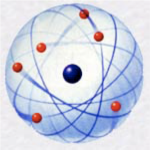
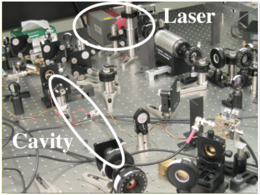
Quantum tomography by regularized linear regressions

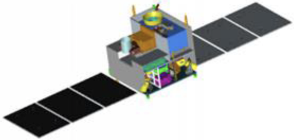
Biqiang Mu[1], Hongsheng Qi[1], Ian R. Petersen[2], and Guodong Shi[3]

1. Academy of Mathmatics and Systems Science, Chinese Academy of Sciences
2. Research School of EEME, The Australian National University
3. Australian Center for Field Robotics, The University of Sydney

The Quantum Leap







Problem Definition and Preliminaries

Problem Formulation

Settings

- $\mathcal{H}$: a d-dimensional Hilbert space that characterizes the state space of a quantum system
- $\mathcal{L}(\mathcal{H})$: the space of linear operators over $\mathcal{H}$
- $\{\mathsf{B}_i\}_{i=1}^{d^2}$: an orthonormal basis of $\mathcal{L}(\mathcal{H})$ with $\mathrm{Tr}(\mathsf{B}_i^\dagger \mathsf{B}_j) = \delta_{ij}$ and $\mathsf{B}_i^\dagger = \mathsf{B}_i$

Quantum State $\rho \in \mathcal{L}(\mathcal{H})$ as a density operator can be expressed by

$$\rho = \sum_{i=1}^{d^2} \theta_i \mathsf{B}_i$$

where $\theta_i = \mathrm{Tr}(\rho \mathsf{B}_i)$.

Problem Formulation

Definition. A positive operator-valued measurement (POVM) over $\mathcal{H}$, denoted by $\{\mathsf{M}_m\}_{m=1}^{M}$ with $\sum_{m=1}^{M} \mathsf{M}_m^\dagger \mathsf{M}_m = \mathsf{I}$.

Then $\mathsf{E}_m \overset{\triangle}{=} \mathsf{M}_m^\dagger \mathsf{M}_m$ can be expressed as

$$\mathsf{E}_m = \sum_{i=1}^{d^2} \beta_{mi} \mathsf{B}_i$$

for each $1 \leq m \leq M$, where $\beta_{mi} = \mathrm{Tr}(\mathsf{E}_m \mathsf{B}_i)$.

When the quantum state ρ is being measured under the POVM $\{\mathsf{M}_m\}_{m=1}^{M}$, the probability of observing outcome m is

$$p_m = \mathrm{Tr}(\mathsf{E}_m \rho) = \boldsymbol{\beta}_m^\top \boldsymbol{\theta},$$

where $\boldsymbol{\beta}_m = [\beta_{m1}, \cdots, \beta_{md^2}]^\top$ and $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_{d^2}]^\top$.

Problem Formulation

Denoting

$$\mathrm{p} = [p_1, \ldots, p_M]^\top \in \mathbb{R}^M$$
$$\mathbf{A} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M]^\top \in \mathbb{R}^{M \times d^2}$$

we have the following fundamental quantum measurement description in the form of a linear algebraic equation:

$$\mathrm{p} = \mathbf{A}\boldsymbol{\theta}.$$

The tomography of an unknown quantum state ρ is equivalent to identifying the vector $\boldsymbol{\theta}$, where $\mathbf{A}$ is known and p is estimated by experimental realizations of measuring ρ from the POVM $\{\mathsf{M}_m\}_{m=1}^{M}$.

Tomography Procedure

1. Prepare n identical copies of an uncertain quantum state ρ;
2. Perform the POVM measurement $\{\mathsf{M}_m\}_{m=1}^{M}$ independently for the n copies;
3. Record the number of times that the outcome m is observed, denoted by $\#m$, from the n experiments for each $1 \leq m \leq M$.

Then

$$\widehat{p}_m = \frac{\#m}{n}$$

is a natural estimator of the probability p_m, leading to

$$\widehat{p}_m = \boldsymbol{\beta}_m^\top \boldsymbol{\theta} + e_m,$$

where $e_m = \widehat{p}_m - p_m$ is the estimation error.

Estimation error

Define i.i.d. Bernoulli random variables $b_l^{(m)}$ for $1 \leq l \leq n$, which takes value 1 with probability p_m and 0 with probability $1 - p_m$. Then there holds

$$e_m = \widehat{p}_m - p_m = \frac{\sum_{l=1}^{n} b_l^{(m)}}{n} - p_m = \sum_{l=1}^{n} \frac{b_l^{(m)} - p_m}{n}.$$

Note that $(b_l^{(m)} - p_m)/n$ takes value $(1 - p_m)/n$ with probability p_m and $-p_m/n$ with probability $1 - p_m$.

