Spectral Thresholding in Quantum State Estimation for Low Rank States

Theo Kypraios
http://www.maths.nottingham.ac.uk/∼tk

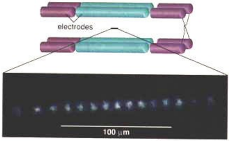
Joint work with:

Mădălin Guţă @ University of Nottingham
Cristina Butucea @ Paris Est

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

Problem: Quantum state estimation

- Quantum computation goal: create specific states of multiple "qubits" (e.g. ions)
- Validation: statistical estimation from measurement outcomes



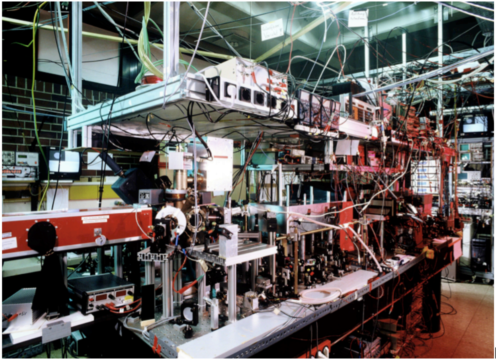
[Häffner *et al*, Nature 2005]

▷ $4^8 - 1 = 65\,535$ parameters

▷ $3^8 \times 100 = 656\,100$ measurements

▷ 10 hours measurement time

▷ days of computer time



Ra

Blatt's Lab, Innsbruck

"quantum computer" with 8 qubits (ions)

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

Quantum states

- Complex Hilbert space of 'wave functions' $\mathcal{H} = \mathbb{C}^d, L^2(\mathbb{R})$...

- State = preparation: complex density matrix ρ on $\mathcal{H}$
  - $\rho = \rho^*$ (selfadjoint)
  - $\rho \geq 0$ (positive)
  - $\mathrm{Tr}(\rho) = 1$ (normalised)

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1} & \rho_{d2} & \cdots & \rho_{dd} \end{pmatrix}$$

- Convex space of states
  - extremals (pure states) : one dimensional projection $P_\psi = |\psi\rangle\langle\psi| = \psi\psi^*$ with $\|\psi\| = 1$
  - Mixed state: convex combination of pure states

$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$$

- Natural distances: $\tau := \rho_1 - \rho_2$

$$\|\tau\| := \lambda_{\max}(|\tau|), \quad \|\tau\|_2^2 := \mathrm{Tr}(|\tau|^2), \quad \|\tau\|_1 := \mathrm{Tr}(|\tau|)$$

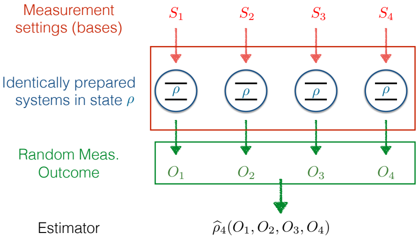
$$h(\rho_1, \rho_2) := 1 - \mathrm{Tr}\left(\sqrt{\rho_1^{1/2}\, \rho_2 \rho_1^{1/2}}\right) \ldots$$

Simple measurements

- Measurement setup given by an orthonormal basis 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

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

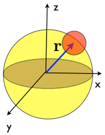
- Quantum state tomography: probe system with sufficient measurements to estimate ρ



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 $\|\mathbf{r}\| \leq 1$

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



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

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

- 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

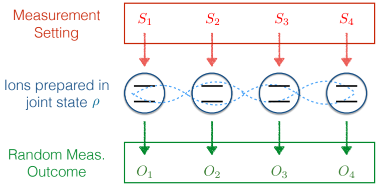
- State space of k two-level systems scales exponentially with k!

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

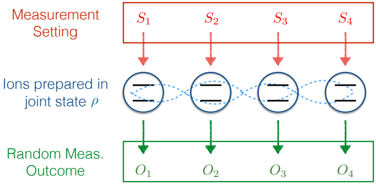
- Joint state of k ions
  - Density matrix ρ has $4^k - 1 = d^2 - 1$ parameters (e.g. $4^8 - 1 = 65535$)
  - Density matrix of rank r has $r(2 \cdot d - r) - 1$ (e.g. $2^8 - 2 = 254$ for a pure state)

- Simultaneous, separate measurements on each ion:
  - 3^k settings $\mathbf{s} = (s_1, \dots s_k) \in \{x, y, z\}^k$
  - 2^k outcomes $\mathbf{o} = (o_1, \dots o_k) \in \{+, -\}^k$
  - 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. 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 $M(\mathbb{C}^{2^k})$!

