Rank-penalized estimation of a quantum system

Cristina Butucea [1]

[1]Laboratoire d'analyse et mathématiques appliquées (LAMA) - Université Paris-Est Marne-la-Vallée

Nottingham, 2012

Outline

joint work with Pierre Alquier, Mohamed Hebiri, Katia Meziani (2012)

- Density matrix $\bar{\rho} \in \mathcal{M}_{2^n \times 2^n}(\mathbb{C})$ of a system of n qubits.
- Measurements: σ_x, σ_y, σ_z (Pauli matrices).
  For each $\mathbf{a} = (a_1, \ldots, a_n)$ in $\{x, y, z\}^n$, we denote by

$$\sigma_{\mathbf{a}} = \sigma_{a_1} \otimes \ldots \otimes \sigma_{a_n}.$$

- The random variable that we observe, $R^{\mathbf{a}}$, takes values $\mathbf{r} = (r_1, ..., r_n)$ in $\{\pm 1\}^n$, with probabilities

$$p(\mathbf{a}, \mathbf{r}) := \mathbb{P}(R^{\mathbf{a}} = \mathbf{r}) = \mathrm{Tr}(\bar{\rho} \cdot P^{\mathbf{a}}_{\mathbf{r}}),$$

where $P^{\mathbf{a}}_{\mathbf{r}} = P^{a_1}_{r_1} \otimes ... \otimes P^{a_n}_{r_n}$ and $P^{a_i}_{r_i}$ denote the projectors on the eigenvectors of the matrix σ_{a_i} associated to the eigenvalue r_i.

- For each measurement $\mathbf{a}$, we observe $R^{\mathbf{a},1}, \ldots, R^{\mathbf{a},m}$, here $m = 100$.
  We estimate $p(\mathbf{a}, \mathbf{r})$, by

$$\hat{p}(\mathbf{a}, \mathbf{r}) = \frac{1}{m} \sum_{i=1}^{m} I(R^{\mathbf{a},i} = \mathbf{r}).$$

- Solving the equation in $\bar{\rho}$

$$\mathrm{Tr}(\bar{\rho} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = p(\mathbf{a}, \mathbf{r})$$

is known as the *inversion method*.

- Solving the equation in $\hat{\bar{\rho}}$

$$\mathrm{Tr}(\hat{\bar{\rho}} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = \hat{p}(\mathbf{a}, \mathbf{r})$$

is known as the *method of moments* .

- Solving the equation in $\bar{\rho}$

$$\mathrm{Tr}(\bar{\rho} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = p(\mathbf{a}, \mathbf{r})$$

is known as the *inversion method*.

- Solving the equation in $\hat{\bar{\rho}}$

$$\mathrm{Tr}(\hat{\bar{\rho}} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = \hat{p}(\mathbf{a}, \mathbf{r})$$

is known as the *method of moments* .

- Identifiability: is there a unique state matrix $\bar{\rho}$ verifying the system of equations:

$$\mathrm{Tr}(\bar{\rho} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = p(\mathbf{a}, \mathbf{r}), \text{ for all } \mathbf{a} \text{ and } \mathbf{r}.$$

- Statistical properties:
-study the behaviour of the linear estimator $\hat{\bar{\rho}}$: bias, variance, error bounds on the risk $\|\hat{\bar{\rho}} - \bar{\rho}\|$ (operator norm)
-choose in an adaptive way (out of data) the finite-rank estimator closest to the linear estimator, by minimization of a risk penalized by the rank of the resulting estimator.

- Final estimator

$$\hat{\hat{\rho}}_\lambda = \arg\inf_R \{\|R - \hat{\rho}\|_F^2 + \lambda \cdot \mathrm{rank}(R)\}$$

and $\hat{k} = \mathrm{rank}(\hat{\hat{\rho}}_\lambda)$.

- By-products:
-estimator of the rank of $\bar{\rho}$, which is consistent asymptotically.
-'effective rank' of the final estimator in the non-asymptotic setup and its relation to the noise of the linear estimator.

-Bunea, She, Wegkamp (2011) Optimal selection of reduced rank estimators of high-dimensional matrices. *Ann. Statist.*
-Koltchinskii (2011): Von Neumann entropy penalization and low rank matrix estimation. *Ann. Statist.*
-Koltchinskii, Lounici, Tsybakov (2011): Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion, *Ann. Statist.*
-Klopp (2011): Rank-penalized estimators for high-dimensional matrices. *Electronic J. Statist.*

-Gross (2011): Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. on Information Theory*
-D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker and J. Eisert (2010): Quantum State Tomography via Compressed Sensing, *Phys. Rev. Lett.*
-Guță, Kypraios, Dryden (2012): Rank-based model selection for multiple ions quantum tomography.

