\mathbf{b}}(\mathbf{o}|\mathbf{s}) A_{\mathbf{b}'}(\mathbf{o}|\mathbf{s}) \cdot \sigma_{\mathbf{b}} \sigma_{\mathbf{b}'} \right\| \\ &\leq \frac{1}{n 2^k} \sum_{\ell=0}^k \sum_{\mathbf{b}: d(\mathbf{b})=\ell} \frac{1}{3^\ell} = \frac{1}{n 2^k} \sum_{\ell=0}^k \binom{k}{\ell} \frac{1}{3^\ell} = \frac{1}{n} \left(\frac{2}{3}\right)^k := W. \end{aligned}$$

Apply the matrix Bernstein inequality in proposition 3 to get, for any $t > 0$:

$$\mathbb{P}_\rho \left(\|\hat{\rho}_n^{(\text{ls})} - \rho\| \geq t \right) \leq 2^{k+1} \exp \left(- \frac{t^2}{1 + 2t/3} \frac{n}{2} \left(\frac{3}{2}\right)^k \right).$$

We choose $t > 0$ such that

$$\varepsilon = 2^{k+1} \exp \left(- \frac{t^2}{1 + 2t/3} \cdot \frac{n}{2} \left(\frac{3}{2}\right)^k \right),$$

which leads to t such that

$$\frac{t^2}{1 + 2t/3} = \frac{2}{n} \left(\frac{2}{3}\right)^k \log \left(\frac{2^{k+1}}{\varepsilon} \right) = \frac{3}{2} \nu(\varepsilon).$$

Then the convenient choice of t is $\nu(\varepsilon) = \nu(\varepsilon)/2 + \sqrt{\nu^2(\varepsilon)/2 + 3/2 \cdot \nu(\varepsilon)}$, which is equivalent to $\sqrt{3/2 \cdot \nu(\varepsilon)}$ when this last term tends to 0. □

A.3. Proof of proposition 1

Proof of proposition 1. Let us denote by $(\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_d)$ the eigenvalues of $|\hat{\rho}_n^{(\text{ls})}|$ and by $\tilde{\lambda}_1, \dots, \tilde{\lambda}_d$ the eigenvalues of the resulting estimator $\hat{\rho}_n^{(\text{ls}, n)}$. It is easy to see that the latter has the same eigenvectors as $\hat{\rho}_n^{(\text{ls})}$.

Therefore

$$(\tilde{\lambda}_1, \dots, \tilde{\lambda}_d) = \underset{\lambda: \sum_{j=1}^d \lambda_j = 1}{\text{argmin}} \sum_{j=1}^d (\lambda_j - \hat{\lambda}_j)^2.$$

This optimization has the explicit solution

$$\tilde{\lambda}_j = \hat{\lambda}_j + \frac{L}{2}, \text{ where } \frac{L}{2} d = 1 - \sum_{j=1}^d \hat{\lambda}_j.$$

Note that

$$\frac{L}{2} = \frac{1}{d} \text{tr}(\rho - \hat{\rho}_n^{(\text{ls})}) \leq \|\hat{\rho}_n^{(\text{ls})} - \rho\|.$$

Therefore, $\|\hat{\rho}_n^{(\text{ls},n)} - \hat{\rho}_n^{(\text{ls})}\| = L/2 \leq \|\hat{\rho}_n^{(\text{ls})} - \rho\|$ and thus

$$\|\hat{\rho}_n^{(\text{ls},n)} - \rho\| \leq \|\hat{\rho}_n^{(\text{ls})} - \rho\| + \|\hat{\rho}_n^{(\text{ls},n)} - \hat{\rho}_n^{(\text{ls})}\| \leq 2\nu(\varepsilon).$$

□

A.4. Proof of corollary 1

Proof of corollary 1. We order the eigenvalues of $\hat{\rho}_n^{(\text{pen})}$ in decreasing order of absolute values ($|\hat{\lambda}_1| \geq \dots \geq |\hat{\lambda}_d|$), and denote by $\tilde{\lambda}_1, \dots, \tilde{\lambda}_d$ the eigenvalues of $\hat{\rho}_n^{(\text{pen},n)}$. As in the previous proof, we can see that both matrices will share the same eigenvectors and that the relation between the eigenvalues is

$$\tilde{\lambda}_j = \hat{\lambda}_j + \frac{L}{2}, \text{ where } \frac{L}{2}d = 1 - \sum_{j=1}^d \hat{\lambda}_j.$$

Thus

$$\|\hat{\rho}_n^{(\text{pen})} - \hat{\rho}_n^{(\text{pen},n)}\|_F^2 = d(L/2)^2 \leq \text{Tr}^2(\hat{\rho}_n^{(\text{pen})} - \rho)/d \leq \|\hat{\rho}_n^{(\text{pen})} - \rho\|_F^2.$$

We deduce that $\|\rho - \hat{\rho}_n^{(\text{pen},n)}\|_F \leq \|\hat{\rho}_n^{(\text{pen})} - \hat{\rho}_n^{(\text{pen},n)}\|_F + \|\hat{\rho}_n^{(\text{pen})} - \rho\|_F \leq 2\|\hat{\rho}_n^{(\text{pen})} - \rho\|_F$.
Therefore

$$\mathbb{P}\left(\|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_F > 8c(\theta)r\nu(\varepsilon)^2\right) < \varepsilon$$

and the previous inequality remains true for all $0 < e \leq \varepsilon < 1$. Let us denote by

$$x(e) = 8r\nu^2(e) = 16\frac{rd}{N} \log\left(\frac{2d}{e}\right),$$

which is a decreasing function of e . This implies that $e = 2d \exp\left(-\frac{Nx(e)}{16rd}\right)$. Thus

$$\begin{aligned} \mathbb{E}_\rho \|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 &= \int_0^\infty \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 > x) dx \\ &= \int_0^{x(\varepsilon)} \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 > x) dx + \int_{x(\varepsilon)}^\infty \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{pen},n)} - \rho\|_2^2 > x) dx \\ &\leq x(\varepsilon) + \int_{x(\varepsilon)}^\infty 2d \exp\left(-\frac{Nx}{16rd}\right) dx \\ &\leq 16\frac{rd}{N} \log\left(\frac{2d}{\varepsilon}\right) + 2d \cdot 16\frac{rd}{N} \exp\left(-\frac{N}{16rd}x(\varepsilon)\right) \\ &\leq 16\frac{rd}{N} \left(\log\left(\frac{2d}{\varepsilon}\right) + \varepsilon\right) \leq C\frac{rd}{N} \log\left(\frac{2d}{\varepsilon}\right). \end{aligned}$$

