Quantum statistics:

Estimation of large dimensional systems

Mădălin Guță

School of Mathematical Sciences
University of Nottingham

C. Butucea, M.G., T. Kypraios, New Journal of Physics, **17**, 113050 (2015)

A. Acharya, T. Kypraios, M.G., New Journal of Physics (2016)

A. Acharya, M.G: 1609.03758

EPSRC

Quantum trajectories, parameter and state estimation

Toulouse 2017

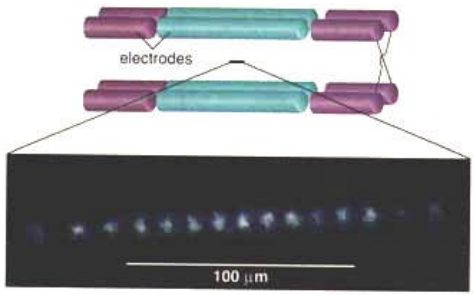
The University of Nottingham

Outline

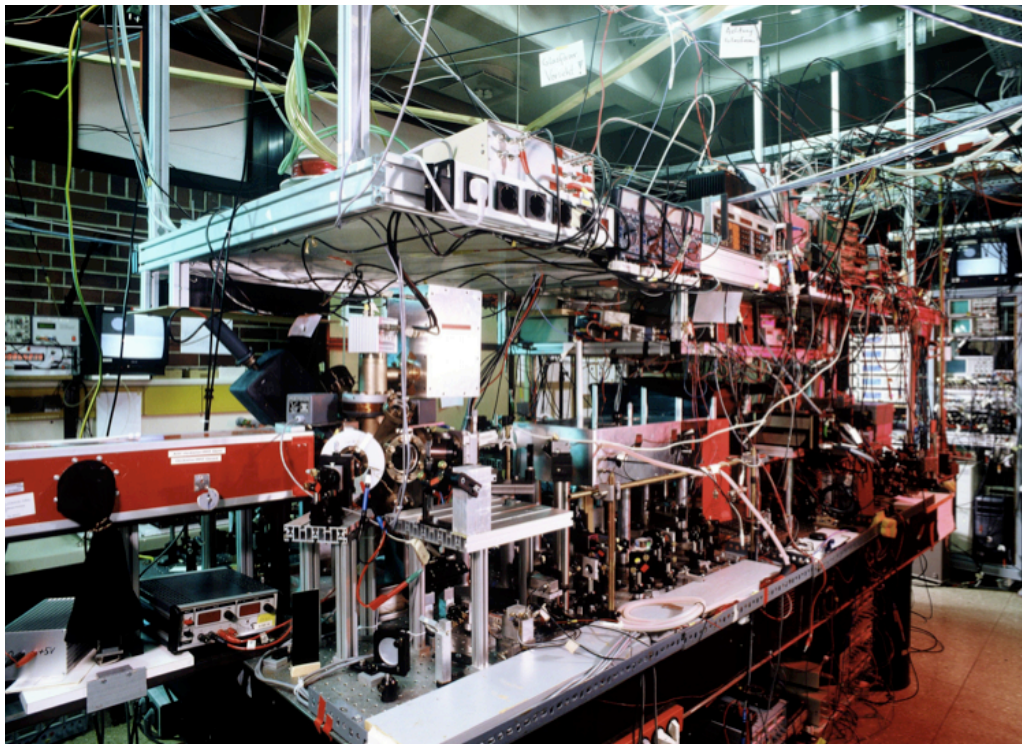
- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Quantum-classical interface is stochastic ⇒ Q. Engineering needs Statistics

- Problem: Quantum state estimation
  - Goal: create a specific state of e.g. 8 ions
  - Validation: statistical estimation from measurement outcomes



[Häffner *et al*, Nature 2005]



Rainer Blatt's Lab, Innsbruck

"quantum computer" with 8 qubits (ions)

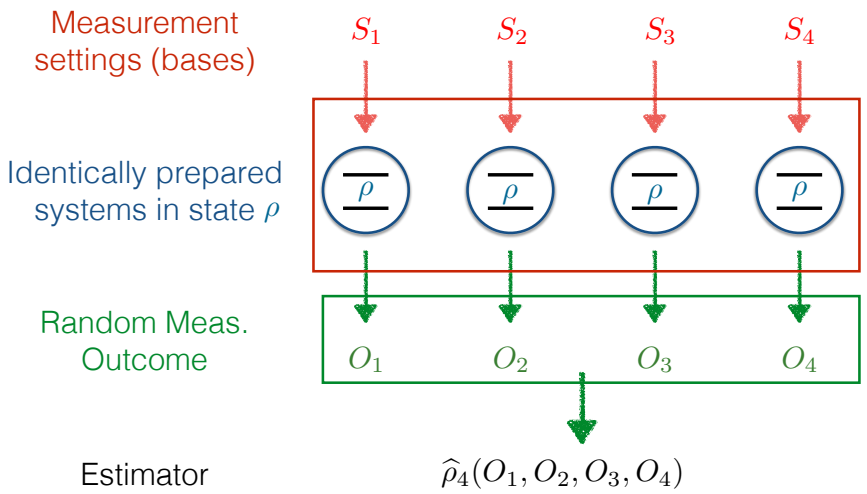
- $4^8 - 1 = 65\,535$ parameters
- $3^8 \times 100 = 656\,100$ measurements
- 10 hours measurement time
- days of computer time

Simple measurements

- Measurement setting given by an orthonormal basis $\mathbf{s} := \{|e_1\rangle, \ldots, |e_d\rangle\}$ in $\mathbb{C}^d$
- Outcome of measurement is a random index of a basis element $O \in \{1, \ldots, d\}$
- Probability distribution: if system is prepared state ρ

$$\mathbb{P}[O = i] = \langle e_i|\rho|e_i\rangle = \rho_{ii}$$

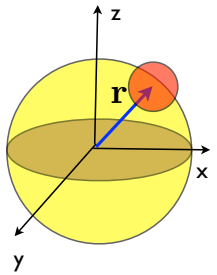
- Quantum state tomography: probe system with sufficient measurements to estimate ρ



Example: spin / two-level ion / qubit tomography

- Any state on $\mathbb{C}^2$ is parametrized by a 3-D Bloch vector $\mathbf{r} = (r_x, r_y, r_z)$ with $\|\mathbf{r}\| \leq 1$

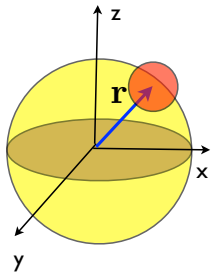
$$\rho_{\mathbf{r}} = \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix}$$



Example: spin / two-level ion / qubit tomography

- Any state on $\mathbb{C}^2$ is parametrized by a 3-D Bloch vector $\mathbf{r} = (r_x, r_y, r_z)$ with $\|\mathbf{r}\| \leq 1$

$$\rho_{\mathbf{r}} = \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix}$$



- 3 standard measurement bases corresponding to $s = x, y, z$ spin observables

$$\underbrace{|e_x^{\pm}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}}_{s=x} \qquad \underbrace{|e_y^{\pm}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix}}_{s=y} \qquad \underbrace{|e_z^{+}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |e_z^{-}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{s=z}$$

- Probability distributions: $\mathbb{P}(o = \pm|s) = \frac{1 \pm r_s}{2}, \qquad s = x, y, z$

- n measurement repetitions $\longrightarrow$ counts $\{N(\pm|x), N(\pm|y), N(\pm|z)\} \longrightarrow$ (LS) estimator

$$\widehat{\rho}_n := \rho_{\hat{\mathbf{r}}}, \qquad \widehat{r}_{x,y,z} := \frac{N(+|x,y,z) - N(-|x,y,z)}{n}$$

Boundary/positivity problem: for pure (rank one) states, estimator may not be physical (positive)

Measuring (correlated) states of multiple ions

- State space of k two-level systems scales exponentially with k !

$$\mathcal{H}_k := \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2 \cong \mathbb{C}^{2^k} = \mathbb{C}^d$$

- Joint state of k ions
  - General density matrix ρ has $4^k - 1 = d^2 - 1$ parameters (e.g. $4^8 - 1 = 65535$)
  - Density matrix of rank r has $r(2 \cdot d - r) - 1$ param. (e.g. $2^8 - 2 = 254$ for a pure state)

Measuring (correlated) states of multiple ions

- State space of k two-level systems scales exponentially with k !