- Let us explicit the solution of

$$\mathrm{Tr}(\bar{\rho} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = p(\mathbf{a}, \mathbf{r}), \text{ for all } \mathbf{a} \text{ and } \mathbf{r}.$$

- 1. Change of variables into the basis of Kronecker products of $\{\sigma_x, \sigma_y, \sigma_z, \sigma_I = I\}$ of $\mathcal{M}_{2^n \times 2^n}(\mathbb{C})$.
  For each $\mathbf{b} = (b_1, ..., b_n)$ in $\{x, y, z, I\}^n$, put

$$\sigma_{\mathbf{b}} = \sigma_{b_1} \otimes ... \otimes \sigma_{b_n}.$$

  Let $\rho = (\rho_{\mathbf{b}})_b$ in $\mathbb{C}^{4^n}$ be the vector of coefficients

$$\rho_{\mathbf{b}} = \mathrm{Tr}(\bar{\rho} \cdot \sigma_{\mathbf{b}}).$$

- Let us explicit the solution of

$$\mathrm{Tr}(\bar{\rho} \cdot P_{\mathbf{a}}^{\mathbf{r}}) = p(\mathbf{a}, \mathbf{r}), \text{ for all } \mathbf{a} \text{ and } \mathbf{r}.$$

- 1. Change of variables into the basis of Kronecker products of $\{\sigma_x, \sigma_y, \sigma_z, \sigma_I = I\}$ of $\mathcal{M}_{2^n \times 2^n}(\mathbb{C})$.
For each $\mathbf{b} = (b_1, ..., b_n)$ in $\{x, y, z, I\}^n$, put

$$\sigma_{\mathbf{b}} = \sigma_{b_1} \otimes ... \otimes \sigma_{b_n}.$$

Let $\rho = (\rho_{\mathbf{b}})_b$ in $\mathbb{C}^{4^n}$ be the vector of coefficients

$$\rho_{\mathbf{b}} = \mathrm{Tr}(\bar{\rho} \cdot \sigma_{\mathbf{b}}).$$

- 2. For each $\mathbf{b}$, denote by $E_{\mathbf{b}}$ the set of positions in $\mathbf{b}$ which are not the identity I and by $d(\mathbf{b})$ the number of indices in $\mathbf{b}$ different from I.
  There exists an explicit matrix V in $\mathcal{M}_{6^n,4^n}(\mathbb{C})$, such that

$$V\rho = p,$$

  that is $\sum_b V_{(\mathbf{a},\mathbf{r}),\mathbf{b}} \rho_{\mathbf{b}} = p(\mathbf{a},\mathbf{r})$ for all $\mathbf{b}$.
- Moreover,

$$[V^\top V]_{\mathbf{b},\mathbf{b}} = 3^{d(\mathbf{b})} 2^n.$$

- 2. For each $\mathbf{b}$, denote by $E_{\mathbf{b}}$ the set of positions in $\mathbf{b}$ which are not the identity I and by $d(\mathbf{b})$ the number of indices in $\mathbf{b}$ different from I.
  There exists an explicit matrix V in $\mathcal{M}_{6^n,4^n}(\mathbb{C})$, such that

  $$V\rho = p,$$

  that is $\sum_b V_{(\mathbf{a},\mathbf{r}),\mathbf{b}} \rho_{\mathbf{b}} = p(\mathbf{a},\mathbf{r})$ for all $\mathbf{b}$.
- Moreover,

  $$[V^\top V]_{\mathbf{b},\mathbf{b}} = 3^{d(\mathbf{b})} 2^n.$$

- 3. Then, the inversion formula is

  $$\rho = (V^\top V)^{-1} V^\top p,$$

  i.e.

  $$\rho_{\mathbf{b}} = \frac{1}{3^{d(\mathbf{b})} 2^n} \sum_{(\mathbf{r},\mathbf{a})} p_{(\mathbf{r},\mathbf{a})} V_{(\mathbf{r},\mathbf{a}),\mathbf{b}}.$$

- 2. For each $\mathbf{b}$, denote by $E_{\mathbf{b}}$ the set of positions in $\mathbf{b}$ which are not the identity I and by $d(\mathbf{b})$ the number of indices in $\mathbf{b}$ different from I.
  There exists an explicit matrix V in $\mathcal{M}_{6^n,4^n}(\mathbb{C})$, such that

$$V\rho = p,$$

  that is $\sum_b V_{(\mathbf{a},\mathbf{r}),\mathbf{b}}\rho_{\mathbf{b}} = p(\mathbf{a},\mathbf{r})$ for all $\mathbf{b}$.
- Moreover,

$$[V^\top V]_{\mathbf{b},\mathbf{b}} = 3^{d(\mathbf{b})}2^n.$$

- 3. Then, the inversion formula is

$$\rho = (V^\top V)^{-1}V^\top p,$$

  i.e.