Tomography Model

Linear regression problem:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{e}$$

where $\mathbf{y} = [\widehat{p}_1, \cdots, \widehat{p}_M]^\top$ and $\mathbf{e} = [e_1, \cdots, e_M]^\top$.

key differences

1. the number M of measurements is fixed
2. the variance of $\mathbf{e}$ decreases as the number n of copies

Natural prior knowledge on the problem

1. Heteroscedasticity

$$\mathbb{E}(e_m) = 0,\ \mathbb{V}(e_m) = \mathbb{E}(e_m)^2 = (p_m - p_m^2)/n$$

2. $\mathrm{Tr}(\rho) = 1$
3. ρ is of low rank

Standard Least Squares

The least squares (LS) solution is

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{\mathrm{LS}} &= \arg\min_{\boldsymbol{\theta}} \; (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \\ &= (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{y}\end{aligned}$$

The $\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}$ admit the following properties:

- $\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}$ is unbiased, namely, $\mathbb{E}\big(\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}\big) = \boldsymbol{\theta}$;
- $$\begin{aligned}\mathbb{MSE}\big(\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}\big) &\triangleq \mathbb{E}(\widehat{\boldsymbol{\theta}}^{\mathrm{LS}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}^{\mathrm{LS}} - \boldsymbol{\theta})^\top \\ &= (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathrm{P}\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\end{aligned}$$

  where $\mathrm{P} = \mathrm{diag}\big([p_1 - p_1^2, \cdots, p_M - p_M^2]\big)/n$.

Remark:

- Standard LS neglects the fact that the e_m have different variances, although they are all zero mean.
- The condition that $\mathbf{A}$ be full column rank means the POVM $\{\mathsf{M}_m\}_{m=1}^{M}$ is informationally complete.

Regularized Linear Regressions

Weighted Linear Regression: Heteroscedasticity

The weighted least squares (WLS) estimate

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} &= \arg\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top \mathrm{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \\ &= (\mathbf{A}^\top \mathrm{W} \mathbf{A})^{-1} \mathbf{A}^\top \mathrm{W} \mathbf{y} \\ \mathrm{W} = \mathrm{P}^{-1} &= n \cdot \mathrm{diag}\big([1/(p_1 - p_1^2), \cdots, 1/(p_M - p_M^2)]\big)\end{aligned}$$

Property of the WLS estimator

- $\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}$ is unbiased, i.e., $\mathbb{E}\big(\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}\big) = \boldsymbol{\theta}$;
- $\mathbb{MSE}\big(\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}\big) = (\mathbf{A}^\top \mathrm{W} \mathbf{A})^{-1}$.

Suppose $\mathrm{rank}(\mathbf{A}) = d^2$ and let $\widehat{\boldsymbol{\theta}}$ be any linear unbiased estimate for $\boldsymbol{\theta}$. Thus we have

$$\mathbb{MSE}(\widehat{\boldsymbol{\theta}}) \geq \mathbb{MSE}(\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}).$$

Weighted Linear Regression

In practice, the matrix W is unknown and a feasible solution is to use the estimate

$$\widehat{\boldsymbol{\theta}}^{\mathrm{AWLS}} = (\mathbf{A}^\top \widehat{\mathrm{W}} \mathbf{A})^{-1} \mathbf{A}^\top \widehat{\mathrm{W}} \mathbf{y},$$

where W is replaced by

$$\widehat{\mathrm{W}} = n \cdot \mathrm{diag}\big([1/(\widehat{p}_1 - \widehat{p}_1^2), \cdots, 1/(\widehat{p}_M - \widehat{p}_M^2)]\big).$$

There holds for large n that

$$\widehat{\boldsymbol{\theta}}^{\mathrm{AWLS}} - \widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} = O_p(1/\sqrt{n})(\mathbf{A}^\top \mathrm{W} \mathbf{A})^{-1} \mathbf{A}^\top \mathrm{W} \mathbf{e}.$$

Constrained Weighted Regression: Unit Trace

The quantum state has an essential requirement

$$\mathrm{Tr}(\rho) = 1.$$

Note that

$$\rho = \sum_{i=1}^{d^2} \theta_i \mathsf{B}_i.$$

This becomes

$$\boldsymbol{\theta}^\top \mathrm{Tr}(\mathsf{B}) = 1$$

where $\mathrm{Tr}(\mathsf{B}) \triangleq [\mathrm{Tr}(\mathsf{B}_1), \cdots, \mathrm{Tr}(\mathsf{B}_{d^2})]^\top$.