□

A.5. Proof upper bound physical estimator

Proof of theorem 2. We recall from proposition 2 that with probability larger than $1 - \varepsilon$, we have $\|\hat{\rho}_n^{(\text{ls})} - \rho\| \leq 2\nu(\varepsilon)$. In particular, by using the Weyl inequality [41], this implies that $|\tilde{\lambda}_k - \lambda_k| \leq 2\nu(\varepsilon)$ for all k from 1 to d , where $\lambda_1, \dots, \lambda_d$ are the eigenvalues of ρ arranged in decreasing order. After a total of $\hat{\ell} = d - \hat{r}$ iterations the algorithm stops and we have [42]

$$\hat{\lambda}_{\hat{r}+1}^{(\text{phys})} = \dots = \hat{\lambda}_d^{(\text{phys})} = 0$$

and

$$\hat{\lambda}_j^{(\text{phys})} := \hat{\lambda}_j^{(\hat{\ell})} = \tilde{\lambda}_j + \frac{1}{\hat{r}} \sum_{k>\hat{r}} \tilde{\lambda}_k \quad j = 1, \dots, \hat{r}.$$

From now on we assume that $\|\hat{\rho}_n^{(\text{ls})} - \rho\| \leq 2\nu(\varepsilon)$ which holds with probability larger than $1 - \varepsilon$, see corollary 3.2. We will show that under this assumption $\hat{r} = r$. For this we consider the two cases $\hat{r} > r$ and $\hat{r} < r$ separately.

Suppose $\hat{r} > r$. For $j \leq \hat{r}$ we have

$$\begin{aligned} |\hat{\lambda}_j^{(\text{phys})} - \lambda_j| &= \left| \tilde{\lambda}_j - \lambda_j + \frac{1}{\hat{r}} \left(1 - \sum_{k \leq \hat{r}} \tilde{\lambda}_k \right) \right| \\ &= \left| \tilde{\lambda}_j - \lambda_j + \frac{1}{\hat{r}} \sum_{k \leq \hat{r}} (\lambda_k - \tilde{\lambda}_k) \right| \\ &\leq |\tilde{\lambda}_j - \lambda_j| + \frac{1}{\hat{r}} \sum_{k \leq \hat{r}} |\tilde{\lambda}_k - \lambda_k| \leq 2\nu + 2\nu = 4\nu. \end{aligned}$$

This implies that $\hat{\lambda}_{\hat{r}}^{(\text{phys})} \leq \lambda_{\hat{r}} + 4\nu = 4\nu$ since $\hat{r} > r$. However, as discussed above, by the definition of the threshold estimator, $\hat{\lambda}_{\hat{r}}^{(\text{phys})} \geq 4\nu$. Therefore \hat{r} cannot be larger than r .

Suppose $\hat{r} < r$. Then $\hat{\lambda}_r^{(\text{phys})} \geq \lambda_r - |\hat{\lambda}_r^{(\text{phys})} - \lambda_r| > 8\nu - 4\nu = 4\nu$. However this is not possible since $\hat{\lambda}_r^{(\text{phys})} = 0$ under the assumption $\hat{r} < r$. Thus, $\hat{r} = r$ with probability larger than $1 - \epsilon$.

We consider now the MSE of the estimator. As shown above, we have $|\hat{\lambda}_j^{(\text{phys})} - \lambda_j| \leq 4\nu$. Let $\rho = UDU^*$ be the diagonalization of ρ , where U is unitary and D is the diagonal matrix of eigenvalues. Similarly, let $\hat{\rho}_n^{(\text{ls})} = \hat{U}_n D_n^{(\text{ls})} \hat{U}_n^*$ and $\hat{\rho}_n^{(\text{phys})} = \hat{U}_n D_n^{(\text{phys})} \hat{U}_n^*$ be the decompositions of the least squares and the physical threshold estimator, where we take into account that the latter two have the same eigenbasis. Then

$$\begin{aligned} \|\hat{\rho}_n^{(\text{phys})} - \rho\|_2 &= \|\hat{U}_n D_n^{(\text{phys})} \hat{U}_n^* - UDU^*\|_2 \\ &\leq \|\hat{U}_n D_n^{(\text{phys})} \hat{U}_n^* - \hat{U}_n D \hat{U}_n^*\|_2 + \|\hat{U}_n D \hat{U}_n^* - UDU^*\|_2 \\ &\leq \|D_n^{(\text{phys})} - D\|_2 + \|\hat{U}_n D \hat{U}_n^* - UDU^*\|_2 \end{aligned} \quad (\text{A.2})$$

The first norm on the right-hand side of the previous inequality is bounded as

$$\|D_n^{(\text{phys})} - D\|_2^2 = \sum_{j \leq \hat{r}} \left(\hat{\lambda}_j^{(\text{phys})} - \lambda_j \right)^2 \leq 16r\nu^2. \quad (\text{A.3})$$

For the second norm we use a similar triangle inequality for the *operator norm*

$$\begin{aligned} \|\hat{U}_n D \hat{U}_n^* - UDU^*\| &\leq \|\hat{U}_n D \hat{U}_n^* - \hat{U}_n D_n^{(\text{ls})} \hat{U}_n^*\| + \|\hat{U}_n D_n^{(\text{ls})} \hat{U}_n^* - UDU^*\| \\ &= \|D - D_n^{(\text{ls})}\| + \|\hat{\rho}_n^{(\text{ls})} - \rho\| \leq \nu + \nu = 2\nu. \end{aligned}$$

The first term is smaller than ν since $|\lambda_i - \hat{\lambda}_i^{(\text{ls})}| \leq \nu$ as it follows from proposition 1, and the Weyl inequality [41]. The second term is also bounded by ν , by proposition 1. Therefore, since $\hat{U}_n D \hat{U}_n^*$ and UDU^* are rank r matrices, the difference is at most rank $2r$ and $\|\hat{U}_n D \hat{U}_n^* - UDU^*\|_2 \leq 2\nu\sqrt{2r}$. By plugging this together with (A.3) into the right side of (A.2) we get

$$\|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 \leq (2\sqrt{2} + 4)^2 r\nu^2,$$

with probability larger than $1 - \epsilon$.

