Hyperaccuracy and Error Scaling in Gate Set Tomography

Kenneth Rudinger[1], Erik Nielsen[1], John K. Gamble[1], and Robin Blume-Kohout[1]

[1]Sandia National Laboratories

October 21, 2014

Standard quantum tomographic techniques present several challenges. Gate characterization relies on the ability to prepare certain known quantum states; however, such states are prepared using the gates that were to be characterized in the first place. Similarly, state characterization relies on the ability to know what measurements (POVMs) are actually being performed on the system, but characterizing implemented POVMs requires knowledge of the states that the POVMs are acting upon. Thus standard tomographic procedures will be limited in their usefulness by errors in the prior knowledge of the implemented POVMs and prepared states.

Gate set tomography (GST) was introduced to solve this problem of self-referential calibration [1]. (See also [2].) GST seeks to simultaneously and self-consistently characterize the set of implemented gates, prepared states, and POVMs (the entire collection of which is referred to as the *gate set*). GST models the quantum device to be characterized as a black box; its behavior can be described fully by: (1) an initial state ρ; (2) some CPTP maps $\{G_k\}$; and (3) a measurement outcome ("effect") E. Every experiment performed on the box is specified as a sequence of operations, beginning with preparation of ρ, followed by the application of a sequence of the CPTP gates G_k, and finally a measurement of $\mathcal{M} = \{E, \mathbb{1} - E\}$. The only additional assumption imposed beyond this structure is that each of these constituent actions is *repeatable*, that is, every time a particular button is pushed on the box, it performs the same state preparation, gate application, or measurement.

While this framework, particularly when used with long gate sequences, looks similar to randomized benchmarking (RB), it differs from RB in several notable regards. First, GST provides high-accuracy estimates of every observable parameter in the gate set, instead of just an overall error rate. Additionally, unlike RB, GST amplifies and detects the effects of both coherent and non-Markovian errors. Lastly, GST works with any informationally complete gate set, as opposed to RB, which requires the usage of approximate Cliffords.

GST incorporates several protocols that 1) specify experiments to be performed (sequences of buttons on the black box to be pushed), and 2) analyze the resulting data to characterize the gate set. The simplest protocol, called *linear gate set tomography* (LGST), analyzes data from short gate sequences, using only linear algebra, and produces a gate set estimate. LGST has been already experimentally demonstrated in an ion trap [1] and in a silicon quantum dot device [3]. As is shown in these experiments, LGST is highly reliable but limited accuracy estimates of gate sets.

This relatively low accuracy is due to finite sample error in the experimental data; this accuracy is comparable to that given by process tomography. Were finite sample error not present (i.e., the experimentalist could repeat each measurement an infinite number of times), LGST would provide an exact reconstruction of the gate set. However, finite sample error will always be present. Standard tomography's accuracy scales as $1/\sqrt{N}$, so even with a dataset of 10^6 samples, gates can only be estimated with an accuracy of $\pm 10^{-3}$. It is generally believed that quantum information processing will demand better error rates.

Fortunately, GST is capable of attaining far higher accuracy by analyzing experiments involving long sequences of gates, chosen to amplify errors so that they can be measured with accuracy proportional to the sequence length (L). However, these long sequence data cannot be analyzed using LGST. Moreover, the ever-popular maximum likelihood estimator fails badly for general GST, because the log-likelihood function over gate sets is highly non-convex and has many local maxima. We have developed two algorithms that leverage LGST to extract high-precision estimates from long-sequence data: *extended linear GST* (eLGST), or *least-squares GST* (LSGST).[1] Such long sequences are chosen in such a way as to amplify deviations from target gates. We demonstrate here, via both numerical simulation and analysis of experimental data, the power of eLGST and LSGST as tomographic tools.

In this poster, we present a detailed analysis of the GST estimators' imperfections, and how they scale with the available parameters of (i) the experiment, and (ii) the algorithm used for GST analysis. By "imperfections", we mean accuracy, consistency, and success rate:

1. **Accuracy** The extent to which a GST estimate is an accurate estimate of the underlying gate set (as measured by, for example, RMS error in the elements of the gate set).

2. **Consistency** The extent to which a GST estimate properly fits the data (that is, goodness of fit, or the χ^2 statistic).

3. **Success rate** The probability of a particular GST algorithm "succeeding", i.e., did the algorithm find and report the best fit to the data?

Our results serve at least three essential purposes. **First**, we establish (from simulations) lower bounds on the experimental resources required to ensure that GST will provide a reliable and useful estimate of the gates. For example, we show that the two iterative algorithms for GST each undergo a phase transition as N (the number of samples for each experiment) is increased, from a regime in which the algorithm usually fails to find the best fit, to a regime in which it almost always succeeds. **Second**, we establish (from simulations) the scaling of GST's accuracy with N (number of samples per experiment), L (maximum length of experiment), and epsilon (rate of incoherent error). For example, we show that inaccuracy scales roughly as $(L\sqrt{N})^{-1}$ for unitary gates, and that GST can be far more accurate than standard tomography, but we also show that the rate of incoherent error sets a floor on the scaling with L. **Third**, we show (from both simulations and experiments) that experiment-by-experiment χ^2 tests are extremely effective at diagnosing inconsistencies in the model caused by non-Markovian noise.