The constrained weighted least squares (CWLS) estimate

$$\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}} = \underset{\boldsymbol{\theta}^\top \mathrm{Tr}(\mathsf{B})=1}{\arg\min} \; (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top \mathrm{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}).$$

Constrained Weighted Regression

Proposition 1. Suppose $\mathrm{rank}(\mathbf{A}) = d^2$. The CWLS estimate $\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}}$ has the following closed-form solution

$$\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}} = \widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} - \frac{C\,\mathrm{Tr}(\mathsf{B})}{\mathrm{Tr}(\mathsf{B})^\top C\,\mathrm{Tr}(\mathsf{B})}\big(\,\mathrm{Tr}(\mathsf{B})^\top \widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} - 1\big)$$

where $C = (\mathbf{A}^\top \mathrm{W} \mathbf{A})^{-1}$ and its MSE matrix is

$$\begin{aligned}\mathbb{MSE}\big(\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}}\big) &\stackrel{\triangle}{=} \mathbb{E}(\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}} - \boldsymbol{\theta})^\top \\ &= F\end{aligned}$$

where

$$F \stackrel{\triangle}{=} C - \frac{C\,\mathrm{Tr}(\mathsf{B})\,\mathrm{Tr}(\mathsf{B})^\top C}{\mathrm{Tr}(\mathsf{B})^\top C\,\mathrm{Tr}(\mathsf{B})}.$$

Constrained Weighted Regression

$\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}}$ is optimal in the sense that

$$\mathbb{MSE}(\widehat{\boldsymbol{\theta}}) \geq \mathbb{MSE}(\widehat{\boldsymbol{\theta}}^{\mathrm{CWLS}}).$$

where $\widehat{\boldsymbol{\theta}}$ is any unbiased estimate for $\boldsymbol{\theta}$ that is affine $\mathbf{y}$ and satisfies the constraint $\boldsymbol{\theta}^\top \operatorname{Tr}(\mathsf{B}) = 1$.

Regularized Weighted Regression

Two motivations

- When the POVM $\{\mathsf{M}_m\}_{m=1}^{M}$ is under-determinate, the matrix $\mathbf{A}$ might not have full column rank;
- ρ would be of low rank.

The nuclear norm of ρ is

$$\|\rho\|_\star \overset{\triangle}{=} \sum_{i=1}^{d} \sigma_i(\rho) = \sum_{i=1}^{d} \sqrt{\lambda_i(\rho^\dagger \rho)} = \sum_{i=1}^{d} \lambda_i(\rho) = \mathrm{Tr}(\rho) = 1$$

The nuclear norm of $\rho^\dagger \rho$ is

$$\begin{aligned} \|\rho^\dagger \rho\|_\star &\overset{\triangle}{=} \sum_{i=1}^{d} \sigma_i(\rho^\dagger \rho) = \mathrm{Tr}(\rho^\dagger \rho) \\ &= \mathrm{Tr}\left[\left(\sum_{i=1}^{d^2} \theta_i \mathsf{B}_i \right)^\dagger \left(\sum_{j=1}^{d^2} \theta_j \mathsf{B}_j \right) \right] = \sum_{i=1}^{d^2} |\theta_i|^2 = \|\boldsymbol{\theta}\|^2 \end{aligned}$$

Regularized Weighted Regression

Consider the following problem

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top \mathrm{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \\ \text{subject to} \quad & \boldsymbol{\theta}^\top \operatorname{Tr}(\mathsf{B}) = 1, \ \|\boldsymbol{\theta}\|^2 \leq c \end{aligned}$$

which is equivalent to the constrained regularized weighted least squares (CRWLS) estimate

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top \mathrm{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) + \gamma \|\boldsymbol{\theta}\|^2 \\ \text{subject to} \quad & \boldsymbol{\theta}^\top \operatorname{Tr}(\mathsf{B}) = 1. \end{aligned}$$

where $\gamma \geq 0$ is a regularization parameter.

Regularized Weighted Regression

The CRWLS estimate is given by

$$\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}} = \widehat{\boldsymbol{\theta}}^{\mathrm{RWLS}} - C\,\mathrm{Tr}(\mathsf{B})\frac{\mathrm{Tr}(\mathsf{B})^\top \widehat{\boldsymbol{\theta}}^{\mathrm{RWLS}} - 1}{\mathrm{Tr}(\mathsf{B})^\top C\,\mathrm{Tr}(\mathsf{B})}$$

where $C = (\mathbf{A}^\top \mathrm{W} \mathbf{A} + \gamma \mathsf{I})^{-1}$ and $\widehat{\boldsymbol{\theta}}^{\mathrm{RWLS}}$ is the regularized weighted least squares (RWLS) estimate

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{\mathrm{RWLS}} &\triangleq \arg\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^\top \mathrm{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) + \gamma \|\boldsymbol{\theta}\|^2 \\ &= C\mathbf{A}^\top \mathrm{W} \mathbf{y}.\end{aligned}$$

Regularized Weighted Regression

The MSE matrix of $\widehat{\boldsymbol{\theta}}^{\text{CRWLS}}$ is

$$\begin{aligned}\mathbb{MSE}(\widehat{\boldsymbol{\theta}}^{\text{CRWLS}}) &\triangleq \mathbb{E}(\widehat{\boldsymbol{\theta}}^{\text{CRWLS}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}^{\text{CRWLS}} - \boldsymbol{\theta})^\top \\ &= F - \gamma F(\mathsf{I} - \gamma \boldsymbol{\theta}\boldsymbol{\theta}^\top)F\end{aligned}$$

where $F = C - \frac{C\,\mathrm{Tr}(\mathsf{B})\,\mathrm{Tr}(\mathsf{B})^\top C}{\mathrm{Tr}(\mathsf{B})^\top C\,\mathrm{Tr}(\mathsf{B})}$. There holds

$$\mathbb{MSE}(\widehat{\boldsymbol{\theta}}^{\text{CRWLS}}) < \mathbb{MSE}(\widehat{\boldsymbol{\theta}}^{\text{CWLS}}),$$

if $0 < \gamma < 2/\big(\|\boldsymbol{\theta}\|^2 - \frac{1}{\|\mathrm{Tr}(\mathsf{B})\|^2}\big)$.

Remark

- $\widehat{\boldsymbol{\theta}}^{\text{CWLS}}$ has the smallest MSE among all the unbiased estimate of $\boldsymbol{\theta}$ affine with $\mathbf{y}$
- $\widehat{\boldsymbol{\theta}}^{\text{CRWLS}}$ has a smaller MSE than $\widehat{\boldsymbol{\theta}}^{\text{CWLS}}$ even if $\widehat{\boldsymbol{\theta}}^{\text{CRWLS}}$ is also affine with $\mathbf{y}$.

Tuning γ: Minimizing risk

Introduce the risk for the estimate $\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}$

$$\begin{aligned} R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}) &\triangleq \mathbb{E}\big(\mathbf{A}\boldsymbol{\theta} - \mathbf{A}\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}\big)^\top \mathrm{W}\big(\mathbf{A}\boldsymbol{\theta} - \mathbf{A}\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}\big) \\ &= \gamma^2 \boldsymbol{\theta}^\top F \mathbf{A}^\top \mathrm{W} \mathbf{A} F \boldsymbol{\theta} + \mathrm{Tr}\left(F\mathbf{A}^\top \mathrm{W} \mathbf{A} F \mathbf{A}^\top \mathrm{W} \mathbf{A}\right) \end{aligned}$$

which is a reference measure to characterize how well the estimate $\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}$ can achieve.

Tune γ by the risk

$$\widehat{\gamma}_R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}) \triangleq \arg\min_{\gamma \geq 0} R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})$$

is the optimal regularization parameter γ for any given data in the risk sense.

Tuning γ: An implementable method

Define the cost function

$$\mathcal{C}(\gamma) \triangleq (\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})^\top \mathrm{W}(\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}) + 2\,\mathrm{Tr}\left(\mathbf{AH}\right)$$

where

$$\mathbf{H} = C\mathbf{A}^\top\mathrm{W} - C\,\mathrm{Tr}(\mathsf{B})\frac{\mathrm{Tr}(\mathsf{B})^\top C\mathbf{A}^\top\mathrm{W}}{\mathrm{Tr}(\mathsf{B})^\top C\,\mathrm{Tr}(\mathsf{B})}$$

$\mathcal{U}(\gamma)$ is an unbiased estimate for the risk measure $R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})$, namely,

$$\mathbb{E}\mathcal{C}(\gamma) = R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})$$

An implementable tuning estimator is

$$\widehat{\gamma}_u(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}) = \underset{\gamma\geq 0}{\arg\min}\ \mathcal{C}(\gamma)$$

An Equivalent Regression Model

An Equivalent Regression Model

Recall the linear model with an equality constraint regression model

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{e}, \text{ subject to } \boldsymbol{\theta}^\top \mathrm{Tr}(\mathsf{B}) = 1$$

