Facilitating the calibration of complex quantum photonic circuits with machine learning assisted gate set tomography

By

Liam Dalgarno

Department of Computer Science

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Engineering in the Faculty of Engineering.

February 9, 2021

Word count: 3214

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MEng in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

SIGNED: ..... DATE: .....

Abstract

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Contents

List of Figures

List of Tables

List of Acronyms

CPTP Completely Positive and Trace-Preserving

GST Gate Set Tomography

LGST Linear Gate Set Tomography

LSGST Long-Sequence Gate Set Tomography

POVM Positive Operator-Valued Measure

Todo list

Chapter 1: Notes

1.1 Gate Set Tomography

Notes taken from [1].
Gate Set Tomography (GST) differs from state and process tomography in that:

- It is almost entirely calibration-free. It does not depend upon a prior description of the measurements used (as in state tomography) or the states that can be prepared (as in process tomography). *These are called "reference frame" operations in the literature?*
- It estimates an entire set of logic operations, rather than a single one.

GST being calibration-free is incredibly important. Both state and process tomography are limited in that they rely on *accurate* characterisation of their "reference frame" operations. Typically, they're either unknown or misidentified.

1.1.1 Mathematical Background

A quantum system is described by a d-dimensional *Hilbert space* $\mathcal{H} = \mathbb{C}^d$, where d is the largest number of outcomes of a repeatable measurement. For a qubit, $d = 2$. GST uses the *Hilbert-Schmidt space*. The Hilbert-Schmidt space is the complex d^2-dimensional vector space of $d \times d$ matrices. We're interested in the d^2-dimensional subspace of Hermitian matrices, denoted $\mathcal{B}(\mathcal{H})$. The basis we use for $\mathcal{B}(\mathcal{H})$ is the set of normalised Pauli matrices $\{\mathbb{I}/\sqrt{2}, \sigma_x/\sqrt{2}, \sigma_y/\sqrt{2}, \sigma_z/\sqrt{2}\}$. This basis has the following properties:

- Hermicity: $B_i = B_i^\dagger$
- Orthonormality: $\mathrm{tr}\,(B_i B_j) = \delta_{ij}$
- Traceless for $i > 0$: $B_0 = \mathbb{I}/\sqrt{d}$ and $\mathrm{tr}\,(B_i) = 0 \ \forall i > 0$.

Elements of $\mathcal{B}(\mathcal{H})$ are represented using an 'extension' of Dirac's bra-ket notation called *super bra-ket notation*. Some element B is represented as a column vector $|B\rangle\!\rangle$, and an element of its dual space by a row vector $\langle\!\langle A|$. Everything works similarly to regular Dirac notation, the main difference is that we can represent everything as vectors in $\mathcal{B}(\mathcal{H})$ rather than the usual matrices.

Measurement of a quantum system yields an outcome from a set of k possibilities. Therefore, the ith outcome can be represented by a dual vector $\langle\!\langle E_i|$, so that $\Pr(i|\rho) = \langle\!\langle E_i|\rho\rangle\!\rangle = \mathrm{tr}\,(E_i\rho)$. Since they represent probabilities, we require that $E_i \geq 0$ and $\sum_i E_i = \mathbb{I}$. The E_i are called *effects*, and the set $\{E_i\}$ is called a Positive Operator-Valued Measure (POVM). Note that since both states and effects are both Hermitian, we can in fact represent them in the d^2-dimensional real subspace of $\mathcal{B}(\mathcal{H})$. Therefore, any reference to $\mathcal{B}(\mathcal{H})$ is referring to the real subspace.