Moreover, the previous inequality remains true for all $0 < e \leq \epsilon < 1$. Let us denote by

$$x(e) = 48r\nu^2(e) = 96 \frac{rd}{N} \log\left(\frac{2d}{e}\right),$$

which is a decreasing function of e . This implies that $e = 2d \exp\left(-\frac{Nx(e)}{96rd}\right)$. Thus

$$\begin{aligned} \mathbb{E}_\rho \|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 &= \int_0^\infty \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 > x) dx \\ &= \int_0^{x(\epsilon)} \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 > x) dx + \int_{x(\epsilon)}^\infty \mathbb{P}_\rho(\|\hat{\rho}_n^{(\text{phys})} - \rho\|_2^2 > x) dx \\ &\leq x(\epsilon) + \int_{x(\epsilon)}^\infty 2d \exp\left(-\frac{Nx}{96rd}\right) dx \\ &\leq 96 \frac{rd}{N} \log\left(\frac{2d}{\epsilon}\right) + 2d \cdot 96 \frac{rd}{N} \exp\left(-\frac{N}{96rd} x(\epsilon)\right) \\ &\leq 96 \frac{rd}{N} \left(\log\left(\frac{2d}{\epsilon}\right) + \epsilon \right) \leq C \frac{rd}{N} \log\left(\frac{2d}{\epsilon}\right). \end{aligned}$$

A.6. The average Fisher information matrix for the full, unconstrained model, with random measurement design

In this section we present a detailed calculation of the average quantum Fisher information at a the rank r state ρ_0 defined in (5.2), for random measurements with uniform distribution over the measurement basis. We consider the full parametrization by $\theta \in \mathbb{R}^{d^2}$ given by equation (5.3). As explained in the proof, the Fisher information matrix for the parametrization $\tilde{\theta}$ of the rotation model $\mathcal{R}_{d,r}$ is a particular block of the larger Fisher matrix computed here. We will come back to this at the end of the computation.

Let

$$\mathbf{B}_U := \{ |\mathbf{o}; U\rangle := U |\mathbf{o}\rangle : \mathbf{o} = 1, \dots, 2^k \}$$

denote the ONB obtained by rotating the standard basis by the unitary U .

The Fisher info $I(\rho|\mathbf{B}_U)$ associated to the measurement with ONB \mathbf{B}_U is given by the matrix

$$I(\rho|\mathbf{B}_U)_{a,b} = \sum_{\mathbf{o}: p_{\rho}(\mathbf{o}|\mathbf{B}_U) > 0} \frac{1}{p_{\rho}(\mathbf{o}|\mathbf{B}_U)} \frac{\partial p_{\rho}(\mathbf{o}|\mathbf{B}_U)}{\partial \theta_a} \frac{\partial p_{\rho}(\mathbf{o}|\mathbf{B}_U)}{\partial \theta_b},$$

where $p_{\rho}(\mathbf{o}|\mathbf{B}_U)$ is the probability of the outcome \mathbf{o}

$$\begin{aligned} p_{\rho}(\mathbf{o}|\mathbf{B}_U) &= \langle \mathbf{o}, U|\rho|\mathbf{o}, U\rangle = \sum_i \rho_{ii} |\langle \mathbf{o}, U|i\rangle|^2 \\ &+ 2 \sum_{i < j} \operatorname{Re}(\rho_{i,j}) \operatorname{Re}(\langle i|\mathbf{o}, U\rangle \langle \mathbf{o}, U|j\rangle) + 2 \sum_{i < j} \operatorname{Im}(\rho_{i,j}) \operatorname{Im}(\langle i|\mathbf{o}, U\rangle \langle \mathbf{o}, U|j\rangle) \end{aligned}$$

and the partial derivatives are given by

$$\begin{aligned} \frac{\partial p_{\rho}(\mathbf{o}|\mathbf{B}_U)}{\partial \rho_{i,i}} &= |\langle \mathbf{o}, U|i\rangle|^2 \\ \frac{\partial p_{\rho}(\mathbf{o}|\mathbf{B}_U)}{\partial \operatorname{Re} \rho_{i,j}} &= 2 \operatorname{Re}(\langle i|\mathbf{o}, U\rangle \langle \mathbf{o}, U|j\rangle) \\ \frac{\partial p_{\rho}(\mathbf{o}|\mathbf{B}_U)}{\partial \operatorname{Im} \rho_{i,j}} &= 2 \operatorname{Im}(\langle i|\mathbf{o}, U\rangle \langle \mathbf{o}, U|j\rangle). \end{aligned}$$

The Fisher information matrix $I(\rho|\mathbf{B}_U)$ has the following block structure

$$I(\rho|\mathbf{B}_U) = \begin{pmatrix} I^{dd}(\rho|\mathbf{B}_U) & I^{dr}(\rho|\mathbf{B}_U) & I^{di}(\rho|\mathbf{B}_U) \\ I^{rd}(\rho|\mathbf{B}_U) & I^{rr}(\rho|\mathbf{B}_U) & I^{ri}(\rho|\mathbf{B}_U) \\ I^{id}(\rho|\mathbf{B}_U) & I^{ir}(\rho|\mathbf{B}_U) & I^{ii}(\rho|\mathbf{B}_U) \end{pmatrix}$$

with superscripts indicating the type of parameter considered: diagonal, real or imaginary part of off-diagonal element. The average Fisher information for a randomly chosen basis is

$$\bar{I}(\rho) := \int \mu(dU) I(\rho|\mathbf{B}_U),$$

where $\mu(dU)$ is the Haar measure over unitaries used for choosing the random basis. Note that by symmetry $\bar{I}(\rho)$ only depends on the spectrum of ρ . We will not compute $\bar{I}(\rho)$ for an arbitrary state ρ but only at $\rho = \rho_0$. The corresponding Fisher information will be denoted $\bar{I} = \bar{I}(\rho_0)$ and is a function of d and r . Below we compute the different blocks of \bar{I} . For the matrix elements we will use a suggestive notation, e.g. $\bar{I}_{ii;jk}^{d,r}$ denotes the element corresponding to the diagonal parameter $\theta_{ii}^d = \rho_{ii}$ and the real part of the off-diagonal element ρ_{jk} , etc.

