SAND2014-17437D

Exceptional service in the national interest

Sandia National Laboratories

Quantum Model Selection:

How big is your system's Hilbert Space?

Travis Scholten (Univ. New Mexico & Sandia), Robin Blume-Kohout (Sandia)

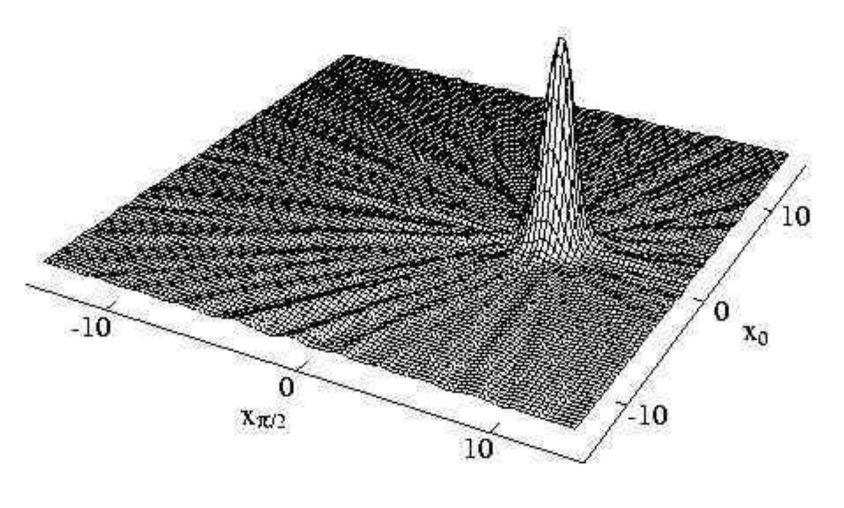


Fig. 1: This Wigner function was reconstructed by straightforward linear inversion. Are the radial wiggles real? or an **overfitting** artifact?

Summary Of Results

- Quantum devices with ill-defined/unknown Hilbert space dimension $\boldsymbol{d}$ pose serious problems for tomography -- what size matrix do you fit?
- We develop a *likelihood ratio* test for $\boldsymbol{d}$ in *continuous-variable* systems.
- A key ingredient in this test is the *null value* of the loglikelihood ratio λ -- its typical value when increasing $\boldsymbol{d}$ provides no advantage.
- Preliminary results: the null value is set by the dimension of the *boundary* of the model ($2\boldsymbol{d}$ - 2), not the dimension of its bulk ($\boldsymbol{d}^2$ - 1).

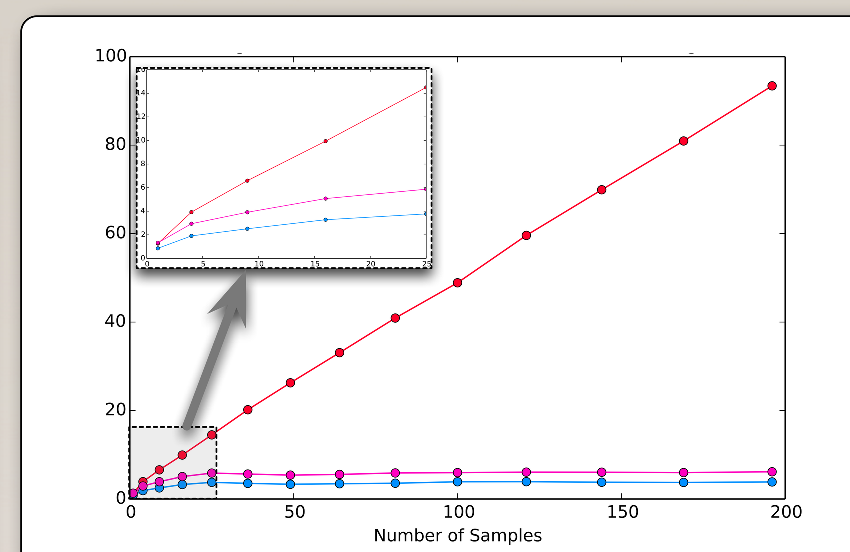


Fig. 2: Loglikelihood ratios for good (d=3,4,5) & bad models (d=2). True state was $|\mathbf{2}\rangle\langle\mathbf{2}|$.

Quantum tomography and Hilbert space dimension

Problem: To characterize a quantum device (e.g. qubit) we use *tomography*:
  (1) measure many identical samples;
  (2) "invert" the observed frequencies to estimate the state, process, gate-set, etc.
However... tomographic inversion protocols require the Hilbert space dimension (d) to be specified *a priori*. Using the wrong d can cause systematic errors (d too small) or overfitting (d too big; Fig 1). Often, d is ill-defined (continuous variables) or unpredictable (leakage, non-Markovian noise).

Solution: We need a method to identify a quantum device's "effective" Hilbert space dimension. It should detect overfitting (d too big) as well as a failure to fit (d too small). Then we can use it to *select* the proper d -- based on the data -- before doing tomographic inversion using that d.