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 T. Monz)]

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

The least squares estimator

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

$$
\begin{aligned}
f_n(\mathbf{o}|\mathbf{s}) &= \frac{N(\mathbf{o}|\mathbf{s})}{n} &&= p(\mathbf{o}|\mathbf{s}) + \epsilon_n(\mathbf{o}|\mathbf{s}) \text{ (“multinomial error”)} \\
\mathbf{f}_n &= \mathbf{p} + \epsilon_n &&= \mathbf{A}\widetilde{\rho} + \epsilon_n
\end{aligned}
$$

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

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

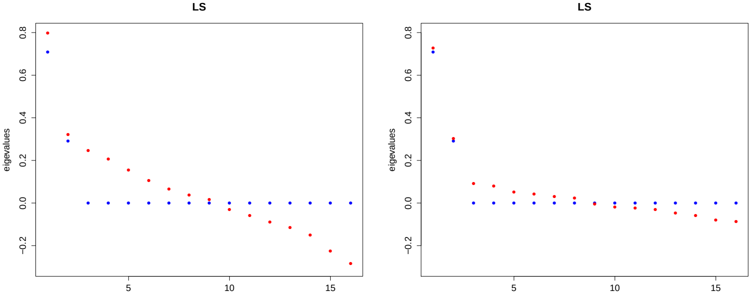
- Disadvantages
  - Least squares estimator is not a density matrix (not positive and trace one)
  - Least squares estimator is too "noisy" for low rank states
  - Least squares estimator minimises prediction rather than estimation error $\mathbb{E}\|\widehat{\rho}_n - \rho\|_2^2$

Eigenvalues distribution for the least squares estimator

- Eigenvalues decomposition for true state and least squares estimator

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

- 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 estimated 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[1]

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

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

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

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

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

$$\mathbb{E} \left\| \widehat{\rho}_n^{(ls)} - \rho \right\|_2^2 \leq C \frac{k}{n} \left(\frac{4}{3}\right)^k \approx C \log(d) \cdot \frac{\#\text{parameters}}{\#\text{samples}}$$

[1] Butucea, Guta & K (2015), arXiv:1504.08295; this improves on the $(4/3)^k$ factor in the upper bound of P. Alquier, C. Butucea, *et al*, Phys. Rev. A (2013)

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

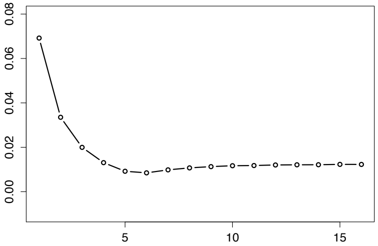
Penalising small eigenvalues

- 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 $k \leq d$

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

Question: how to choose the truncation rank ?

- Oracle estimator has rank which minimises the norm-two error E(k):=$\|\widehat{\rho}_n(k) - \rho\|_2^2$



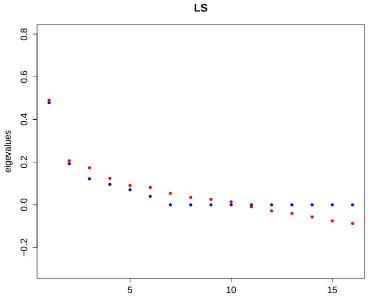
$E(k)$ for a state of rank $r = 6$, as function of truncation rank k

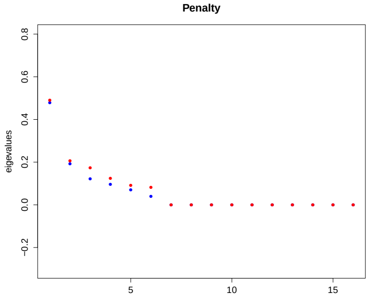
Penalised least squares estimator

- 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

$$\widehat{k} = \arg\min_k \left[\|\widehat{\rho}_n(k) - \widehat{\rho}_n\|_2^2 + k \cdot \nu_n^2\right] = \arg\min_k \left[\sum_{i=k+1}^{d} |\widehat{\lambda}_i|^2 + \nu_n^2 \cdot k\right]$$