(A) *Diagonal–diagonal block.*

$$I(\rho|\mathbf{B}_U)_{ii;jj}^{dd} = \sum_{\mathbf{o}: p_{\rho}(\mathbf{o}|\mathbf{B}_U) > 0} \frac{1}{p_{\rho}(\mathbf{o}|\mathbf{B}_U)} |\langle \mathbf{o}, U|i\rangle|^2 |\langle \mathbf{o}, U|j\rangle|^2.$$

By integrating over unitaries we obtain the corresponding matrix element of the average Fisher matrix \bar{I} . Since $p_{\rho_0}(\mathbf{o}|\mathbf{B}_U) > 0$ is true for all \mathbf{o} , with probability one with respect to $\mu^d(dU)$, we drop the condition from the sum. At the state $\rho = \rho_0$ defined in (5.2), we have

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= r \sum_{\mathbf{o}} \int \frac{|\langle \mathbf{o}, U|i \rangle|^2 |\langle \mathbf{o}, U|j \rangle|^2}{\sum_{k=1}^r |\langle \mathbf{o}, U|k \rangle|^2} \mu^d(dU) \\ &= r \cdot d \int \frac{|\langle \mathbf{1}, U|i \rangle|^2 |\langle \mathbf{1}, U|j \rangle|^2}{\sum_{k=1}^r |\langle \mathbf{1}, U|k \rangle|^2} \mu^d(dU) \\ &= r \cdot d \int \frac{|\langle \psi|i \rangle|^2 |\langle \psi|j \rangle|^2}{\sum_{k=1}^r |\langle \psi|k \rangle|^2} \nu^d(d\psi), \end{aligned}$$

where $\nu^d(d\psi)$ is the uniform measure over the projective space on \mathbb{C}^d . To compute the integral we decompose $|\psi\rangle$ as

$$|\psi\rangle = q|\psi_1\rangle + \sqrt{1 - q^2}|\psi_2\rangle$$

with $|\psi_1\rangle \in \mathcal{H}_r := \text{Span}\{|1\rangle, \dots, |r\rangle\}$ and $|\psi_2\rangle \in \mathcal{H}_r^\perp$ normalized (orthogonal) vectors, and $0 \leq q \leq 1$. The uniform measure can be expressed as

$$\nu^d(d\psi) = m^{d,r}(dq) \times \nu^r(d\psi_1) \times \nu^{d-r}(d\psi_2),$$

where $m^{d,r}(dq)$ is the distribution of the length of the projection of a random vector in \mathbb{C}^d onto an r -dimensional subspace. With this notation we have

$$\bar{I}_{ii;jj}^{dd} = r \cdot d \int \int \frac{1}{q^2} |\langle \psi|i \rangle|^2 |\langle \psi|j \rangle|^2 m^{d,r}(dq) \nu^r(d\psi_1) \nu^{d-r}(d\psi_2). \tag{A.4}$$

We distinguish four sub-cases depending on whether each of the indices i, j belongs to $\{1, \dots, r\}$ or $\{r + 1, \dots, d\}$.

Sub-case 1: $i, j \leq r$. In this case (A.4) becomes

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= r \cdot d \int \frac{1}{q^2} q^4 m^{d,r}(dq) \times \int |\langle \psi_1|i \rangle|^2 |\langle \psi_1|j \rangle|^2 \nu^r(d\psi_1) \\ &= r \cdot d \int \frac{1}{q^2} q^4 m^{d,r}(dq) \times \int U_{1,i}^* U_{i,1} U_{1,j}^* U_{j,1} \mu^r(dU). \end{aligned} \tag{A.5}$$

In the last line we have re-written $|\psi_1\rangle$ as $U|1\rangle$ in order to use the existing formulas for integrals of monomials over the unitary group [43]. Since we will use these formulas repeatedly, we recall that

$$\int U_{i_1,j_1} U_{i_2,j_2} U_{i'_1,i'_1}^* U_{i'_2,i'_2}^* \mu^l(dU) = \sum_{\sigma, \tau \in S_2} \delta_{i_1, \tau \sigma_1} \delta_{i_2, \tau \sigma_2} \delta_{j_1, \sigma_1} \delta_{j_2, \sigma_2} \text{Wg}^l(\sigma \tau^{-1}), \tag{A.6}$$

where σ and τ are permutations in $S_2 = \{(1, 1), (2)\}$ and $\text{Wg}^l(\cdot)$ is the Weingarten function over S_2 :

$$\text{Wg}^l(1, 1) = \frac{1}{l^2 - 1}, \quad \text{Wg}^l(2) = \frac{-1}{l(l^2 - 1)}.$$

The integral over q can be easily evaluated as

$$\int q^2 m^{d,r}(dq) = \sum_{k=1}^r \int |\langle k|\psi \rangle|^2 \mu^d(d\psi) = \frac{r}{d}. \tag{A.7}$$

By inserting (A.6) and (A.7) into (A.5) we get

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= r \cdot d \cdot \frac{r}{d} \left(\frac{1}{r^2 - 1} - \frac{1}{r(r^2 - 1)} \right) = \frac{r}{r + 1}, \quad i \neq j \\ \bar{I}_{ii;ii}^{dd} &= r \cdot d \cdot \frac{r}{d} \cdot 2 \left(\frac{1}{r^2 - 1} - \frac{1}{r(r^2 - 1)} \right) = \frac{2r}{r + 1}. \end{aligned}$$

Sub-case 2: $i \leq r$ and $j > r$. From (A.4) we get

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= r \cdot d \int \frac{1}{q^2} q^2 (1 - q^2) m^{d,r}(dq) \times \int |\langle \psi_1|i \rangle|^2 \nu^r(d\psi_1) \times \int |\langle \psi_2|j \rangle|^2 \nu^{d-r}(d\psi_2) \\ &= r \cdot d \left(1 - \frac{r}{d} \right) \cdot \frac{1}{r} \cdot \frac{1}{d - r} = 1. \end{aligned}$$

Sub-case 3: $i, j > r$. Similarly, in this case

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= r \cdot d \int \frac{1}{q^2} (1 - q^2)^2 m^{d,r}(dq) \times \int |\langle \psi_2 | i \rangle|^2 |\langle \psi_2 | j \rangle|^2 \nu^{d-r}(d\psi_2) \\ &= r \cdot d \int \frac{1}{q^2} (1 - q^2)^2 m^{d,r}(dq) \times \int U_{1,i}^* U_{i,1} U_{1,j}^* U_{j,1} \mu^{d-r}(dU). \end{aligned} \tag{A.8}$$

We first simplify the integral on the right side

$$\int \frac{1}{q^2} (1 - q^2)^2 m^{d,r}(dq) = \int \frac{1}{q^2} m^{d,r}(dq) - 2 + \frac{r}{d}. \tag{A.9}$$

To evaluate the remaining integral, we consider a multivariate Gaussian random variable $c = (c_1, \dots, c_{2d}) \sim N(0, I_{2d})$, and denote by $g^{2d}(dc)$ its probability distribution. From this we construct the complex vector with uniform distribution over the unit ball in \mathbb{C}^d

$$|\psi\rangle := \frac{1}{\|c\|} \sum_{i=1}^d (c_i + ic_{d+1}) |i\rangle, \quad \|c\|^2 = \sum_{l=1}^{2d} c_l^2.$$

We can now write

$$\begin{aligned} \int \frac{1}{q^2} m^{d,r}(dq) &= \int g^{2d}(dc) \frac{\sum_{l=1}^{2d} c_l^2}{\sum_{l=1}^{2r} c_l^2} = 1 + \int g^{2d}(dc) \frac{\sum_{l=2r+1}^{2d} c_l^2}{\sum_{l=1}^{2r} c_l^2} \\ &= 1 + \int \|c^1\|^2 g^{2(d-r)}(dc^1) \cdot \int \frac{1}{\|c^2\|^2} g^{2r}(dc^2) \\ &= 1 + 2(d-r) \cdot \frac{1}{2r-2} = \frac{d-1}{r-1}. \end{aligned} \tag{A.10}$$

Above we used the fact that $\|c^2\|^2$ is a χ^2 variable with $2r$ degrees of freedom, so the second integral is the mean of its inverse. By inserting (A.10) into (A.9) we obtain

$$\int \frac{1}{q^2} (1 - q^2)^2 m^{d,r}(dq) = -2 + \frac{r}{d} + \frac{d-1}{r-1} = \frac{(d-r)(d-r+1)}{d(r-1)}. \tag{A.11}$$

Finally, by inserting (A.11) into (A.8) and applying (A.6) we obtain

$$\begin{aligned} \bar{I}_{ii;jj}^{dd} &= rd \cdot \frac{(d-r)(d-r+1)}{d(r-1)} \cdot \left(\frac{1}{(d-r)^2 - 1} - \frac{1}{(d-r)((d-r)^2 - 1)} \right) \\ &= \frac{r}{r-1}, \quad i \neq j \\ \bar{I}_{ii;ii}^{dd} &= \frac{2r}{r-1}. \end{aligned}$$

Note that for $r = 1$ these matrix elements are infinite. This is due to large contributions from measurements which have one basis vector close to being orthogonal to the one-dimensional vector state.. This is somewhat akin to what happens in the case of a Bernoulli variable (coin toss), in the case when the probability is close to zero or one. While this phenomenon is interesting, it does not play any role in our analysis for which the diagonal matrix elements are not relevant parameters.

(B) Diagonal-real block.

$$I(\rho | \mathbf{B}_U)_{ii;jk}^{dr} = 2 \sum_{\mathbf{o}: p_{\mathbf{o}}(\mathbf{o} | \mathbf{B}_U) > 0} \frac{1}{p_{\mathbf{o}}(\mathbf{o} | \mathbf{B}_U)} |\langle \mathbf{o}, U | i \rangle|^2 \cdot \text{Re}(\langle j | \mathbf{o}, U \rangle \langle \mathbf{o}, U | k \rangle).$$

As before, the corresponding entry of the average Fisher information matrix is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int \frac{|\langle \psi | i \rangle|^2 \cdot \text{Re}(\langle j | \psi \rangle \langle \psi | k \rangle)}{\sum_{k=1}^r |\langle \psi | k \rangle|^2} \nu^d(d\psi) \\ &= 2 \cdot d \cdot r \iiint \frac{1}{q^2} |\langle \psi | i \rangle|^2 \cdot \text{Re}(\langle j | \psi \rangle \langle \psi | k \rangle) m^{d,r}(dq) \nu^r(d\psi_1) \nu^{d-r}(d\psi_2). \end{aligned}$$

Sub-case 1: $i, j, k \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int q^2 m^{d,r}(dq) \int |\langle \psi_1 | i \rangle|^2 \operatorname{Re}(\langle j | \psi_1 \rangle \langle \psi_1 | k \rangle) \nu^r(d\psi_1) \\ &= d \cdot r \cdot \frac{r}{d} \int \mu^r(dU) U_{i1} U_{1i}^* (U_{j1} U_{1k}^* + U_{k1} U_{1j}^*) = 0, \end{aligned}$$

where we applied formula (A.6) for the integrals over the unitaries.

Sub-case 2: $i, j \leq r$ and $k > r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int q \sqrt{1 - q^2} m^{d,r}(dq) \\ &\times \operatorname{Re} \int |\langle \psi_1 | i \rangle|^2 \langle j | \psi_1 \rangle \nu^r(d\psi_1) \int \langle \psi_2 | k \rangle \nu^{d-r}(d\psi_2) = 0. \end{aligned}$$

Sub-case 3: $i \leq r$ and $j, k > r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int (1 - q^2) m^{d,r}(dq) \\ &\times \int |\langle \psi_1 | i \rangle|^2 \nu^r(d\psi_1) \cdot \operatorname{Re} \int \langle j | \psi_2 \rangle \langle \psi_2 | k \rangle \nu^{d-r}(d\psi_2) = 0. \end{aligned}$$

Sub-case 4: $i > r$ and $j, k \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int (1 - q^2) m^{d,r}(dq) \\ &\times \int |\langle \psi_2 | i \rangle|^2 \nu^{d-r}(d\psi_2) \cdot \operatorname{Re} \int \langle j | \psi_1 \rangle \langle \psi_1 | k \rangle \nu^r(d\psi_1) = 0. \end{aligned}$$

Sub-case 5: $i > r$ and $j, k \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int (1 - q^2) m^{d,r}(dq) \\ &\times \int |\langle \psi_2 | i \rangle|^2 \nu^{d-r}(d\psi_2) \cdot \operatorname{Re} \int \langle j | \psi_1 \rangle \langle \psi_1 | k \rangle \nu^r(d\psi_1) = 0. \end{aligned}$$

Sub-case 6: $i, k > r$ and $j \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int q \sqrt{1 - q^2} m^{d,r}(dq) \\ &\times \operatorname{Re} \int |\langle \psi_2 | i \rangle|^2 \langle \psi_2 | k \rangle \nu^{d-r}(d\psi_2) \cdot \int \langle j | \psi_1 \rangle \nu^r(d\psi_1) = 0. \end{aligned}$$

Sub-case 7: $i, j, k > r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ii;jk}^{dr} &= 2 \cdot d \cdot r \int \frac{(1 - q^2)^2}{q^2} m^{d,r}(dq) \\ &\times \int |\langle \psi_2 | i \rangle|^2 \operatorname{Re} \langle j | \psi_2 \rangle \langle \psi_2 | k \rangle \nu^{d-r}(d\psi_2) = 0. \end{aligned}$$

The last integral is zero for the same reason as in sub-case 1.

In conclusion all diagonal-real matrix elements are zero

$$\bar{I}_{ii;jk}^{dr} = 0, \quad \text{for all } i = 1, \dots, 2^k, \text{ and } 1 \leq j < k \leq 2^k.$$

(C) Diagonal-imaginary block.