Test case: We are testing and applying *model selection* methods for continuous variable (CV) tomography, where the Hilbert space -- $L^2(\mathbf{R})$ -- is infinite-dimensional. We consider heterodyne tomography (i.e., the coherent-state POVM, or direct sampling from the Husimi distribution). We assume that the true state is energy-limited -- i.e., it is supported on the Fock subspace given by $\mathscr{H}_d = \mathbf{Span}(|\mathbf{0}\rangle...|\boldsymbol{d}\text{-}\mathbf{1}\rangle)$ -- but we don't know d and have to choose it based on the data.

Model selection and CV tomography

Heterodyne CV tomography: Outcomes of the measurement are labeled by phase space points $\boldsymbol{\alpha} = \boldsymbol{x} + i\boldsymbol{p}$. Datasets are lists $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_N\}$ of N observations.

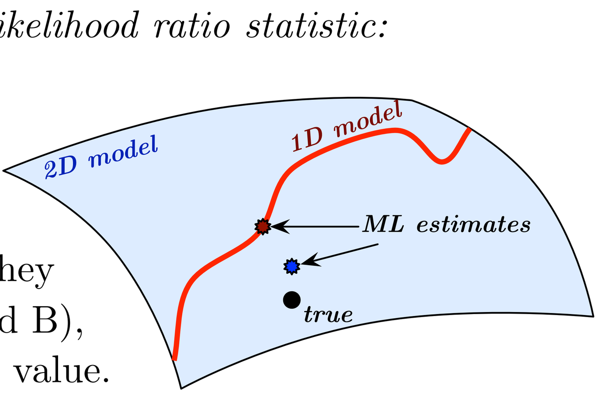
Maximum likelihood tomography: Given data $\mathbf{D}$, find the $d\times d$ state maximizing $\mathscr{L} = \Pr(\mathbf{D}|\rho)$.

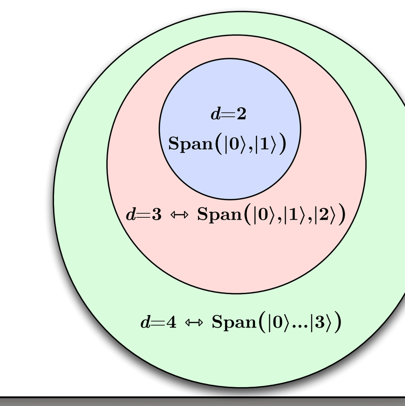
Model selection: Given data $\mathbf{D}$, find the dimension d that best describes the data. Many methods exist! We apply & investigate the *likelihood-ratio test* -- a powerful hypothesis test.

LR test paradigm: *Nested models*: for each d, the model = set of all density matrices on $\mathscr{H}_d$.
We compare models A and B with the *log likelihood ratio statistic:*

$$\lambda = -2\log\frac{\max_{\rho\in A}\mathcal{L}(\rho)}{\max_{\rho\in B}\mathcal{L}(\rho)}$$

If A is closer to the truth than B, then λ will grow **linearly** with N (Fig 2). But if they are equally good (e.g. truth is in both A and B), then λ will asymptote to a **constant** "null" value.





Numerical methods and results: toward a reliable loglikelihood test for dim($\mathscr{H}$)

We want to use λ (the loglikelihood ratio statistic) as a criterion for selecting an effective Hilbert space dimension (d). This requires a knowing how λ will behave when *both* the smaller and the larger model are equally good (and therefore the larger model is *overfitting* and should be discarded). A simple theory (the Wilks Theorem) predicts $\lambda_{\text{avg}} = n(d_{\text{big}})$ - $n(d_{\text{small}})$, where n(model) is the number of parameters in the model.
However, it's not clear how many effective parameters the quantum models have (due to boundary issues), so our goal is to determine this *null value* by studying the behavior of λ numerically.

We generate simulated heterodyne data by rejection sampling, then maximize $\mathscr{L}$ over d-dimensional state spaces for d=2,3,4,5,6 using gradient ascent.

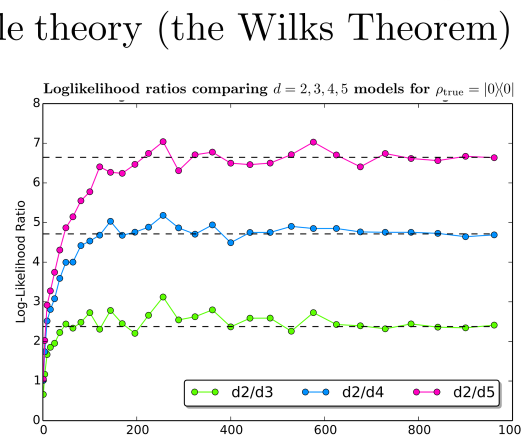


Fig 3a: When the true state lies in $\mathscr{H}_2$, (ρ=$|\mathbf{0}\rangle\langle\mathbf{0}|$), fitting data with larger (d=3,4,5) models yields only a small constant (w/r.t. N) increase in λ.