$$\rho_{\mathbf{b}} = \frac{1}{3^{d(\mathbf{b})}2^n}\sum_{(\mathbf{r},\mathbf{a})} p_{(\mathbf{r},\mathbf{a})}V_{(\mathbf{r},\mathbf{a}),\mathbf{b}}.$$

The linear estimator is

$$\hat{\rho} = (V^\top V)^{-1} V^\top \hat{p}.$$

i.e.

$$\hat{\rho}_{\mathbf{b}} = \frac{1}{3^{d(\mathbf{b})} 2^n} \sum_{(\mathbf{r},\mathbf{a})} \hat{p}_{(\mathbf{r},\mathbf{a})} V_{(\mathbf{r},\mathbf{a}),\mathbf{b}}.$$

Note that larger is $d(\mathbf{b})$, better is the accuracy for estimating the coefficient $\rho_{\mathbf{b}}$.
Indeed, more measurements will bring information on $\rho_{\mathbf{b}}$ in this case.

The final estimator is defined as

$$\hat{\hat{\rho}}_\lambda = \arg\min_R \left[\left\| R - \hat{\hat{\rho}} \right\|_F^2 + \lambda \cdot \mathrm{rank}(R) \right],$$

where the minimum is taken over all Hermitian matrices R.

Remark for computational use:

$$\min_R \left[\left\| R - \hat{\hat{\rho}} \right\|_F^2 + \lambda \cdot \mathrm{rank}(R) \right] = \min_k \min_{R:\mathrm{rank}(R)=k} \left[\left\| R - \hat{\hat{\rho}} \right\|_F^2 + \lambda \cdot k \right].$$

Proposition The estimator $\hat{\rho}$ of ρ has the following properties:

1. it is unbiased, that is $\mathbb{E}[\hat{\rho}] = \rho$;
2. it has variance bounded as follows
$$Var(\hat{\rho}_{\mathbf{b}}) \leq \frac{1}{3^{d(\mathbf{b})} 4^n m};$$
3. for any $\varepsilon > 0$,
$$\mathbb{P}\left(\|\hat{\bar{\rho}} - \bar{\rho}\| \leq 4\sqrt{2\left(\frac{4}{3}\right)^n \frac{n\log(2) - \log(\varepsilon)}{m}} \right) \geq 1 - \varepsilon.$$

(use matrix-Hoeffding inequality - Tropp 2011).

Theorem For any $\theta > 0$ put $c(\theta) = 1 + 2/\theta$. We have on the event $\{\lambda \geq (1+\delta)\|\hat{\bar{\rho}} - \bar{\rho}\|^2\}$ that

$$\|\hat{\bar{\rho}}_\lambda - \bar{\rho}\|_F^2 \leq \min_k \left\{ c^2(\theta) \sum_{j>k} \lambda_j^2(\bar{\rho}) + 2c(\theta)\lambda k \right\},$$

where $\lambda_j(\bar{\rho})$ for $j = 1, \ldots, 2^n$ are the eigenvalues of $\bar{\rho}$ ordered decreasingly.

Choice of the penalty term For any $\theta > 0$ put $c(\theta) = 1 + 2/\theta$ and for some small $\varepsilon > 0$ choose

$$\lambda(\theta, \varepsilon) = 32(1+\theta)\left(\frac{4}{3}\right)^n \frac{n\log(2) - \log(\varepsilon)}{m}.$$

Then, we have

$$\|\hat{\bar{\rho}}_{\lambda(\theta,\varepsilon)} - \bar{\rho}\|_F^2 \leq \min_k \left\{ c^2(\theta) \sum_{j>k} \lambda_j^2(\bar{\rho}) + 2c(\theta)\lambda k \right\},$$

with probability larger than $1 - \varepsilon$.

Note that, if $\mathrm{rank}(\bar{\rho}) = d$, the previous theorem implies that

$$\|\hat{\bar{\rho}}_\lambda - \bar{\rho}\|_F^2 \leq 2c(\theta)\lambda d.$$

With the particular choice of the penalty, we have

$$\frac{1}{2^n}\|\hat{\bar{\rho}}_{\lambda(\theta,\varepsilon)} - \bar{\rho}\|_F^2 \leq 64c(\theta)(1+\theta)d\left(\frac{2}{3}\right)^n \frac{n\log(2) - \log(\varepsilon)}{m}.$$

($d(2/3)^n$ is the number of parameters of the state divided by the number of measurements).

Optimality of these bounds?!

The next result will state properties of $\hat{k}$, the rank of the final estimator $\hat{\hat{\rho}}_\lambda$.