Similarly to the case of diagonal-real elements, we obtain that all diagonal-imaginary matrix elements are zero

$$\bar{I}_{ii;jk}^{di} = 0, \quad \text{for all } i = 1, \dots, 2^k, \text{ and } 1 \leq j < k \leq 2^k.$$

(D) Real-real block.

$$I(\rho | \mathbf{B}_U)_{ij;kl}^{rr} = 4 \sum_{\mathbf{o}: p_\rho(\mathbf{o} | \mathbf{B}_U) > 0} \frac{1}{p_\rho(\mathbf{o} | \mathbf{B}_U)} \operatorname{Re}(\langle i | \mathbf{o}, U \rangle \langle \mathbf{o}, U | j \rangle) \cdot \operatorname{Re}(\langle k | \mathbf{o}, U \rangle \langle \mathbf{o}, U | l \rangle).$$

As before the corresponding entry of the average Fisher information matrix is

$$\begin{aligned} \bar{I}_{ij;kl}^{rr} &= 4 \cdot d \cdot r \int \frac{\operatorname{Re}(\langle i|\psi\rangle\langle\psi|j\rangle) \cdot \operatorname{Re}(\langle k|\psi\rangle\langle\psi|l\rangle)}{\sum_{k=1}^r |\langle\psi|k\rangle|^2} \nu^d(d\psi) \\ &= 4 \cdot d \cdot r \iiint \frac{1}{q^2} \operatorname{Re}(\langle i|\psi\rangle\langle\psi|j\rangle) \cdot \operatorname{Re}(\langle k|\psi\rangle\langle\psi|l\rangle) m^{d,r}(dq) \nu^r(d\psi_1) \nu^{d-r}(d\psi_2). \end{aligned}$$

The same reasoning as before can be applied to transform the integral into a single integral, or a product of integrals over unitaries. If a single index is larger than r while the other three are smaller, or conversely a single index is smaller than r while the other are larger, then we have two integrals over unitaries which are zero since the monomial is of odd order. Therefore the following cases remain to be analysed.

Sub-case 1: $i, j, k, l \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ij;kl}^{rr} &= 4 \cdot d \cdot r \int q^2 m^{d,r}(dq) \int \operatorname{Re}(\langle i|\psi_1\rangle\langle\psi_1|j\rangle) \cdot \operatorname{Re}(\langle k|\psi_1\rangle\langle\psi_1|l\rangle) \nu^r(d\psi_1) \\ &= d \cdot r \cdot \frac{r}{d} \cdot \int \mu^r(dU) (U_{i1}U_{1j}^* + U_{j1}U_{1i}^*) (U_{k1}U_{1l}^* + U_{l1}U_{1k}^*). \end{aligned}$$

Using formula (A.6) we find that the integral over unitaries is zero unless we deal with a diagonal matrix element of I , i.e. $i = k$ and $j = l$. For the latter we have

$$\begin{aligned} \bar{I}_{ij;ij}^{rr} &= d \cdot r \cdot \frac{r}{d} \cdot \int \mu^r(dU) (U_{i1}U_{1j}^* + U_{j1}U_{1i}^*) (U_{i1}U_{1j}^* + U_{j1}U_{1i}^*) \\ &= 2 \cdot d \cdot r \cdot \frac{r}{d} \left(\frac{1}{r^2 - 1} - \frac{1}{r(r^2 - 1)} \right) = \frac{2r}{r + 1}. \end{aligned}$$

Sub-case 2: $i, j, k, l > r$. In this case, the integral is

$$\begin{aligned} \bar{I}_{ij;kl}^{rr} &= 4 \cdot d \cdot r \int \frac{(1 - q^2)^2}{q^2} m^{d,r}(dq) \int \operatorname{Re}(\langle i|\psi_2\rangle\langle\psi_2|j\rangle) \cdot \operatorname{Re}(\langle k|\psi_2\rangle\langle\psi_2|l\rangle) \nu^{d-r}(d\psi_2) \\ &= d \cdot r \cdot \frac{(d - r)(d - r + 1)}{d(r - 1)} \cdot \int \mu^{d-r}(dU) (U_{i1}U_{1j}^* + U_{j1}U_{1i}^*) (U_{k1}U_{1l}^* + U_{l1}U_{1k}^*), \end{aligned}$$

where we have used formula (A.11) for the integral over q . A similar calculation as above shows that all off-diagonal elements are zero and

$$\begin{aligned} \bar{I}_{ij;ij}^{rr} &= 2d \cdot r \cdot \frac{(d - r)(d - r + 1)}{d(r - 1)} \cdot \left(\frac{1}{(d - r)^2 - 1} - \frac{1}{(d - r)((d - r)^2 - 1)} \right) \\ &= \frac{2r}{r - 1}. \end{aligned}$$

Sub-case 3: ($i, j \leq r$ and $k, l > r$) or ($i, j > r$ and $k, l \leq r$). In this case we deal with a product of integrals over unitaries of the form

$$\int \mu^r(dU) U_{i1}U_{1j}^* = 0.$$

So all matrix elements are zero.

Sub-case 4: $i, k \leq r$ and $j, l > r$. Again, the off-diagonal elements are zero and

$$\begin{aligned} \bar{I}_{ij;ij}^{rr} &= 4 \cdot d \cdot r \int (1 - q^2) m^{d,r}(dq) \int \operatorname{Re}(\langle i|\psi_1\rangle\langle\psi_2|j\rangle) \cdot \operatorname{Re}(\langle i|\psi_1\rangle\langle\psi_2|j\rangle) \nu^r(d\psi_1) \nu^r(d\psi_2) \\ &= 2d \cdot r \cdot \left(1 - \frac{r}{d} \right) \cdot \frac{1}{r} \frac{1}{d - r} = 2. \end{aligned}$$

In conclusion, the only non-zero elements of the real–real block are on the diagonal and are given by

$$\begin{aligned} \bar{I}_{ij;ij}^{rr} &= \frac{2r}{r + 1}, & 1 \leq i < j \leq r, \\ \bar{I}_{ij;ij}^{rr} &= 2, & i \leq r \text{ and } r < j \leq 2^k, \\ \bar{I}_{ij;ij}^{rr} &= \frac{2r}{r - 1}, & r < i < j \leq 2^k. \end{aligned}$$

(E) *Imaginary–imaginary block.* This block is similar to the real–real one. All off diagonal elements are zero, and the diagonal ones are

$$\begin{aligned} \bar{I}_{ij;ij}^{ii} &= \frac{2r}{r+1}, & 1 \leq i < j \leq r, \\ \bar{I}_{ij;ij}^{ii} &= 2, & i \leq r \text{ and } r < j \leq 2^k, \\ \bar{I}_{ij;ij}^{ii} &= \frac{2r}{r-1}, & r < i < j \leq 2^k. \end{aligned}$$

(F) *Real-imaginary block.* Next we show that all real-imaginary off-diagonal elements are equal to zero.