Construct an orthogonal matrix $\mathbf{Q}$ of size $d^2 \times d^2$ as follows. The first row of $\mathbf{Q}$ is

$$\mathrm{Tr}(\mathsf{B})^\top / \|\mathrm{Tr}(\mathsf{B})\|$$

and the remaining rows are chosen such that $\mathbf{Q}$ is orthogonal. Thus, we have

$$\mathbf{y} = \underbrace{\mathbf{A}\mathbf{Q}^\top}_{\mathbf{D}} \underbrace{\mathbf{Q}\boldsymbol{\theta}}_{\boldsymbol{\beta}} = \mathbf{D}\boldsymbol{\beta} + \mathbf{e}$$

$$\mathbf{D} \triangleq \mathbf{A}\mathbf{Q}^\top = [\mathbf{d}, \mathbf{K}], \ \mathbf{d} = \mathbf{A}\mathrm{Tr}(\mathsf{B}) / \|\mathrm{Tr}(\mathsf{B})\|$$

$$\boldsymbol{\beta} \triangleq \mathbf{Q}\boldsymbol{\theta} = [\beta_1, \boldsymbol{\alpha}^\top]^\top, \ \beta_1 = 1/\|\mathrm{Tr}(\mathsf{B})\|.$$

An Equivalent Regression Model

The unconstrained linear model

$$\mathbf{z} = \mathbf{y} - \frac{1}{\|\mathrm{Tr}(\mathsf{B})\|}\mathbf{d} = \mathbf{K}\boldsymbol{\alpha} + \mathbf{e}.$$

The RWLS estimate for the equivalent model is defined as

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}} &= \arg\min_{\boldsymbol{\alpha}}(\mathbf{z}-\mathbf{K}\boldsymbol{\alpha})^\top \mathrm{W}(\mathbf{z}-\mathbf{K}\boldsymbol{\alpha})+\gamma\|\boldsymbol{\alpha}\|^2 \\ &= \mathbf{U}\mathbf{z}\end{aligned}$$

where

$$\mathbf{U} \triangleq \mathbf{V}\mathbf{K}^\top \mathrm{W}, \quad \mathbf{V} \triangleq (\mathbf{K}^\top \mathrm{W}\mathbf{K} + \gamma \mathsf{I})^{-1}.$$

An Equivalent Regression Model

Intuitively, for an estimate $\widehat{\boldsymbol{\alpha}}$ of the unconstrained linear model, the vector defined by

$$\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}) \triangleq \mathbf{Q}^\top \begin{bmatrix} \frac{1}{\|\mathrm{Tr}(\mathrm{B})\|} \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix}$$

should be the corresponding estimate for constrained linear model and independent of the choice of $\mathbf{Q}$.

For any regularization parameter $\gamma \geq 0$, there holds

$$\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) = \widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}.$$

Moreover,

$$\begin{aligned} \mathbb{MSE}\big(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}(\gamma)\big) &\triangleq \mathbb{E}(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}} - \boldsymbol{\alpha})(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}} - \boldsymbol{\alpha})^\top \\ &= \gamma^2 \mathbf{V}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\mathbf{V} + \mathbf{V}\mathbf{K}^\top \mathrm{W}\mathbf{K}\mathbf{V}. \end{aligned}$$

Asymptotically Optimal Regularization Gain

The risk for the estimate $\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}$ can be similarly defined as

$$R(\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}) \triangleq \mathbb{E}\big(\mathbf{K}\boldsymbol{\alpha} - \mathbf{K}\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}\big)^\top \mathrm{W}\big(\mathbf{K}\boldsymbol{\alpha} - \mathbf{K}\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}\big)$$

and the resulting optimal regularization parameter is

$$\widehat{\gamma}_R(\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}) \triangleq \arg\min_{\gamma \geq 0} R(\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}(\gamma)).$$

Let us construct an unbiased estimate for $R(\widehat{\boldsymbol{\alpha}}^{\text{RWLS}})$

$$\mathcal{C}_u(\gamma) \triangleq (\mathbf{z} - \mathbf{K}\widehat{\boldsymbol{\alpha}}^{\text{RWLS}})^\top \mathrm{W}(\mathbf{z} - \mathbf{K}\widehat{\boldsymbol{\alpha}}^{\text{RWLS}}) + 2\,\mathrm{Tr}\,\big(\mathbf{KU}\big)$$

$$\mathbb{E}\mathcal{C}_u(\gamma) = R(\widehat{\boldsymbol{\alpha}}^{\text{RWLS}})$$