1.1.2 Quantum Logic Gates

An *ideal* quantum logic gate is *reversible* and corresponds to a unitary transform of $\mathcal{H}$. Such a gate would transform ρ as $\rho \to U\rho U^\dagger$ for some unitary matrix U. This is a linear transformation from $\mathcal{B}(\mathcal{H})$ to itself; the linear transformation $\rho \to U\rho U^\dagger$ is called a *superoperator*. In reality, logic gates are not perfectly reversible. These superoperators are known as quantum processes or quantum channels. We can represent any superoperator Λ as a $d^2 \times d^2$ matrix, which acts on $|\rho\rangle\!\rangle \in \mathcal{B}(\mathcal{H})$ by left multiplication. This representation is called the *transfer matrix* of Λ, and is denoted by S_Λ. Thus,

$$\Lambda : |\rho\rangle\!\rangle \mapsto S_\Lambda\,|\rho\rangle\!\rangle \tag{1.1}$$

If Λ is performed on some input state ρ, then the probability of outcome E_i is therefore

$$p_i = \langle\langle E_i | S_\Lambda | \rho \rangle\rangle = \mathrm{tr}\left(E_i S_\Lambda \rho\right) \tag{1.2}$$

Not all superoperators describe physical operations. To be physically possible, they must be:

- *Trace-preserving*: $\mathrm{tr}\left(\Lambda(\rho)\right)$ must equal 1 for all ρ.
- *Completely Positive*: when Λ acts on part of a larger system, it must preserve positivity for the entire system. A superoperator is *positive* iff. $\Lambda(\rho) \geq 0$ for all ρ. A superoperator is *completely positive* iff. $\Lambda \otimes \mathbb{I}_{\mathcal{A}}$ is positive for any auxiliary state space $\mathcal{A}$.

This Completely Positive and Trace-Preserving (CPTP) constraint alone is sufficient – any CPTP superoperator can be physically implemented. The TP condition corresponds to $\langle\langle \mathbb{I}| S_\Lambda = \langle\langle \mathbb{I}|$. Since our basis is traceless for $i > 0$, then Λ is TP iff. the first row of S_Λ is $[1, 0, \ldots, 0]$. The CP condition is a lot more tricky to describe. We first rewrite S_Λ in the operator-sum representation:

$$\Lambda : \rho \mapsto \sum_{ij} \chi^\Lambda_{ij} B_i \rho B_j^\dagger \tag{1.3}$$

where $\{B_i\}$ is a basis, and χ^Λ_{ij} is a matrix of coefficients called the "Choi process matrix" which represents Λ. Similarly to the chi matrix representation from earlier, this completely describes Λ (read this answer for more about their relationship). The mapping between S_Λ and χ^Λ_{ij} is known as the Choi-Jamiołkowski isomorphism:

$$\chi^\Lambda = d(S_\Lambda \otimes \mathbb{I}) \left|\Pi_{\mathrm{EPR}}\right\rangle\!\rangle \tag{1.4}$$

where $\left|\Pi_{\mathrm{EPR}}\right\rangle\!\rangle$ is a maximally entangled state. Note that the equality above really means element-wise equality in a consistent basis. This is all quite complex. Fortunately, the process is simple once we have the Choi process matrix: Λ is CP iff. χ is positive semidefinite.

TODO: Research Choi representation vs. Chi representation

1.1.3 Gate Sets

A quantum processor's capabilities can be specified with a *gate set*. Consider a processor that can perform:

- N_G distinct gates,
- N_ρ distinct state preparations,
- N_M distinct measurements, with $N_E^{(m)}$ distinct outcomes.

We can use these to construct a gate set

$$\mathcal{G} = \left\{ \left\{ \left|\rho^{(i)}\right\rangle\!\right\rangle \right\}_{i=1}^{N_\rho} ; \{G_i\}_{i=1}^{N_G} ; \left\{ \langle\!\langle E_i^{(m)} | \right\}_{m=1,i=1}^{N_m, N_E^{(m)}} \right\} \tag{1.5}$$

where

$$\begin{aligned} G_i : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}) \quad &\text{for} \quad i = 1, \ldots, N_G \\ \left|\rho^{(i)}\right\rangle\!\right\rangle \in \mathcal{B}(\mathcal{H}) \quad &\text{for} \quad i = 1, \ldots, N_\rho \\ \langle\!\langle E_i^{(m)}| \in \mathcal{B}(\mathcal{H})^* \quad &\text{for} \quad m = 1, \ldots, N_M, i = 1, \ldots, N_E^{(m)} \end{aligned} \tag{1.6}$$

In other words, we can describe the capabilities of a quantum processor by collecting together the possible 'input states', the quantum gates, and the possible measurement outcomes for each input state.