$$I(\rho | \mathbf{B}_U)_{ij;kl}^{ri} = -4 \cdot \sum_{\mathbf{o}: p_\rho(\mathbf{o} | \mathbf{B}_U) > 0} \frac{1}{P_\rho(\mathbf{o} | \mathbf{B}_U)} \operatorname{Re}(\langle i | \mathbf{o}, U \rangle \langle \mathbf{o}, U | j \rangle) \operatorname{Im}(\langle k | \mathbf{o}, U \rangle \langle \mathbf{o}, U | l \rangle).$$

By integration we obtain the corresponding matrix element of the average Fisher information matrix

$$\begin{aligned} \bar{I}_{ij;kl}^{ri} &= -4 \cdot d \cdot r \int \frac{\operatorname{Re}(\langle i | \psi \rangle \langle \psi | j \rangle) \cdot \operatorname{Im}(\langle k | \psi \rangle \langle \psi | l \rangle)}{\sum_{k=1}^r |\langle \psi | k \rangle|^2} \nu^d(d\psi) \\ &= -4 \cdot d \cdot r \iiint \frac{1}{q^2} \operatorname{Re}(\langle i | \psi \rangle \langle \psi | j \rangle) \cdot \operatorname{Im}(\langle k | \psi \rangle \langle \psi | l \rangle) m^{d,r}(dq) \nu^r(d\psi_1) \nu^{d-r}(d\psi_2). \end{aligned}$$

Again, if one index is smaller than r while the other three are larger, or otherwise, then the matrix element is zero. We analyse the remaining cases.

Sub-case 1: $i, j, k, l \leq r$. In this case the integral is

$$\begin{aligned} \bar{I}_{ij;kl}^{ri} &= -4 \cdot d \cdot r \int q^2 m^{d,r}(dq) \int \operatorname{Re}(\langle i | \psi_1 \rangle \langle \psi_1 | j \rangle) \operatorname{Im}(\langle k | \psi_1 \rangle \langle \psi_1 | l \rangle) \nu^r(d\psi_1) \\ &= i \cdot d \cdot r \cdot \frac{r}{d} \int \mu^r(dU) (U_{i1} U_{1j}^* + U_{j1} U_{1i}^*) (U_{k1} U_{1l}^* - U_{l1} U_{1k}^*). \end{aligned}$$

Using formula (A.6) we find that the integral over unitaries is zero unless $i = k$ and $j = l$. However even in this case, two of the four terms are zero, and the other two cancel each other.

Sub-case 2: $i, j, k, l > r$. Here the integrals over the unitaries are similar as in *case 1* above, with the difference that they taken with respect to $\mu^{d-r}(dU)$. Therefore the matrix elements are zero.

Sub-case 3: ($i, j \leq r$ and $k, l > r$) or ($i, j > r, k, l \leq r$). In this case we deal with a product of integrals over unitaries of the form

$$\int \mu^r(dU) U_{i1} U_{1j}^* = 0.$$

so all matrix elements are zero.

Sub-case 4: $i, k \leq r$ and $j, l > r$. Again, the off-diagonal elements are zero and

$$\begin{aligned} \bar{I}_{ij;ij}^{ri} &= -4 \cdot d \cdot r \int (1 - q^2) m^{d,r}(dq) \int \operatorname{Re}(\langle i | \psi_1 \rangle \langle \psi_2 | j \rangle) \cdot \operatorname{Im}(\langle i | \psi_1 \rangle \langle \psi_2 | j \rangle) \nu^r(d\psi_1) \nu^r(d\psi_2) \\ &= i \cdot d \cdot r \cdot \left(1 - \frac{r}{d}\right) \int \int (U_{i1} V_{1j}^* + V_{j1} U_{1i}^*) (U_{i1} V_{1j}^* - V_{j1} U_{1i}^*) \mu^r(dU) \mu^{d-r}(dV) \\ &= 0. \end{aligned}$$

In the last integral, two terms are zero and two have different signs and cancel each other. In conclusion all elements of the real-imaginary block are equal to zero.

Summary of the computation of \bar{I} . We found that all off-diagonal blocks of \bar{I} are zero

$$\bar{I}^{dr} = \bar{I}^{di} = \bar{I}^{ri} = \bar{I}^{rd} = \bar{I}^{id} = \bar{I}^{dr} = 0.$$

Moreover the real and imaginary diagonal blocks are diagonal, equal to each other and have three distinct values depending on the position of the indices $i < j$ with respect to r :

$$\bar{I}^{rr/ii} = \operatorname{Diag} \begin{cases} \bar{I}_{ij;ij}^{rr/ii} = \frac{2r}{r+1}, & 1 \leq i, j \leq r \\ \bar{I}_{ij;ij}^{rr/ii} = 2, & 1 \leq i \leq r, \text{ and } r < j \leq 2^k \\ \bar{I}_{ij;ij}^{rr/ii} = \frac{2r}{r-1} & r < i < j \leq 2^k. \end{cases}$$

Finally, the $d \times d$ block \tilde{I}^{dd} is not diagonal but has a simple form

$$\tilde{I}_{ii;jj}^{dd} = \begin{cases} \tilde{I}_{ii;ii}^{dd} = \frac{2r}{r+1}, & 1 \leq i \leq r \\ \tilde{I}_{ii;ii}^{dd} = \frac{2r}{r-1}, & r < i \leq 2^k \\ \tilde{I}_{ii;jj}^{dd} = \frac{r}{r+1}, & i, j \leq r, \text{ and } i \neq j \\ \tilde{I}_{ii;jj}^{dd} = \frac{r}{r+1}, & 1 \leq i \leq r \text{ and } r < j \leq 2^k \\ \tilde{I}_{ii;jj}^{dd} = \frac{r}{r+1}, & 1 \leq j \leq r \text{ and } r < i \leq 2^k \\ \tilde{I}_{ii;jj}^{dd} = \frac{r}{r-1} & r < i, j, \text{ and } i \neq j. \end{cases}$$

The model $\mathcal{R}_{d,r}$ can be seen (locally) as the restriction of the full unconstrained model $\mathcal{S}_{r,d}$ parametrized by θ , to the subset of parameters $\tilde{\theta}$ which are real and imaginary parts of matrix elements ρ_{ij} with $i \leq r < j$. Therefore the corresponding average Fisher information matrix is equal to the corresponding block of \tilde{I} , i.e. $\tilde{I} = 2\mathbb{1}_{2r(d-r)}$. □