- 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[2]

- 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

$$\|\widehat{\rho}_n^{(pen)} - \rho\|_2^2 \leq C \frac{r}{n} \left(\frac{2}{3}\right)^k \log\left(\frac{2^{k+1}}{\varepsilon}\right)$$

- *Concentration inequality gives upper bound for the MSE*

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

[2] Butucea, Guta & K (2015), arXiv:1504.08295

Outline

- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

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:

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

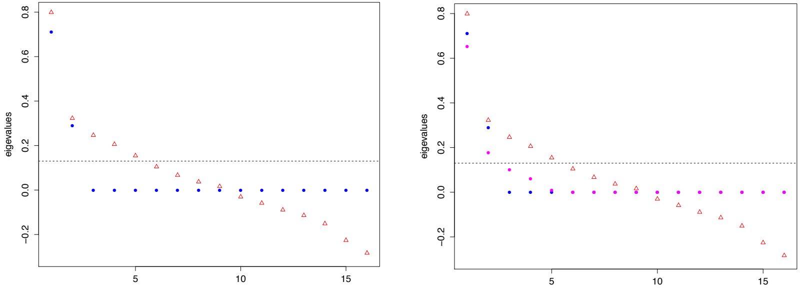
  - $\widetilde{\rho}_n^{(ls)}$ is the "normalised LS estimator" s.t. $\mathrm{Tr}\widetilde{\rho}_n^{(ls)} = 1$
  - $\mathcal{S}(\nu_n)$ denotes the set of states at noise level ν_n

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

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

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

$$\left\|\widehat{\rho}_n^{(phys)} - \rho\right\|_2^2 \leq C\frac{r}{n}\left(\frac{2}{3}\right)^k \log\left(\frac{2^{k+1}}{\varepsilon}\right)$$

- *Concentration inequality gives upper bound for the MSE*

$$\mathbb{E}\left\|\widehat{\rho}_n^{(ls)} - \rho\right\|_2^2 \leq C\frac{r\cdot k}{n}\left(\frac{2}{3}\right)^k \approx C\log(d)\cdot\frac{\#\text{parameters}(\text{rank}=r)}{\#\text{samples}}$$

Outline

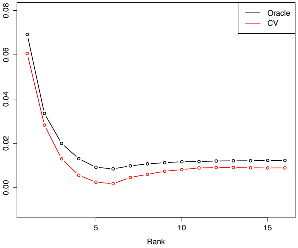
- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

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 $\widehat{\rho}_{n;j}^{(ls)}$ and $\widehat{\rho}_{n;-j}^{(ls)}$ 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)

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

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

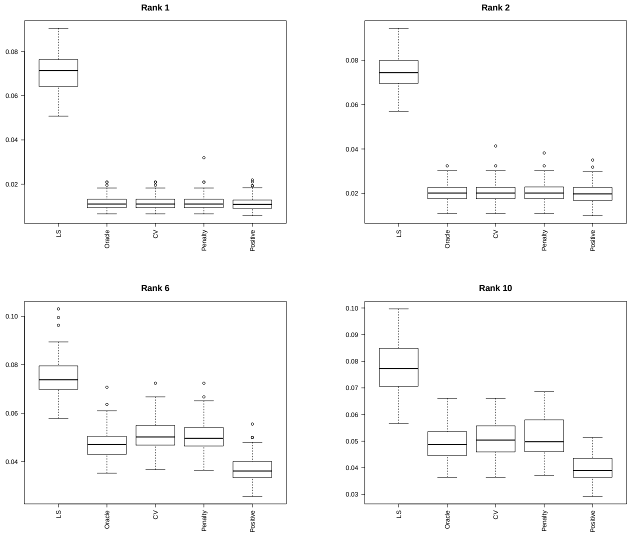


E(k) (black) and CV(k) (red) for one dataset from a rank 6 state with $n = 500$ repetitions