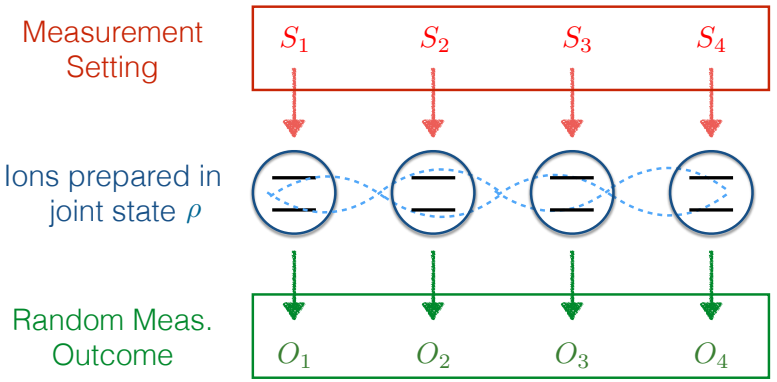
$$\mathcal{H}_k := \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2 \cong \mathbb{C}^{2^k} = \mathbb{C}^d$$

- Joint state of k ions
  - General density matrix ρ has $4^k - 1 = d^2 - 1$ parameters (e.g. $4^8 - 1 = 65535$)
  - Density matrix of rank r has $r(2 \cdot d - r) - 1$ param. (e.g. $2^8 - 2 = 254$ for a pure state)

- Simultaneous, separate measurements on each ion:
  - 3^k settings $\mathbf{s} = (s_1, \ldots, s_k) \in \{x, y, z\}^k \longrightarrow$ product basis $|e_{s_1}^{o_1} \otimes \cdots \otimes e_{s_k}^{o_k}\rangle$
  - 2^k outcomes $\mathbf{o} = (o_1, \ldots, o_k) \in \{+, -\}^k$
  - probabilities

$$\mathbb{P}_\rho(\mathbf{o}|\mathbf{s}) = \left\langle e_{s_1}^{o_1} \otimes \cdots \otimes e_{s_k}^{o_k} \right| \rho \left| e_{s_1}^{o_1} \otimes \cdots \otimes e_{s_k}^{o_k} \right\rangle$$

Measurement procedure and statistical model[1]



1. For each ion choose a spin direction to measure basis $s \in \{x, y, z\}$
2. measure each ion and obtain outcome $\mathbf{o} := (o_1, \ldots, o_k) \in \{1, -1\}^k$
3. Repeat n times and collect counts of outcomes $\{N_{\mathbf{o},\mathbf{s}} : \mathbf{o} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho\left(\{N(\mathbf{o}|\mathbf{s}) : \mathbf{o} \in \{1, -1\}^k\}\right) = \frac{n!}{\prod_o N(\mathbf{o}|\mathbf{s})!} \prod_{\mathbf{o}} \mathbb{P}_\rho(\mathbf{o}|\mathbf{s})^{N(\mathbf{o}|\mathbf{s})}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

[1] statistical model based on counts is different from that of compressed sensing D. Gross et al, Phys. Rev. Lett. 2010

Measurement data

- 3^{k} columns of length 2^{k}
- one column for each measurement setting
- each column contains the counts totalling $n = 100$, of the $2^{\mathrm{k}} = 16$ possible outcomes
- frequencies of outcomes are bad estimates of probabilities, but overall info is high

1	2	11	11	11	21	5	16	21	19	11	16	2	26	15	5
2	19	10	6	15	4	22	10	3	12	8	16	18	5	14	16
3	30	12	15	9	10	18	14	3	6	11	4	4	2	1	5
4	0	4	15	10	17	2	4	14	13	0	4	8	5	1	3
5	21	13	12	7	6	5	14	12	8	12	7	19	3	8	3
6	1	12	14	0	1	1	0	6	6	12	8	2	6	2	7
7	1	2	0	19	7	12	14	6	7	14	7	9	23	15	34
8	0	1	1	0	4	8	0	6	6	0	7	12	4	15	5
9	21	17	8	10	7	7	14	9	8	15	6	9	6	3	0
10	2	16	15	0	12	9	0	3	4	1	7	3	0	4	6
11	0	0	1	17	9	2	14	12	7	0	1	0	5	5	2
12	1	1	1	0	2	8	0	4	3	0	1	0	0	3	1
13	1	0	1	1	0	0	0	0	0	14	9	7	6	2	4
14	0	1	0	0	0	1	0	0	1	1	5	6	0	2	2
15	1	0	0	1	0	0	0	0	0	1	2	0	9	6	3
16	0	0	0	0	0	0	0	1	0	0	0	1	0	4	4

[Dataset 4 ions (from Blatt group, Innsbruck)]

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Linear regression and least squares

Problem: linear regression

estimate the unknown vector $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Linear regression and least squares

Problem: linear regression

estimate the unknown vector $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Least squares: Find $\hat{x}$ which minimises

$$\sum_i |Y_i - \sum_j A_{ij} \hat{x}_j|^2 = (\mathbf{Y} - A\hat{\mathbf{X}})^T (\mathbf{Y} - A\hat{\mathbf{X}})$$

Linear regression and least squares

Problem: linear regression

estimate the unknown vector $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Least squares: Find $\hat{x}$ which minimises

$$\sum_i |Y_i - \sum_j A_{ij} \hat{x}_j|^2 = (\mathbf{Y} - A\hat{\mathbf{X}})^T (\mathbf{Y} - A\hat{\mathbf{X}})$$

Explicit solution coinciding with maximum likelihood estimator

$$\hat{\mathbf{X}} = (A^T A)^{-1} A^T \mathbf{Y}$$

Covariance matrix of $\hat{X}$

$$Var(\hat{\mathbf{X}}) = \sigma^2 (A^T A)^{-1}$$

The least squares estimator

- For large n frequencies are close to the probabilities of the corresponding outcome

$$f_n(\mathbf{o}|\mathbf{s}) = \frac{N(\mathbf{o}|\mathbf{s})}{n} = p_\rho(\mathbf{o}|\mathbf{s}) + \epsilon_n(\mathbf{o}|\mathbf{s}) \quad (\text{“multinomial error”})$$

$$\mathbf{f}_n = \mathbf{p}_\rho + \epsilon_n = \mathbf{A}\widetilde{\rho} + \epsilon_n$$

- State estimation as a "linear regression" problem: least squares estimator

$$\widehat{\rho}_n^{(ls)} = \arg\min_\tau \|\mathbf{A}\widetilde{\tau} - \mathbf{f}_n\|_2^2 = (\mathbf{A}^t\mathbf{A})^{-1} \cdot \mathbf{A}^t \cdot \mathbf{f}_n$$