Asymptotically Optimal Regularization Gain

Tune γ by

$$\widehat{\gamma}_u(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) \stackrel{\triangle}{=} \arg\min_{\gamma \geq 0} \mathcal{C}_u(\gamma)$$

There hold

$$\widehat{\gamma}_R(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) = \widehat{\gamma}_R(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})$$
$$\widehat{\gamma}_u(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) = \widehat{\gamma}_u(\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}})$$

Asymptotically Optimal Regularization Gain

Denote

$$\Sigma \triangleq \mathbf{K}^\top \mathrm{diag}\big([p_1 - p_1^2, \cdots, p_M - p_M^2]\big)\mathbf{K}$$
$$\Upsilon \triangleq \mathbf{A}^\top \mathrm{diag}\big([p_1 - p_1^2, \cdots, p_M - p_M^2]\big)\mathbf{A}.$$

Suppose $\mathrm{rank}(\mathbf{A}) = d^2$. The limits take place as the sample size $n \to \infty$ by

$$\widehat{\gamma}_R(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) \to \gamma^\star \text{ deterministically}$$
$$\widehat{\gamma}_u(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) \to \gamma^\star \text{ almost surely}$$

Asymptotically Optimal Regularization Gain

- The limit

$$\gamma^\star = \frac{\operatorname{Tr}\left(\Sigma^{-1}\right)}{\boldsymbol{\alpha}^\top \Sigma^{-1} \boldsymbol{\alpha}} = \frac{\operatorname{Tr}\left(\Upsilon^{-1}\right) - \frac{\operatorname{Tr}(\mathrm{B})^\top \Upsilon^{-2} \operatorname{Tr}(\mathrm{B})}{\operatorname{Tr}(\mathrm{B})^\top \Upsilon^{-1} \operatorname{Tr}(\mathrm{B})}}{\boldsymbol{\theta}^\top \Upsilon^{-1} \boldsymbol{\theta} - \frac{\boldsymbol{\theta}^\top \Upsilon^{-1} \operatorname{Tr}(\mathrm{B}) \operatorname{Tr}(\mathrm{B})^\top \Upsilon^{-1} \boldsymbol{\theta}}{\operatorname{Tr}(\mathrm{B})^\top \Upsilon^{-1} \operatorname{Tr}(\mathrm{B})}}.$$

- There hold as $n \to \infty$

$$n\left(\widehat{\gamma}_R(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) - \gamma^\star\right) \to \frac{3\gamma^\star\left(\gamma^\star \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}^{-\mathbf{2}} \boldsymbol{\alpha} - \operatorname{Tr}\left(\Sigma^{-2}\right)\right)}{\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{\alpha}}$$

  deterministically and

$$\sqrt{n}\left(\widehat{\gamma}_u(\widehat{\boldsymbol{\alpha}}^{\mathrm{RWLS}}) - \gamma^\star\right) \to \mathcal{N}\left(0, \frac{4(\gamma^\star)^2 \boldsymbol{\alpha}^\top \Sigma^{-3} \boldsymbol{\alpha}}{\left(\boldsymbol{\alpha}^\top \Sigma^{-1} \boldsymbol{\alpha}\right)^2}\right)$$

  in distribution.

Numerical Examples

Example 1

We consider the following quantum Werner state tomography for a two-qubit system:

$$\rho_q = q|\Psi^-\rangle\langle\Psi^-| + \frac{1-q}{4}\mathsf{I}$$

where $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ and $q \in [0,1]$ is a parameter associated with the state.

We take an orthonormal basis $\{\mathsf{B}_i\}_{i=1}^{16}$ as

$$\mathsf{B}_i = \frac{1}{\sqrt{2}}\sigma_j \otimes \frac{1}{\sqrt{2}}\sigma_k, \quad i = 4j + k + 1$$

for $j, k = 0, 1, 2, 3$ from standard computational basis, where

$$\sigma_0 = I_2, \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Example 1

Let

$$|\varphi_1\rangle = \frac{1}{\sqrt{6}}[1,1]^\top, |\varphi_2\rangle = \frac{1}{\sqrt{6}}[1,-1]^\top, |\varphi_3\rangle = \frac{1}{\sqrt{6}}[1,i]^\top,$$
$$|\varphi_4\rangle = \frac{1}{\sqrt{6}}[1,-i]^\top, |\varphi_5\rangle = \frac{1}{\sqrt{3}}[1,0]^\top, |\varphi_6\rangle = \frac{1}{\sqrt{3}}[0,1]^\top.$$

Then

$$\mathsf{E}_m = |\varphi_j\rangle\langle\varphi_j| \otimes |\varphi_k\rangle\langle\varphi_k|, \quad m = 6(j-1)+k,$$

for $j,k = 1,2,\dots,6$ form our measurement basis $\{\mathsf{M}_m\}_{m=1}^{36}$ with $\mathsf{M}_m = |\varphi_j\rangle \otimes |\varphi_k\rangle$.