The operations that a gate set describes are related to each other. In fact, the specification given above is an overspecification of the gate set. Consider a transformation of the gate set that acts as

$$\begin{aligned} \langle\langle E_i^{(m)}| &\to \langle\langle E_i^{(m)}| \, M^{-1} \\ |\rho^{(i)}\rangle\rangle &\to M \, |\rho^{(i)}\rangle\rangle \\ G_i &\to M G_i M^{-1} \end{aligned} \tag{1.7}$$

where M is an invertible superoperator. Although this changes the gate set, it does not change the observed probabilities (see Equation 1.10). This means that gate sets in fact describe a family of equivalent gate sets. This degeneracy is known as *gauge freedom*.

1.1.4 Circuits

The term 'quantum circuit' can refer to different things. There are two related but distinct types of quantum circuit that we're interested in.

- **Fixed-Input, Classical-Output (FI/CO)**: a quantum circuit where each experiment is described by a quantum circuit that begins by initialising and ends by measuring all of the qubits. A FI/CO circuit describes a probability distribution over classical bit strings.
- **Quantum-Input, Quantum-Output (QI/QO)**: a quantum circuit which is an arrangement of unitary logic gates, with no explicit initialisation or measurement. This may be inserted into a large quantum circuit as a 'sub-routine'.

QI/QO circuits are defined as a sequence of layers $S = (\gamma_1, \gamma_2, \ldots, \gamma_L)$. Each layer corresponds to applying some superoperator G_{γ_i}. The entire QI/QO circuit S also corresponds to applying a superoperator. We denote the transfer matrix for S by $\tau(S)$, where

$$\tau(S) = \tau((\gamma_1, \gamma_2, \ldots, \gamma_L)) = G_{\gamma_L} \ldots G_{\gamma_2} G_{\gamma_1} \tag{1.8}$$

which is formed by composing the elements from each layer. Exponentiation of a circuit describes repetition, and since $S^2 = SS$, it follows that

$$\tau(S^n) = \tau(S)^n \tag{1.9}$$

For a set of FI/CO circuits, we can generate data sets by repeating each one N times, and recording the results. The results are summarized by observed frequencies, $f_k = n_k/N$, which approaches the corresponding probabilities

$$f_k \approx \langle\langle E_k^{(m)}| G_{\gamma_L} \ldots G_{\gamma_2} G_{\gamma_1} |\rho^{(i)}\rangle\rangle \tag{1.10}$$

and can be used to estimate them. This is not the only way to do so, but it illustrates that we can infer some of $|\rho^{(i)}\rangle\rangle$, $\langle\langle E_k^{(m)}|$, and G_j based on observed frequencies. This is called tomography. *Each type of tomography treats some of the operations as known, and uses them as the reference frame to estimate the others.*

1.1.5 Tomography

Quantum State Tomography

In order to perform *any* type of tomography, we must have a *fiducial* and *informationally complete* set. "Fiducial" means that it is accepted as a fixed basis of reference, and "informationally complete" means that it will uniquely identify any target information.

In the context of state tomography, we're aiming to describe some unknown quantum state ρ, given some fiducial and informationally complete measurements. In other words, the measurements are a frame of reference and must uniquely identify ρ. This implies that the set of measurements $\{E_i^{(m)}\}_{m,i}$ must span the entire space of effects. To perform state tomography, many copies of ρ are

made available, and divided into M pools. The mth fiducial measurement is applied to all the copies in the mth pool, and used to estimate the probabilities,

$$p_i^{(m)}(\rho) = \mathrm{tr}\left(\rho E_i^{(m)}\right) \tag{1.11}$$

which should uniquely identify ρ. In general, this is only true if we have infinite copies of ρ. In practice, we have limited numbers of ρ, and therefore $\hat{p}_i = f_i$ where f_i are the frequencies for each measurement. This often yields an estimate $\hat{\rho}$ which is not positive. Nevertheless, it is still useful.