We find that LSGST (when it succeeds) gives greater accuracy, but that eLGST requires fewer measurement samples to succeed. Both LSGST and eLGST can, with high probability, "lock on" to the correct gate set with surprisingly few measurements per gate sequence, sometimes as few as 16. We also address a critical experimental question: whether the best possible GST estimate (within a given statist ical model) is consistent with the data. Experimental qubits often experience non-Markovian noise that violates GST's assumption of repeatability, and can produce significant systematic errors in the estimate. We show that such inconsistencies can be detected using a collection of χ^2 tests, one for each individual experiment performed (see Figure 1). When the

[1] eLGST and LSGST both produce gate set estimates by each finding a gate set which minimizes a particular objective function. eLGST seeks to minimize the squared distance of the estimate and LGST estimates of progressively longer gate sequences, while LSGST seeks to minimize the error in the goodness of fit (χ^2) of the gate set estimate given the underlying experimental data.

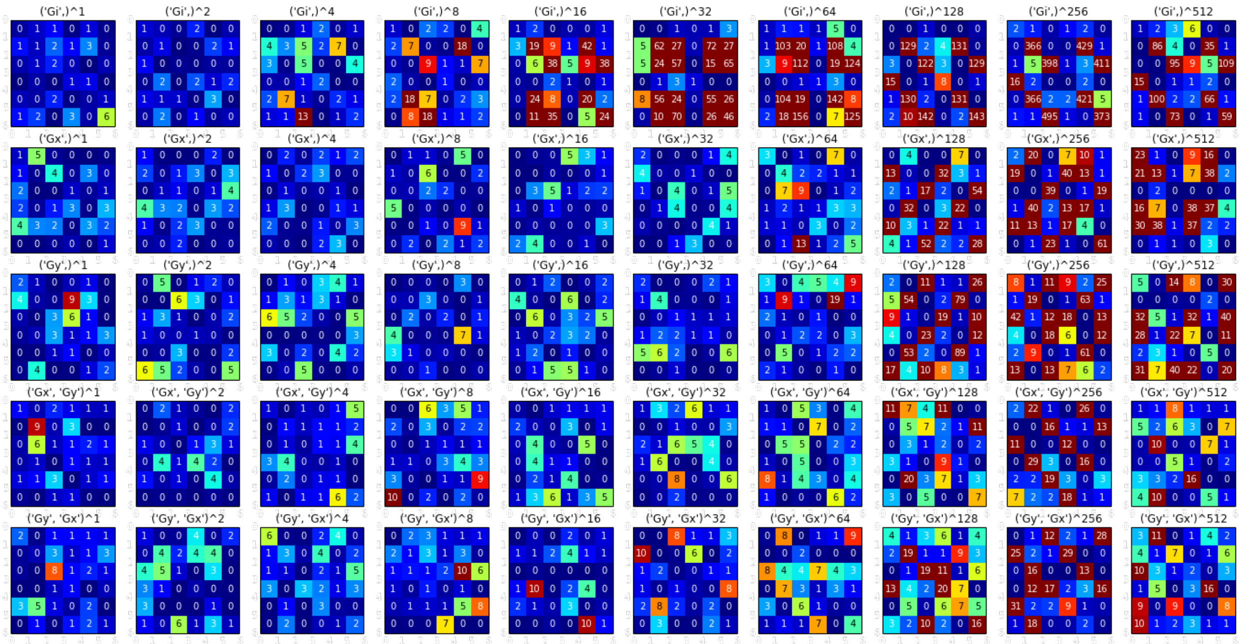

Figure 1: This plot shows χ^2 goodness-of-fit statistics for each of 1800 experiments used to characterize a set of 3 gates ($X_\pi/2$, $Y_\pi/2$, Identity) on a trapped-ion qubit. Five different short "germ" sequences are repeated 1,2,4,8,...,512 times, and the resulting processes are sandwiched between 36 pairs of short "fiducial" sequences to obtain complete tomographic information. If the gates are perfectly Markovian, each of the 1800 χ^2 values should be randomly selected from a χ^2_1 distribution (with mean value 1). The experimental results are color-coded; red squares ($\chi^2 > 9$) should occur roughly 1 time in 1000. This sample dataset shows sparse but experimentally significant inconsistencies, which indicate exactly which gates experience non-Markovian errors of what kind.

gates are perfectly Markovian, we find (by analysis of simulated data generated by realistic but perfectly Markovian gatesets) that the χ^2 statistics behave in accordance with Wilks' Theorem. Analysis of experimental data, however, shows extremely significant deviations from theory. These deviations serve to diagnose non-Markovian errors, and also provide extensive information about the kind of non-Markovian errors. In particular, such GST analysis can be used to determine when non-Markovian behavior becomes experimentally relevant.

Acknowledgements– This work was supported in part by the Laboratory Directed Research and Development program at Sandia National Laboratories and ARO (W911NF-12-1-0607). 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.

References

[1] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D. Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. arxiv:1310.4492, 2013.

[2] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, S. Poletto, A. D. Crcoles, B. R. Johnson,

Colm A. Ryan, and M. Steffen. Self-consistent quantum process tomography. arxiv:1211.0322, 2012.

[3] Dohun Kim, D. R. Ward, C. B. Simmons, John King Gamble, Robin Blume-Kohout, Erik Nielsen, D. E. Savage, M. G. Lagally, Mark Friesen, S. N. Coppersmith, and M. A. Eriksson. Microwave-driven coherent operations of a semiconductor quantum dot charge qubit. arxiv:1407.7607, 2014.