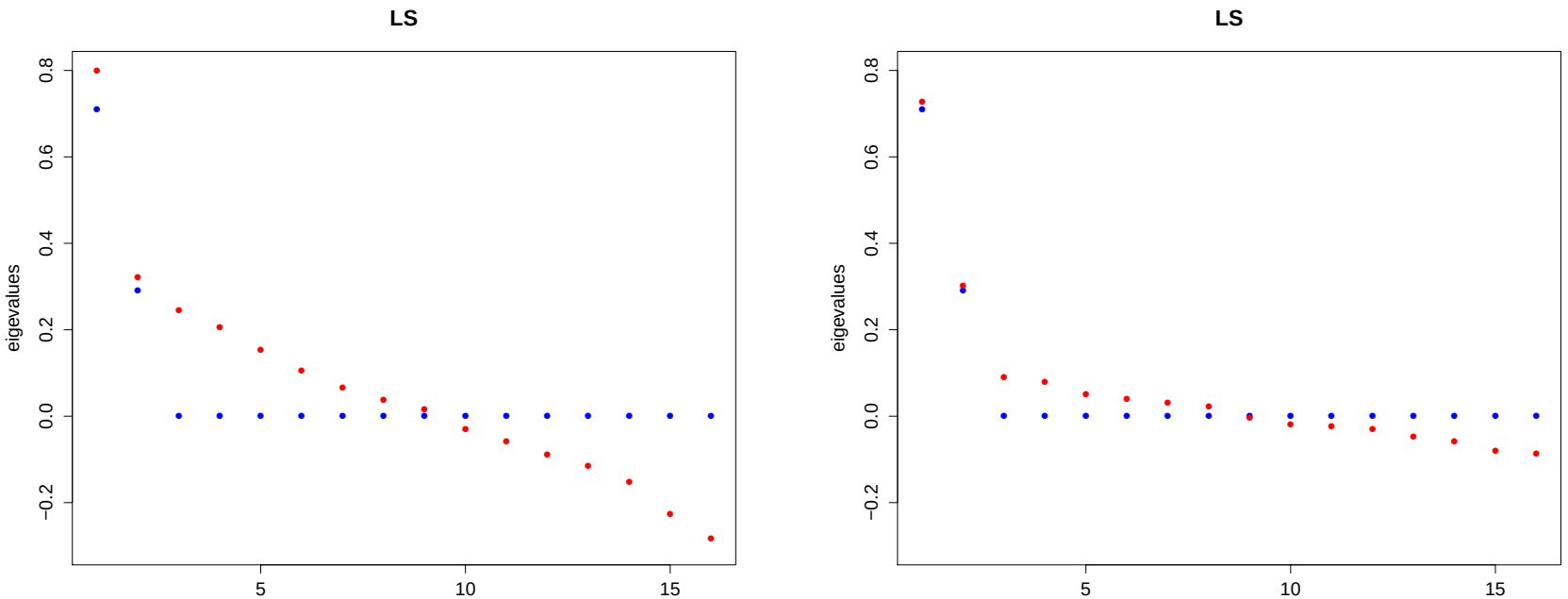
- Disadvantages
  - Least squares estimator is not a density matrix (not positive and trace one)
  - Least squares estimator is too "noisy" for low rank states
  - Least squares estimator minimises prediction rather than estimation error $\mathbb{E}\|\widehat{\rho}_n - \rho\|_2^2$

Eigenvalues distribution for the least squares estimator

- Eigenvalues decomposition for true state and least squares estimator

$$\rho = \sum_{i=1}^{r} \lambda_i |\psi_i\rangle\langle\psi_i| \qquad \widehat{\rho}_n^{(ls)} = \sum_{i=1}^{d} \widehat{\lambda}_i |\hat{\psi}_i\rangle\langle\hat{\psi}_i|$$

- If $r \ll d = 2^k$, the $MSE = \mathbb{E}\|\widehat{\rho}_n^{(ls)} - \rho\|_2^2$ is large due to variance contributions from many eigenvalues $\widehat{\lambda}_i$ which estimate zero eigenvalues of ρ



Eigenvalues of true state of rank 2 (blue) versus least squares estimator (red)

LEFT: $n = 20$ repetitions RIGHT: $n = 100$ repetitions

Norm-error upper bound for the least squares estimator[2]

- operator-norm distance $\|\rho - \tau\| = |\lambda_{max}(\Delta)|, \qquad \Delta := \rho - \tau$
- norm-two distance $\|\rho - \tau\|_2^2 = \sum_i |\lambda_i(\Delta)|^2 \leq d \cdot \|\rho - \tau\|^2 \qquad (*)$

Theorem

For any $\varepsilon > 0$ small enough the following inequality holds with probability larger than $1 - \varepsilon$

$$\left\| \widehat{\rho}_n^{(ls)} - \rho \right\| \leq \nu_n(\varepsilon),$$

where the rate $\nu_n(\varepsilon)^2$ is

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

with $N := n \cdot 3^k$ the total number of measurements.

- *Concentration inequality and $(*)$ give upper bound for the MSE*

$$\mathbb{E} \left\| \widehat{\rho}_n^{(ls)} - \rho \right\|_2^2 \leq C \frac{6^k \cdot k}{N} \approx k \cdot \left(\frac{3}{2}\right)^k \cdot \frac{\#\text{parameters}}{\#\text{samples}}$$

[2] C. Butucea, M.G. and T. Kypraios, New Journal of Physics, **17**, 113050 (2015)
this improves on the $4^k/N$ factor in the upper bound of P. Alquier, C. Butucea, *et al*, Phys. Rev. A (2013)

Idea of the proof

- Write

$$\widehat{\rho}_n^{(ls)} - \rho = \sum_{\mathbf{s}} \sum_{i} W_{\mathbf{s},i}$$

where $W_{\mathbf{s},i}$ are i.i.d. centred random matrices

- Use matrix Bernstein inequality[3] for i.i.d. Hermitian matrices

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

where $\|Y_j\| \leq V$ and $\|\sum_j \mathbb{E}(Y_j^2)\| \leq W$

[3] J. A. Tropp, *Found Comput Math* **12** 389-434 (2012)

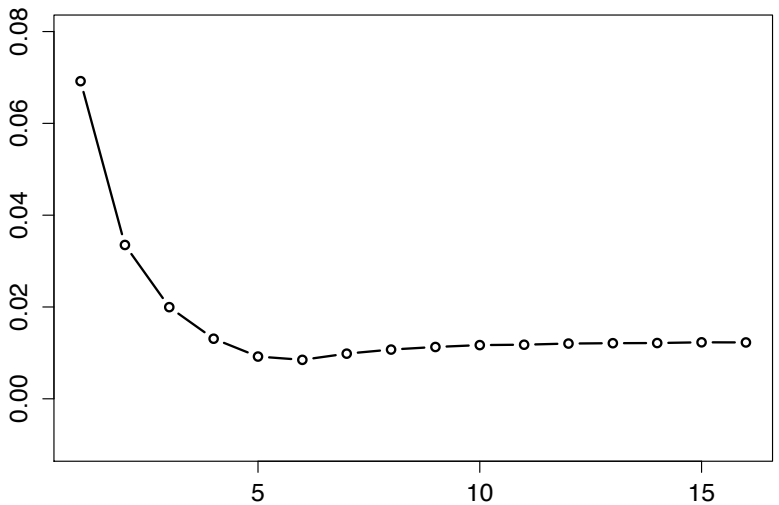
Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Penalising small eigenvalues

- Assume true state ρ of low rank: $\rho = \sum_{i=1}^{r} \lambda_i |\psi_i\rangle\langle\psi_i|$ with $r \ll d$
- Idea: $\|\widehat{\rho}_n^{(ls)} - \rho\| \approx \nu_n \Rightarrow$ eigenvalues of $\widehat{\rho}_n^{(ls)}$ s.t. $|\widehat{\lambda}_i| \leq \nu_n$ may be "statistical noise"
- Truncated versions of the LS estimator: order $|\widehat{\lambda}_i| \geq \cdots \geq |\widehat{\lambda}_d|$ and for each $k \leq d$

$$\widehat{\rho}_n^{(ls)} = \sum_{i=1}^{d} \widehat{\lambda}_i |\hat{\psi}_i\rangle\langle\hat{\psi}_i| \longrightarrow \widehat{\rho}_n(k) = \sum_{i=1}^{k} \widehat{\lambda}_i |\hat{\psi}_i\rangle\langle\hat{\psi}_i|$$



Norm-two error $E(k) := \|\widehat{\rho}_n(k) - \rho\|_2^2$ for a state of rank $r = 6$, as function of truncation rank k

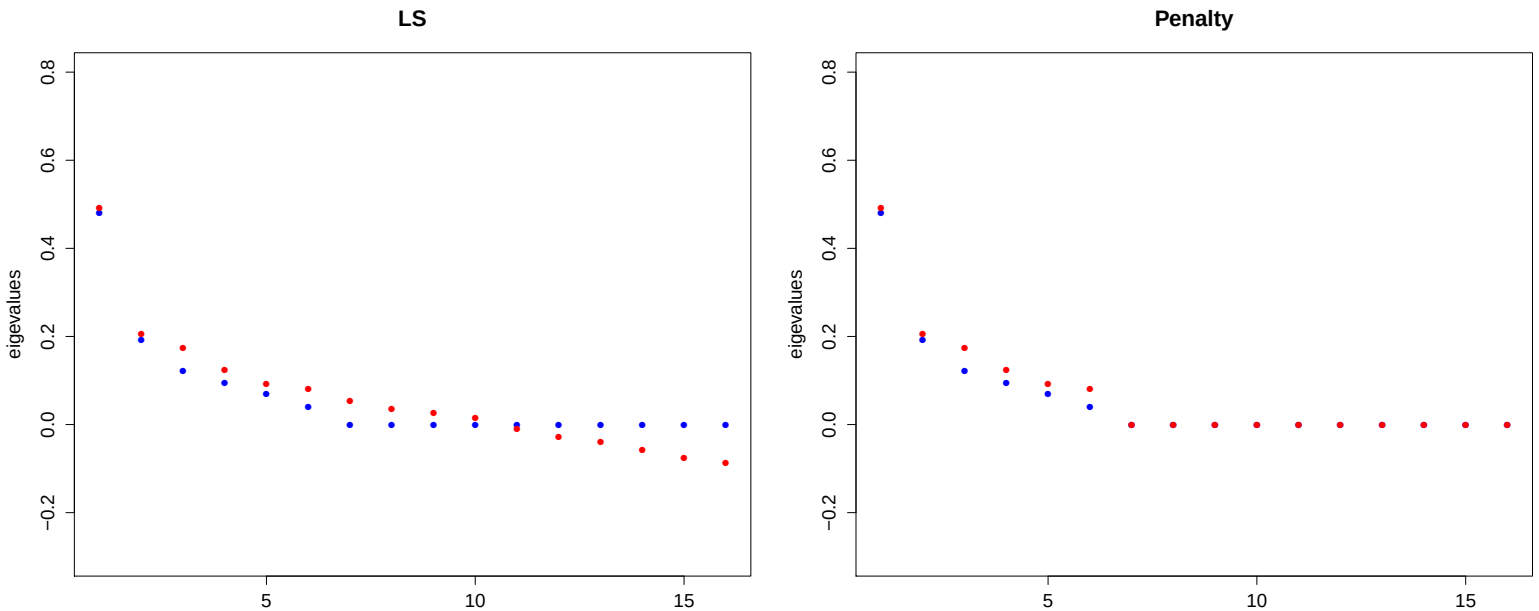
Question: how to choose the truncation rank ?

Penalised least squares estimator

- Choose rank $\widehat{k} := \max\{k : \widehat{\lambda}_k^2 \geq \nu_n^2\}$ with $|\widehat{\lambda}_1| \geq |\widehat{\lambda}_2| \geq \cdots \geq |\widehat{\lambda}_d|$
- Equivalently, $\widehat{k}$ minimises the rank-penalised distance to the least squares

$$\widehat{k} = \arg\min_k \left[\|\widehat{\rho}_n(k) - \widehat{\rho}_n\|_2^2 + k \cdot \nu_n^2 \right]$$

- Penalised estimator: $\widehat{\rho}_n^{(pen)} := \widehat{\rho}_n(\widehat{k})$



Eigenvalues of true state ρ (blue) versus: LS (red) on LEFT and penalised estimator (red) RIGHT

for a rank 6 state with $n = 100$ repetitions

MSE upper bound for the penalised estimator[4]

- Penalised estimator: with $\widehat{k} := \max\{k : \widehat{\lambda}_k^2 \geq \nu_n^2\}$

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

Theorem

Let ρ be a state of unknown rank r.
Let $\varepsilon > 0$ be a small parameter. Then with probability larger than $1 - \varepsilon$, we have

$$\|\widehat{\rho}_n^{(pen)} - \rho\|_2^2 \leq C \cdot r \cdot \nu_n(\epsilon)^2 = C \left(\frac{3}{2}\right)^k \log\left(\frac{d}{\varepsilon}\right) \frac{dr}{N}$$

- *Concentration inequality gives upper bound for the MSE*

$$\mathbb{E} \left\| \widehat{\rho}_n^{(pen)} - \rho \right\|_2^2 \leq Ck \cdot \left(\frac{3}{2}\right)^k \cdot \frac{\#\text{parameters(rank = r)}}{\#\text{samples}}$$

[4] C. Butucea, M.G. and T. Kypraios, New Journal of Physics, **17**, 113050 (2015)

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Physical estimator

- Disadvantage of penalised estimator: $\widehat{\rho}_n^{(pen)}$ may not be a state (positive, trace-one matrix)
- Physical estimator $\widehat{\rho}_n^{(phys)}$ exploits the positivity properties of ρ:

$$\widehat{\rho}_n^{(phys)} = \underset{\sigma \in \mathcal{S}(\nu_n)}{\arg\min} \left\| \sigma - \widetilde{\rho}_n^{(ls)} \right\|_2^2,$$

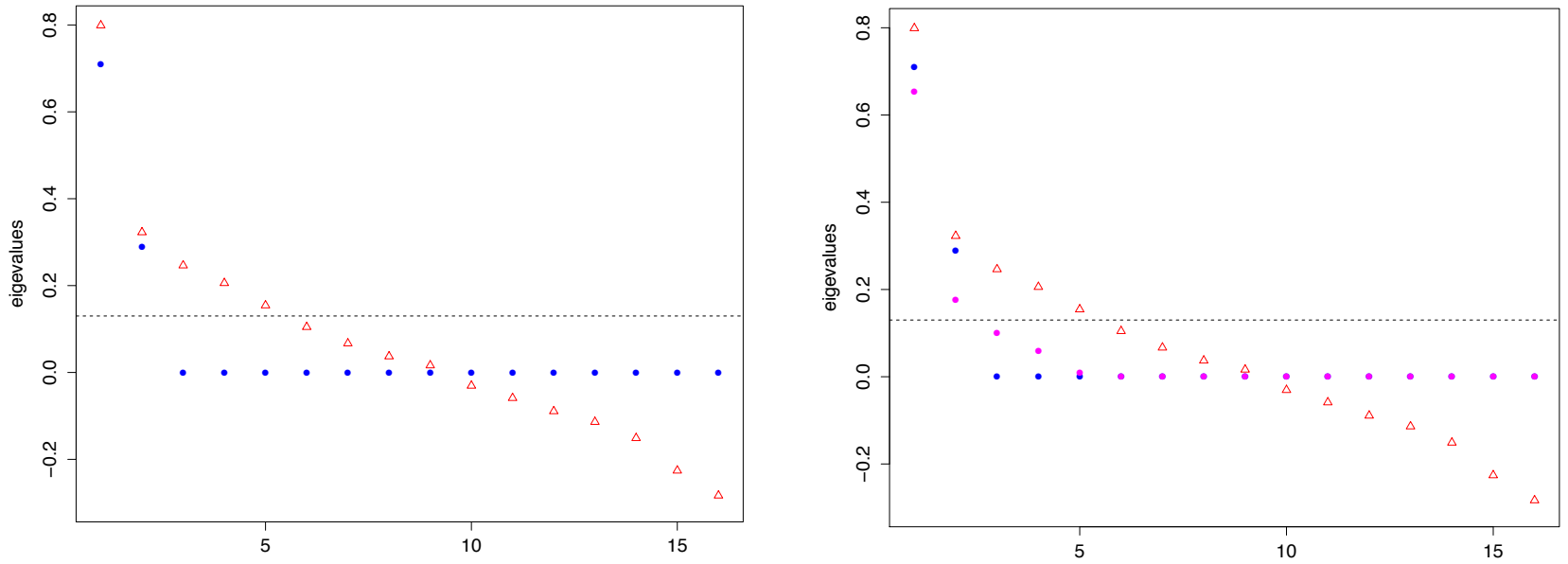
  - $\widetilde{\rho}_n^{(ls)}$ is the "normalised LS estimator" s.t. $\mathrm{Tr}\widetilde{\rho}_n^{(ls)} = 1$
  - Set of states at noise level ν_n

$$\mathcal{S}(\nu_n) = \{\sigma \ : \ \text{state with eigenvalues } \lambda_j \in \{0\} \cup (4\nu_n, 1],\ j = 1, ..., d\}.$$

Questions: can it be computed efficiently, and what is its MSE?

Physical estimator: implementation

- Optimisation: solution is a truncated LS matrix $\widehat{\rho_n}(\widehat{k}) = \sum_{i=1}^{\widehat{k}} \widehat{\lambda}_i |\hat{\psi}_i\rangle\langle\hat{\psi}_i|$
- Truncation rank: simple iterative algorithm on eigenvalues arranged as $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_d$ selects maximum k for which all eigenvalues of $\widehat{\rho_n}(\widehat{k})$ are larger than threshold after being normalised by shifting with constant



Eigenvalues of true state ρ (blue circles) versus LS (red triangles) on LEFT vs. eigenvalues of physical estimator (rose) on RIGHT for a rank 2 state with $n = 20$ repetitions

Physical estimator: upper bound[5]

Theorem

Let ρ be a state of unknown rank r.
Let $\varepsilon > 0$ be a small parameter, and assume that $\lambda_r > 8\nu_n(\varepsilon)$. Then, with probability larger than $1-\varepsilon$ we have

$$\left\| \widehat{\rho}_n^{(phys)} - \rho \right\|_2^2 \leq C \left(\frac{3}{2}\right)^k \log\left(\frac{d}{\varepsilon}\right) \frac{dr}{N}$$

- *Concentration inequality gives upper bound for the MSE*

$$\mathbb{E} \left\| \widehat{\rho}_n^{(phys)} - \rho \right\|_2^2 \leq Ck \cdot \left(\frac{3}{2}\right)^k \cdot \frac{\#\text{parameters}(\text{rank} = r)}{\#\text{samples}}$$

[5] C. Butucea, M.G. and T. Kypraios, ArXiv:1504.08295

Outline

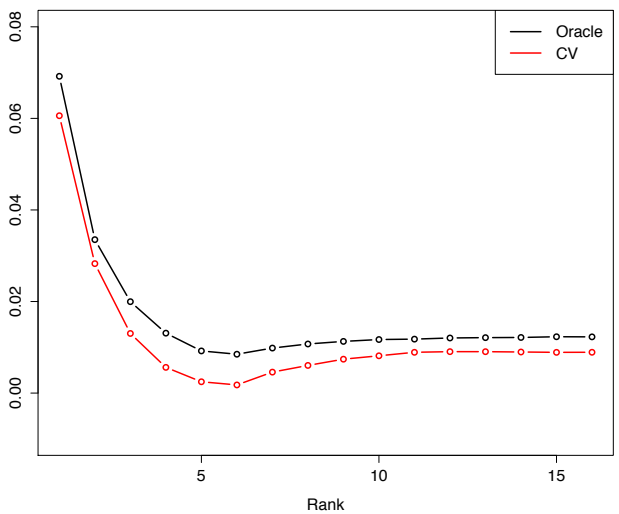
- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Choosing truncation rank by cross-validation

- Norm-two error E(k):$=\|\widehat{\rho}_n(k) - \rho\|_2^2$ minimised by oracle estimator
- Cross-validation:
  - Split dataset in 5 independent batches and compute $\widehat{\rho}^{(ls)}_{n;j}$ and $\widehat{\rho}^{(ls)}_{n;-j}$ on batch j and respectively all-but-j batches, for $j = 1, \ldots, 5$.
  - Replace $E(k)$ by unbiased estimator (up to constant independent of k)

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

  - Cross-validation estimator: $\widehat{\rho}^{(cv)}_n := \widehat{\rho}_n(\widehat{k})$ where $\widehat{k}$ is the minimiser of $CV(k)$.

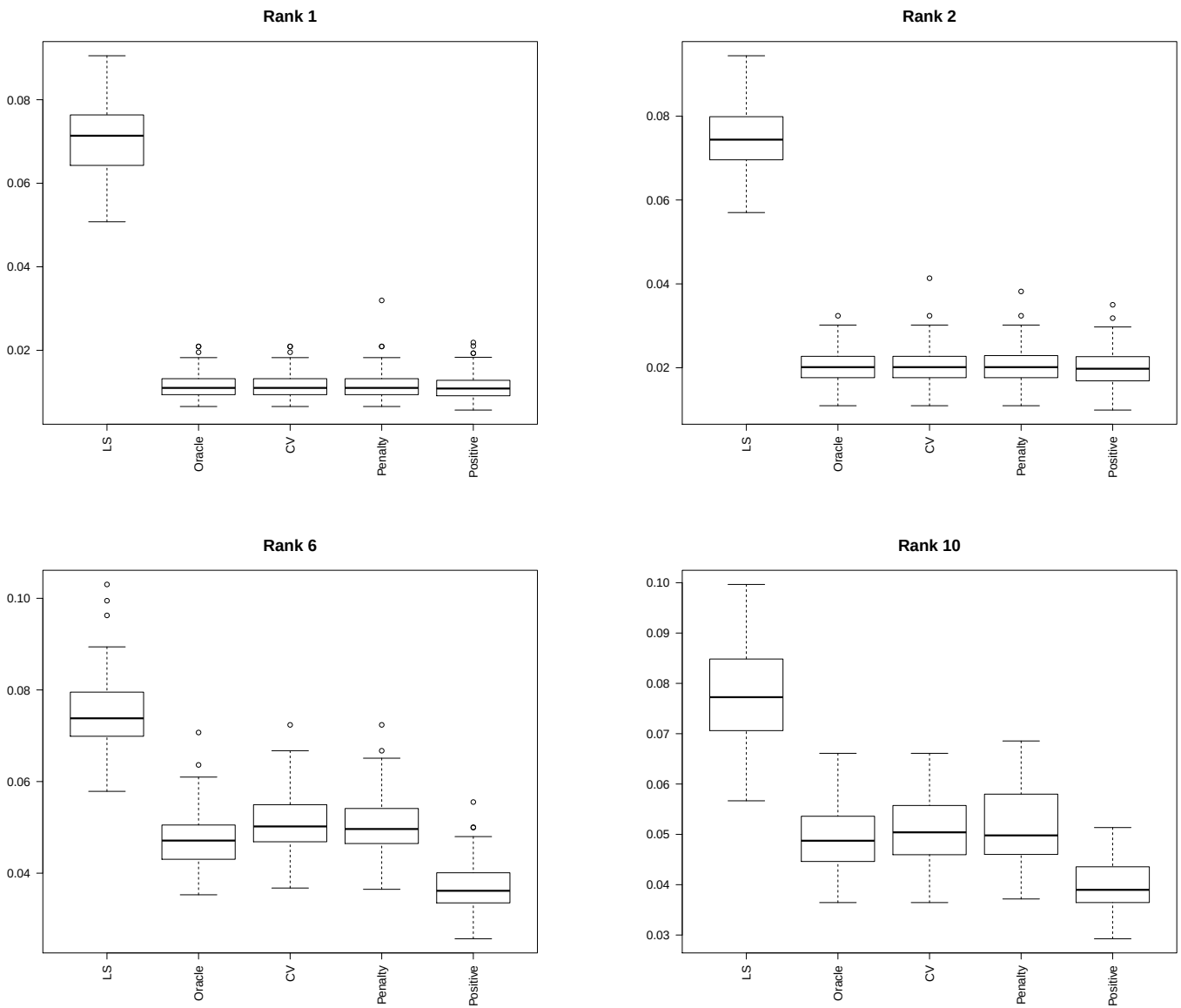


E(k) (black) and CV(k) (red) for one dataset from a rank 6 state with $n = 500$ repetitions

Outline

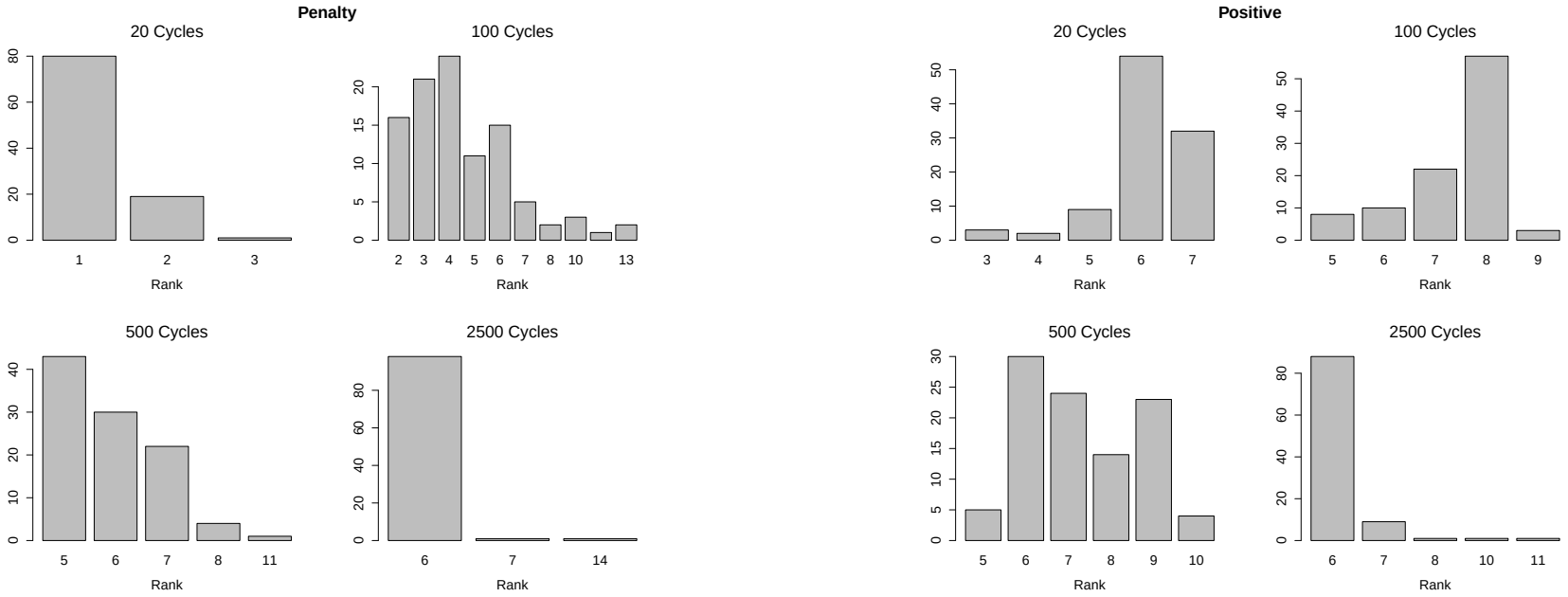
- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Comparison of estimators: SEs for different states, with $n = 100$



Boxplots of norm-two errors $\|\widehat{\rho_n} - \rho\|_2^2$ of different estimators for states of ranks $1, 2, 6, 10$ with $n = 100$ repetitions (computed from 100 datasets)

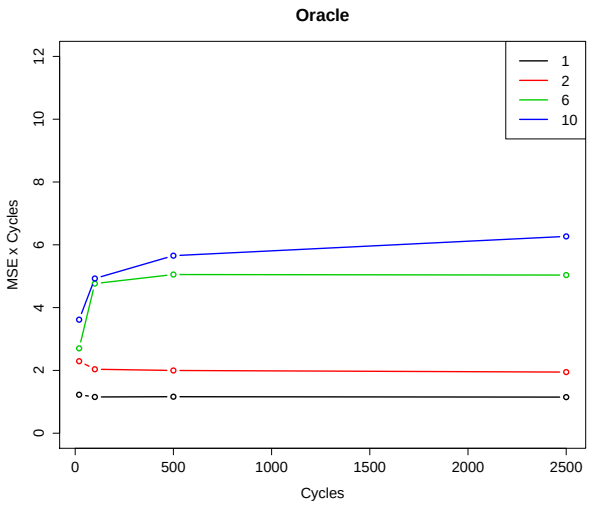
Comparison of estimators: Empirical distribution of chosen rank

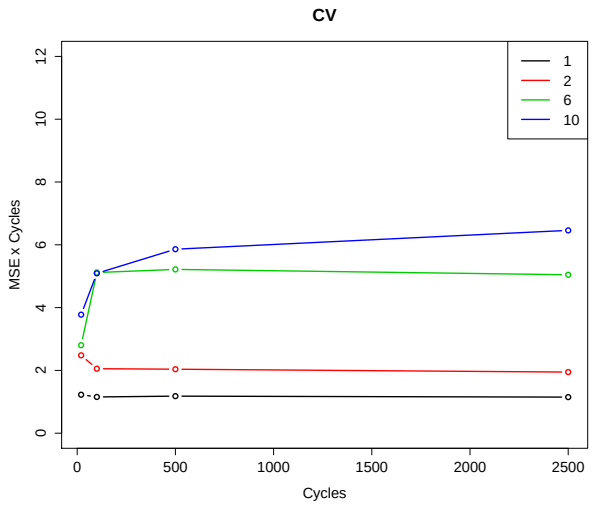


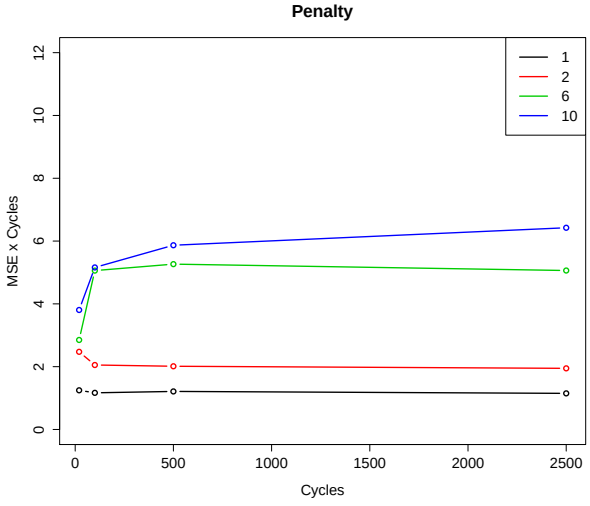
Empirical distributions of the chosen rank for a state of rank $r = 6$

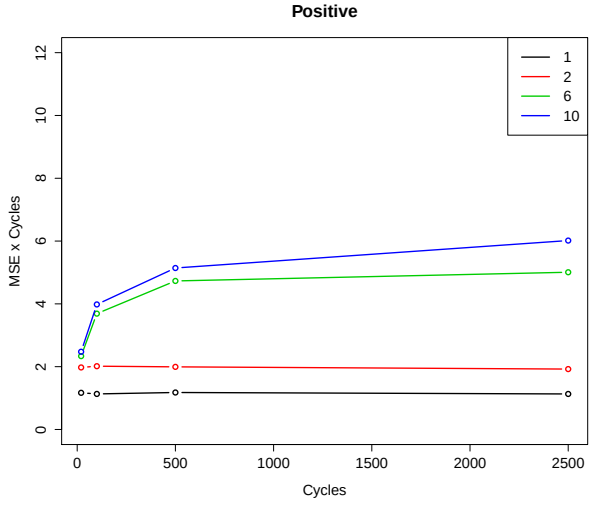
Left: penalised estimator & Right: physical estimator