Let's assume that we have the exact probabilities for each measurement outcome for the state ρ. We can in fact ignore the measurements themselves, what is important is the list of effects, so we can write the set simply as $\{E_j : j = 1, \ldots, N_{f1}\}$ where $f1$ is the total number of distinct measurement outcomes. Additionally, we can represent these as dual vectors $\langle\langle E_j|$ in $\mathcal{B}(\mathcal{H})$ just like before. To recover ρ, we can write Born's rule as an inner product,

$$\begin{aligned} p_j &= \mathrm{tr}\left(E_j \rho\right) \\ &= \langle\langle E_j | \rho \rangle\rangle \end{aligned} \tag{1.12}$$

Then, we can stack all of the effects into a single $N_{f1} \times d^2$ matrix

$$A = \begin{bmatrix} \langle\langle E_1| \\ \langle\langle E_2| \\ \vdots \\ \langle\langle E_{N_{f1}}| \end{bmatrix} \tag{1.13}$$

which gives $\vec{p} = A\,|\rho\rangle\rangle$. If A is square, then we can recover ρ with:

$$|\rho\rangle\rangle = A^{-1}\vec{p} \tag{1.14}$$

If N_{f1} is greater than d^2, making A non-square, we must solve with a pseudo-inverse:

$$\begin{aligned} (A^T A)^{-1} A^T \vec{p} &= (A^T A)^{-1} A^T A\,|\rho\rangle\rangle \\ &= (A^T A)^{-1} (A^T A)\,|\rho\rangle\rangle \\ &= I\,|\rho\rangle\rangle \\ &= |\rho\rangle\rangle \end{aligned} \tag{1.15}$$

Quantum Process Tomography

In the context of process tomography, we're aiming to describe some quantum process (e.g. a quantum gate), given an informationally complete set of known fiducial states. Broadly, we prepare many copies of them, pass them through the target process, and perform state tomography on the output states.

Let G be the superoperator representing the process we want to reconstruct. If state ρ_i is prepared, G is applied, and measurement is performed with possible outcomes $\{E_j\}$, then the probability of observing E_j is

$$\begin{aligned} Pj, i &= \mathrm{tr}\left(E_j G[\rho_i]\right) \\ &= \langle\langle E_j | G | \rho_i \rangle\rangle \end{aligned} \tag{1.16}$$

We can then define a $d^2 \times N_{f2}$ matrix B, similarly to A, which represents the fiducial states $|\rho_i\rangle\rangle$:

$$B = \begin{bmatrix} |\rho_1\rangle\rangle & |\rho_2\rangle\rangle & \ldots & |\rho_{N_{f2}}\rangle\rangle \end{bmatrix} \tag{1.17}$$

which gives the $N_{f1} \times N_{f2}$ matrix $P = AGB$. Similarly to before, we can recover G with (pseudo)-inverses. If P is square, we can recover G with:

$$G = A^{-1} P B^{-1} \tag{1.18}$$

If P is non-square, then again we use a pseudo-inverse:

$$G = (A^T A)^{-1} A^T P B^T (B B^T)^{-1} \tag{1.19}$$

Calibration

The requirements of fiducial states/measurements in state and process tomography show why GST being calibration-free is important. In practice, we never have access to perfectly known states/measurements, and they're also not noiseless. In order to identify the exact fiducial measurements for state tomography, we would need perfectly known states, which would require state tomography. Similarly, process tomography relies on fiducial states and measurements which are almost always produced by applying quantum logic gates. Identifying these would again require process tomography. It's an endless loop of self-referentiality. In realistic scenarios, errors in state preparation and measurement (SPAM) dominate inaccuracy in process tomography.

1.1.6 Linear Gate Set Tomography (LGST)

Linear Gate Set Tomography (LGST) looks very much like process tomography, but is doing something significantly different. Unlike process tomography, it reconstructs the entire gate set up to *gauge freedom* as shown in Equation 1.7. Quantum operations are usually described relative to an implicit and absolute reference frame. But in most experiments, no such reference frame is available. GST characterises all of these operations *relative to each other*, and *estimates* every property of a gate set that can be measured without a reference frame. Those that cannot be measured without a reference frame correspond to gauge degrees of freedom.

Because of gauge freedom, the representation produced by LGST is generally not unique. Another shortcoming of LGST is its similarity to process tomography. N trials of an event with probability p generally yields $\hat{p} = p \pm O(1/\sqrt{N})$, meaning that errors scale with $O(1/\sqrt{N})$ like process tomography. Therefore, in order to estimate a gate set to within $\pm 10^{-5}$ would require repeating each circuit $N \approx 10^{10}$ times, which is impractical.

LGST Algorithm

We make some assumptions first, which we relax later:

- We can create informationally complete sets of fiducial states and measurement effects. However, we do not know them.
- We ignore finite sample error.
- We assume that the fiducial states and effects are *exactly* informationally complete, giving $N_{f1} = N_{f2} = d^2$.

Similarly to process tomography, to reconstruct the set of gates $\{G_k\}$, we require a matrix P_k for each gate:

$$[P_k]_{i,j} = \langle\langle E'_i | G_k | \rho'_j \rangle\rangle \tag{1.20}$$

We don't know what ρ'_j and E'_i are, but we are able to prepare them, meaning that we can measure the probabilities that they produce. As before, we have

$$P_k = AG_kB \tag{1.21}$$

but *we do not know what A or B are*. So it may seem that we cannot solve for G_k. Instead, we measure some additional probabilities that correspond to tomography on the *null operation* which will take our ignorance about ρ'_j and E'_i into account. We arrange these into a Gram matrix

$$\tilde{\mathbb{1}}_{i,j} = \langle\langle E'_i | \rho'_j \rangle\rangle \tag{1.22}$$

which is in fact $\tilde{\mathbb{1}} = AB$. Then we can solve for G_k with

$$G_k = B\tilde{\mathbb{1}}^{-1}P_kB^{-1} \tag{1.23}$$

To reconstruct the states $\rho^{(l)}$ and measurement effects $\{E_l^{(m)}\}$, we can similarly construct vectors of observable probabilities:

$$\begin{aligned}\left[\vec{R}^{(l)}\right]_j &= \langle\langle E'_j|\rho^{(l)}\rangle\rangle \\ \left[\vec{Q}^{(m)}\right]_j &= \langle\langle E_l^{(m)}|\rho'_j\rangle\rangle\end{aligned} \tag{1.24}$$

Measuring these probabilities corresponds to state tomography on each native state $\rho^{(l)}$, and measurement tomography on every native effect $\{E_l^{(m)}\}$. They can be written in terms of A and B as

$$\begin{aligned}\vec{R}^{(l)} &= A\,|\rho^{(l)}\rangle\rangle \\ \vec{Q}_l^{(m)T} &= \langle\langle E_l^{(m)}|\rho'_j\rangle\rangle\end{aligned} \tag{1.25}$$

which we can sub $\tilde{\mathbb{1}} = AB$ into to get:

$$G_k = B\tilde{\mathbb{1}}^{-1}P_kB^{-1} \tag{1.26}$$

$$|\rho^{(l)}\rangle\rangle = B\tilde{\mathbb{1}}^{-1}\vec{R}^{(l)} \tag{1.27}$$

$$\langle\langle E_l^{(m)}| = \vec{Q}_l^{(m)T}B^{-1} \tag{1.28}$$

This has recovered the original gate set up to gauge freedom! The best choice for B requires *a posteriori* gauge-fixing.

TODO: Look at gauge-fixing

Over-Completeness

We assumed that $N_{f1} = N_{f2} = d^2$, but this isn't always the case. This means that A, B and P_k are generally not square and invertible. Additionally, due to finite sample errors,

$$\begin{aligned}P_k &= AG_kB \\ \tilde{\mathbb{1}} &= AB\end{aligned} \tag{1.29}$$

may not have exact solutions. Instead we can find an approximate solution with a least-squares estimator, i.e. we want to minimise $|P_k - AG_kB|^2$ and $|\tilde{\mathbb{1}} - AB|^2$. For $|\tilde{\mathbb{1}} - AB|^2$, using differentiation yields

$$A = \tilde{\mathbb{1}}B^T(BB^T)^{-1} \tag{1.30}$$

We can do the same for $|P_k - AG_kB|^2$ and solve for G_k with pseudo-inverses, yielding

$$G_k = (A^TA)^{-1}A^TP_kB^T(BB^T)^{-1} \tag{1.31}$$

and substituting A from before yields

$$G_k = B\left[B^T\left(B\tilde{\mathbb{1}}^T\tilde{\mathbb{1}}\right)^{-1}B\right]\tilde{\mathbb{1}}^TP_k\left[B^T(BB^T)^{-1}\right] \tag{1.32}$$

This is just a generic version of the formulation from before, *however*, B is no longer assumed to be square and invertible. In this case, this isn't true, which means that B does affect the probability estimates, but *only through its support*. B has dimensions $d^2 \times N_{f1}$, but since the fiducial states are informationally complete, then its rows only span a d^2-dimensional subspace of the space of observable probabilities.

We can write B as $B = B_0\Pi$ where Π is a $d^2 \times N_{f1}$ matrix and B_0 is a $d^2 \times d^2$ matrix. Here, B_0 determines *only* the gauge, while Π has a real effect – we can choose B_0 arbitrarily, but not Π. To choose an optimal Π, first we rewrite out AB using Equation 1.30 as

$$AB = \tilde{\mathbb{1}}B^T(BB^T)^{-1}B = \tilde{\mathbb{1}}\Pi^T\Pi \tag{1.33}$$

and then defining the complement projector $\Pi_c = \mathbb{I} - \Pi^T\Pi$ gives

$$|\tilde{\mathbb{1}} - AB|^2 = \mathrm{tr}\left(\Pi_c\tilde{\mathbb{1}}^T\tilde{\mathbb{1}}\Pi_c\right) \tag{1.34}$$

This is uniquely minimised by choosing Π to be the projector onto the d^2 right singular vectors of $\tilde{\mathbb{1}}$ with the largest singular values. *I think this is referring to decomposing $\tilde{\mathbb{1}} = U\Sigma V^*$ using SVD, and projecting onto the first d^2 vectors of V^*. I am not sure why this minimises $|\tilde{\mathbb{1}} - AB|^2$.*

TODO: Find out why tr $\left(\Pi_c \tilde{\mathbb{1}}^T \tilde{\mathbb{1}} \Pi_c\right)$ is minimised by using SVD here. But is it actually important?

Subsituting $B = B_0\Pi$ into the estimates for native states, gates, and effects yields

$$|\rho^{(l)}\rangle\!\rangle = B_0 \left(\Pi\tilde{\mathbb{1}}^T\tilde{\mathbb{1}}\Pi^T\right)^{-1} \Pi\tilde{\mathbb{1}}^T\vec{R}^{(l)} \tag{1.35}$$

$$G_k = B_0 \left(\Pi\tilde{\mathbb{1}}^T\tilde{\mathbb{1}}\Pi^T\right)^{-1} \left(\Pi\tilde{\mathbb{1}}^T P_k \Pi^T\right) B_0^{-1} \tag{1.36}$$

$$\langle\!\langle E_l^{(m)}| = \left[\vec{Q}_l^{(m)}\right]^T \Pi^T B_0^{-1} \tag{1.37}$$

Fiducial States

We also assumed that informationally complete sets of fudicial states and effects were available. *Most processors admit just one native state preparation and measurement.* Therefore, fiducial states and measurements must be implemented using gates from the gate set itself. To do this, we define two sets of QI/QO fiducial circuits. Each fiducial state is prepared by applying one of the preparation fiducial circuits to a native state, and each fiducial measurement is performed by applying one of the measurement fiducial circuits before a native measurement. We can represent this as

$$\langle\!\langle E_i'| = \langle\!\langle E_{t(i)}^{(m(i))}| \, \tau\left(H_{h(i)}\right) \tag{1.38}$$

$$|\rho_j'\rangle\!\rangle = \tau\left(F_{f(j)}\right) |\rho^{r(j)}\rangle\!\rangle \tag{1.39}$$

which looks very complicated. *Recall that $\tau(S)$ is the transfer matrix of S.* Notation:

- F_k: preparation fiducial circuits
- H_k: measurement fiducial circuits
- $r(j)$: native preparation index
- $m(i)$: native measurement index
- $f(j), h(i)$: fiducial index

Consequences:

- Every observable probability required can be obtained by running a specific circuit.
- It reduces the number of free parameters in the model, because fiducial states are not entirely independent.
- It places the burden of informational completeness on the choice of fiducial circuits, which requires the gates to be minimally erroneous. Errors can be checked by checking if the d^2 largest singular values of $\tilde{\mathbb{1}}$ are sufficiently large.

MLE

This is more of an aside? Nothing concrete was given here, but might be worth mentioning in background chapter?

The LGST algorithms described aim to minimise the objective function

$$f(\text{gate set}) = \sum_j \left(p_j^{\text{gate set}} - f_j^{\text{observed}}\right)^2 \tag{1.40}$$

but actually, not really. It doesn't minimise the average squared error over the entire dataset because it doesn't take other observed probabilities into account, e.g. those in P_k. A better objective function is the likelihood function

$$\mathcal{L}(\text{gate set}) = \Pr(\text{data}|\text{gate set}) \tag{1.41}$$

Just know that least-squares does work, but MLE is better.

1.1.7 Long-Sequence Gate Set Tomography (LSGST)

If the A and B matrices are well-conditioned – having condition number $O(1)$ – then each element of G_k is close to a linear combination of observed probabilities. The condition number shows how strongly small changes in an input are magnified in the output. If each circuit is performed N times, then we know that

$$\hat{p} = p \pm \frac{O(1)}{\sqrt{N}} \tag{1.42}$$

where the accuracy is limited by SPAM noise. We can break this boundary by using *deep circuits*, where a gate may appear many times. e.g. The probabilities for

$$\langle\langle E|G_k G_k G_k G_k|\rho\rangle\rangle \tag{1.43}$$

are four times as sensitive as $\langle\langle E|G_k|\rho\rangle\rangle$, which therefore allows four times the precision in estimating some aspects of G_k. This is the basis of Long-Sequence Gate Set Tomography (LSGST).

Broadly, LSGST circuits have three parts:

1. Prepare a state $|\rho'_k\rangle\rangle$ by performing a native operation followed by a fiducial circuit.
2. Perform p repetitions of a short circuit g.
3. Perform a particular measurement $\langle\langle E_i'^{(m)}|$ by performing a fiducial circuit and then a native POVM measurement.

which we can use to estimate probabilities like

$$p = \langle\langle E_i'^{(m)}|\tau(g_j^p)|\rho'_k\rangle\rangle \tag{1.44}$$

We call the short circuit g a *germ* and p the *germ power*. The idea here is that each repetition will amplify particular errors in g. e.g. Suppose we have $g = G$ where G is a single unitary gate which rotates by θ; tomography on g^p measures $p\theta$ to a precision of $\pm\epsilon$, which means we have measured θ to a precision of $\pm\epsilon/p$. Two kinds of parameters cannot be amplified: gauge parameters cannot be measured at all, and SPAM operations cannot be amplified because they only occur once.

A set of germs which amplifies all possible variation in the gate set is called *amplificationally complete*. The set of germs which are amplificationally complete varies depending on the target gate set, so they must be determined for each new gate set. *A set of germs $\{g_j\}$ is amplificationally complete iff. the union of the error subspaces amplified by each g_j span the complement of the subspace of gauge variations.*

Once a set of germs is selected, a set of *base circuits* must also be constructed. Each circuit is then raised to several powers p. Using several powers prevents aliasing issues, e.g. repeating an over-rotation of $\theta = \pi/16$ by $p = 32$ would look like no error at all. What is important here is the depth l of the circuit, not the number of repetitions p itself. The optimal choices for l are logarithmically spaced $l = 1, m, m^2, \ldots$. Empirically, $m = 2$, i.e. $l = 1, 2, 4, 8, \ldots$ was found to be a good starting point, but others may work. The depth of a "depth l" circuit is actually approximately l: a germ of depth 5 would appear at first at $l = 8, p = 1$, then again at $l = 16, p = 3$. Note that the maximum circuit depth L is configurable, and should be carefully. Increasing L yields more precision, but increases the time taken to analyse the data. Increasing L beyond a certain point is useless since decoherence and stochastic errors will dominate. If the rate of decoherence is η, then little can be learned from circuits of depth $L > O(1)/\eta$.