The measurement set $\{\mathsf{M}_m\}_{m=1}^{36}$ is overcomplete and the matrix $\mathbf{A} = (\beta_1,\dots,\beta_{36})^\top \in \mathbb{R}^{36\times 16}$ has full column rank.

Small Sample Size

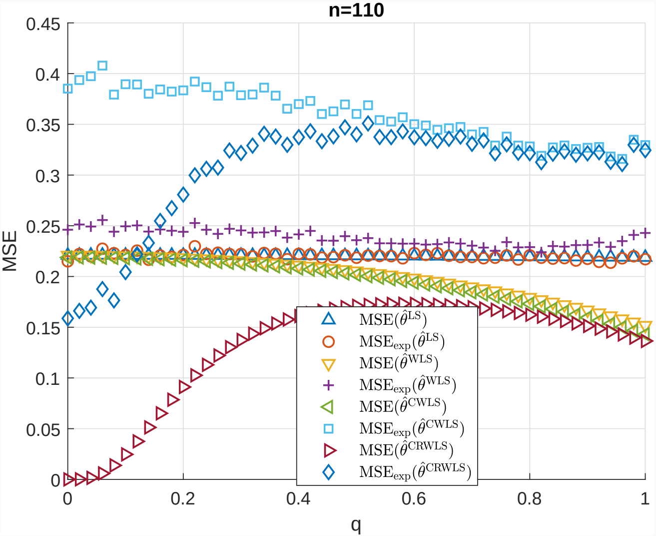


Figure 1: MSEs for estimating Werner states with $n = 110$ copies.

Medium Sample Size

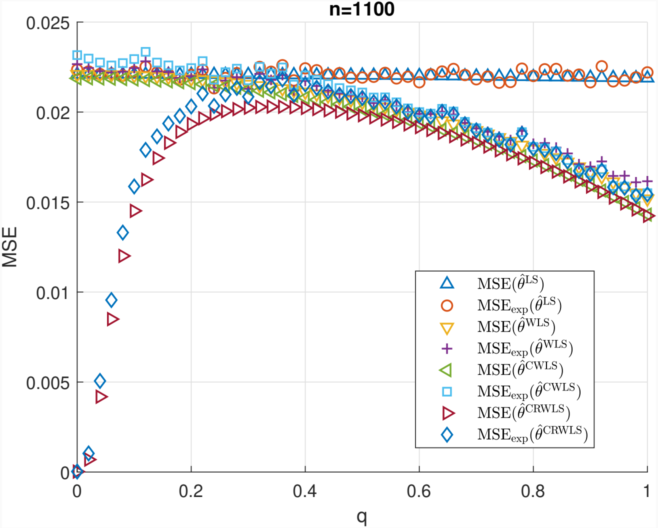


Figure 2: MSEs for estimating Werner states with $n = 1100$ copies.

Large Sample Size

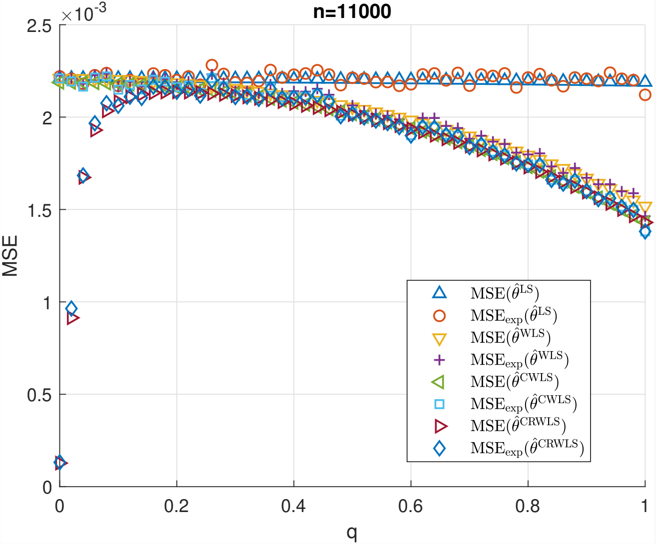


Figure 3: MSEs for estimating Werner states with $n = 11000$ copies.