Comparison of estimators: MSE for different states and repetitions n









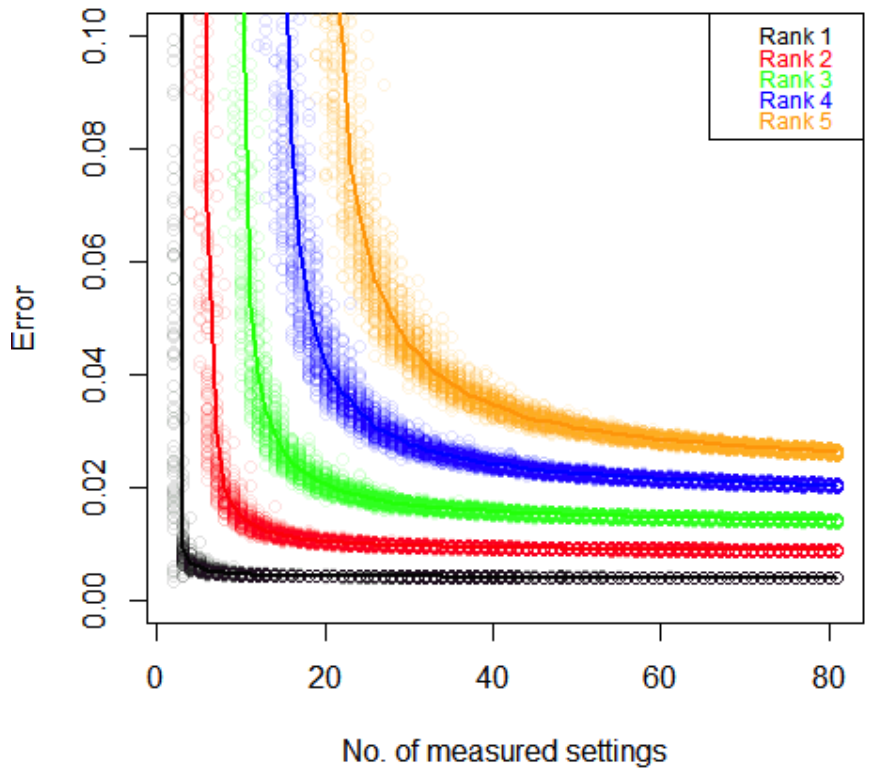
Renormalised MSEs $= n \cdot \mathbb{E}\|\widehat{\rho_n} - \rho\|_2^2$ as a function of number of repetitions n
for states with different ranks: 1(black), 2 (red), 6 (green), 10 (blue) .

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Spectral thresholding estimators
  - Penalised estimator
  - Physical estimator
  - Cross-validation estimator
  - Simulation results
- Efficient estimation with reduced measurement settings

Can we estimate low rank states with reduced measurement settings ?[6]

- Counting parameters: rank r state $\longrightarrow r \cdot d$ parameters $\longrightarrow \approx r$ settings ($\ll 3^k$)
- Random measurement design:
  choose m random settings $\mathcal{S} := \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ and measure each setting $n = \frac{N}{m}$ times
- Mean square error of MLE is stable for a large range of number of settings m



Mean square error $\mathbb{E}\|\hat{\rho}^{(ml)} - \rho\|_2^2$ for 4 ions states of ranks 1-5 and randomly chosen settings

[6] similar to "compressed sensing" D. Gross, *et al*, Phys. Rev. Lett. (2010) but uses "raw" rather than "coarse grained" data

Concentration for Fisher information matrix[7]

- More randomness helps: consider measurements w.r.t. random bases (Haar measure)
- Asymptotics: for large n mean square error of ML estimator scales as in Cramér-Rao bound

$$\|\hat{\rho}^{(ml)} - \rho\|_2^2 \approx \frac{1}{N}\mathrm{Tr}(I(\rho|\mathcal{S})^{-1}G(\rho))$$

- Fisher information matrix (per setting) converges to average

$$I(\rho|\mathcal{S}) = \frac{1}{m}\sum_{i=1}^{m} I(\rho|\mathbf{s}_i) \longrightarrow \bar{I}(\rho) = \int I(\rho|\mathbf{s})d\mathbf{s}$$

Theorem (Fisher info & MSE concentrate with $r \cdot \log rd$ settings)

Let ρ be rank r state with spectrum $(1/r, \ldots, 1/r, 0, \ldots, 0)$.

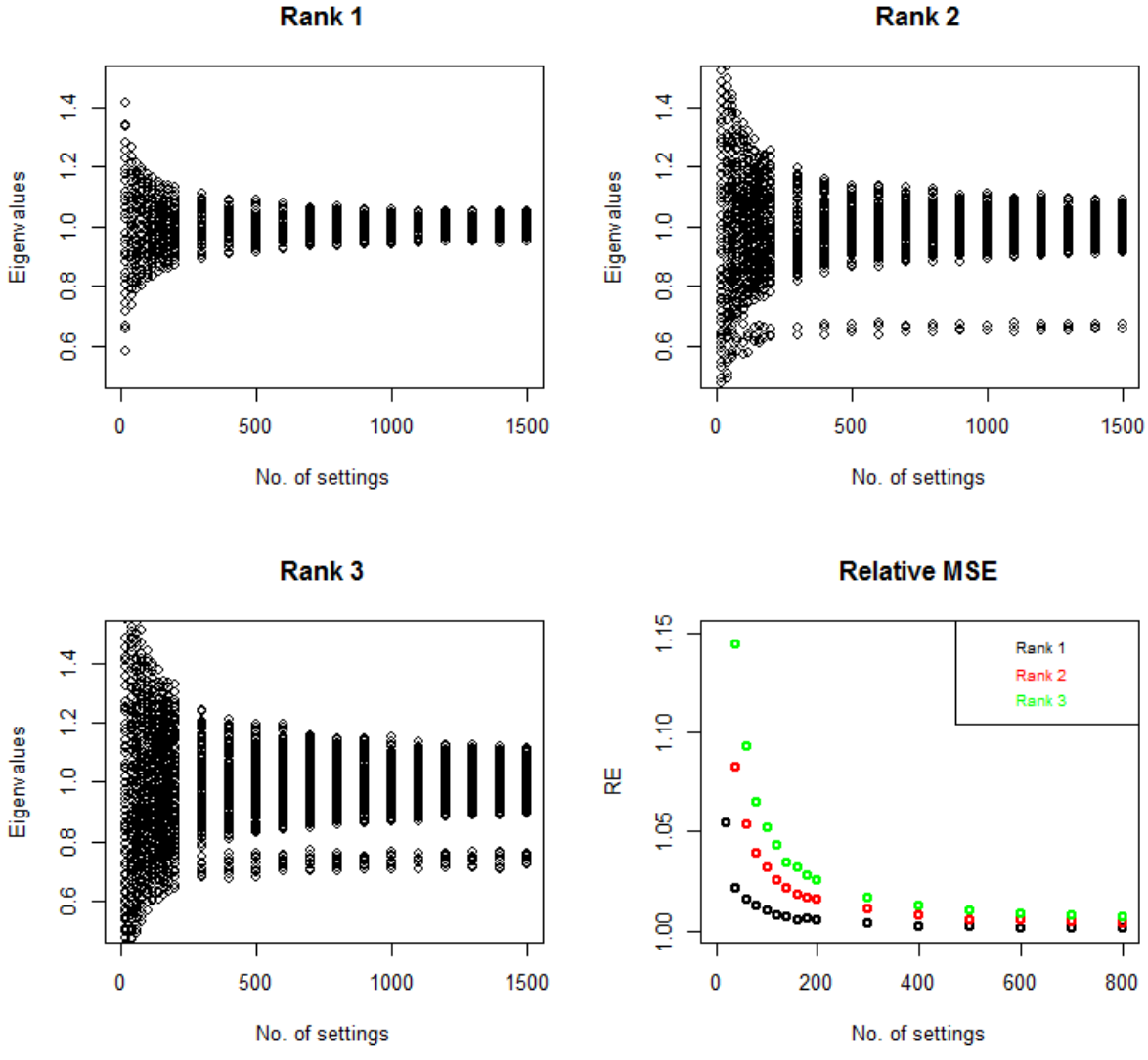
If $m = C(r+1)\log(2(2rd - r^2 - 1)/\delta\epsilon^2)$ then the bounds hold with probability $1 - \delta$

$$(1-\epsilon)\overline{I}(\rho) \leq \quad I(\rho|\mathcal{S}) \quad \leq (1+\epsilon)\overline{I}(\rho)$$

$$(1-\epsilon)\mathrm{Tr}\left[\overline{I}(\rho)^{-1}G(\rho)\right] \leq \quad \mathrm{Tr}\left[I(\rho|\mathcal{S})^{-1}G(\rho)\right] \quad \leq (1+\epsilon)\mathrm{Tr}\left[\overline{I}(\rho)^{-1}G(\rho)\right]$$