Proposition

If there exists k such that $\lambda_k(\bar{\rho}) > (1+\delta)\sqrt{\lambda}$ and $\lambda_{k+1}(\bar{\rho}) < (1-\delta)\sqrt{\lambda}$ for some δ in $(0,1]$, then

$$\mathbb{P}(\hat{k} = k) \geq 1 - \mathbb{P}(\|\hat{\hat{\rho}} - \bar{\rho}\| \geq \delta\sqrt{\lambda}).$$

1. If the linear estimator is consistent asymptotically, $\hat{k}$ is consistent estimator of the true rank of $\bar{\rho}$.
2. In a finite sample approach, the procedure will select the eigenvalues above the noise $\|\hat{\hat{\rho}} - \bar{\rho}\|$ of the linear estimator.

Input: Random variables $R^{\mathbf{a},i}$, for each $\mathbf{a}$ in $\{x,y,z\}^n$ and i between 1 and m. Penalty term $\lambda > 0$.
Algorithm:

1. compute $\hat{p} = (\hat{p}(\mathbf{a},\mathbf{r}))_{\mathbf{a},\mathbf{r}}$ and the matrix V;
2. compute $\hat{\rho}$ and change the variables into $\hat{\hat{\rho}}$;
3. compute the eigenvectors $W = [w_1, ..., w_{2^n}]$ of $\hat{\hat{\rho}}^* \hat{\hat{\rho}}$; $U = \hat{\hat{\rho}} W$;
4. let U_k and W_k be the restriction to the first k columns of U and W, respectively;
5. $R_k = U_k W_k^*$ and $C_k = \|R_k - \hat{\hat{\rho}}\|_F^2 + \lambda \cdot k$, k in $\{1, ..., 2^n\}$;
6. choose the minimum between $\{C_k\}_k$.

Output: $\hat{k}$ and $\hat{\hat{\rho}}_\lambda = R_{\hat{k}}$.

Figure: Frequency of good selection of the true rank d with respect to d, with $\lambda = \lambda_n^{(1)}$ (green squares) and with $\lambda = \lambda_n^{(2)}$ (blue stars). The results are established on 20 repetitions. A value equal to 1 in the y-axis means that the method always selects the good rank, whereas 0 means that it always fails. Left: $m = 50$ measurements – Right: $m = 100$ measurements

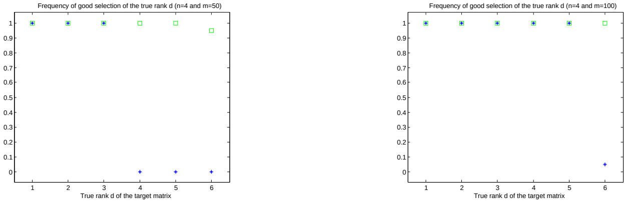

Figure: Evaluation of the operator norm $\sqrt{\lambda_n^{(1)}} = \|\hat{\bar{\rho}} - \bar{\rho}\|$. The results are established on 20 repetitions. $n = 4$, $m = 50$ repetitions of the measurements ; we compare the errors when d takes values between 1 and 6 –

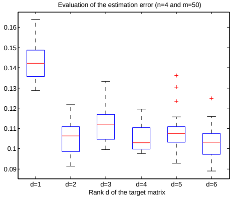

Figure: Evaluation of the operator norm $\sqrt{\lambda_n^{(1)}} = \|\hat{\bar{\rho}} - \bar{\rho}\|$. The results are established on 20 repetitions. $n = 5$, $m = 100$; we compare the errors when d takes values between 1 and 6 –

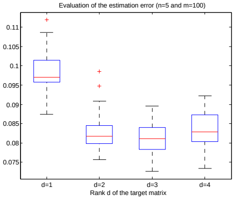

Figure: Evaluation of the operator norm $\sqrt{\lambda_n^{(1)}} = \|\hat{\bar{\rho}} - \bar{\rho}\|$. The results are established on 20 repetitions. The rank equals $d = 4$. Compare the error for $m = 50$ and 100.

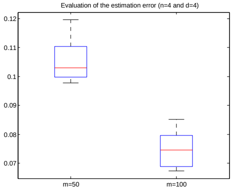

Figure: Eigenvalues of the linear estimator in increasing order and the penalty choice; $m = 100$ and $n = 4$.

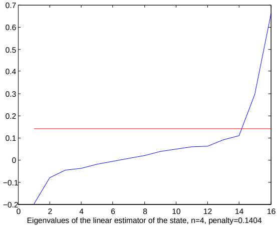

Figure: Eigenvalues of the linear estimator in increasing order and the penalty choice; $m = 100$ and $n = 5$.

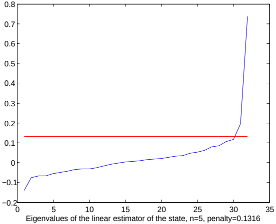

Figure: Eigenvalues of the linear estimator in increasing order and the penalty choice; $m = 100$ and $n = 6$.