Outline

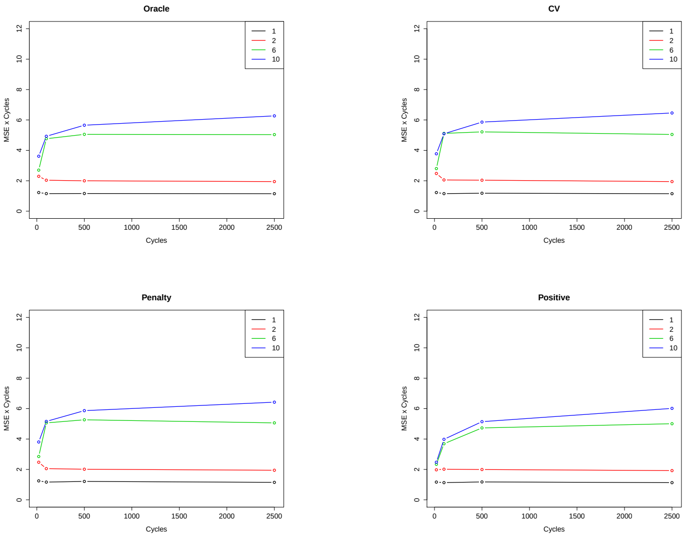
- Statistical model for multiple ions tomography
- Least-squares estimator
- Penalised estimator
- Physical estimator
- Cross-validation estimator
- Simulation Results

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



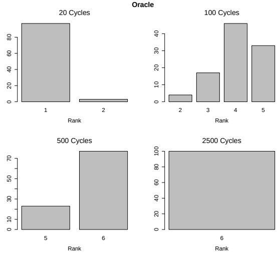
Boxplots of norm-two errors $\|\widehat{\rho}_n - \rho\|_2^2$ of different estimators for states of ranks $1, 2, 6, 10$ with $n = 100$ repetitions (computed from 100 datasets)

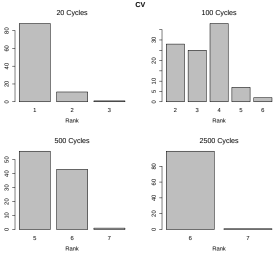
Comparison of estimators: MSE for different states and repetitions n

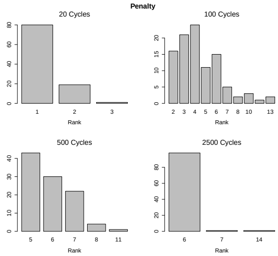


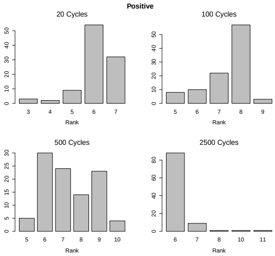
Renormalised MSEs $= n \cdot \mathbb{E}\|\widehat{\rho}_n - \rho\|_2^2$ as a function of number of repetitions n for states with different ranks: 1(black), 2 (red), 6 (green), 10 (blue) .

Comparison of estimators: Empirical distribution of chosen rank





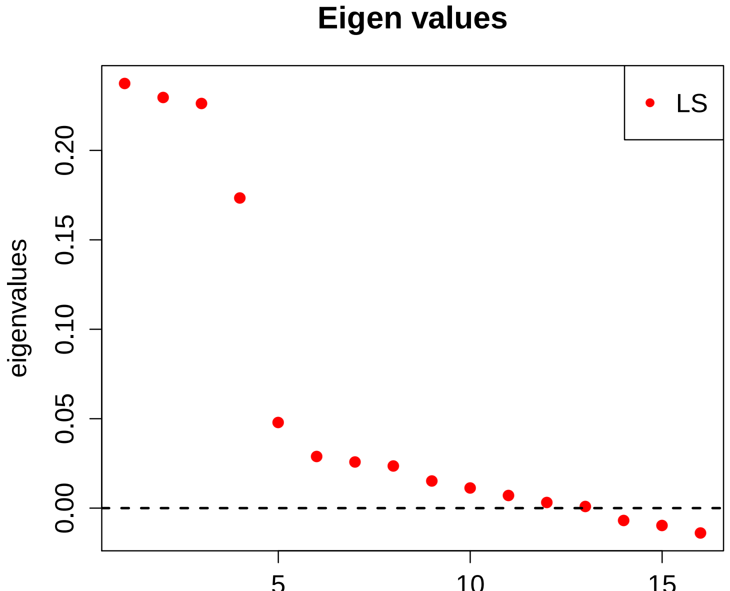




Histograms of the probabilities of the chosen rank for different estimators, as function of the number of repetitions n for a state of rank $r = 6$