[7] A. Acharya, T. Kypraios, M.G., New Journal of Physics (2016)

Eigenvalues and MSE concentration



Concentration of the eigenvalues of Fisher information matrix and the MSE for 4 ions states of ranks 1,2,3

Proof

- Matrix Chernoff bound[8]

$$(1-\epsilon)\bar{I}(\rho) \leq I(\rho|\mathcal{S}) \leq (1+\epsilon)\bar{I}(\rho)$$

- Number of settings required (up to log factors)

$$m \approx \frac{\lambda_{\max}}{\lambda_{\min}} := \frac{\max_{\mathbf{s}} \lambda_{\max} I(\rho|\mathbf{s})}{\lambda_{\min}(\bar{I})}$$

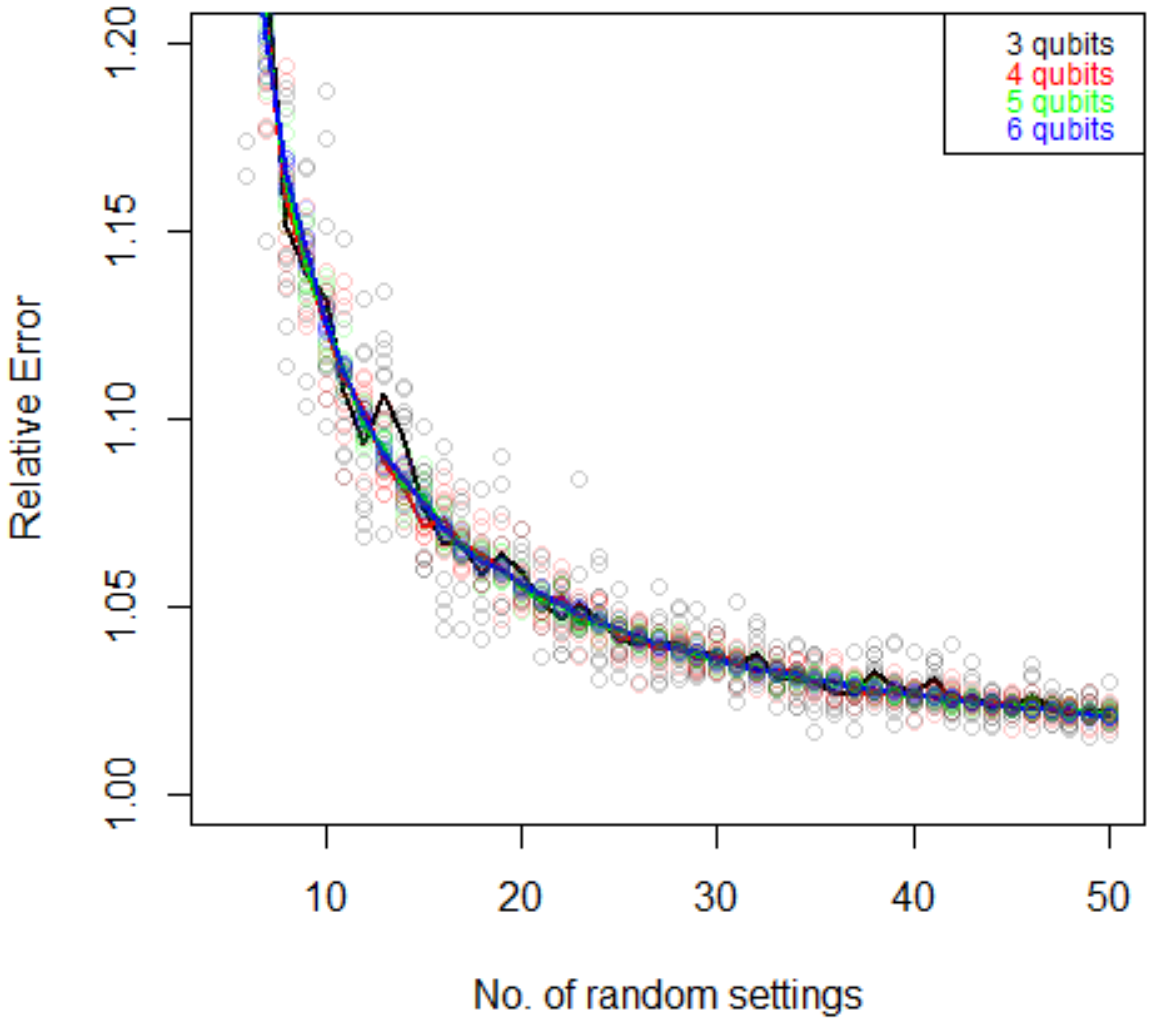
- $\bar{I}$ can be computed explicitly $\longrightarrow \lambda_{\min}(\bar{I}) = r/(r+1)$

- Quantum Cramér-Rao bound

$$I(\rho|\mathbf{s}) \leq F(\rho) \quad \longrightarrow \quad \lambda_{\max} I(\rho|\mathbf{s}) \leq \lambda_{\max} F(\rho) = 2r$$

[8] Ahlswede R. and Winter A., IEEE Transactions Information Theory **48** 569-579 (2002)

Log factors may not be necessary



Relative error w.r.t. asymptotic MSE for random settings, and pure states of 3-6 qubits.

Error upper bound for "compressive measurements"[9]

- when $\lambda_{\min}(\rho) \to 0$ the Fisher information matrix does not concentrate
- number of settings needed $m = \frac{C}{\lambda_{\min}(\rho)} \log\left(\frac{2(2rd-r^2-1)}{\delta}\right)$
- interested only in the behaviour of the asymptotic MSE $\mathrm{Tr}\left(I(\rho|\mathcal{S})^{-1}G(\rho)\right)$

Theorem (compressed sensing of rank r states)

Let ρ be a rank r state. If the number of settings is $m = Cr\log(2(2rd-r^2-1)/\delta)$, the asymptotic MSE satisfies

$$\mathrm{Tr}\left(I(\rho|\mathcal{S})^{-1}G(\rho)\right) \leq C(2rd-r^2-1)$$

with probability $1-\delta$.

[9] A. Acharya, M.G., arxiv:1609.03758

Lower bound

Question: are the proposed estimators "optimal" ?

- Asymptotic minimax risk over states $\rho \in \mathcal{S}_{d,r}$ of rank r

$$R_{minmax}(r) := \liminf_{n\to\infty} \inf_{\widehat{\rho}_n} \sup_{\rho\in\mathcal{S}_{d,r}} N \cdot \mathbb{E}\left(\|\widehat{\rho}_n - \rho\|_2^2\right)$$

Theorem

The following lower bound holds for the asymptotic minimax risk

$$R_{minmax}(r) \geq 2r(d-r).$$

- no estimation method can have rate faster than $\frac{\#\text{parameters(rank}=r)}{\#\text{samples}}$
- ratio between penalised and physical upper bound and minimax lower bound: $k\left(\frac{3}{2}\right)^k$

Idea of the proof

- Minimax rate in terms of Fisher information

$$R_{minmax}(r) = 3^k \sup_{\rho \in \mathcal{S}_{d,r}} \mathrm{Tr}\left(I^{-1}(\rho)G(\rho)\right)$$

- Minimax risk is larger than Bayes risk with uniform prior over matrices with spectrum $(1/r, \ldots, 1/r, 0, \ldots 0)$

$$R_{minmax}(r) \geq R_\pi(r,k) := 3^k \int \pi(d\rho)\mathrm{Tr}(G(\rho)^{1/2} I^{-1}(\rho) G(\rho)^{1/2})$$

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

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

- Due to the rotation symmetry the integral can be computed explicitly using Weingarten formulas

Outlook

- New class of estimators based on spectral truncation of the LS estimator
  - Can LS be replaced by a better linear estimators as starting point ?
  - Better understanding of the role of positivity (e.g. LS with positivity constraints)
  - Confidence intervals / regions
- MSE concentration for random measurements settings design
  - Concentration for random Pauli bases
  - Behaviour near boundary (very small non-zero eigenvalues)
  - Choosing number of settings for states with unknown rank