$$p_{abij} = \langle\langle E_{t(a)}^{m(a)} | \tau \left(H_{h(a)} \right) \tau(g_i)^{p_{i,j}} \tau \left(F_{f(b)} \right) | \rho^{r(b)} \rangle\rangle \tag{1.45}$$

With these changes, LSGST experiments are constructed as follows:

1. Select a set of amplificationally complete germs.
2. Select a set of base circuits given by $\mathcal{O} = \{g_i^{p_{i,j}}\}_{i,j}$ where i indexes a germ and $p_{i,j}$ is the j-th power applied to the i-th germ.
3. Perform the circuits in 1.45 to obtain probabilities, which corresponds to:
   a) Prepare the $r(b)$-th state.
   b) Perform the circuit $F_{f(b)} g_i^{p_{i,j}} H_{h(a)}$.
   c) Measure using the $m(a)$-th type of measurement.
4. Estimate p_{abij} with $\hat{p}_{abij} = f_{t(a)} = n_{t(a)}/N$, where $n_{t(a)}$ is the number of times the $t(a)$-th outcome was observed after repeating the circuit N times.

Estimating Gate Set Parameters with LSGST

Previously, we saw that MLE is a better choice of objective function over least-squares, although we'll be using the log-likelihood since it is simpler. MLE answers the question "given a set of data, which set of parameters is most likely to produce it?" To answer this, we must create a *model* for a gate set. $\mathcal{G}$ denotes a gate set model which maps a vector of parameters $\vec{\theta}$ to concrete gate set. In other words, the gate set is decided entirely by the set of parameters $\vec{\theta}$ and the choice of mapping. LSGST is agnostic to the choice of model, but usually the model imposes constraints such as CPTP. *Would the parameters here be the calibration of the Mach-Zehender interferometers*?

We're aiming to fit the data gathered from LSGST experiments as shown in the previous section. The log-likelihood of a single circuit is the multinomial likelihood for an m_s-outcome Bernoulli scheme

$$\log \mathcal{L}_s = N_s \sum_{\beta_s} f_{s,\beta_s} \log(p_{s,\beta_s}) \tag{1.46}$$

where

- s indexes the circuits,
- N_s is the total number of times circuit s was repeated,
- m_s is the number of outcomes of s,
- N_{s,β_s} is the number of times outcome β_s was observed,
- p_{s,β_s} is the true probability *predicted* by $\mathcal{G}$ of getting outcome β_s,
- f_{s,β_s} is the observed frequency for outcome β_s.

For the entire experiment, we simply sum over all circuits

$$\log \mathcal{L} = \sum_s \log \mathcal{L}_s = \sum_{s,\beta_s} N_s f_{s,\beta_s} \log(p_{s,\beta_s}) \tag{1.47}$$

Estimating the maximum likelihood is actually pretty hard, so a good 'first guess' is important. LGST provides a good starting point for the parameters, which can then be refined by incorporating the LSGST data in order to maximise the log-likelihood. Broadly, it works as:

1. Generate an LSGST dataset $\mathcal{D}_0$.

2. Calculate an initial set of parameters $\vec{\theta} \leftarrow \vec{\theta}_0$ using LGST.

3. For each circuit depth $L \in 1, m, m^2, \ldots$

   a) Take the subset $\mathcal{D} \subseteq \mathcal{D}_0$ corresponding to circuits whose germ-power has depth $\leq L$.

   b) Update the current estimate $\vec{\theta}$ according to $\arg\min(\chi^2, \mathcal{G}, \mathcal{D}, \vec{\theta})$, which returns the local optimum for $\chi^2(\mathcal{G}(\vec{\theta}), \mathcal{D})$.

4. Calculate the final estimate $\vec{\theta}$ according to $\arg\min(-\log\mathcal{L}, \mathcal{G}, \mathcal{D}_0, \vec{\theta})$.

The basic idea here is that each estimate gets closer and closer to the true MLE, whilst being much easier to calculate. The χ^2 statistic is a local approximation to the negative log-likelihood, and can be computed faster.

Bibliography

[1] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, "Gate Set Tomography," *arXiv:2009.07301 [quant-ph]*, Sept. 2020.