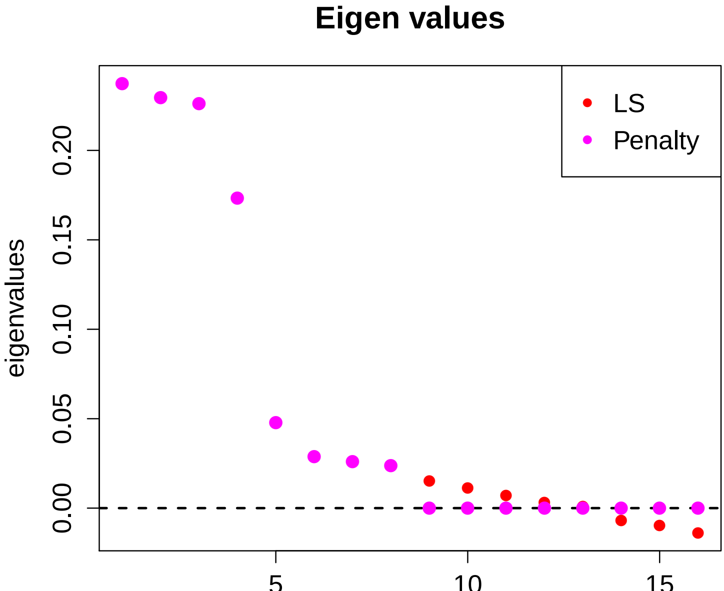
Real Data: 4 ions (thanks to Thomas Monz, Innsbruck)

4 atoms, 4800 total cycles



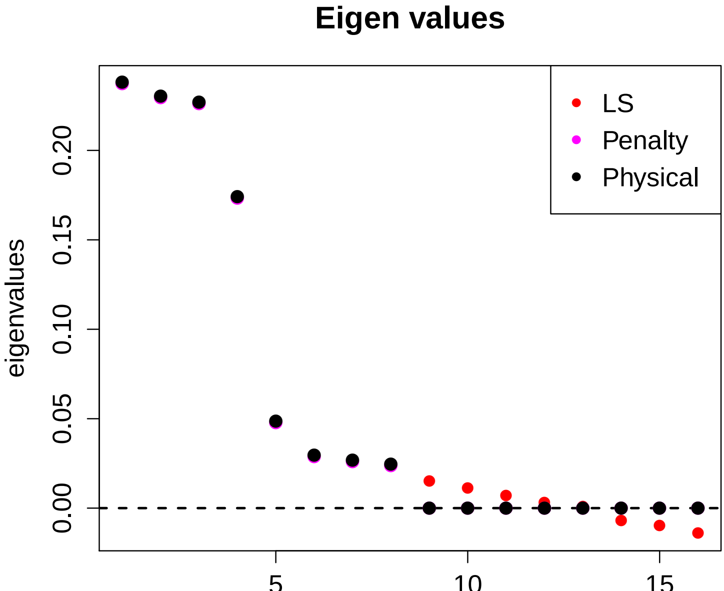
Real Data: 4 ions (thanks to Thomas Monz, Innsbruck)

4 atoms, 4800 total cycles



Real Data: 4 ions (thanks to Thomas Monz, Innsbruck)

4 atoms, 4800 total cycles



Real Data: 4 ions (thanks to Thomas Monz, Innsbruck)

The results (ie chosen ranks) are comparable to the ones obtained by more traditional rank (model) selection methods in Guta, K & Dryden (2012):

- Akaike Information Criterion: rank $= 9$
- Bayesian Information Criterion: rank $= 6$

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
- Measurement design with reduced measurement settings:
  - new threshold estimators
  - Behaviour of MSE with number of settings