References

- [1] Häffner H et al 2005 Scalable multipartite entanglement of trapped ions *Nature* **438** 643
- [2] Blume-Kohout R 2010 Optimal, reliable estimation of quantum states *New J. Phys.* **12** 043034
- [3] Khoo H and Englert Ng and B-G 2012 A simple minimax estimator for quantum states *Int. J. Quantum Inf.* **10** 1250038
- [4] Smolin J A, Gambetta J M and Smith G 2012 Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise *Phys. Rev. Lett.* **108** 070502
- [5] Heinosaari T, Mazzarella L and Wolf M M 2013 Quantum tomography under prior information *Commun. Math. Phys.* **318** 355–74
- [6] Teo Y S, Stoklasa B, Englert B-G, Řeháček J and Hradil Z 2012 Incomplete quantum state estimation: a comprehensive study *Phys. Rev. A* **85** 042317
- [7] Guță M, Kypraios T and Dryden I 2012 Rank based model selection for multiple ions quantum tomography *New J. Phys.* **14** 105002
- [8] Alquier P, Butucea C, Hebiri M, Meziani K and Morimae T 2013 Rank penalized estimation of a quantum system *Phys. Rev. A* **88** 032113
- [9] Smith G A, Silberfarb A, Deutsch I H and Jessen P S 2006 Efficient quantum-state estimation by continuous weak measurement and dynamical control *Phys. Rev. Lett.* **97** 180403
- [10] Merkel S T, Riofrío C A, Flammia S T and Deutsch I H 2010 Random unitary maps for quantum state reconstruction *Phys. Rev. A* **81** 032126
- [11] Nunn J, Smith B J, Puentes G, Walmsley I A and Lundeen J S 2010 Optimal experiment design for quantum state tomography: fair, precise, and minimal tomography *Phys. Rev. A* **81** 042109
- [12] Rahimi-Keshari S, Scherer A, Mann A, Rezakhani A T, Lvovsky A and Sanders B C 2011 Process tomography of ion trap quantum gates *New J. Phys.* **13** 013006
- [13] Lundeen J S, Feito A, Coldenstrodt-Ronge H, Pregnell K L, Silberhorn C, Ralph T C, Eisert J, Plenio M B and Walmsley I A 2010 Random unitary maps for quantum state reconstruction *Phys. Rev. A* **81** 032126
- [14] Blume-Kohout R 2012 Robust error bars for quantum tomography arXiv:1202.5270
- [15] Audenaert K M R and Scheel S 2009 Quantum tomographic reconstruction with error bars: a Kalman filter approach *New J. Phys.* **11** 023028
- [16] Christandl M and Renner R 2012 Reliable quantum state tomography *Phys. Rev. Lett.* **109** 120403
- [17] Jupp P E, Kim P T, Koo J-Y and Pasiaka A 2012 Testing quantum states for purity *J. R. Stat. Soc. Ser. C* **61** 753763
- [18] Temme K and Verstraete F 2015 Quantum chi-squared and goodness of fit testing *J. Math. Phys.* **56** 012202
- [19] Landon-Cardinal O and Poulin D 2012 Practical learning method for multi-scale entangled states *New J. Phys.* **14** 085004
- [20] Kahn J and Guță M 2009 Local asymptotic normality for finite dimensional quantum systems *Commun. Math. Phys.* **289** 597–652
- [21] Hayashi M and Matsumoto K 2008 Asymptotic performance of optimal state estimation in qubit system *J. Math. Phys.* **49** 102101
- [22] Audenaert K M R, Nussbaum M, Szkola A and Verstraete F 2008 Asymptotic error rates in quantum hypothesis testing *Commun. Math. Phys.* **279** 251283
- [23] Guță M and Kiukas J 2015 Equivalence classes and local asymptotic normality in system identification for quantum Markov chains *Commun. Math. Phys.* **335** 1397–428
- [24] Monz T, Schindler P, Barreiro J T, Chwalla M, Nigg D, Coish W A, Harlander W, Haensel M, Hennrich M and Blatt R 2011 14-qubit entanglement: creation and coherence *Phys. Rev. Lett.* **106** 130506
- [25] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 Quantum state tomography via compressed sensing *Phys. Rev. Lett.* **105** 150401
- [26] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 Quantum tomography via compressed sensing: error bounds, sample complexity, and efficient estimators *New J. Phys.* **14** 095022
- [27] Cramer M, Plenio M B, Flammia S T, Gross D, Bartlett S D, Somma R, Landon-Cardinal O, Liu Y-K and Poulin D 2010 Efficient quantum state tomography *Nat. Commun.* **1** 149
- [28] Carpentier A, Eisert J, Gross D and Nickl R 2015 Uncertainty quantification for matrix compressed sensing and quantum tomography problems arXiv:1504.03234v1
- [29] Hradil Z, Řeháček J, Fiurášek J and Ježek M 2004 Maximum-likelihood methods in quantum mechanics *Quantum State Estimation* ed M G A Paris and J Řeháček vol 649 (Berlin: Springer) pp 59–112
- [30] Claeskens G and Hjort N L 2008 *Model Selection and Model Averaging* (Cambridge: Cambridge University Press)

- [31] van der Vaart A W 1998 *Asymptotic Statistics* (Cambridge: Cambridge University Press)
- [32] Koltchinskii V 2013 A remark on low rank matrix recovery and noncommutative Bernstein type inequalities *IMS Collect.* **9** 213–26
- [33] Gross D 2011 Recovering low-rank matrices from few coefficients in any basis *IEEE Trans. Inf. Theory* **57** 1548–66
- [34] Rehacek J, Hradil Z, Knill E and Lvovsky A I 2007 Diluted maximum-likelihood algorithm for quantum tomography *Phys. Rev. A* **75** 042108
- [35] Bunea F, She Y and Wegkamp M H 2011 Optimal selection of reduced rank estimators of high-dimensional matrices *Ann. Stat.* **39** 1282–309
- [36] Harris J 1995 *Algebraic Geometry: A First Course* (Berlin: Springer)
- [37] Gill R D and Levit B Y 1995 Applications of the van trees inequality: a bayesian Cramér–Rao bound *Bernoulli* **1** 59–79
- [38] Gill R D and Massar S 2000 State estimation for large ensembles *Phys. Rev. A* **61** 042312
- [39] Ahlswede R and Winter A 2002 Strong converse for identification via quantum channels *IEEE Trans. Inf. Theory* **48** 569–79
- [40] Tropp J A 2012 User-friendly tail bounds for sums of random matrices *Found. Comput. Math.* **12** 389–434
- [41] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
- [42] Smolin J A, Gambetta J M and Smith G 2012 Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise *Phys. Rev. Lett.* **108** 070502
- [43] Collins B and Sniady P 2006 Integration with respect to the Haar measure on unitary, orthogonal and symplectic group *Commun. Math. Phys.* **264** 773–95