- The experimental estimates are approaching the theoretical ones as the number of samples n grows large for all four estimates, LS, WLS, CWLS, and CRWLS, which validates the theoretical results;
- For small sample size ($n = 110$), the WLS, CWLS, and CRWLS are apparently producing worse experimental mean-square error compared to LS;
- For relatively larger sample size ($n = 11000$), the WLS, CWLS, and CRWLS all provide significant improvments compared to LS.

Example 2

Consider exactly the same quantum state and tomography setup as in Example 1. Let $\mathrm{W} = \mathrm{I}$ in $\widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}$ so that we define

$$\widehat{\boldsymbol{\theta}}^{\mathrm{CRLS}} = \widehat{\boldsymbol{\theta}}^{\mathrm{CRWLS}}\Big|_{\mathrm{W}=\mathrm{I}}$$

as the unweighted CRLS estimate. The regularization gain γ is selected under the optimal value $\widehat{\gamma}_R$ in the risk sense and its unbiased estimate $\widehat{\gamma}_u$ from, under which for any ρ_q we carry out the tomography procedure for 1000 rounds based on $n = 110, 1100$ copies, respectively.

Small Sample Size

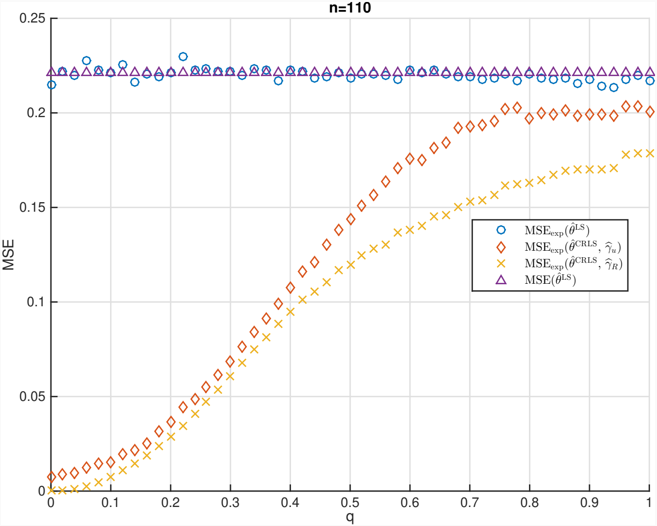


Figure 4: CRLS vs. LS estimates for Werner states with $n = 110$ copies.

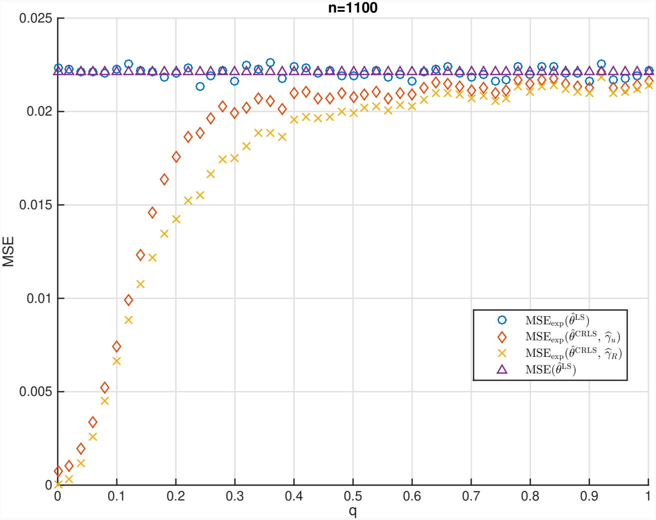


Figure 5: CRLS vs. LS estimates for Werner states with $n = 1100$ copies.

Example 2: Remarks

- With $n = 110$, the regularizer for $\widehat{\boldsymbol{\theta}}^{\mathrm{CRLS}}$ significantly improves the estimation accuracy compared to $\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}$ under both $\widehat{\gamma}_R$ and $\widehat{\gamma}_u$.;
- While with $n = 1100$, for relatively large q, the advantage of $\widehat{\boldsymbol{\theta}}^{\mathrm{CRLS}}$ is no longer obvious compared to $\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}$ since in this case, the use of the weight W becomes essential for the performance.

Conclusions

Conclusions

- We have studied a series of linear regression methods for quantum state tomography based on regularization.
- With complete or over-complete measurement bases, the empirical data was shown to be useful for the construction of a weighted LSE from the measurement outcomes of an unknown quantum state.
- For general measurement bases, either complete or incomplete, we prove that ℓ_2-regularization with proper regularization parameter could yield even lower mean-square error under a penalty in bias.
- An explicit formula was established for the regularization parameter under an equivalent regression model, which is asymptotic optimal as the number of samples grows to infinity for both theoretical and practical risk metrics.

Thanks for your listening

Questions?