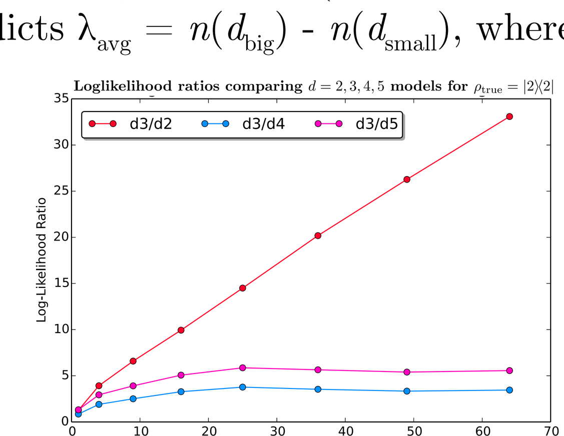


Fig 3b: When the true state is *not* in $\mathscr{H}_2$ (ρ=$|\mathbf{2}\rangle\langle\mathbf{2}|$), λ rises linearly w/N and the d=2 model can be confidently rejected after as few as N=20 samples.

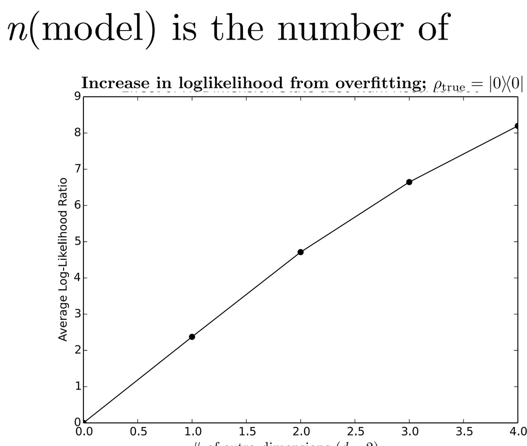


Fig 3c: The average gain in λ due to overfitting (see Fig 3a), vs. the # of unneeded Hilbert space dim. Results consistent with the # of parameters in *pure* state manifolds.

Frameworks for Model Selection: HT vs IC

Our central physics problem is clearly addressed by model selection -- but there are multiple tools/techniques/frameworks for model selection. Here, we have used a *hypothesis testing* concept (loglikelihood ratio tests). One obvious alternative is *information criteria*, e.g. Akaike's AIC or the Bayesian BIC.

Hypothesis testing is motivated by finding the *true* model, and fits naturally in a frequentist context. This is quite artificial (we have no reason to believe that a "true" model exists in the lab!), but also provides a clean and simple sandbox for developing ideas and tests.

Information criteria are motivated by finding the *best* model -- i.e., the one that, when its parameters are fit, will predict future data most accurately (and generate the greatest *utility*). This matches the needs of quantum information technology well... at the cost of simplicity.

The IC approach is ultimately better, but needs more work. We do not yet have ICs whose derivation is valid in tomography! For example, the AIC assumes that *future* samples (to be predicted) come from the same measurement that generated *past* samples (data). This is not generally true in tomography -- we may do heterodyne tomography, but use the results to predict measurements in the Fock basis.

Loglikelihood ratios appear everywhere -- not just in hypothesis testing, but also in most of the known ICs. For example, the standard AIC looks nearly identical to the LR test (except for a factor of 2). Thus, results from this investigation will be applicable in the IC framework with minimal work, once somebody develops ICs that are suitable for quantum tomography problems.

Road map -- where this research goes from here

The research presented here is explicitly preliminary -- it represents "reconnaissance by force" to identify specific problems.

Model selection for identifying effective Hilbert space dimension is a big deal. Most (all?) experimental qubits really do have ill-defined dimension. However, in practice this is more critical for *process* (or gate-set) tomography than state tomography. We will extend the techniques developed in this sandbox to that (more vital) context.

Continuous-variable tomography is significant, especially for QKD and photonic QIP. It desperately needs effective-dimension estimation; direct Wigner function estimation is probably overfitting wildly. We will expand our focus to homodyne tomography.

Other techniques for model selection are a major focus of near-future work. The loglikelihood ratio statistic is central to many of them (including the AIC, BIC, and indirectly the χ^2 test), so this study of LR tests is foundational for further exploration. We believe that the information criterion paradigm is an ultimately better approach, but also presents greater conceptual challenges.

Linear inversion tomography and χ^2 tests are intimately related, and can be dramatically faster and easier to implement than likelihood-based methods. We will determine whether (and when) these linear methods provide equally powerful model selection.

Conclusions

- Likelihood ratio tests can be used to regularize CV tomography and avoid overfitting.
- Our results are portable to the information criterion framework (when it gets developed).

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
SAND No. 2011-XXXXP

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT