Projected Least-Squares Quantum Process Tomography

Trystan Surawy-Stepney[1], Jonas Kahn[2], Richard Kueng[3], and Madalin Guţă[1]

[1]School of Mathematical Sciences, University of Nottingham, United Kingdom
[2]Institut de Mathématiques de Toulouse and ANITI, Université de Toulouse, France
[3]Institute for Integrated Circuits, Johannes Kepler University Linz, Austria

We propose and investigate a new method of quantum process tomography (QPT) which we call projected least squares (PLS). In short, PLS consists of first computing the least-squares estimator of the Choi matrix of an unknown channel, and subsequently projecting it onto the convex set of Choi matrices. We consider four experimental setups including direct QPT with Pauli eigenvectors as input and Pauli measurements, and ancilla-assisted QPT with mutually unbiased bases (MUB) measurements. In each case, we provide a closed form solution for the least-squares estimator of the Choi matrix. We propose a novel, two-step method for projecting these estimators onto the set of matrices representing physical quantum channels, and a fast numerical implementation in the form of the hyperplane intersection projection algorithm. We provide rigorous, non-asymptotic concentration bounds, sampling complexities and confidence regions for the Frobenius and trace-norm error of the estimators. For the Frobenius error, the bounds are linear in the rank of the Choi matrix, and for low ranks, they improve the error rates of the least squares estimator by a factor d^2, where d is the system dimension. We illustrate the method with numerical experiments involving channels on systems with up to 7 qubits, and find that PLS has highly competitive accuracy and computational tractability.

1   Introduction

Quantum process tomography (QPT) – the task of estimating an unknown quantum transformation completely from measurement data – is a powerful, but resource-intensive, tool for quantum technology. It has been applied in a variety of experimental contexts [1, 2, 3, 4, 5, 6]. However, the ongoing growth in size of quantum hardware, such as quantum circuits and simulators, motivates the design of QPT techniques that use relevant resources as sparingly as possible.

The earliest QPT proposals [7, 8] used linear inversion techniques to reconstruct quantum channels from informationally complete datasets obtained by feeding different (known) input states into the process and performing full quantum state tomography on the resulting outputs. Later, it was realised that by applying the channel-state duality uncovered by Choi [9] and Jamiołkowski [10], one could perform QPT by simply applying the channel to a maximally entangled state on the system and an ancilla, and using quantum state tomography to reconstruct the channel's Choi matrix directly [11, 12]. Other statistical methods developed in the context of state estimation have been adapted for QPT, including maximum-likelihood [13, 14, 15, 16] and Bayesian inference [17, 18, 19, 20] techniques. In practical applications, point estimators of the quantum process need to be accompanied by error bars [21], hence the importance of developing confidence regions methodology [22, 23, 24].

A major challenge in current applications is the experimental and computational cost of estimating high dimensional systems. The 'low-rank' paradigm developed in the context of compressed sensing [25, 26, 27] partly addresses this challenge by reducing the number of measurements required to estimate channels with a small number of Kraus operators, or equivalently, a Choi matrix

---

Trystan Surawy-Stepney: corresponding author: eetss@leeds.ac.uk

that is low-rank. Other types of latent structures, such as matrix product states with small bond dimension, have been exploited in quantum tomography [28, 29, 30] and more recently extended to quantum state and process tomography in conjunction with machine learning [31, 32]. Techniques like randomized benchmarking can reliably certify unitary channels with limited resource consumption and a robustness against state preparation and measurement errors [33, 34]. An alternative approach is to estimate the fidelity with respect to a known target process [35, 36].

In this work we go back to the traditional 'full-tomography' problem and propose a computationally effective and statistically tractable QPT method, based on an extension of the *projected least-squares* (PLS) method of state estimation proposed in [37]. The method is particularly effective in estimating low rank channels as a proxy for noisy unitary transformations realised in near-term quantum architectures. We provide theoretical concentration bounds for the Frobenius and trace-norm errors and and present simulation results for channels with up to 7 qubits (this corresponds to a Choi matrix with $2^{28} \approx 2.684 \times 10^8$ degrees of freedom).

Let us consider an unknown channel $\mathcal{C} : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$, and represent this by its Choi matrix $\Phi = \mathcal{C} \otimes \mathcal{I}_d(\Omega)$, where Ω is a maximally entangled state on $\mathbb{C}^d \otimes \mathbb{C}^d$. We gather data by repeatedly measuring either the output of the channel with known inputs (direct QPT), or the Choi matrix as output of $\mathcal{C} \otimes \mathcal{I}_d$ (ancilla-assisted QPT). The PLS estimator $\hat{\Phi}_{PLS}$ is obtained by first computing the least squares estimator $\hat{\Phi}_{LS}$ of Φ based on the measurement data, and then projecting this on the convex space of Choi matrices, the latter being the intersection between the cone of positive matrices $\mathcal{CP}$ and the hyperplane $\mathcal{TP}$ determined by the linear constraints $\mathrm{Tr}_s(\Phi) = \mathbb{1}_d/d$. We note that the idea of projecting the least-squares estimator of a Choi matrix onto the intersection of $\mathcal{CP}$ and $\mathcal{TP}$ has been previously considered in [38]. However, this differs in several respects to our implementation: we consider specific choices of input states and measurement settings for which we provide explicit expressions for the LS estimator, and more significantly, we use a different projection algorithm. These result in a protocol of significantly lower computational complexity for which we are able to provide precise statistical convergence results.

We implement the projection using several methods including Dykstra's algorithm [39] and our own *hyperplane intersection projection* (HIP). Dykstra's algorithm involves an iteration of alternating projections onto $\mathcal{CP}$ and $\mathcal{TP}$ with certain adjustments that make sure that the limit is the closest point with respect to the Frobenius distance. While the projection onto $\mathcal{TP}$ is fast, the projection onto the cone $\mathcal{CP}$ requires matrix diagonalisation and is the slowest subroutine of the algorithm. The HIP algorithm also alternates between those projections, but uses them to approximate $\mathcal{CP}$ and directly compute the projection on the intersection between this approximation and $\mathcal{TP}$. This vastly decreases the required number of iterations and leads to a significant speed-up in the computation time, while the statistical errors remain comparable to those of PLS using Dykstra's algorithm. To illustrate the versatility of PLS we estimated a 7-qubit noisy version of the quantum Fourier transform, with Pauli measurements (cf. Section 3.1 for details of the experimental setup). The channel is obtained by performing the quantum Fourier transform and then measuring the first qubit in the z direction with probability 1/4. Figure 1 contains a summary of the results, showing increasing accuracy in trace-norm error, as well as eigenvalues and eigenvectors estimators, with sample size. While sample sizes of order 10^{12} are prohibitive for experiments, we note that important features become visible for more moderate sample sizes of around 10^8. For a rank-one channel, 10^7 would arguably be enough, and is significantly less than the number of measurement settings.

The proposed methods are accompanied by mathematically rigorous performance guarantees. They are summarised in Theorem 1 and Corollary 1 which establish concentration bounds for estimators resulting from four possible experimental scenarios detailed in Section 3. In particular, we show that the number of samples required to reconstruct a rank r Choi matrix to accuracy ϵ with respect the the Frobenius distance is rd^2/ϵ^2 (for trace-norm distance it becomes r^2d^2/ϵ^2). This shows that that PLS is particularly well-suited for estimating approximately unitary channels ($r \approx 1$), for which it achieves error reduction by a factor d^2 compared to the ordinary LS estimator. In Theorem 2 we provide a recipe for constructing confidence regions without any prior assumption about the Kraus rank r.
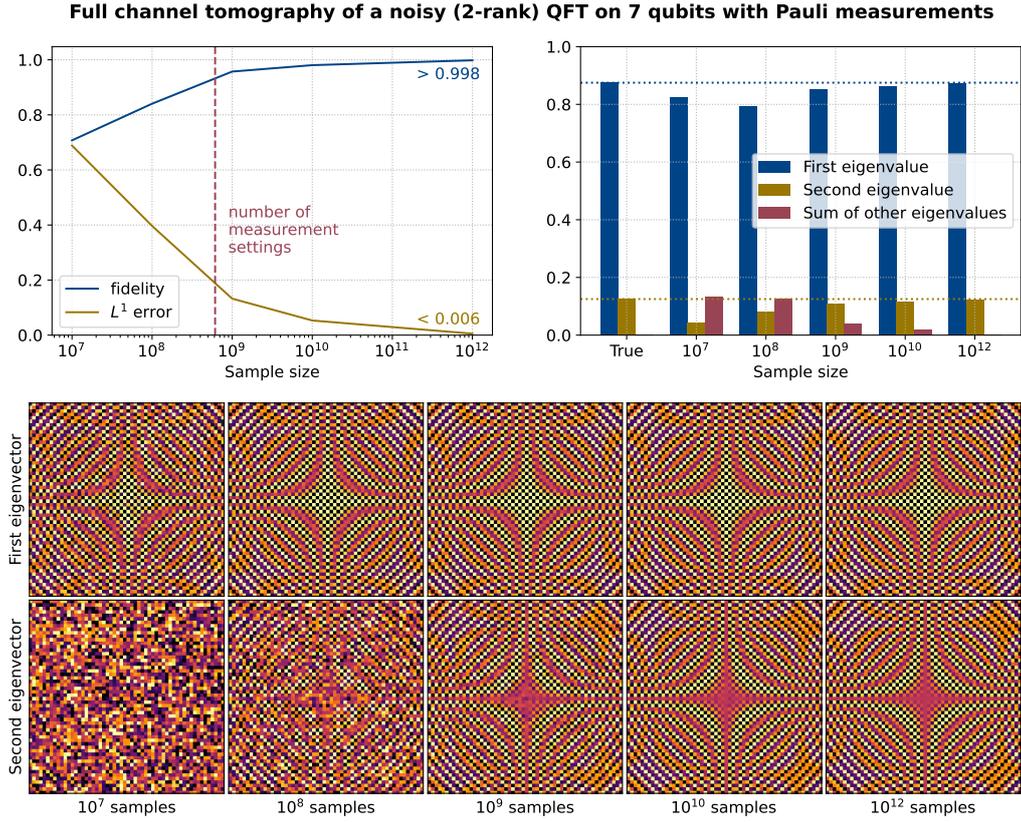
Figure 1: PLS QPT of a 7-qubit noisy quantum Fourier transform (QFT) channel of rank 2 with product input states and Pauli measurements. The top left panel shows the trace-norm error and fidelity as function of sample size, with reasonable errors for one repetition for each combination of input state and measurement setting. The top right panel shows the true and estimated eigenvalues of the Choi matrix as function of the sample size. The bottom panels show details of the real part of the first Kraus operator, and the imaginary part of the second one for different sample sizes. The structure of the first eigenvector becomes visible at 10^7 samples while that of the second at 10^9.

Roadmap   Section 2 summarises important background, in particular the LS estimator. Section 3 describes the 4 experimental procedures we analyse. Each procedure concerns either direct or ancilla-assisted QPT and involves either Pauli measurements or mutually unbiased bases measurements. Explicit expressions for each LS estimator are also provided. Section 4 introduces the PLS estimator and its different implementations. Section 5 contains the main theoretical results described above. Section 6 details the numerical algorithms and present numerical results comparing different projection methods, the dependence of errors on rank, and that of the computation time on dimension. Detailed proofs of concentration bounds for LS and PLS estimators can be found in the appendices.

2   Background

In quantum mechanics, the state of a d-dimensional quantum system is represented by a density matrix, i.e. a positive operator $\rho \in M(\mathbb{C}^d)$ of trace one ($\mathrm{Tr}(\rho) = 1$). This state characterizes all properties and correlations of the underlying quantum system, but is not directly accessible. One can gain information about it by measuring the system. A measurement with a finite number of outcomes, say $\{1, 2, \ldots, s\}$, is described by a positive operator valued measure (POVM). This is a set of positive operators $\{M_i\}_{i=1}^s$ on $\mathbb{C}^d$, such that $\sum_i M_i = \mathbb{1}_d$, where $\mathbb{1}_d$ is the identity operator. The measurement outcome is random and the underlying probability distribution p is given by

Born's rule:
$$p = \{p_i\}_{i=1}^s, \qquad \text{with } p_i = \text{Tr}(\rho M_i).$$

In quantum state tomography (QST) one aims to estimate an unknown state ρ by measuring a (large) ensemble of independent state copies prepared in the state ρ. Among various measurement and estimation techniques described in the literature [40], the projected least squares (PLS) method [41, 37] is particularly suitable for estimating states which are well approximated by low rank operators. In a nutshell, PLS combines an initial 'quick and dirty' least squares estimation step with a 'noise reduction' step where the unphysical estimator is projected onto the space of quantum states. For low-rank states, the projection step reduces the estimation error by a factor of order d, and PLS was shown to achieve optimal error rates when the measurements are chosen from a sufficiently generic ensemble (2-design).

In this paper we address the related problem of quantum process tomography (QPT). Rather than estimating a quantum state, we want to estimate a physical evolution, such as that implemented by a noisy quantum circuit. Such a transformation is described by a quantum channel, i.e. a completely positive, trace preserving map

$$\mathcal{C} : M(\mathbb{C}^d) \to M(\mathbb{C}^d).$$

In QPT we assume that the channel is unknown, and we would like to estimate it by probing many identical realisations of $\mathcal{C}$ with known input states and performing measurements on the outputs.

A particularly useful representation of $\mathcal{C}$ is the *Choi matrix* [9]. It arises from maximally entangling the input intended for $\mathcal{C}$ with an equally-sized quantum memory, also called an ancilla.

$$\Omega := |\omega\rangle\langle\omega|, \text{ where} |\omega\rangle := \frac{1}{\sqrt{d}} \sum_{q=1}^d |q\rangle \otimes |q\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$$

is a maximally entangled state on two d-dimensional quantum systems. Here, $\{|q\rangle\}$ denote the computational basis vectors in $\mathbb{C}^d$. By applying the channel $\mathcal{C}$ to the system and leaving the quantum memory untouched, we obtain the Choi matrix of the channel

$$\Phi = \Phi_{\mathcal{C}} = (\mathcal{C} \otimes \mathcal{I}_d)(\Omega).$$

Here, $\mathcal{I}_d$ denotes the identity operation on the quantum memory (do nothing). The Choi matrix describes a bipartite quantum state which satisfies the constraint $\text{Tr}_s(\Phi) = \mathbb{1}_d/d$, where $\text{Tr}_{s/a}$ denotes the partial trace over the system or ancilla respectively. Conversely, any positive operator satisfying the above constraint is the Choi matrix of a quantum channel. The action of the quantum channel $\mathcal{C}$ on a state ρ can be expressed in terms of its Choi matrix as

$$\mathcal{C}(\rho) = \text{Tr}_a(\Phi \,(\mathbb{1}_d \otimes \rho^\top)), \tag{1}$$

where $\top$ denotes the transpose.

Quantum channel and Choi matrix are two mathematically equivalent descriptions of the underlying process. Moving from one to the other allows us to recast QPT as an instance of QST. Meaningful distance measures between channels, say $\mathcal{C}_1$ and $\mathcal{C}_2$, can also be expressed in terms of differences between the Choi matrices Φ_1 and Φ_2. This includes the squared Frobenius distance

$$\|\Phi_1 - \Phi_2\|_2^2 = \text{Tr}\left[(\Phi_1 - \Phi_2)^2\right],$$

as well as trace distance

$$\|\Phi_1 - \Phi_2\|_1 = \text{Tr}(|\Phi_1 - \Phi_2|), \quad |X| = \sqrt{X^*X}.$$

The latter is related to the diamond distance between the underlying channels – arguably one of the strongest and most widely used distance measures for quantum channels, see e.g. [42, Proposition 50].

We restrict our attention to independent and sequential uses of the channel. This is an actual restriction, as there are powerful techniques for QST [43, 44, 45], as well as quantum enhanced

metrology [46, 47], which require the parallel application of the channel to different parts of a global quantum system. We do, however, investigate both ancilla-free and ancilla-assisted strategies.

In the first case the system is measured after an appropriate input state ρ has been passed through the channel $\mathcal{C}$. In this case the probabilities associated to an output measurement with POVM $\{M_i\}_{i=1}^s$ are

$$p_i = \text{Tr}(\mathcal{C}(\rho)M_i) = \text{Tr}\left(\Phi(M_i \otimes \rho^\top)\right),$$

where the second equality uses Eq. (1). This procedure is illustrated in Figure 4, Section 3.2.

In the ancilla-assisted case, we input one half of a maximally entangled state Ω into the unknown channel, the other half is left unchanged in a quantum memory. This procedure is illustrated in Figure 3, Section 3.1 and produces an output quantum state described by the Choi matrix Φ. Subsequently, we perform quantum measurements. The outcome probabilities associated with POVM $\{M_i\}_{i=1}^s$ are

$$p_i = \text{Tr}(\Phi M_i).$$

A common feature of both setups described above is the existence of a positive linear transformation

$$\mathcal{A} : M(\mathbb{C}^d) \otimes M(\mathbb{C}^d) \to \mathbb{C}^s$$

which maps the Choi matrix Φ into the probability vector of the measurement outcomes. That is, $p = \mathcal{A}(\Phi) \in \mathbb{C}^s$, where s counts the number of measurement outcomes. This follows from the fact that the measured states depend linearly on the channel which, in turn, is in a linear, one-to-one correspondence with its Choi matrix.

After performing several independent preparation-measurement rounds, the data is collected as a *vector of frequencies* $f \in \mathbb{R}^s$, so that for for large sample sizes $f \approx p$. The resulting estimation problem can be recast as linear regression

$$f = p + \epsilon = \mathcal{A}(\Phi) + \epsilon,$$

where ϵ is the statistical noise due to finite sample size. The simplest estimator for Φ is the least squares (LS) estimator:

$$\hat{\Phi}_{LS} = \arg\min_\tau \|\mathcal{A}(\tau) - f\| = (\mathcal{A}^\dagger \mathcal{A})^{-1}\mathcal{A}^\dagger(f), \tag{2}$$

where $\mathcal{A}^\dagger$ is the adjoint of $\mathcal{A}$. The minimum is taken over all matrices τ on $\mathbb{C}^d \otimes \mathbb{C}^d$ with compatible dimension, and the loss function $\|\cdot\|$ is the ℓ_2-norm.

The solution to Problem (2) is easy to compute and produces a self-adjoint matrix with unit trace. However, $\hat{\Phi}_{LS}$ is typically not a Choi matrix of a channel. A quantum channel must be *trace-preserving* and *completely positive*; properties which provide additional constraints on the estimators of the Choi matrix. To obtain estimators representing physical quantum channels, we propose to project the LS estimators onto the set of matrices that satisfy the given constraints.

By Choi's theorem [9], the complete positivity condition for the quantum channel is equivalent to the *positive semidefiniteness* of the $d^2 \times d^2$ Choi matrix. The convex set of positive-semidefinite matrices forms a pointed cone in the space of complex square matrices. We call this cone $\mathcal{CP}$, because it contains matrices that represent completely positive maps. The trace-preserving property of the quantum channel is equivalent to a partial trace condition for the Choi matrix: $\text{Tr}_s \Phi = \mathbb{1}_d/d$. This provides a further affine constraint involving d^2 elements. Hence, the trace-preserving matrices lie on a flat plane of dimension $d^4 - d^2$. We call this set $\mathcal{TP}$.

The intersection of $\mathcal{CP}$ and $\mathcal{TP}$ forms the set of matrices that represent *physical quantum channels*. Accordingly, we call this intersection $\mathcal{CPTP}$ and refer to Figure 2 for a geometric illustration.

For concreteness, we focus on reconstructing physical evolutions of multi-qubit systems. That is, the unknown channel $\mathcal{C}$ acts on a quantum system comprised of k qubits, i.e. $d = 2^k$. This will allow us to discuss different scenarios depending on the choice of measurement and input states. We will focus on two types of measurement setups: single-qubit Pauli measurements and (global) mutually unbiased basis measurements. The different choices of experimental strategies (ancila-free or ancilla-assisted) and measurement setups (single-qubit Pauli or mutually unbiased bases) produce four scenarios which are discussed in detail in the following section. For each scenario, we give a closed form solution to Eq. (2).
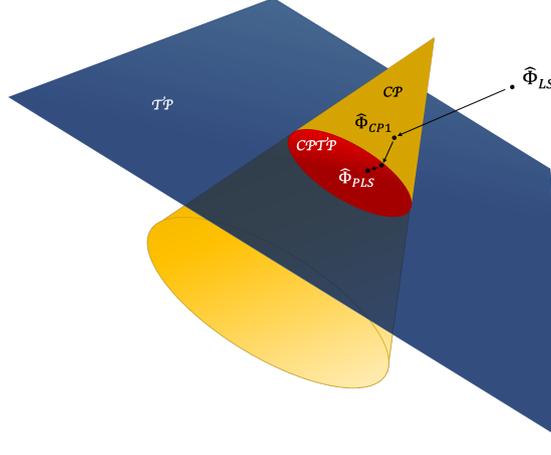
Figure 2: The set of CP maps on $M(\mathbb{C}^d)$ are represented via the Choi isomorphism by a pointed cone embedded in $M(\mathbb{C}^{d^2})$. The set of trace preserving maps forms a hyperplane with codimension d^2, whose intersection with the CP cone is the set of Choi matrices. We also show a representation of the projection of the LS estimator onto the set of quantum states, giving $\hat{\Phi}_{CP1}$, followed by the projection onto $\mathcal{CPTP}$, including mixing with $\mathbb{1}_d/d$, with solution $\hat{\Phi}_{PLS}$.

3 Proposed experimental procedures and least-squares estimators

3.1 Scenario 1: ancilla-assisted QPT with single-qubit Pauli measurements

We first propose an ancilla-assisted protocol. The experimenter prepares many independent copies of the maximally entangled state Ω and acts with the quantum channel on the right subsystem consisting of k qubits, before performing state tomography on the collection of output states which are instances of the Choi matrix Φ (Figure 3).
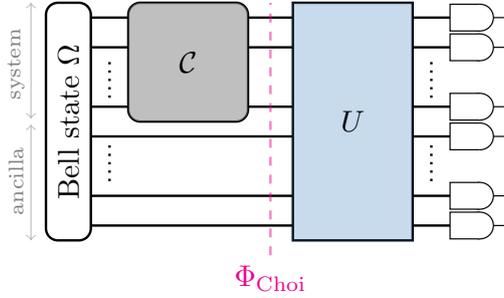
Figure 3: Ancilla-assisted QPT.

The experimenter performs single-qubit Pauli measurements on each of the $2k$ qubits. That is, each qubit is measured either in the (Pauli) x-, y- or z-basis. Let $\{|0, s\rangle, |1, s\rangle\}$ denote the eigenbasis of the Pauli matrix σ_s, i.e. $\sigma_s|o, s\rangle = (-1)^o|o, s\rangle$ for $s = x, y, z$ (Pauli matrix) and $o = -1, +1$ (eigenvalue). In particular, $|+1, z\rangle = |0\rangle$ and $|-1, z\rangle = |1\rangle$ (computational basis), while $|+1, x\rangle = |+\rangle$ and $|-1, x\rangle = |-\rangle$ (Hadamard basis). To measure the Pauli observable σ_s we apply a single-qubit basis change $U_s = \sum_{o=0}^1 |o, s\rangle\langle o, s|$ and, subsequently, measure the qubit in the standard basis. For a given collective setting $\mathbf{s} = (s_1, \ldots, s_{2k}) \in \{x, y, z\}^{2k}$, the corresponding $2k$-qubit unitary factorizes nicely into tensor products: $U = \otimes_i U_{s_i}$, see Figure 3. The outcome of such a measurement is a sequence $\mathbf{o} = (o_1, \ldots, o_{2k}) \in \{0, 1\}^{2k}$, and the corresponding POVM is the set of one dimensional projections $P_{\mathbf{o}}^{\mathbf{s}} = \otimes_{i=1}^{2k} |o_i, s_i\rangle\langle o_i, s_i|$.

Two schemes for generating measurement settings come to mind: a 'fixed' one and a 'random' one. In the fixed scheme, we cycle through all 3^{2k} possible measurement settings a total of ν

times. There, sample size N transforms to $\nu 3^{2k}$ cycles. In the random scheme, we choose measurement settings uniformly at random, and write $\nu = N/3^{2k}$ for the mean number of times each measurement setting is used. In the random scheme ν may be non-integer and strictly smaller than one.

In both cases, the frequency $f_{\mathbf{o}}^{\mathbf{s}}$ is equal to the number of times the outcome $\mathbf{o}$ has occurred when measuring in setting $\mathbf{s}$, divided by ν.

These frequencies are unbiased estimators of, and tend to in the limit of large N, the probabilities of observing outcome $\mathbf{o}$ when measuring with setting $\mathbf{s}$, given by Born's rule:

$$p_{\mathbf{o}}^{\mathbf{s}} = \mathrm{Tr}(\Phi P_{\mathbf{o}}^{\mathbf{s}}). \tag{3}$$

The formula of the least-squares estimator can be adapted from [37] which deals with QST from single-qubit Pauli measurements:

$$\hat{\Phi}_{LS} = \frac{1}{3^{2k}} \sum_{\mathbf{s},\mathbf{o}} f_{\mathbf{o}}^{\mathbf{s}} \bigotimes_{i=1}^{2k} (3|o_i, s_i\rangle\langle o_i, s_i| - \mathbb{1}_2). \tag{4}$$

Here, the $|o_i, s_i\rangle$ are single-qubit eigenstates of a Pauli matrix, and $\mathbb{1}_2$ is the 2-dimensional identity operator.

3.2 Scenario 2: Direct QPT with single-qubit Pauli measurements

Compared to preparing a large number of maximally entangled states of $2k$ qubits, individual k-qubit (pure) product states are a much more tractable proposition. Our second proposed experimental setup, illustrated in Figure 4, takes this into account.
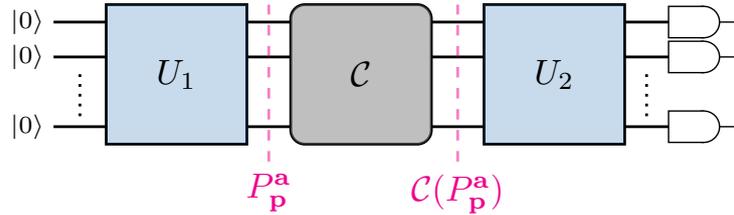
Figure 4: Direct QPT (without ancilla)

The experimenter prepares a state from the set of pure product states $\{P_{\mathbf{p}}^{\mathbf{a}\top}\}$ where $\mathbf{a} \in \{x, y, z\}^k$ and $\mathbf{p} \in \{0, 1\}^k$, and $\top$ denotes the transpose. This is achieved by initialising the qubits in the product state $|\mathbf{0}\rangle = \otimes_{i=1}^{k} |0\rangle$ and applying a unitary $U_{p_i}^{a_i}$ to qubit i such that

$$U_p^x |0\rangle \mapsto |p, x\rangle,\ U_p^y |0\rangle \mapsto |p \oplus 1, y\rangle,\ U_p^z |0\rangle \mapsto |p, z\rangle$$

The overall product unitary is $U = \otimes_{i=1}^{k} U_{b_i}^{a_i}$.

The quantum channel $\mathcal{C}$ is applied to the prepared state producing the output state $\mathcal{C}(P_{\mathbf{p}}^{\mathbf{a}\top})$. Using another product unitary U_2 as described in the previous section, the experimenter performs a Pauli measurement with measurement setting $\mathbf{b} \in \{x, y, z\}^k$, to obtain an outcome $\mathbf{q} \in \{0, 1\}^k$. Similarly to scenario 1, either we repeat the procedure a fixed number of times ν for each choice of indices $\mathbf{a}, \mathbf{b}, \mathbf{p}$, so that the sample size is $N = \nu 3^{2k} 2^k$, or we choose the indices at random for each of the N samples and set $\nu = N/(3^{2k} 2^k)$ for the mean number of times each setting is used. The $\mathbf{q}$-th frequency entry $\{f_{\mathbf{pq}}^{\mathbf{ab}}\}_{\mathbf{q}}$ counts the number of times outcome $\mathbf{q}$ is observed having measured the state $\mathcal{C}(P_{\mathbf{p}}^{\mathbf{a}\top})$ with setting $\mathbf{b}$, divided by ν. These frequencies are unbiased estimators of the probabilities

$$p_{\mathbf{pq}}^{\mathbf{ab}} = \mathrm{Tr}(\mathcal{C}(P_{\mathbf{p}}^{\mathbf{a}\top}) P_{\mathbf{q}}^{\mathbf{b}}) \tag{5}$$

and converge to them in the limit of large N.

Proposition 1. *For direct QPT with single-qubit Pauli inputs and single-qubit Pauli measurements, the least-squares estimator for the Choi matrix Φ of a k-qubit quantum channel takes the following form:*

$$\hat{\Phi}_{LS} = \frac{1}{3^{2k}d} \sum_{abpq} f_{qp}^{ab} \bigotimes_{i=1}^{k} M_{p_i}^{\ b_i} \bigotimes_{j=1}^{k} M_{q_j}^{\ a_j}, \quad where \tag{6}$$

$$M_{p_i}^{\ b_i} = (3|p_i, b_i\rangle\langle p_i, b_i| - \mathbb{1}_2), \tag{7}$$

$$M_{q_j}^{\ a_j} = (3|q_j, a_j\rangle\langle q_j, a_j| - \mathbb{1}_2). \tag{8}$$

We refer to Appendix 8.1 for a proof.

3.3 Scenario 3: Ancilla-assisted QPT with mutually unbiased basis measurements

The set-up is similar to that of ancilla-assisted QPT with single-qubit Pauli measurements, though we now use the POVM consisting of a maximal set of mutually unbiased bases (MUB) [48, 49] to measure instances of the Choi matrix. Such measurements require entangling gates to implement, but they also have a rich, global structure that turns out to produce sampling complexities that are essentially optimal [37]. A MUB POVM on $2k$ qubits consists of $m = d^2(d^2 + 1)$ rank-one operators $\{M_k = |v_k\rangle\langle v_k|/(d^2 + 1)\}_{k=1}^{m}$ in $d^2 = 2^{2k}$ dimension. Each vector (range) $|v_k\rangle$ belongs to one of $d^2 + 1$ orthonormal bases, and any pair belonging to different bases satisfies the unbiasedness condition $|\langle v_i | v_j \rangle|^2 = d^{-2}$. The following 'near isotropy' property is shared with a larger class of 2-design POVMs, see e.g. [50, Lemma 8], and plays an important role in proving our results:

$$\sum_{k=1}^{m} |v_k\rangle\langle v_k| \text{Tr}(|v_k\rangle\langle v_k|A) = A + \text{Tr}(A)\mathbb{1}_d^{\otimes 2},$$

for all operators A with compatible dimensions. Our proofs rely mainly on the near isotropy property and can therefore be readily generalized to other (approximate) 2-design constructions. Concrete examples are symmetric informationally complete (SIC) POVMs [51], complete sets of stabiliser states [52] and local random circuits of depth (order) k [53, 54, 55].

Let us now return to the actual protocol. The system-ancilla input state Ω is passed through the channel $\mathcal{C} \otimes \mathcal{I}_d$ and produces a system characterized by the Choi matrix Φ as output. This system is then measured with a MUB POVM as described above. The frequency f_i indicates the number of times outcome i is observed, divided by the total number of measurements N. These frequencies are unbiased estimators of, and tend to in the limit of large N, the probabilities

$$p_i = \text{Tr}(\Phi M_i). \tag{9}$$

As with single-qubit Pauli measurements, the expression of the least squares estimator is found in [37]

$$\hat{\Phi}_{LS} = (d^2 + 1) \sum_{i=1}^{m} f_i |v_i\rangle\langle v_i| - \mathbb{1}_{d^2}. \tag{10}$$

3.4 Scenario 4: Direct QPT with mutually unbiased basis measurements

Our final experimental method is similar to direct QPT with single-qubit Pauli measurements, in that we do not use an ancillary system. However, we use a different POVM and a different set of input states.

As input states we use the transposes of one dimensional projections $|w_k\rangle\langle w_k|$ onto vectors of a set of $d + 1$ MUBs. On the respective outputs, we measure a MUB POVM whose elements $|v_l\rangle\langle v_l|$ may be different from those of the input. As with the previous scenario, similar results can be obtained with other 2-designs.

Here, frequencies f_l^k indicate the number of times outcome $l \in \{1, \dots, m\}$ is observed having measured the state $\mathcal{C}(|w_k\rangle\langle w_k|^\top)$, divided by the (mean) number of repetitions $\nu = \frac{N}{m}$. These frequencies are unbiased estimators of the probabilities

$$p_l^k = \text{Tr}(\mathcal{C}(|w_k\rangle\langle w_k|^\top) M_l), \tag{11}$$

with $M_l = \frac{1}{d+1}|v_l\rangle\langle v_l|$ and converge to them in the limit of large N.

Proposition 2. *For direct QPT with (transposed) MUB inputs $(|w_i\rangle\langle w_i|)^\top$ and MUB measurements $\frac{d}{m}|v_l\rangle\langle v_l|$, the least-squares estimator for the Choi matrix Φ of a k-qubit quantum channel takes the following form:*

$$
\begin{aligned}
\hat{\Phi}_{LS} &= \frac{d+1}{d}\sum_{l,k=1}^{m} f_l^k |v_l\rangle\langle v_l| \otimes |w_k\rangle\langle w_k| \\
&\quad -\frac{1}{d}\sum_{l,k=1}^{m} f_l^k (|v_l\rangle\langle v_l| \otimes \mathbb{1}_d + \mathbb{1}_d \otimes |w_k\rangle\langle w_k|) \\
&\quad +\mathbb{1}_d \otimes \mathbb{1}_d.
\end{aligned}
\tag{12}
$$

We refer to Appendix 8.2 for a proof.

We conclude this section by emphasizing that all estimators found from (2) are linear and unbiased, given that the frequencies observed are unbiased estimators of true probabilities. Hence those estimators discussed so far, $\hat{\Phi}_{LS}$, are unbiased estimators of the underlying Choi matrix Φ.

4 Methods of projection

We turn our attention now to the task of finding our final estimator, $\hat{\Phi}_{PLS} \in \mathcal{CPTP}$, from the least-squares estimator $\hat{\Phi}_{LS}$. We do this in the following way (see Figure 2 for an illustration of the geometry):

1. Project $\hat{\Phi}_{LS}$ onto the set of $\mathcal{CP}$ matrices of trace one (the quantum states) - we call this set $\mathcal{CP}1$ and the resulting estimator $\hat{\Phi}_{CP1}$.

2. 'Project' $\hat{\Phi}_{CP1}$ onto $\mathcal{CPTP}$ to obtain $\hat{\Phi}_{PLS}$.

For reference, the projection of a matrix M onto a non-empty closed convex set $\mathcal{S}$ with respect to a norm $\|\cdot\|_\alpha$ is defined by: $M_{proj} = \operatorname{argmin}_{M'\in\mathcal{S}} \|M - M'\|_\alpha$.

In order for our theoretical results to hold, the above two procedures need not constitute *true* projections. For our purposes an adequate 'projection' would be to find points in $\mathcal{CP}1$ and $\mathcal{CPTP}$ respectively that satisfy the following two key properties:

$$
\|\Phi - \hat{\Phi}_{CP1}\|_\infty \leq 2\|\Phi - \hat{\Phi}_{LS}\|_\infty,
\tag{13}
$$

$$
\|\Phi - \hat{\Phi}_{PLS}\|_2 \leq \|\Phi - \hat{\Phi}_{CP1}\|_2,
\tag{14}
$$

where Φ is the Choi matrix we are trying to estimate and here, and henceforth, $\|\cdot\|_\infty$ indicates the operator norm. While a direct, Frobenius projection of $\hat{\Phi}_{LS}$ onto $\mathcal{CPTP}$ (e.g. using Dykstra's algorithm) may seem less involved than the above procedure, our method turns out to be more amenable to deriving *a priori* error bounds that are capable of resolving latent structure in the form of low rank. In addition, the advantage of working with more relaxed definitions is that we obtain 'projections' that are easier to compute numerically, and better-behaved in practice. For the rest of the paper we focus on the two-step method, but some results also hold for the direct projection method (see Appendix 12). We give the general description of steps 1 and 2 and point to the numerics Section 6 for the detailed algorithms.

Step 1. The projection onto $\mathcal{CP}1$ is implemented by an eigenvalue thresholding algorithm (cf. Section 6.1 for the details and the proof of property (13)). This step constitutes a true projection.

Step 2. Finding $\hat{\Phi}_{PLS}$ from $\hat{\Phi}_{CP1}$ is the problem of projecting a matrix onto the intersection of two convex sets ($\mathcal{CP}$ and $\mathcal{TP}$) and is the most involved part of the PLS algorithm. We sketch 3 possible, iterative procedures, including our preferred HIP method, and refer to Section 6.2 for detailed algorithms and the proof of property (14). The key idea behind the contraction property ((14)) is contained in the following lemma whose proof can be found a number of textbooks, e.g. [56], Prop 4.16.

Lemma 1. *Let S be a closed, non-empty, convex subset of $M(\mathbb{C}^{d^2})$ and let $\mathcal{P}_S : M(\mathbb{C}^{d^2}) \to M(\mathbb{C}^{d^2})$ be the projection onto S with respect to the Frobenius distance. Then the projection of a matrix A onto S is closer to every point B in S than A, i.e.*

$$\|\mathcal{P}_S(A) - B\|_2 \leq \|A - B\|_2.$$

Using Lemma 1, one may construct a 'projection' of a matrix X onto $\mathcal{CPTP}$, satisfying property (14), by repeatedly projecting X onto a set of convex sets containing $\mathcal{CPTP}$. One such choice is to project X alternately onto $\mathcal{TP}$ and $\mathcal{CP}$. We refer to this as alternating projections (AP). AP can be made into a true projection onto $\mathcal{CPTP}$ by adding a correction term orthogonal to $\mathcal{CP}$ after each $\mathcal{CP}$ projection – this is Dykstra's algorithm [39]. In this work, we define our own fast, generalised algorithm which makes use of Lemma 1 – called *hyperplane intersection projection* (HIP), a discussion of which is deferred to Section 6.2. Again, this does not constitute a true projection but satisfies property (14). We find this algorithm to be significantly faster than AP or Dykstra's algorithm, with continued convergence at large iterations (see Section 6.3.).

Whichever of the above iterative methods we chose, after a number of iterations we need to ensure that $\hat{\Phi}_{PLS}$ exactly satisfies the conditions to be an element of $\mathcal{CPTP}$. For this, we mix the current estimator with the maximally mixed state. Concretely, after near-convergence of the projection algorithm, we have an estimator $\hat{\Phi}'_{PLS}$ in $\mathcal{TP}$ but which may have very small negative eigenvalues. We call the largest magnitude of these $\lambda_{\min}$. For a system of dimension d, we thus then take our final estimator to be

$$\hat{\Phi}_{PLS} = (1-p)\hat{\Phi}'_{PLS} + \frac{p}{d^2}\mathbb{1}_d \tag{15}$$

where p is the solution to $(1-p)\lambda_{\min} + p/d^2 = 0$. In this way, we ensure the estimator is positive-semidefinite while simultaneously preserving the partial trace condition. Mixing with the maximally mixed state adds an error proportional to $\|\lambda_{\min}\|$, hence this additional error can be made arbitrarily small by stopping the algorithm when $\lambda_{\min}$ is sufficiently small in magnitude. The convergence rate of the algorithm is hence defined by how quickly $\|\lambda_{\min}\|$ falls below the required threshold.

Closed form expressions for projections onto $\mathcal{CP}$, $\mathcal{TP}$ and $\mathcal{CP}1$ are required for the implementation of any protocol discussed in this work. These are provided and proven in Appendix 13.

5 Error bounds and sampling complexities

In this section we provide rigorous error bounds on the PLS estimator in the four experimental scenarios described in Section 3. We first state our results in terms of concetration bounds for the squared Frobenius norm distance $\|\hat{\Phi} - \Phi\|_2^2$ and the trace norm (L^1) distance $\|\hat{\Phi} - \Phi\|_1$, and then discuss how such results can be used to construct confidence regions. A discussion of different channel distance measures, and how they compare to each other, can be found in [57], see also [58, 59, 60].

Theorem 1. *Let $\mathcal{C}$ be a k-qubit channel and assume that its Choi matrix Φ has rank r. The squared Frobenius norm error of the estimators $\hat{\Phi}_{PLS}$ derived from the LS estimators $\hat{\Phi}_{LS}$ in scenarios 3.1, 3.2, 3.3, and 3.4 satisfy the following error bounds: for $\epsilon \in (0,1)$,*

$$\Pr[\|\hat{\Phi}_{PLS} - \Phi\|_2 \geq \epsilon] \leq 2^{2k}\exp\left(-\frac{3N\epsilon^2}{8}\frac{g(k)}{8r}\right), \tag{16}$$

and, respectively,

$$\Pr[\|\hat{\Phi}_{PLS} - \Phi\|_1 \geq \epsilon] \leq 2^{2k}\exp\left(-\frac{3N\epsilon^2}{32}\frac{g(k)}{8r^2}\right). \tag{17}$$

Here, $g(k)$ is a function that depends on the number of qubits k and the type of measurement scenario:

$$g(k) = \begin{cases} \frac{1}{3^{2k}} & \text{Scenarios 3.1, 3.2,} \\ \frac{1}{2}\frac{1}{2^{2k}} & \text{Scenario 3.3,} \\ \frac{1}{4}\frac{1}{2^{2k}} & \text{Scenario 3.4.} \end{cases} \quad (18)$$

Theorem 1 shows that the PLS estimators can be brought arbitrarily close to the true Choi matrix, provided that we perform sufficiently many measurements. In particular, the Frobenius square error rates scale as $O(r/(Ng(k)))$ up to logarithmic factors, and we note that scenarios 3.1 and 3.2 exhibit error bounds which are larger by a factor $(3/2)^{2k}$ than those of scenarios 3.3 and 3.4. This is consistent with the fact that MUB measurements are more 'informative' than local Pauli measurements that necessarily factorize in to tensor products [61, 62, 63]. Another interesting observation is that the error rates for ancilla-assisted versus ancilla-free settings are the same in the case of Pauli measurements and differ only by a factor 2 in the case of MUB measurements.

The next corollary provides the sampling complexities derived from the above the theorem.

Corollary 1. *Fix* $\epsilon, \eta \in (0, 1)$. *To achieve Frobenius accuracy*

$$\Pr[\|\hat{\Phi}_{PLS} - \Phi\|_2 \geq \epsilon] \leq \eta, \quad (19)$$

one requires a sample size of

$$N(k, r) \geq \frac{32r}{g(k)}\frac{8}{3\epsilon^2}\log(\frac{2^{2k}}{\eta}). \quad (20)$$

Up to logarithmic factors, this results in sampling complexity

$$N(k, r) = \begin{cases} \mathcal{O}(\frac{1}{\epsilon^2}d^{3.17}r) & \text{for Scenarios 3.1, 3.2,} \\ \mathcal{O}(\frac{1}{\epsilon^2}d^2r) & \text{for Scenarios 3.3, 3.4.} \end{cases} \quad (21)$$

These observations showcase that the PLS method for QPT yields sampling complexities which increase linearly for the Frobenius error (quadratically for norm-one error) with the rank of the channel, so that low rank channels have lower errors than full rank ones. This behavior is very similar to compressed sensing estimators that are designed to exploit (approximate) low rank [25, 26, 27]. In contrast, the LS estimator shows weak dependence on rank and for low rank states, its Frobenius error is $O(d^2)$ larger that that of PLS, as shown in Corollary 2 in Appendix 9.5 .

Although the concentration bounds characterise the PLS error behaviour in terms of rank, it is unlikely that an experimenter knows the rank of the Choi matrix of the quantum process under investigation. Therefore, the bounds cannot be used to construct *confidence regions* unless the rank is known. We now state a more general result which allows us to define confidence regions without prior knowledge of the rank of Φ, but rather in terms of properties of the estimator $\hat{\Phi}_{CP1}$.

To state the result, we first formalise a notion of being close to rank r:

Definition 1. *A state* ρ *is* δ-*almost rank* r *if there is a rank* r *state* $\rho_{(r)}$ *such that*

$$\|\rho - \rho_{(r)}\|_\infty \leq \delta. \quad (22)$$

Theorem 2 below provides confidence balls for the PLS estimator for both the Frobenius and the trace-norm distance, in terms of computable properties of the intermediary estimator $\hat{\Phi}_{CP1}$.

Theorem 2. *Let* Φ *be the Choi matrix of a channel and* $\hat{\Phi}_{PLS}$ *our estimator, generated with any 'projections' satisfying Properties* (13) *and* (14). *Suppose that* $\hat{\Phi}_{CP1}$ *is* δ-*almost rank* r *for some* (r, δ). *Then, with probability (at least)* $1 - \epsilon$,

$$\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \leq 2r\left(\delta + 2\sqrt{\frac{8\ln(2^{2k}/\epsilon)}{3Ng(k)}}\right)^2. \quad (23)$$

The trace distance error $\|\hat{\Phi}_{PLS} - \Phi\|_1$ *is instead bounded by*

$$r\left((4\sqrt{2}+2)\delta + (4+8\sqrt{2})\sqrt{\frac{8\ln(2^{2k}/\epsilon)}{3Ng(k)}}\right). \tag{24}$$

In both cases, $g(k)$ has been defined in (18).

The theorem can be applied by choosing the pair (r, δ) which provides the tightest bound and constitutes the confidence ball. All the results of this section follow from Theorem 4 - proven in appendix 11.

We finally note that the constants given in Theorems 1 and 2 and Corollary 1 are likely to the pessimistic and the true error probabilities may be significantly smaller. However, the advantage is that the bounds hold exactly for any channel and any N. It is certainly possible to get tighter asymptotic bounds for large N.

6 Numerical experiments

In this section we outline the implementation and results of several computer simulations to illustrate the proposed methods.

All experiments presented in this section were run on a quad-core Intel Core i3-8100 with 16 GB memory (computation times given in Figure 8). Mainly because of memory requirements, the experiments on 7 qubits in Figure 1 were done on 64 cores of Intel Xeon E7-8890 with 512 GB memory, needing around three days for each.

The code is available on GitHub [64]. The repository also contains up-to-date information on the future article on the theoretical underpinnings and heuristics behind HIP [65].

6.1 Thresholded projection on trace-one CP operators

As described in Section 4, the first step of the PLS method is the projection of $\hat{\Phi}_{LS}$ onto the space of quantum states $\mathcal{CP}1$. This is implemented using the following algorithm which describes a general thresholded projection process.

Algorithm 1 Thresholded Projection on trace-one $\mathcal{CP}$

1: **function** PROJCP(Φ matrix with trace 1, τ =threshold)
2:     $\lambda_1 \leq \cdots \leq \lambda_d \leftarrow$ eigenvalues of Φ
3:     $v_1, \ldots, v_n \leftarrow$ corresponding eigenvectors of Φ
4:     **for** $1 \leq i \leq d$ **do**
5:         **if** $\lambda_i \leq \tau$ **then**
6:             $\mu_i \leftarrow 0$
7:         **else**
8:             $\mu_i \leftarrow \lambda_i + \tau$
9:     **if** $\sum \mu_i \geq 1$ **then**
10:         Find x_0 such that $\sum(\mu_i - x_0)_+ = 1$
11:         $\forall i, \ \mu_i \leftarrow (\mu_i - x_0)_+$
12:     **else**
13:         find j such that
14:         $\sum_{j+1}^d \lambda_i + (d-j-1)\tau < 1 \leq \sum_j^d \lambda_i + (d-j)\tau$
15:         $\forall i \geq j, \ \mu_i \leftarrow \lambda_i + \tau$
16:         $\mu_{j-1} \leftarrow 1 - \sum_j^d \lambda_i + (d-j)\tau$
17:         $\forall i < j-1, \ \mu_i \leftarrow 0$
18:     **return** operator $\sum_i \mu_i |v_i\rangle\langle v_i|$

The algorithm is separated into two parts. The first, up to and including line 8, is a thresholded projection onto $\mathcal{CP}$, while the second (lines 9 to 17) ensures unit trace and completes the thresholded projection onto $\mathcal{CP}1$. Note that performing the first part of the algorithm while setting $\tau = 0$

produces the direct Frobenius projection onto $\mathcal{CP}$ (see appendix 13.2 for the proof). We choose, for projection onto $CP1$, $\tau = -\lambda_{\min}(\hat{\Phi}_{LS})$ (the sign-flipped, least eigenvalue of $\hat{\Phi}_{LS}$), though we note that the output satisfies Property (13) regardless of the value of τ (Lemma 2 in the appendix 10).

For numeric implementations, we use the full eigenvalue decomposition (EVD) but acknowledge that iterative, randomised implementations or the use of partial EVDs can have a lower theoretical complexity with lower memory requirements - especially in the estimation of high-dimensional, low-rank processes.

6.2 Hyperplane intersection projection

We now move on to discuss the numerical implementation of the second step in the PLS projection, that of $\hat{\Phi}_{CP1}$ onto $\mathcal{CPTP}$. In order to solve this problem, we have developed a fast, generalised 'projection' algorithm to find a point in the intersection of an affine space and a convex set satisfying Property (14). Here we give a high-level description of the *hyperplane intersection projection* and prove that it satisfies Property (14) (cf. Lemma 3 in appendix 10).

We know how to efficiently project (in Euclidean distance) onto $\mathcal{CP}$ and $\mathcal{TP}$, though the projection $\texttt{proj}_{\mathcal{CP}}$ on $\mathcal{CP}$ is more costly. (Closed form expressions for these projections are given in appendix 13, and we do not provide explicit algorithms for them, given their simplicity.)

In the hyperplane intersection projection (HIP) algorithm, we switch between the AP regime (projecting alternately onto $\mathcal{CP}$ and $\mathcal{TP}$) and *HIP mode*, based on a criterion discussed below. In HIP mode, we keep a list of half-spaces that contain $\mathcal{CP}$ - defined by the estimator after each iteration - and project on the intersection of a large subset of these half-spaces and $\mathcal{TP}$. The key idea that allows efficient computations is that it is easy to do the following two things: project onto the intersection of hyperplanes, and check that the projection onto the intersection of hyperplanes is the same as the projection onto the intersection of half-spaces. We select a large subset of half-spaces that ensures that this equality holds, before projecting onto the intersection of $\mathcal{TP}$ and the associated set of hyperplanes. See algorithms 2 and 3 for pseudocode respectively showing construction of the set of hyperplanes, and the HIP algorithm.

Algorithm 2 Choosing hyperplanes

1: **function** HIP_INNER($\texttt{w\_list}$ list of half-spaces, Φ state)
2:     $\texttt{w\_active} \leftarrow$ empty list
3:     **for** $\texttt{w} \in \texttt{w\_list}$ **do**
4:         $\triangleright$ The following test can be checked by looking at whether some coefficients are all non-negative.
5:         **if** the projection of Φ on the intersection of $\mathcal{TP}$ and the half-spaces of $\texttt{w\_active}$ and $\texttt{w}$ is equal to the projection of Φ on the intersection of $\mathcal{TP}$ and the corresponding hyperplanes **then**
6:             Append $\texttt{w}$ to $\texttt{w\_active}$
7:     $\Phi_{new} \leftarrow$ the projection of Φ on the intersection of $\mathcal{TP}$ and the half-spaces in $\texttt{w\_active}$
8:     **return** $\texttt{w\_active}, \Phi_{new}$

Whatever the criterion for switching between AP and HIP mode, Property 14 is always satisfied. An easy choice that works well in practice is to take a small fixed number of steps (say six) in AP mode, and a bigger fixed number of steps (say thirty) in HIP mode. The criterion we use in the experiments is more convoluted, and can be read in the implementation [64].

6.3 Comparison of projection algorithms

Here, we compare the performance of several variants of the HIP algorithm discussed in Section 6.2, with AP and Dykstra's algorithm.

It turns out that the largest reduction in error during the projection of $\hat{\Phi}_{LS}$ onto $\mathcal{CPTP}$ occurs during the initial thresholded projection onto $\mathcal{CP}1$ discussed in Section 6.1. After near-convergence

---

Algorithm 3 Hyperplane Intersection Projection

1: **function** HIP(Φ initial state, ϵ tolerance)
2:     mode $\leftarrow$ AP
3:     **while** $\lambda_{\min}(\Phi) < -\epsilon$ **do**
4:         **while** mode $=$ AP **do**
5:             $\Phi \leftarrow \text{proj}_{\mathcal{CP}}(\Phi)$
6:             $\Phi \leftarrow \text{proj}_{\mathcal{TP}}(\Phi)$
7:             **if** SwitchCondition_toHIP **then**
8:                 mode $\leftarrow$ HIP
9:                 w_active $\leftarrow$ empty list
10:         **while** mode $=$ HIP **do**
11:             $\Phi_{\mathcal{CP}} \leftarrow \text{proj}_{\mathcal{CP}}(\Phi)$
12:             w $\leftarrow$ half-space containing $\mathcal{CP}$ orthogonal at $\Phi_{\mathcal{CP}}$ to $\Phi_{\mathcal{CP}} - \Phi$
13:             Add w as first element of w_active
14:             w_active, $\Phi \leftarrow$ HIP_inner(w_active, Φ)
15:             **if** SwitchCondition_toAP **then**
16:                 mode $\leftarrow$ AP
17:     **return** Φ

---

of any of the subsequent algorithms discussed, before mixing with the depolarizing channel, the L^1 distance $\|\hat{\Phi}_{PLS} - \Phi\|_1$ is the same as that of $\|\hat{\Phi}_{CP1} - \Phi\|_1$, up to a relative change of five percent. Hence the important question regarding the choice of algorithm is *which converges the fastest?*

We test convergence times for 'projections' of $\hat{\Phi}_{LS}$ for a 5-qubit quantum Fourier transform, with 5 different algorithmic implementations: AP, Dykstra, HIPswitch, OneHIP (HIP without any memory of past hyperplanes or switches to AP), and PureHIP (no switch to AP). (HIPswitch, OneHIP and PureHIP are variations on Algorithm 3, and we provide their explicit algorithmic implementations in Appendix 14.) We do so primarily to test the convergence time of our main algorithm HIPswitch compared to that of AP and Dykstra. Additionally, by turning on parts of HIPswitch in turn (OneHIP, PureHIP) we explore the efficacy of the algorithm. Note also in Figure 5, that we display the results of a dual-approach to the projection of $\hat{\Phi}_{CP1}$ onto $\mathcal{CPTP}$; we defer a discussion of this to Section 6.7.
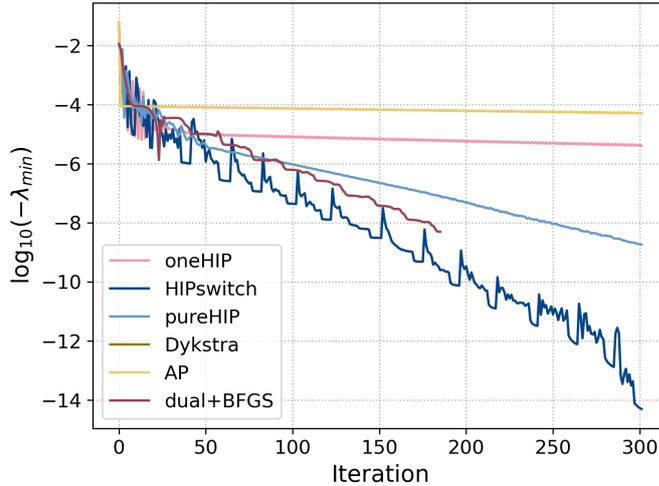
Figure 5: Convergence of different projections for the quantum Fourier transform on 5 qubits, sample size 10^7. The plot shows minus the log of the least eigenvalue of $\hat{\Phi}_{\mathcal{TP}}$ as function of number of iterations for different implementations. The projection algorithm based on the dual problem is discussed in section 6.7.

In Figure 5, we see that:

- Dykstra's algorithm and AP are almost indistinguishable, and do not converge in reasonable time.

- The simplified version oneHIP of HIP, without memory of past hyperplanes or AP, does a little better, but also fails to converge rapidly.

- The simplified version of HIP 'pureHIP', in which we always stay in HIP mode, improves linearly in this setting.

- The main algorithm HIPswitch converges faster than pureHIP. The flat low points correspond to the AP mode. At the end, we might have hit the quadratic convergence regime. HIPswitch and pureHIP attain this faster on easier problems (lower dimension, higher rank) for convergence.

In all subsequent experiments discussed in this article, we use the algorithm HIPswitch (Algorithm 3). We stop when the effect of adding the depolarizing channel changes the trace-norm distance by at most 2×10^{-3}, corresponding to a least eigenvalue of 10^{-7} for 5 qubits, i.e. -7 on Figure 5.

6.4 Effect of sample size
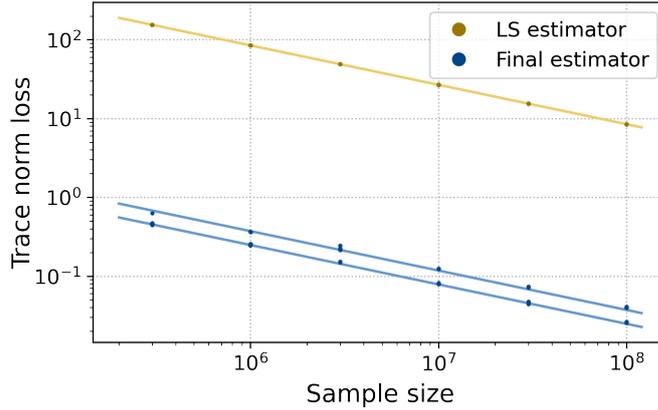
Figure 6: Trace-norm error as function of sample size for PLS (blue lines) and LS (yellow line) of a QFT on 5 qubits. The slope of the lines corresponds to a $\sqrt{N}$ rate. The two blue lines correspond to two regimes determined by the rank of the first projection $\hat{\Phi}_{CP1}$, and are roughly $4/d^2$ times the least-squares line.

We now look at how the performance of our QPT method depends on the number of measurements made. We repeat ten times the estimation of the QFT on 5 qubits, for sample sizes of $N = 3e5, 1e6, 3e6, 1e7, 3e7, 1e8$. We can see on Figure 6 that:

- unsurprisingly, the least-squares estimator error scales as $N^{-1/2}$.

- The least-squares estimator always has almost the same error for the same sample size: each set of ten points is indistinguishable on the figure.

- Our final estimator has two possible regimes. It depends on the rank of the first projection $\hat{\Phi}_{CP1}$. For a given rank, there is visible but low variation of the loss.

- Most importantly, the final estimator also scales as $N^{-1/2}$.

- For this rank-one channel, the final estimator divides the loss of the least-squares estimator by $d^2/3$ or $d^2/5$. That is, for 5 qubits, we improve by a factor of 300.

6.5 Effect of rank
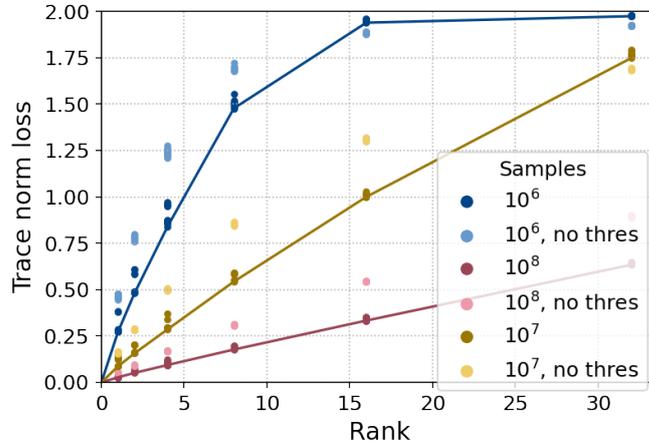
Figure 7: Trace norm error for 5-qubit channels of different ranks, with different sample sizes. The dots corrsespond to ten repetitions for each experiment. Light colored dots correspond to intial projection on $\mathcal{CP}1$ with no thresholding.

In this section, we estimate sums of orthogonal unitary channels, with the ranks $1, 2, 4, 8, 16, 32$, and equal eigenvalues. We include a comparison of the performance of the projected estimator when using the two methods of initial projection onto $\mathcal{CP}1$: our default thresholded projection, and Frobenius projection, using the same data. We test for sample sizes $N = 1e6, 1e7, 1e8$. Each experiment is reproduced ten times. In Figure 7, we see that:

- the loss is slightly less than linear in the rank, until it is close to the maximum of 2.

- The thresholded projection is much better than the standard projection, unless the loss is already close to the maximum of 2, in which case it is worse.

- The effect of getting the wrong rank with the thresholded projection decreases with the rank.

6.6 Effect of dimension

We now test the estimation of the Quantum Fourier Transform using Pauli measurements on one to six qubits, and for mutually unbiased measurements in dimensions $3, 7, 11, 17, 31, 67$. To get comparable losses, we copy the scaling of Theorem 1, and set the respective sample sizes as 10×9^k for k qubits, and $100 \times d^2$, equivalent to 100×4^k, for MUBs.

In Figure 8, we show the mean trace norm loss and time requirements over ten experiments. For the latter, we show the time needed for one loop of algorithm 3, and the time needed for the whole estimation. (We do not include the time taken for the generation of the least-squares estimator, as it is included in the data generation.) We can see that:

- the scaling of Theorem 1 seems comparable to what we get in practice.

- The time for each loop does not depend on the setting. Slightly more loops are usually needed for MUBs than for Pauli.

- We see that the scaling in time gets steeper: possibly for low dimension, bigger arrays are computed faster per entry. The most time-consuming operation is the diagonalisation in proj_{CP}, and its implementation will dictate the complexity. Going from five to six qubits makes loops forty times slower, and the whole computation ninety times slower, hence more loops are required.
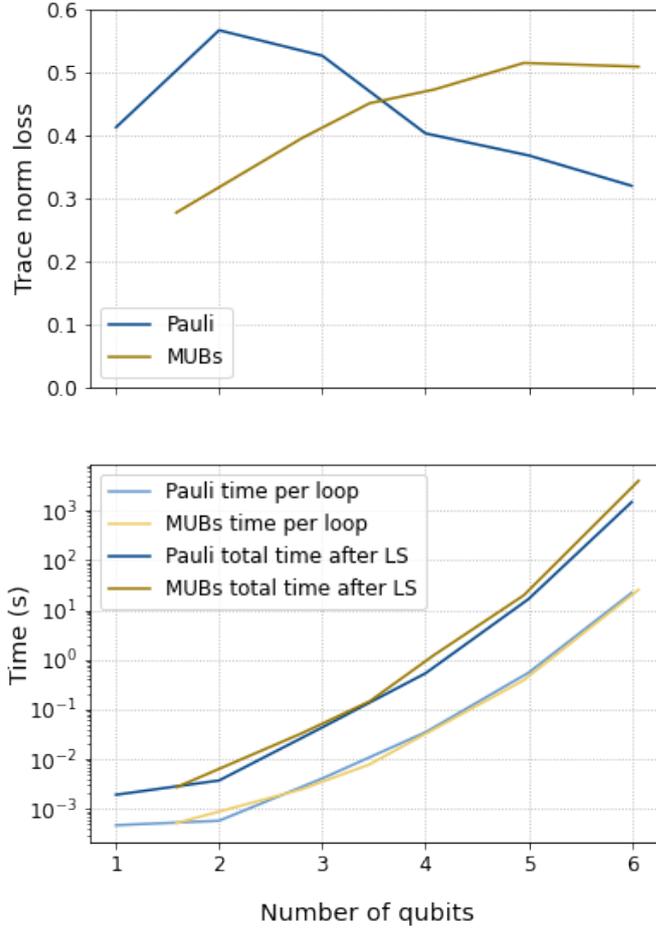
Figure 8: Sample size: 10×9^k for Pauli, 100×4^k for MUBs.

- Experiments on six qubits can be done in half an hour on a personal computer (plus 15 minutes to generate the data); on 5 qubits in less than twenty seconds.

The memory requirements are at the very minimum of d^4, the size of the Choi matrix. With Algorithm 3, we need to keep one such matrix in memory for each hyperplane, which can multiply the needs by 20. However, the limiting factor for the simulations in the Pauli setting is the data generation, where a careful reasonably fast implementation requires 24^k complex entries.

6.7 An alternative optimisation procedure

We thank an anonymous referee for pointing out that the dual optimisation problem for finding the projection onto $\mathcal{CPTP}$ may be more well-behaved than the primal one. Starting from that point, it is possible to devise a simpler projection algorithm that yields the Euclidean projection, is well-rooted in optimisation theory, and has speeds comparable to HIP - based on the few experiments we tested it on. In this section we outline the method and present a preliminary result but we postpone a more in-depth analysis and comparison to HIP for the future publication [65].

The direct projection problem (without projection of $\hat{\Phi}_{LS}$ onto $\mathcal{CP}1$) may be written as follows:

$$\hat{\Phi}_{PLS} = \operatorname*{argmin}_{\Phi \in \mathcal{CP}, \Phi \in \mathcal{TP}} \left\| \Phi - \hat{\Phi}_{LS} \right\|_2. \tag{25}$$

The constraint $\Phi \in \mathcal{CP}$ can be made explicit as $\Phi \geq 0$, that is, a set of linear inequality constraints, and the constraint $\Phi \in \mathcal{TP}$ can be made explicit as $\mathrm{Tr}_s(\Phi) = \frac{1}{d}\mathbb{1}_d$, that is, a set of linear equality constraints. Hence, we can look at the following dual problem. We relax the trace-preserving constraint and form the Lagrangian:

$$\mathcal{L}(\Phi, \nu) = \|\Phi - \hat{\Phi}_{LS}\|_2 + \langle \nu | \mathrm{Tr}_s(\Phi) - \frac{1}{d}\mathbb{1}_d \rangle_F, \tag{26}$$

with $\nu \in \mathrm{Herm}_{n \times n}$ our dual variable. Let the solution to this relaxed problem be:

$$\Phi_{rel}(\nu) = \operatorname*{argmin}_{\Phi \in \mathcal{CP}} \mathcal{L}(\Phi, \nu). \tag{27}$$

The associated dual function is

$$q(\nu) = \min_{\Phi \in \mathcal{CP}} \mathcal{L}(\Phi, \nu), \tag{28}$$

and the optimal value of ν is found by solving the dual problem:

$$\nu_{opt} = \operatorname*{argmax}_{\nu \in \mathrm{Herm}_{n \times n}} q(\nu). \tag{29}$$

The totally mixed state $\Phi = \frac{1}{d^2}\mathbb{1}_{d^2}$ strictly satisfies the inequalities, hence Slater's conditions [66, 67] hold, and the solution of the primal problem (25) is the same as the one yielded by the dual problem: $\Phi_{rel}(\nu_{opt}) = \hat{\Phi}_{PLS}$.

This is beneficial as the solution to the relaxed problem is easy to compute:

$$\begin{aligned}
\mathcal{L}(\Phi, \nu) &= \left\| \Phi - \hat{\Phi}_{LS} \right\|_2 + \langle \nu | \mathrm{Tr}_s(\Phi) - \frac{1}{d}\mathbb{1}_d \rangle_F \\
&= \left\| \Phi - \hat{\Phi}_{LS} \right\|_2 + \langle \mathbb{1}_d \otimes \nu | \Phi - \frac{1}{d^2}\mathbb{1}_{d^2} \rangle_F \\
&= \left\| \Phi - (\hat{\Phi}_{LS} - \frac{1}{2}\mathbb{1}_d \otimes \nu) \right\|_2 - f(\hat{\Phi}_{LS}, \nu),
\end{aligned}$$

for a function f which we do not need to make explicit, so that

$$\Phi_{rel}(\nu) = \mathrm{proj}_{\mathcal{CP}}(\hat{\Phi}_{LS} - \frac{1}{2}\mathbb{1}_d \otimes \nu). \tag{30}$$

This allows simple computation of the Lagrangian. We can also check that the gradient is $\nabla q(\nu) = \mathrm{Tr}_s(\Phi(\nu)) - \frac{1}{d}\mathbb{1}_d$.

This dual problem is well-behaved: to start with, there are only d^2 real parameters, instead of d^4. Moreover, the function is much smoother: the gradient is Lipschitz, so that we can use fast optimisation algorithms. We used BFGS ([68, 69, 70, 71]), since it is available in scipy, but options such as Nesterov's accelerated gradient [72] could be considered.

Some simple simulations show that the speeds obtained using this algorithm are comparable with HIP (Figure 5), with again most of the time spent in $\mathrm{proj}_{\mathcal{CP}}$. We need slightly more loops than for HIP in the five qubit case, as in Figure 5, but there is likely to be room for improvement. Notice that the line corresponding to the dual+BFGS algorithm stops around 10^{-8}. This is because this implementation of the optimisation algorithm stops for lack of precision at that point.

We can also give an interesting interpretation of the dual variable ν: when we use alternate projections, we project on $\mathcal{CP}$, then correct by projecting on $\mathcal{TP}$, and so on. By contrast, in the dual optimisation, we look for an element in the orthogonal space to $\mathcal{TP}$ to be added to $\hat{\Phi}_{LS}$, so that the projection on $\mathrm{proj}_{\mathcal{CP}}$ belongs to $\mathcal{TP}$, too.

7 Conclusions and outlook

In this work we proposed and investigated a computationally efficient and statistically tractable quantum process tomography method called projected least squares (PLS), inspired by previous work on quantum state tomography [37]. The estimator is constructed by first computing the LS estimator and then projecting this on the space of physical Choi matrices. We propose this projection be carried out in two steps: a projection onto the set of quantum states followed by a projection onto the set of Choi matrices. Additionally, we proposed a fast numerical implementation of this projection – hyperplane intersection projection – which accelerates an alternating projections procedure by employing fast projections onto the intersection of several tangent spaces and the hyperplane of linear constraints.

We studied four experimental scenarios for data generation. We consider the most experimentally feasible to be ancilla-free tomography with single-qubit Pauli measurements (scenario 3.2), while the most statistically efficient is the ancilla-assisted setup with mutually unbiased bases measurements (scenario 3.3). For each scenario we provided explicit expressions for the LS estimator (Section 3), and non-asymptotic concentration bounds for the PLS estimator with respect to the Frobenius and trace norm error of the associated Choi matrix, cf. Theorem 1.

Via the two-step projection procedure, all proposed methods are able to exploit latent structure in the form of low rank. Low rank channels are particularly interesting as a class of transformations modelling noisy implementations of unitary channels, which are relevant for quantum technology and quantum computing. Our sampling complexity bounds for the Frobenius error depend linearly on the rank r of the channel and are smaller than those of the LS estimator by a factor r/d^2. This qualitative behaviour is confirmed by numerical simulations which show that the PLS errors are significantly lower than those of LS for low rank channels.

We carried out several numerical studies involving 5 to 7 qubit channels, including a comparison between the two projection methods, a study on the error dependence on rank, and an analysis of how the computational time scales with system dimension. The code is available on GitHub [64]. The repository also contains up-to-date information on the future article on the theoretical underpinnings and heuristics behind HIP [65]. Following a suggestion by an anonymous referee, we implemented an alternative projection algorithm based on the dual optimisation problem, which shows a similar behaviour to HIP and will be analysed in more detail in [65].

We leave as a topic of future investigations the extension of these results to quantum instruments described in terms of non trace-preserving quantum operations. In a different direction, we would like to further investigate the efficacy and possible improvement of the confidence regions prescribed by Theorem 2. Research should also be carried out to find improvements to the algorithmic implementation of the projections of $\hat{\Phi}_{LS}$ onto $\mathcal{CPTP}$ to decrease computational complexity and memory demands. These could arise from "matrix-free" representations of the least-squares estimator and from exploiting the low rank of many high-dimensional processes of interest [73].

References

[1] M. Riebe, K. Kim, P. Schindler, T. Monz, P. O. Schmidt, T. K. Körber, W. Hänsel, H. Häffner, C. F. Roos, and R. Blatt. "Process tomography of ion trap quantum gates". Phys. Rev. Lett. **97**, 220407 (2006).

[2] Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, S. Lloyd, and D. G. Cory. "Quantum process tomography of the quantum fourier transform". The Journal of Chemical Physics **121**, 6117–6133 (2004).

[3] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. "Quantum process tomography of a controlled-not gate". Phys. Rev. Lett. **93**, 080502 (2004).

[4] L. A. Pachón, A. H. Marcus, and A. Aspuru-Guzik. "Quantum process tomography by 2d fluorescence spectroscopy". The Journal of Chemical Physics **142**, 212442 (2015).

[5] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, M. Steffen, and et al. "Quantum process tomography of a universal entangling gate implemented with josephson phase qubits". Nature Physics **6**, 409–413 (2010).

[6] M. Howard, J. Twamley, C. Wittmann, T. Gaebel, F. Jelezko, and J. Wrachtrup. "Quantum process tomography and linblad estimation of a solid-state qubit". New Journal of Physics **8**, 33–33 (2006).

[7] I. L. Chuang and M. A. Nielsen. "Prescription for experimental determination of the dynamics of a quantum black box". Journal of Modern Optics **44**, 2455–2467 (1997).

[8] J. F. Poyatos, J. I. Cirac, and P. Zoller. "Complete characterization of a quantum process: The two-bit quantum gate". Phys. Rev. Lett. **78**, 390–393 (1997).

[9] M. D. Choi. "Completely positive linear maps on complex matrices". Linear Algebra and its Applications **10**, 285 – 290 (1975).

[10] A. Jamiołkowski. "Linear transformations which preserve trace and positive semidefiniteness of operators". Reports on Mathematical Physics **3**, 275 – 278 (1972).

[11] D. W. Leung. "Towards robust quantum computation". Thesis (PhD). STANFORD UNIVERSITY, Source DAI-B 61/11, p. 5911, 225 pages (2000). arXiv:cs/0012017.

[12] G. M. D'Ariano and P. Lo Presti. "Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation". Phys. Rev. Lett. **86**, 4195–4198 (2001).

[13] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. "Measurement of qubits". Phys. Rev. A **64**, 052312 (2001).

[14] A. I. Lvovsky. "Iterative maximum-likelihood reconstruction in quantum homodyne tomography". Journal of Optics B: Quantum and Semiclassical Optics **6**, S556–S559 (2004).

[15] R. Blume-Kohout. "Hedged maximum likelihood quantum state estimation". Phys. Rev. Lett. **105**, 200504 (2010).

[16] J. A. Smolin, J. M. Gambetta, and G. Smith. "Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise". Phys. Rev. Lett. **108**, 070502 (2012).

[17] L. Granade, J. Combes, and D. G. Cory. "Practical bayesian tomography". New Journal of Physics **18**, 033024 (2016).

[18] M. Christandl, R. König, and R. Renner. "Postselection technique for quantum channels with applications to quantum cryptography". Phys. Rev. Lett. **102**, 020504 (2009).

[19] R. Blume-Kohout. "Optimal, reliable estimation of quantum states". New Journal of Physics **12**, 043034 (2010).

[20] C. Granade, C. Ferrie, and S. T. Flammia. "Practical adaptive quantum tomography". New Journal of Physics **19**, 113017 (2017).

[21] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt. "Scalable multiparticle entanglement of trapped ions". Nature **438**, 643–646 (2005).

[22] M. Christandl and R. Renner. "Reliable quantum state tomography". Phys. Rev. Lett. **109**, 120403 (2012).

[23] P. Faist and R. Renner. "Practical and reliable error bars in quantum tomography". Phys. Rev. Lett. **117**, 010404 (2016).

[24] L. P. Thinh, P. Faist, J. Helsen, D. Elkouss, and S. Wehner. "Practical and reliable error bars for quantum process tomography". Phys. Rev. A **99**, 052311 (2019).

[25] S. T. Flammia, D. Gross, Y. K. Liu, and J. Eisert. "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators". New Journal of Physics **14**, 095022 (2012).

[26] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch. "Recovering quantum gates from few average gate fidelities". Phys. Rev. Lett. **121**, 170502 (2018).

[27] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. "Guaranteed recovery of quantum processes from few measurements". Quantum **3**, 171 (2019).

[28] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y. K. Liu. "Efficient quantum state tomography". Nature communications **1**, 149 (2009).

[29] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio. "Scalable reconstruction of density matrices". Phys. Rev. Lett. **111**, 020401 (2013).

[30] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempe, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos. "Efficient tomography of a quantum many-body system". Nature Physics **13**, 1158–1162 (2017).

[31] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, , and G. Carleo. "Neural network quantum state tomography". Nature Physics **14**, 447–450 (2018).

[32] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita. "Quantum process tomography with unsupervised learning and tensor networks" (2020). arXiv:2006.02424.

[33] J. M. Chow, J. M. Gambetta, L. Tornberg, Jens Koch, Lev S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. "Randomized benchmarking and process tomography for gate errors in a solid-state qubit". Phys. Rev. Lett. **102**, 090502 (2009).

[34] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. "Randomized benchmarking of quantum gates". Physical Review A**77** (2008).

[35] L. Steffen, M. P. da Silva, A. Fedorov, M. Baur, and A. Wallraff. "Experimental monte carlo quantum process certification". Phys. Rev. Lett. **108**, 260506 (2012).

[36] M. P. da Silva, O. Landon-Cardinal, and D. Poulin. "Practical characterization of quantum devices without tomography". Phys. Rev. Lett. **107**, 210404 (2011).

[37] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp. "Fast state tomography with optimal error bounds". Journal of Physics A: Mathematical and Theoretical **53**, 204001 (2020).

[38] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger. "Quantum process tomography via completely positive and trace-preserving projection". Phys. Rev. A **98**, 062336 (2018).

[39] R. L. Dykstra. "An algorithm for restricted least squares regression". Journal of the American Statistical Association **78**, 837–842 (1983). url: http://www.jstor.org/stable/2288193.

[40] M. Paris and J. Rehacek, editors. "Quantum state estimation". Volume 649 of Lecture Notes in Physics. Springer. (2004).

[41] T. Sugiyama, P. S. Turner, and M. Murao. "Precision-guaranteed quantum tomography". Phys. Rev. Lett. **111**, 160406 (2013).

[42] M. Kliesch and I. Roth. "Theory of quantum system certification". PRX Quantum **2**, 010201 (2021).

[43] J. Kahn and M. Guţă. "Local asymptotic normality for finite dimensional quantum systems". Communications in Mathematical Physics **289**, 597–652 (2009).

[44] R. O'Donnell and J. Wright. "Efficient quantum tomography". In Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing. Pages 899–912. STOC '16New York, NY, USA (2016). ACM.

[45] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. "Sample-optimal tomography of quantum states". IEEE T. Inform. Theory **63**, 5628–5641 (2017).

[46] V. Giovannetti, S. Lloyd, and L. Maccone. "Advances in quantum metrology". Nature Photonics **5**, 222–229 (2011).

[47] S. Zhou and L. Jiang. "Asymptotic theory of quantum channel estimation". PRX Quantum **2**, 010343 (2021).

[48] J. Schwinger. "Unitary operator bases". Proc. Nat. Acad. Sci. U.S.A. **46**, 570–579 (1960).

[49] W. K. Wootters and B. D. Fields. "Optimal state-determination by mutually unbiased measurements". Ann. Physics **191**, 363–381 (1989).

[50] D. Gross, F. Krahmer, and R. Kueng. "A partial derandomization of phaselift using spherical designs". J. Fourier Anal. Appl. **21**, 229–266 (2015).

[51] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. "Symmetric informationally complete quantum measurements". Journal of Mathematical Physics **45**, 2171–2180 (2004).

[52] C. Dankert, R. Cleve, J. Emerson, and E. Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". Phys. Rev. A **80**, 012304 (2009).

[53] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki. "Local random quantum circuits are approximate polynomial-designs". Comm. Math. Phys. **346**, 397–434 (2016).

[54] N. Hunter-Jones. "Unitary designs from statistical mechanics in random quantum circuits" (2019). arXiv:1905.12053.

[55] Jonas Haferkamp and Nicholas Hunter-Jones. "Improved spectral gaps for random quantum circuits: Large local dimensions and all-to-all interactions". Phys. Rev. A **104**, 022417 (2021).

[56] H. H. Bauschke, P. L. Combettes, et al. "Convex analysis and monotone operator theory in hilbert spaces". Volume 408. Springer. (2011).

[57] A. Gilchrist, N. K. Langford, and M. A. Nielsen. "Distance measures to compare real and ideal quantum processes". Physical Review A**71** (2005).

[58] J. Watrous. "Unital channels and majorization". Page 201–249. Cambridge University Press. (2018).

[59] J. J. Wallman. "Bounding experimental quantum error rates relative to fault-tolerant thresholds" (2015).

[60] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia. "Comparing experiments to the fault-tolerance threshold". Phys. Rev. Lett. **117**, 170502 (2016).

[61] R. Kueng, H. Rauhut, and U. Terstiege. "Low rank matrix recovery from rank one measurements". Appl. Comput. Harmon. Anal. **42**, 88–116 (2017).

[62] H. Y. Huang, R. Kueng, and J. Preskill. "Predicting many properties of a quantum system from very few measurements". Nat. Phys. **16**, 1050—1057 (2020). url: doi.org/10.1038/s41567-020-0932-7.

[63] D. S. França, F. G. S. L. Brandão, and R. Kueng. "Fast and Robust Quantum State Tomography from Few Basis Measurements". In Min-Hsiu Hsieh, editor, 16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021). Volume 197 of Leibniz International Proceedings in Informatics (LIPIcs), pages 7:1–7:13. Dagstuhl, Germany (2021). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[64] J. Kahn. "Hyperplane intersection projection". https://github.com/Hannoskaj/Hyperplane_Intersection_Projection (2021).

[65] J. Kahn, M Guţă, R. Kueng, and T. Surawy-Stepney. "Hyperplane intersection projection" (To be written).

[66] M. Slater. "Lagrange multipliers revisited". Cowles Commission Discussion Paper No. 403 (1950). url: cowles.yale.edu/sites/default/files/files/pub/d00/d0080.pdf.

[67] S. Boyd and L. Vandenberghe. "Convex optimization". Cambridge University Press. (2009).

[68] Charles George Broyden. "The convergence of a class of double-rank minimization algorithms 1. general considerations". IMA Journal of Applied Mathematics **6**, 76–90 (1970).

[69] David F Shanno. "Conditioning of quasi-newton methods for function minimization". Mathematics of computation **24**, 647–656 (1970).

[70] Roger Fletcher. "A new approach to variable metric algorithms". The computer journal **13**, 317–322 (1970).

[71] Donald Goldfarb. "A family of variable-metric methods derived by variational means". Mathematics of computation **24**, 23–26 (1970).

[72] Yurii Nesterov. "A method for unconstrained convex minimization problem with the rate of convergence o $(1/k\hat{\ }2)$". In Doklady an ussr. Volume 269, pages 543–547. (1983). url: cir.nii.ac.jp/crid/1570572699326076416.

[73] Alp Yurtsever, Joel A. Tropp, Olivier Fercoq, Madeleine Udell, and Volkan Cevher. "Scalable semidefinite programming". SIAM Journal on Mathematics of Data Science **3**, 171–200 (2021).

[74] P. Tol. "Colour schemes" (2018).

[75] M. A. Nielsen and I. L. Chuang. "Quantum computation and quantum information: 10th anniversary edition". Cambridge University Press. (2010).

[76] J. A. Tropp. "User-friendly tail bounds for sums of random matrices". Foundations of Computational Mathematics **12**, 389–434 (2012).

[77] J. L. W. V. Jensen. "Sur les fonctions convexes et les inégalités entre les valeurs moyennes". Acta Math. **30**, 175–193 (1906).

Appendix

8 Form of the LS estimators in Section 3

8.1 Proof of Proposition 1

We start with a restatement of the form of the least-squares estimator for the Choi matrix, found using measurement scenario 2.

Proposition 1. *For direct QPT with single-qubit Pauli inputs and single-qubit Pauli measurements, the least-squares estimator for the Choi matrix Φ of a k-qubit quantum channel takes the following form:*

$$\hat{\Phi}_{LS} = \frac{1}{3^{2k}d} \sum_{\boldsymbol{abpq}} f_{\boldsymbol{qp}}^{\boldsymbol{ab}} \bigotimes_{i=1}^{k} M_{\ p_i}^{\ b_i} \bigotimes_{j=1}^{k} M_{\ q_j}^{\ a_j}, \quad where \tag{31}$$

$$M_{\ p_i}^{\ b_i} = (3|p_i, b_i\rangle\langle p_i, b_i| - \mathbb{1}_2), \tag{32}$$

$$M_{\ q_j}^{\ a_j} = (3|q_j, a_j\rangle\langle q_j, a_j| - \mathbb{1}_2). \tag{33}$$

Proof. The proof of this result comes from relating the frequency measurements made in scenarios 3.1 and 3.2. It relies on the following identity relating the output state $\mathcal{C}(\rho)$ of a channel $\mathcal{C}$ to the Choi matrix $\Phi = \mathcal{C}(|\Omega\rangle\langle\Omega|)$

$$\mathcal{C}(\rho) = d \times \mathrm{Tr}_A(\Phi\,(\mathbb{1} \otimes \rho^\top)). \tag{34}$$

Using this result, we have that the probability of observing the result $\mathbf{p} \in \{0,1\}^k$ when measuring with setting $\mathbf{b}$, having passed in the pure state $P_{\mathbf{q}}^{\mathbf{a}\top}$ is:

$$p_{\mathbf{q}\,\mathbf{p}}^{\mathbf{a}\,\mathbf{b}} = d \times \mathrm{Tr}(\Phi\, P_{\mathbf{p}}^{\mathbf{b}} \otimes P_{\mathbf{q}}^{\mathbf{a}}). \tag{35}$$

Note that up to the factor d, these coincide with the probabilities (3) from scenario 3.1, with the identifications $(\mathbf{a}\,\mathbf{b}) = \mathbf{o}$ and $(\mathbf{q}\,\mathbf{p}) = \mathbf{s}$. The difference is due to the fact that while in scenario 3.1 both parts $(\mathbf{a}\,\mathbf{b})$ of the outcome $\mathbf{o}$ are random, in scenario 3.2 a single dataset is generated from random outcomes $\mathbf{b}$ for each *deterministically* chosen state with labels $\mathbf{a}$ and $\mathbf{q}$.

Since the maps from the Choi matrix to probabilities are the same in the two scenarios, up to the factor d, the expression (4) of the LS estimator in scenario 3.1 can be transferred to scenario 3.2 by replacing the frequencies $f_{\mathbf{o}}^{\mathbf{s}}$ with $f_{\mathbf{qp}}^{\mathbf{ab}}/d$ which gives us (31). $\qquad\square$

In the rest of this section we give a longer, but perhaps more transparent proof of the equality

$$p_{\mathbf{o}}^{\mathbf{s}} = \frac{1}{d} p_{\mathbf{q}\,\mathbf{p}}^{\mathbf{a}\,\mathbf{b}}$$

which may be of interest in its own right. We make the following comparisons between scenarios 3.1 and 3.2.

In scenario 3.2, we have that the probability of observing the result $\mathbf{p} \in \{\pm\}^k$ when measuring with setting $\mathbf{b}$, having passed in the projector state $P_{\mathbf{q}}^{\mathbf{a}}$ is:

$$p_{\mathbf{q}\,\mathbf{p}}^{\mathbf{a}\,\mathbf{b}} = \mathrm{Tr}(\mathcal{C}(P_{\mathbf{q}}^{\mathbf{a}\top})P_{\mathbf{p}}^{\mathbf{b}}). \tag{36}$$

If we choose to prepare input state with indices $(\mathbf{a}, \mathbf{q})$ and measure with each setting $\mathbf{b}$ an equal number of times, the construction of the estimator (31) involves sampling equally from the $3^{2k} \times 2^k$ corresponding distributions.

Alternatively, we could assign a probability of $\frac{1}{d}$ to choosing index $\mathbf{q}$ while the setting indices $(\mathbf{a}, \mathbf{b})$ are deterministic as above; in this case the probability for recording joint outcome $(\mathbf{q}, \mathbf{p})$ for an input/measurement setting of $(\mathbf{a}, \mathbf{b})$ is:

$$\tilde{p}_{\mathbf{q}\,\mathbf{p}}^{\mathbf{a}\,\mathbf{b}} = \frac{1}{d} \mathrm{Tr}(\mathcal{C}(P_{\mathbf{q}}^{\mathbf{a}\top})P_{\mathbf{p}}^{\mathbf{b}}). \tag{37}$$

In this case we are sampling from the 3^{2k} distributions corresponding to each setting $(\mathbf{a}, \mathbf{b})$.

Consider now scenario 3.1. Observing the joint outcome $\mathbf{o} = (\mathbf{q}, \mathbf{p})$ when measuring the Choi matrix with setting $\mathbf{s} = (\mathbf{a}, \mathbf{b})$ is equivalent to first measuring the ancilla qubits with setting $\mathbf{a}$, recording the outcome $\mathbf{q}$, and then measuring the conditional system state with setting $\mathbf{b}$ and recording outcome $\mathbf{p}$. The probability of observing outcome $\mathbf{q}$ when measuring with setting $\mathbf{a}$ is always $\frac{1}{d}$, since the marginal state of the ancilla is always the maximally mixed state.

The deferred measurement principle in quantum information theory [75] implies that the this set up is equivalent to making a measurement on the marginal ancilla state *before* acting with $\mathcal{C} \otimes \mathcal{I}_d$ on the joint system. Defining $\Omega_{\mathbf{q}}^{\mathbf{a}}$ as the marginal system state conditional on the measurement on the ancilla side of the maximally entangled state Ω, we will now show that the probabilities in scenario 1 (3) can be expressed as:

$$p_{\mathbf{o}}^{\mathbf{s}} = \tilde{p}_{\mathbf{q}\,\mathbf{p}}^{\mathbf{a}\,\mathbf{b}} = \frac{1}{d} \operatorname{Tr}(\mathcal{C}(\Omega_{\mathbf{q}}^{\mathbf{a}}) \, P_{\mathbf{p}}^{\mathbf{b}}). \tag{38}$$

Let $\{|\mathbf{q}\rangle = \bigotimes_{i=1}^{k/2} |q_i\rangle\}$ be the standard basis in $\mathbb{C}^d$. For each different measurement setting $\mathbf{a}$, the corresponding Pauli basis $\{|\mathbf{q}, \mathbf{a}\rangle\}$, is related to $\{|\mathbf{q}\rangle\}$ by a unitary transformation $U^{\mathbf{a}}$

$$U^{\mathbf{a}}|\mathbf{q}\rangle = |\mathbf{q}, \mathbf{a}\rangle \tag{39}$$

and the associated projection operators transform as

$$P_{\mathbf{q}}^{\mathbf{a}} = U^{\mathbf{a}} P_{\mathbf{q}} (U^{\mathbf{a}})^*. \tag{40}$$

The maximally entangled input state $|\omega\rangle$ in the standard basis is:

$$|\omega\rangle = \frac{1}{\sqrt{d}} \sum_{\mathbf{p}} |\mathbf{p}\rangle \otimes |\mathbf{p}\rangle. \tag{41}$$

Therefore, after measuring the ancilla qubits with setting $\mathbf{a}$ on the joint state $|\omega\rangle$ and obtaining outcome $\mathbf{q}$, the projected state is

$$\begin{aligned}
\mathbb{1}_d \otimes P_{\mathbf{q}}^{\mathbf{a}} |\omega\rangle &= (\mathbb{1}_d \otimes (U^{\mathbf{a}} P_{\mathbf{q}} (U^{\mathbf{a}})^*))|\omega\rangle \\
&= (\mathbb{1}_d \otimes (U^{\mathbf{a}} P_{\mathbf{q}} (U^{\mathbf{a}})^*) \bar{U}^{\mathbf{a}} \otimes U^{\mathbf{a}}|\omega\rangle \\
&= (\bar{U}^{\mathbf{a}} \otimes U^{\mathbf{a}} P_{\mathbf{q}})|\omega\rangle \\
&= (\bar{U}^{\mathbf{a}} \otimes U^{\mathbf{a}}) \frac{1}{\sqrt{d}} |\mathbf{q}\rangle \otimes |\mathbf{q}\rangle.
\end{aligned} \tag{42}$$

Therefore, the conditional state of the system after making the above measurement on the ancilla qubits is $\bar{U}^{\mathbf{a}}|\mathbf{q}\rangle$. We conclude, then, that

$$\Omega_{\mathbf{q}}^{\mathbf{a}} = \bar{U}^{\mathbf{a}} |\mathbf{q}\rangle\langle\mathbf{q}| (\bar{U}^{\mathbf{a}})^*. \tag{43}$$

But $|\mathbf{q}\rangle\langle\mathbf{q}| = P_{\mathbf{q}} = P_{\mathbf{q}}^{\top}$, so:

$$\Omega_{\mathbf{q}}^{\mathbf{a}} = P_{\mathbf{q}}^{\mathbf{a}\,\top} \tag{44}$$

Therefore the probabilities obtained making measurements in scenario 1 are:

$$p_{\mathbf{o} = \mathbf{q}\,\mathbf{p}}^{\mathbf{s} = \mathbf{a}\,\mathbf{b}} = \frac{1}{d} \operatorname{Tr}(\mathcal{C}(P_{\mathbf{q}}^{\mathbf{a}\,\top}) \, P_{\mathbf{p}}^{\mathbf{b}}). \tag{45}$$

which coincides with equation (38) corresponding to the alternative scenario 2, where the input indices $\mathbf{q}$ were chosen randomly with uniform probability. $\qquad \square$

8.2 Proof of Proposition 2

Proposition 2. *For direct QPT with (transposed) MUB inputs $(|w_i\rangle\langle w_i|)^\top$ and MUB measurements $\frac{d}{m}|v_l\rangle\langle v_l|$, the least-squares estimator for the Choi matrix Φ of a k-qubit quantum channel takes the following form:*

$$
\begin{aligned}
\hat{\Phi}_{LS} &= \frac{d+1}{d} \sum_{l,k=1}^{m} f_l^k |v_l\rangle\langle v_l| \otimes |w_k\rangle\langle w_k| \\
&\quad - \frac{1}{d} \sum_{l,k=1}^{m} f_l^k (|v_l\rangle\langle v_l| \otimes \mathbb{1}_d + \mathbb{1}_d \otimes |w_k\rangle\langle w_k|) \\
&\quad + \mathbb{1}_d \otimes \mathbb{1}_d.
\end{aligned}
\tag{46}
$$

Proof. We apply the relation

$$
\mathcal{C}(\rho) = d \times \mathrm{Tr}_A(\Phi\,(\mathbb{1} \otimes \rho^\top)).
$$

to express the probability distribution for a give input state $|w_k\rangle\langle w_k|^\top$ and measurement with POVM elements $M_l = \frac{d}{m}|v_l\rangle\langle v_l|$ as

$$
p_l^k = \mathrm{Tr}(\mathcal{C}(|w_k\rangle\langle w_k|^\top)\,M_l) = \frac{d^2}{m}\mathrm{Tr}(\Phi\,(P_l \otimes Q_k))
\tag{47}
$$

where $P_l = |v_l\rangle\langle v_l|, Q_k = |w_k\rangle\langle w_k|$. The linear map from Choi matrices to the collection of probability distributions is then

$$
\begin{aligned}
\mathcal{A} : M(\mathbb{C}^d) &\rightarrow \mathbb{C}^m \otimes \mathbb{C}^m \cong \mathbb{C}^{m^2} \tag{48} \\
\mathcal{A} : \Phi &\mapsto p := \{p_l^k : l, k = 1, \ldots, m\} \tag{49}
\end{aligned}
$$

It's adjoint is

$$
\mathcal{A}^\dagger : q \mapsto \frac{d^2}{m} \sum_{l,k=1}^{m} q_l^k\, P_l \otimes Q_k.
$$

The general form of the LS estimator, assuming $\mathcal{A}^\dagger \mathcal{A}$ is invertible, is given by

$$
\hat{\Phi} = (\mathcal{A}^\dagger \mathcal{A})^{-1} \mathcal{A}^\dagger f
\tag{50}
$$

where $f := \{p_l^k : l, k = 1, \ldots, m\}$ is the vector of empirical frequencies. To find a more explicit expression and prove the validity of equation (50), we analyse $\mathcal{A}^\dagger \mathcal{A}$ in more detail. For this we use the relations characteristic for 2-designs

$$
\sum_{k=1}^{m} \mathrm{Tr}(XP_l)P_l = X + \mathrm{Tr}(X)\mathbb{1}_d
$$

$$
\sum_{k=1}^{m} \mathrm{Tr}(XQ_k)Q_k = X + \mathrm{Tr}(X)\mathbb{1}_d
$$

We now show that $\mathcal{A}^\dagger \mathcal{A}$ has 3 eigenvalues with eigenspaces consisting of linear spans of the following 4 types of operators:

1. Let $\Lambda_1 = X \otimes Y$ with $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = 0$. In this case

$$
\begin{aligned}
\mathcal{A}^\dagger \mathcal{A}(\Lambda_1) &= \frac{d^4}{m^2} \sum_{l,k=1}^{m} \mathrm{Tr}(Z(P_l \otimes Q_k))P_l \otimes Q_k \\
&= \frac{d^4}{m^2} \left(\sum_{l=1}^{m} \mathrm{Tr}(XP_l)P_l \right) \left(\sum_{k=1}^{m} \mathrm{Tr}(YQ_k)Q_k \right) \\
&= \frac{d^4}{m^2} X \otimes Y = \frac{d^2}{(d+1)^2}\Lambda_1
\end{aligned}
$$

2. Let $\Lambda_2 = X \otimes \mathbb{1}_d$ with $\mathrm{Tr}(X) = 0$. In this case

$$
\begin{aligned}
\mathcal{A}^\dagger \mathcal{A}(\Lambda_2) &= \frac{d^4}{m^2} \sum_{l,k=1}^m \mathrm{Tr}(\Lambda_2(P_l \otimes Q_k)) P_l \otimes Q_k \\
&= \left(\sum_{l=1}^m \mathrm{Tr}(X P_l) P_l \right) \left(\sum_{k=1}^m \mathrm{Tr}(Q_k) Q_k \right) \\
&= \frac{d^4(d+1)}{m^2} X \otimes \mathbb{1}_d = \frac{d^2}{d+1} \Lambda_2
\end{aligned}
$$

where in the last step we used that $\sum_k Q_k = (d+1)\mathbb{1}_d$.

3. A similar equality holds for $\Lambda_3 = \mathbb{1}_d \otimes Y$ with $\mathrm{Tr}(Y) = 0$.

4. Let $\Lambda_4 = \mathbb{1}_d \otimes \mathbb{1}_d$. Here, we have

$$
\begin{aligned}
\mathcal{A}^\dagger \mathcal{A}(\Lambda_4) &= \frac{d^4}{m^2} \sum_{l,k=1}^m \mathrm{Tr}(Z(P_l \otimes Q_k)) P_l \otimes Q_k \\
&= \frac{d^4}{m^2} \left(\sum_{l=1}^m P_l \right) \left(\sum_{k=1}^m Q_k \right) \\
&= \frac{d^4(d+1)^2}{m^2} \mathbb{1}_d \otimes \mathbb{1}_d = d^2 \Lambda_4
\end{aligned}
$$

Writing $P_l = \tilde{P}_l + \mathbb{1}_d/d$ with $\mathrm{Tr}(\tilde{P}_l) = 0$, and $Q_k = \tilde{Q}_k + \mathbb{1}_d/d$ with $\mathrm{Tr}(\tilde{Q}_k) = 0$, we now have

$$
\begin{aligned}
\left(\mathcal{A}^\dagger \mathcal{A} \right)^{-1} \mathcal{A}^\dagger f &= \left(\mathcal{A}^\dagger \mathcal{A} \right)^{-1} \frac{d^2}{m} \sum_{l,k=1}^m f_l^k P_l \otimes Q_k \\
&= \left(\mathcal{A}^\dagger \mathcal{A} \right)^{-1} \frac{d^2}{m} \sum_{l,k=1}^m f_l^k (\tilde{P}_l + \frac{\mathbb{1}_d}{d}) \otimes (\tilde{Q}_k + \frac{\mathbb{1}_d}{d}).
\end{aligned}
$$

By expanding the brackets and using the fact that each term is an eigenvector of $\mathcal{A}^\dagger \mathcal{A}$, taking the form of one of $Z_{(1,2,3,4)}$ above, we get

$$
\begin{aligned}
\left(\mathcal{A}^\dagger \mathcal{A} \right)^{-1} \mathcal{A}^\dagger f &= \frac{d^2}{m} \frac{(d+1)^2}{d^2} \sum_{l,k=1}^m f_l^k \tilde{P}_l \otimes \tilde{Q}_k \\
&+ \frac{d^2}{m} \frac{d+1}{d^2} \sum_{l,k=1}^m f_l^k \tilde{P}_l \otimes \frac{\mathbb{1}_d}{d} \\
&+ \frac{d^2}{m} \frac{d+1}{d^2} \sum_{l,k=1}^m f_l^k \frac{\mathbb{1}_d}{d} \otimes \tilde{Q}_k \\
&+ \frac{d^2}{m} \frac{1}{d^2} \sum_{l,k=1}^m f_l^k \frac{\mathbb{1}_d}{d} \otimes \frac{\mathbb{1}_d}{d}
\end{aligned}
$$

By expressing $\tilde{P}_l = P_l - \mathbb{1}_d/d$ and $\tilde{Q}_k = Q_k - \mathbb{1}_d/d$ we get

$$
\begin{aligned}
\left(\mathcal{A}^\dagger \mathcal{A} \right)^{-1} \mathcal{A}^\dagger f &= \frac{d+1}{d} \sum_{l,k=1}^m f_l^k (P_l - \frac{\mathbb{1}_d}{d}) \otimes (Q_k - \frac{\mathbb{1}_d}{d}) \\
&+ \frac{1}{d} \sum_{l,k=1}^m f_l^k \left[(P_l - \frac{\mathbb{1}_d}{d}) \otimes \frac{\mathbb{1}_d}{d} + \frac{\mathbb{1}_d}{d} \otimes (Q_k - \frac{\mathbb{1}_d}{d}) \right] \\
&+ \frac{\mathbb{1}_d}{d} \otimes \frac{\mathbb{1}_d}{d}
\end{aligned}
$$

Finally, by rearranging terms we get

$$\left(\mathcal{A}^\dagger \mathcal{A}\right)^{-1} \mathcal{A}^\dagger f \quad = \frac{d+1}{d} \sum_{l,k=1}^{m} f_l^k P_l \otimes Q_k$$

$$-\frac{1}{d} \sum_{l,k=1}^{m} f_l^k (P_l \otimes \mathbb{1}_d + \mathbb{1}_d \otimes Q_k)$$

$$+\mathbb{1}_d \otimes \mathbb{1}_d$$

$\square$

9 Concentration bounds for the LS estimators

In order to prove Theorem 1, we first find error bounds for the least-squares estimators for each of the scenarios 3.1- 3.4 in Frobenius norm distance, in sections 9.1-9.4. For this we will be applying a matrix generalisation of the classical scalar Bernstein inequality, see Ref. [76].

Theorem 3. *Consider a sequence of N independent, Hermitian, random matrices $\{A_i\}_{i=1}^N \in \mathbb{C}^{n \times n}$ satisfying*

$$\mathbb{E}[A_k] = 0 \text{ and } \|A_k\|_\infty \le R. \tag{51}$$

Then, for $t \ge 0$:

$$\Pr[\|\sum_{i=1}^N A_i\|_\infty \ge t] \le N \exp\left(-\frac{t^2/3}{\sigma^2 + Rt/3}\right), \tag{52}$$

where $\sigma^2 = \|\sum_{i=1}^N \mathbb{E}[(A_i)^2]\|_\infty$ captures the norm of the total variance.

9.1 Scenario 3.1

Proposition 3. *The operator norm distance of the least-squares estimator from the true Choi matrix for scenario 3.1 satisfies the following bound:*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_\infty \ge \tau] \le d^2 \exp\left(-\frac{3N\tau^2}{8 \times 3^{2k}}\right) \tag{53}$$

for a $2k$-qubit joint system-ancilla system, where $d = 2^k$, N is the total number of measurements, $\hat{\Phi}_{LS}$ is the least-squares estimator for the true state Φ, and $\tau \in [0, 1]$.

Proof. We construct this bound by borrowing the result for the 'Pauli basis measurements' case given in [37] and adapting it for a larger system. The proof is an application of the matrix Bernstein Inequality and we sketch it here, though the reader is directed to [37] for details.

We can write the LS estimator (4) in the following way:

$$\hat{\Phi}_{LS} = \sum_{\mathbf{s}} \frac{1}{N} \sum_i^{N/3^{2k}} X_i^{\mathbf{s}} \tag{54}$$

where, for every $\mathbf{s} \in \{x, y, z\}^{2k}$, the $X_i^{\mathbf{s}}$ are independent instances of the random matrices:

$$X^{\mathbf{s}} : \mathbf{o} \mapsto \overset{2k}{\underset{i=1}{\otimes}} (3|o_i, s_i\rangle\langle o_i, s_i| - \mathbb{1}_{d^2}) \tag{55}$$

with probability distribution over $\mathbf{o} \in \{0, 1\}^{2k}$

$$p_{\mathbf{o}}^{\mathbf{s}} = \operatorname{Tr}(\Phi \, P_{\mathbf{o}}^{\mathbf{s}}). \tag{56}$$

We can then write

$$\|\hat{\Phi}_{LS} - \Phi\|_\infty = \left\| \sum_{\mathbf{s}} \frac{1}{N} \sum_{i}^{N/3^{2^k}} (X_i^{\mathbf{s}} - \mathbb{E}[X^{\mathbf{s}}]) \right\|_\infty. \tag{57}$$

If we define the matrices:

$$A_i^{\mathbf{s}} = \frac{1}{N}(X_i^{\mathbf{s}} - \mathbb{E}[X_i^{\mathbf{s}}]), \tag{58}$$

equation (57) becomes

$$\|\hat{\Phi}_{LS} - \Phi\|_\infty = \| \sum_{\mathbf{s}} \sum_{i}^{N/3^{2^k}} A_i^{\mathbf{s}} \|_\infty. \tag{59}$$

Thus we can apply the matrix Bernstein inequality to bound $\|\hat{\Phi}_{LS} - \Phi\|_\infty$. We have that:

$$\sigma^2 = \left\| \sum_{\mathbf{s}} \frac{1}{N^2} \sum_{i}^{N/3^{2^k}} \mathbb{E}[(X_i^{\mathbf{s}} - \mathbb{E}[X^{\mathbf{s}}])^2] \right\|_\infty \tag{60}$$

and

$$R = \max\{\frac{1}{N}\|X_i^{\mathbf{s}} - \mathbb{E}[X_i^{\mathbf{s}}]\|_\infty\}. \tag{61}$$

Using the following inequality:

$$\mathbb{E}[(X_i^{\mathbf{s}} - \mathbb{E}[X_i^{\mathbf{s}}])^2] \leq \mathbb{E}[(X^{\mathbf{s}})^2] \tag{62}$$

and Jensen's inequality [77], we can solve (60) and (61) to find:

$$\sigma^2 = \frac{3^k}{N} \quad \text{and} \quad R = \frac{2^{k+1}}{N}. \tag{63}$$

Applying the matrix Bernstein inequality then gives us equation (53), provided that τ is not too large. $\qquad\square$

We now convert this bound into one containing the Frobenius norm.

Proposition 4. *The following bound holds*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_2^2 \geq \delta^2] \leq d^2 \exp(-\frac{3N\delta^2}{8 \times d^2 \times 3^{2^k}}) \tag{64}$$

for $\delta^2 \in [0, 1]$.

Proof. Taking (66) as our starting point, we can make use of the following inequality involving the Frobenius norm and operator norm of a Hermitian matrix:
Let $A \in \text{Herm}_{n \times n}$ be an arbitrary Hermitian matrix, then

$$\|A\|_2^2 \leq d^2 \|A\|_\infty^2. \tag{65}$$

$\hat{\Phi}_{LS} - \Phi$ is clearly Hermitian since it is constructed as a sum over a set of random Hermitian matrices.

If $\|\hat{\Phi}_{LS} - \Phi\|_\infty \leq \tau$ with some probability, it must be that $\|\hat{\Phi}_{LS} - \Phi\|_2^2 \leq d^2\tau^2$ with at least that probability, by equation (65). Therefore, setting $\delta^2 = d^2\tau^2$ we arrive at (64). This is valid for $\delta^2 \in [0, d^2]$, so is too in the region $[0, 1]$ that we are interested in.

$\qquad\square$

9.2 Scenario 3.2

Proposition 5. *The LS estimator in scenario 3.2 satisfies the concentration bound*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_\infty \geq \tau] \leq d^2 \exp(-\frac{3N\tau^2}{8 \times 3^{2k}}) \tag{66}$$

where N is the total number of measurements, $\hat{\Phi}_{LS}$ is the least-squares estimator for the true state Φ, and $\tau \in [0, 1]$.

Proof. Using Proposition 1 we can express the least squares estimator in scenario 3.2 as a sum of random matrices, similarly to scenario 3.1.

$$\hat{\Phi}_{LS} = \sum_{\mathbf{a,b,q}} \frac{1}{N} \sum_i^{N/s} X_{\mathbf{q},i}^{\mathbf{a,b}}$$

where $s = 3^{2k} \cdot 2^k$, and $\mathbf{a}, \mathbf{b} \in \{x, y, z\}^k$, $\mathbf{q} \in \{0, 1\}^k$. Here, $X_{\mathbf{q},i}^{\mathbf{a,b}}$ are independent instances of the random matrices:

$$X_{\mathbf{q}}^{\mathbf{a,b}} : \mathbf{p} \mapsto \bigotimes_{i=1}^k M^{\,b_i}_{\,p_i} \bigotimes_{j=1}^k M^{\,a_j}_{\,q_j}$$

where

$$M^{\,b_i}_{\,p_i} = (3|p_i, b_i\rangle\langle p_i, b_i| - \mathbb{1}_2), \quad M^{\,a_j}_{\,q_j} = (3|q_j, a_j\rangle\langle q_j, a_j| - \mathbb{1}_2).$$

with probability distribution over $\mathbf{p} \in \{0, 1\}^k$

$$p_{\mathbf{q\,p}}^{\mathbf{a\,b}} = \mathrm{Tr}(\mathcal{C}(P_{\mathbf{q}}^{\mathbf{a}\top})P_{\mathbf{p}}^{\mathbf{b}}).$$

Then

$$\hat{\Phi}_{LS} - \Phi = \sum_{\mathbf{a,b,q}} \sum_i^{N/s} A_{\mathbf{q},i}^{\mathbf{a,b}}$$

where $A_{\mathbf{q},i}^{\mathbf{a,b}} = \frac{1}{N}(X_{\mathbf{q},i}^{\mathbf{a,b}} - \mathbb{E}(X_{\mathbf{q},i}^{\mathbf{a,b}}))$ and we can apply the concentration bound from Theorem 3 similarly to Proposition 3. Since $X_{\mathbf{q}}^{\mathbf{a,b}}$ takes values in the same set as the random matrix $X^{\mathbf{s}}$ in Proposition 3 (scenario 3.1), we obtain

$$R = \max_{\mathbf{a,b,q}} \frac{1}{N}\|X_{\mathbf{q}}^{\mathbf{a,b}} - \mathbb{E}(X_{\mathbf{q}}^{\mathbf{a,b}})\|_\infty = \frac{2^{k+1}}{N}.$$

We now consider the variance

$$\sigma^2 = \left\| \sum_{\mathbf{a,b,q}} \frac{1}{N^2} \sum_i^{N/s} \mathbb{E}[(X_{\mathbf{q},i}^{\mathbf{a,b}} - \mathbb{E}[X_{\mathbf{q}}^{\mathbf{a,b}}])^2] \right\|_\infty$$

Using $\mathbb{E}[(X = \mathbb{E}(X))^2] \leq \mathbb{E}[X^2]$ we get

$$\sigma^2 \leq \left\| \sum_{\mathbf{a,b,q}} \frac{1}{sN} \mathbb{E}[(X_{\mathbf{q}}^{\mathbf{a,b}})^2] \right\|_\infty$$

Now using the relation between scenario 3.1 and 3.2 probabilities $p_{\mathbf{q,p}}^{\mathbf{a,b}} = d p_{\mathbf{o}}^{\mathbf{s}}$ where $\mathbf{s} = (\mathbf{a,b})$ and $\mathbf{o} = (\mathbf{q,p})$, we have

$$\begin{aligned}
\sum_{\mathbf{a,b,q}} \frac{1}{sN} \mathbb{E}[(X_{\mathbf{q}}^{\mathbf{a,b}})^2] &= \frac{1}{sN} \sum_{\mathbf{a,b,q}} \sum_{\mathbf{p}} p_{\mathbf{q,p}}^{\mathbf{a,b}}(X_{\mathbf{q}}^{\mathbf{a,b}}(p))^2 \\
&= d \sum_{\mathbf{o,s}} p_{\mathbf{o}}^{\mathbf{s}}(X^{\mathbf{s}}(\mathbf{o}))^2 \\
&= d \sum_{\mathbf{s}} \mathbb{E}[(X^{\mathbf{s}})^2].
\end{aligned}$$

Taking into account that $s = d3^{2k}$ we arrive at the same upper bound expression as found in scenario 3.1

$$\begin{aligned} \sigma^2 \quad &\leq \|\frac{d}{sN}\sum_{\mathbf{s}}\mathbb{E}[(X^{\mathbf{s}})^2]\|_\infty \\ &= \frac{1}{3^{2k}N}\|\sum_{\mathbf{s}}\mathbb{E}[(X^{\mathbf{s}})^2]\|_\infty \\ &= \frac{3^{2k}}{N}. \end{aligned}$$

$\square$

Proposition 6. *The following bound holds*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_2^2 \geq \delta^2] \leq d^2\exp(-\frac{3N\delta^2}{8 \times d^2 \times 3^{2k}}) \tag{67}$$

for $\delta^2 \in [0,1]$.

9.3 Scenario 3.3

Proposition 7. *The operator norm error of the least-squares estimator in scenario 3.3 satisfies the following bound:*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_\infty \geq \tau] \leq d^2\exp(-\frac{3N\tau^2}{16}\frac{1}{d^2}). \tag{68}$$

This is a direct application of the result for 2-design POVMs in [37] to a system of dimension of d^2. It serves no purpose to repeat the proof - another application of the matrix-Bernstein inequality - here, and the reader is directed to [37] for an exposition.

Proposition 8.

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_2^2 \geq \delta^2] \leq d^2\exp(-\frac{3N\delta^2}{16}\frac{1}{d^4}). \tag{69}$$

Proof. As with scenario 3.1, we use the relation (65) to derive the extra factor of $\frac{1}{d^2}$ in the exponential, and arrive at (69). $\square$

9.4 Scenario 3.4

Proposition 9. *The operator norm error of the least-squares estimator in scenario 3.4 satisfies the following bound:*

$$\Pr[\|\hat{\Phi}_{LS} - \Phi\|_\infty \geq \tau] \leq d^2\exp\left(-\frac{3N\tau^2}{32d^2}\right) \tag{70}$$

Proof. We return to another application of the concentration inequality (52).

Once again, define $P_l = |v_l\rangle\langle v_l|$ and $Q_k = |w_k\rangle\langle w_k|$. Define also $Y_l^k = \frac{d+1}{d}P_l \otimes Q_k$ and $Z_l^k = \frac{1}{d}(P_l \otimes \mathbb{1}_d + \mathbb{1}_d \otimes Q_k)$ for future use.

For each $k \in [m]$ we define the random matrices

$$X^k : l \mapsto \frac{d+1}{d}P_l \otimes Q_k - \frac{1}{d}(P_l \otimes \mathbb{1}_d + \mathbb{1}_d \otimes Q_k) \tag{71}$$

with probability distribution over $l \in \{1, ..., m\}$

$$p_l^k = \text{Tr}(\mathcal{C}(Q_k^\top)M_l). \tag{72}$$

We denote by X_i^k the independent samples of X^k corresponding to the measurement outcomes for input state $Q_k^\top$, with $i = 1, ..., N/m$.

We then write the LS estimator (46) in the following way:

$$\hat{\Phi}_{LS} = \sum_{k=1}^{(d+1)d} \frac{m}{N} \sum_{i=1}^{N/m} X_i^k + \mathbb{1}_d \otimes \mathbb{1}_d, \tag{73}$$

assuming again that the N measurements are distributed uniformly among the different input states.

Then, we can write the least-squares estimator operator-norm distance as

$$
\begin{aligned}
\|\hat{\Phi}_{LS} - \Phi\|_\infty &= \left\| \sum_{k=1}^{(d+1)d} \frac{m}{N} \sum_{i=1}^{N/m} X_i^k - \sum_{k=1}^{(d+1)d} \mathbb{E}[X^k] \right\|_\infty \\
&= \left\| \sum_{k=1}^{(d+1)d} \sum_{i=1}^{N/m} A_i^k \right\|_\infty
\end{aligned}
\tag{74}
$$

i.e. the operator norm of a sum over independent centred random matrices:

$$A_i^k = \frac{m}{N}(X_i^k - \mathbb{E}[X^k]); \ \ i \in [N], \ k \in [(d+1)d]. \tag{75}$$

The upper bound, R, for the operator norm of each contribution to the sum (see equation (52)) can be found in the following way:

$$
\begin{aligned}
\|A_i^k\|_\infty &\le \frac{m}{N}(\|X_i^k\|_\infty + \|\mathbb{E}[X_i^k]\|_\infty) \\
&\le \frac{m}{N}(\|X_i^k\|_\infty)(1 + \max\{p_l^k\}) \\
&\le \frac{2m}{N}(\|Y_i^k\|_\infty + \|Z_i^k\|_\infty) \\
&= \frac{2(d+1)(d+3)}{N} =: R.
\end{aligned}
\tag{76}
$$

And the variance term:

$$
\begin{aligned}
\sigma^2 &= \left\| \sum_{k=1}^{m} \sum_{i=1}^{N/m} \mathbb{E}[(A_i^k)^2] \right\|_\infty \\
&= \left\| \sum_{k=1}^{m} \frac{m^2}{N^2} \sum_{i=1}^{N/m} \mathbb{E}[(X_i^k - \mathbb{E}[x_i^k])^2] \right\|_\infty \\
&= \frac{m}{N} \left\| \sum_{k=1}^{m} \mathbb{E}[(X^k - \mathbb{E}[X^k])^2] \right\|_\infty \\
&\le \frac{m}{N} \left\| \sum_{k=1}^{m} \mathbb{E}[(X^k)^2] \right\|_\infty
\end{aligned}
\tag{77}
$$

where the inequality is with respect to Loewner ordering. Now

$$
\begin{aligned}
\sum_{k=1}^{m} \mathbb{E}[(X^k)^2] &= \sum_{k,l=1}^{m} p_l^k (Y_l^k - Z_l^k)^2 \\
&= \frac{1}{d^2} \sum_{k,l=1}^{m} p_l^k (d(d-1) P_l \otimes Q_k - d X_l^k) \\
&= \frac{d-1}{d} \sum_{k,l=1}^{m} p_l^k P_l \otimes Q_k - \frac{1}{d} \sum_{k=1}^{m} \mathbb{E}[X^k].
\end{aligned}
\tag{78}
$$

Given that (46) is unbiased, we have that $\sum_{k=1}^{m} \mathbb{E}[X^k] = \Phi - \mathbb{1}_d \otimes \mathbb{1}_d$. Focussing on the first term:

$$
\begin{aligned}
\sum_{k,l=1}^{m} p_l^k P_l \otimes Q_k &= \sum_{k,l=1}^{m} \text{tr}(\frac{d}{m} P_l \mathcal{C}(Q_k^\top)) P_l \otimes Q_k \\
&= \frac{d^2}{m} \sum_{k,l=1}^{m} \text{tr}(P_l \otimes Q_k \Phi) P_l \otimes Q_k
\end{aligned}
\tag{79}
$$

by equation (34).

We now use the fact that the P_l and Q_k form spherical 2-designs. The near-isotropy of such 2-designs ensures the following for arbitrary selfadjoint operators A, B:

$$
\begin{aligned}
\frac{d^2}{m} \sum_{k,l=1}^{m} \text{tr}((P_l \otimes Q_k)(A \otimes B)) P_l \otimes Q_k &= d^3(d+1)(\frac{1}{m} \sum_{l=1}^{m} P_l \text{tr}(P_l A))(\frac{1}{m} \sum_{l=1}^{m} Q_k \text{tr}(Q_k B)) \\
&= \frac{d}{d+1}(A + \text{tr}(A)\mathbb{1}_d)(B + \text{tr}(B)\mathbb{1}_d) \\
&= \frac{d}{d+1}(A \otimes B + A \otimes \text{tr}(B)\mathbb{1}_d + \text{tr}(A)\mathbb{1}_d \otimes B + \text{tr}(A)\text{tr}(B)\mathbb{1}_d \otimes \mathbb{1}_d).
\end{aligned}
\tag{80}
$$

This can be extended linearly to all selfadjoint $\Phi \in M(\mathbb{C}^d)^{\otimes 2}$ such that:

$$
\frac{d^2}{m} \sum_{k,l=1}^{m} \text{tr}(P_l \otimes Q_k \Phi) P_l \otimes Q_k = \frac{d}{d+1}(\Phi + \text{tr}_A(\Phi) \otimes \mathbb{1}_d + \mathbb{1}_d \otimes \text{tr}_S(\Phi) + \text{tr}(\Phi)\mathbb{1}_d \otimes \mathbb{1}_d).
\tag{81}
$$

Thus:

$$
\begin{aligned}
\sigma^2 &\leq \frac{m}{N} \| \sum_{k=1}^{m} \mathbb{E}[(X^k)^2] \|_\infty \\
&\leq \left\| \frac{d-1}{d+1}(\Phi + \text{tr}_A(\Phi) \otimes \mathbb{1}_d + \mathbb{1}_d \otimes \text{tr}_S(\Phi) \right. \\
&\qquad\qquad + \text{tr}(\Phi)\mathbb{1}_d \otimes \mathbb{1}_d)) \\
&\qquad\qquad \left. - \frac{1}{d}(\Phi + \mathbb{1}_d \otimes \mathbb{1}_d)) \right\|_\infty \\
&\leq \frac{m}{N}\left(\frac{4(d-1)}{d+1} + \frac{2}{d}\right) \\
&\leq \frac{4d^2}{N}
\end{aligned}
\tag{82}
$$

Substituting our upper bound of σ^2 into the matrix Berstein inequality yields the claim, provided that τ does not become too large. $\qquad \square$

Proposition 10. *The following bound holds*

$$
\Pr[\|\hat{\Phi}_{LS} - \Phi\|_2^2 \geq \delta^2] \leq d^2 \exp\left(-\frac{3N\delta^2}{32d^4}\right).
\tag{83}
$$

Proof. As with scenario (3.1), we use the relation (65) to arrive at (83). $\qquad \square$

9.5 Summary of LS concentration bounds

The LS error bounds derived in sections 9.1, 9.2, 9.3 and 9.4 can be summarised as follows

Corollary 2.

$$
\Pr[\|\hat{\Phi}_{LS} - \Phi\|_\infty \geq \tau] \leq d^2 \exp(-\frac{3N\tau^2}{8}g(k))
\tag{84}
$$

$$
\Pr[\|\hat{\Phi}_{LS} - \Phi\|_2^2 \geq \delta^2] \leq d^2 \exp(-\frac{3N\delta^2}{8}\frac{g(k)}{d^2})
\tag{85}
$$

where:

$$g(k) = \begin{cases} \frac{1}{3^{2k}} & \text{Scenarios 3.1, 3.2} \\ \frac{1}{2}\frac{1}{2^{2k}} & \text{Scenario 3.3} \\ \frac{1}{4}\frac{1}{2^{2k}} & \text{Scenario 3.4} \end{cases} \tag{86}$$

for τ and $\delta^2 \in [0, 1]$.

10 Key properties for two-step projection

In this section, we prove that the implementations of our projection procedures for computing $\hat{\Phi}_{PLS}$ satisfy the two properties (13) and (14).

We start with the following lemma relating the operator norm distances of Φ_{LS} and Φ_{CP1} from Φ, that is, ensuring Property (13).

Lemma 2. *For any non-negative threshold τ that is less than $\|\hat{\Phi}_{LS} - \Phi\|_\infty$, the first thresholded CP1 estimator Φ^τ_{CP1} defined in Algorithm 1 satisfies:*

$$\|\hat{\Phi}^\tau_{CP1} - \Phi\|_\infty \le 2\|\hat{\Phi}_{LS} - \Phi\|_\infty. \tag{87}$$

In particular, if τ is zero (projection on CP) or if $\tau = -\lambda_{\min}(\hat{\Phi}_{LS})$ (our main implementation), the condition is satisfied.

Proof. We write λ_i for the eigenvalues of $\hat{\Phi}_{LS}$, as in Algorithm 1.

First, since Φ is positive semi-definite, all negative eigenvalues of $\hat{\Phi}_{LS}$ have smaller absolute value than $\|\hat{\Phi}_{LS} - \Phi\|_\infty$. In particular $\tau = -\lambda_{\min}(\hat{\Phi}_{LS})$ satisfies the conditions of the lemma.

The operators $\hat{\Phi}^\tau_{CP1}$ and $\hat{\Phi}_{LS}$ are diagonal in the same basis. By the triangle inequality, we just have to prove that no eigenvalue gets changed by more than $\|\hat{\Phi}_{LS} - \Phi\|_\infty$.

If the condition on line 9 is met, then all eigenvalues between $\lambda_{\min}(\hat{\Phi}_{LS})$ and τ are set to zero, and all the other eigenvalues μ_i are set to $(\lambda_i + \tau - x_0)_+$. So that the small eigenvalues change by at most $\tau \vee |\lambda_{\min}(\hat{\Phi}_{LS})| \le \|\hat{\Phi}_{LS} - \Phi\|_\infty$, and the big eigenvalues change by at most $\tau \vee |\tau - x_0|$. We must then bound $|\tau - x_0|$. First notice that $x_0 \ge 0$ since $\sum \mu_i \ge 1$ and $\sum(\mu_i - x_0)_+ = 1$. Our worst case is then if $x_0 > 2\tau$. But in that case, all final eigenvalues can be seen as $(\lambda_i + \tau - x_0)_+$. Since $\langle v_i|\Phi|v_i\rangle \ge (\lambda_i - \|\hat{\Phi}_{LS} - \Phi\|_\infty)_+$, and $\text{Tr}\,\Phi = 1$, we deduce $x_0 - \tau \le \|\hat{\Phi}_{LS} - \Phi\|_\infty$, as wished.

On the other hand, if the condition on line 9 is not met, all eigenvalues bigger than the critical eigenvalue λ_j are increased by τ, all smaller eigenvalues are set to zero, and the final eigenvalue j satisfies $0 \le \mu_j \le \lambda_j + \tau$. So that we only have to prove that $0 \le \lambda_j \le \tau$. Since the trace of $\hat{\Phi}_{LS}$ is one, the left inequality on line 14 ensures that $\lambda_j > 0$. On the other hand, if $\lambda_j > \tau$, then the right inequality on line 14 would ensure that the the condition on line 9 was met. The contradiction ends the proof. □

We now prove that the second 'projection' does not increase the Frobenius distance, that is, Property (14).

Lemma 3. *Assume that $\hat{\Phi}_{PLS}$ is, given the input $\hat{\Phi}_{CP1}$, the output of either Dykstra's algorithm at convergence, AP algorithm for any tolerance, or hyperplane intersection projections algorithm with any switch conditions for any tolerance. Then*

$$\|\hat{\Phi}_{PLS} - \Phi\|_2 \le \|\hat{\Phi}_{CP1} - \Phi\|_2 \tag{88}$$

Proof. By Lemma 1, we merely have to prove that all our 'projections' can be seen as a succession of usual projections in Frobenius norm on convex sets that contain $\mathcal{CPTP}$.

Dykstra's algorithm directly outputs the Frobenius projection on $\mathcal{CPTP}$.

AP are a succession of Frobenius projections on $\mathcal{CP}$, and on $\mathcal{TP}$, both of which are convex set that include $\mathcal{CPTP}$.

Hyperplane intersection projections in AP mode is just AP. In HIP mode, we project on the intersection of: $\mathcal{TP}$, and a set of half-spaces that contain $\mathcal{CP}$. All those sets are convex sets that contain $\mathcal{CPTP}$, hence their intersection also is. □

11 Proof of Theorem 1

We now return to the proof of Theorem 1 and use the concentration bounds (84) for the LS estimators in the 4 scenarios to obtain bounds for PLS.

We will prove a stronger version, Theorem 4 which also covers the proof of Theorem 2.

To state it, we recall the notion of being close to rank r: a state ρ is δ-almost rank r if there is a rank r state $\rho_{(r)}$ such that

$$\|\rho - \rho_{(r)}\|_\infty \leq \delta. \tag{89}$$

Theorem 4. *Let Φ be the Choi matrix of a channel and $\hat{\Phi}_{PLS}$ our estimator, generated with any 'projections' satisfying Properties (13) and (14). Then, with probability at least $1 - \eta$, the bounds (90), (91) and (92) below hold simultaneously true, for all r and corresponding values of δ.*

Assume that either Φ or $\hat{\Phi}_{CP1}$ is δ-almost rank r. With probability at least $1 - \eta$:

$$\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \leq 2r \left(\delta + 2\sqrt{\frac{8\ln(d^2/\eta)}{3Ng(k)}} \right)^2, \tag{90}$$

If it was Φ that was δ-almost rank r, then:

$$\|\hat{\Phi}_{PLS} - \Phi\|_1 \leq r \left((4\sqrt{2} + 2)\delta + 4\sqrt{2}\sqrt{\frac{8\ln(d^2/\eta)}{3Ng(k)}} \right), \tag{91}$$

If it was $\hat{\Phi}_{CP1}$ that was δ-almost rank r, then:

$$\|\hat{\Phi}_{PLS} - \Phi\|_1 \leq r \left((4\sqrt{2} + 2)\delta + (4 + 8\sqrt{2})\sqrt{\frac{8\ln(d^2/\eta)}{3Ng(k)}} \right). \tag{92}$$

As a remark, Theorem 4 reduces to Theorem 1 when Φ is δ-almost rank r, with $\delta = 0$, i.e. it is exactly rank r. The formulation in Theorem 4 can be seen by taking ϵ to equal the right hand side of the equations, solving for η and inverting the probabilities.

First, Property (13) relates the L^∞ norms of the errors of $\hat{\Phi}_{LS}$ and $\hat{\Phi}_{CP1}$:

$$\|\hat{\Phi}_{CP1}^\tau - \Phi\|_\infty \leq 2\|\hat{\Phi}_{LS} - \Phi\|_\infty.$$

The next step is to relate the Frobenius norm of the distance between $\hat{\Phi}_{CP1}$ and Φ to their operator norm distance.

Lemma 4. *Assume either that $\hat{\Phi}_{CP1}$ or Φ is δ-almost rank r. Then*

$$\|\hat{\Phi}_{CP1} - \Phi\|_2^2 \leq 2r(\|\hat{\Phi}_{CP1} - \Phi\|_\infty + \delta)^2. \tag{93}$$

Proof. A key argument is presented in [37] for the following inequality relating the 1-norm to the operator-norm of two trace-1 positive-semidefinite matrices ρ_1 (of definite rank r) and ρ_2:

$$\|\rho_1 - \rho_2\|_1 \leq 2r\|\rho_1 - \rho_2\|_\infty. \tag{94}$$

Thus we have:

$$\|\hat{\Phi}_{CP1} - \Phi\|_1 \leq \|\hat{\Phi}_{CP1} - \Phi^r\|_1 + \|\Phi^r - \Phi\|_1 \tag{95}$$

$$\leq 2r(\|\hat{\Phi}_{CP1} - \Phi^r\|_\infty + \|\Phi^r - \Phi\|_\infty) \tag{96}$$

$$\leq 2r(\delta + \|\hat{\Phi}_{CP1} - \Phi\|_\infty + \delta). \tag{97}$$

There exists the following inequality relating the Frobenius, 1- and operator-norms:
Let $A \in \mathrm{Herm}_{n\times n}$ be an arbitrary Hermitian matrix, then

$$\|A\|_2^2 \leq \|A\|_\infty \|A\|_1, \tag{98}$$

which we can make use of since $\hat{\Phi}_{CP1} - \Phi$ is the difference of two quantum states, and hence must be Hermitian.

Using this, we can say:

$$\|\hat{\Phi}_{CP1} - \Phi\|_2^2 \leq 2r(\|\hat{\Phi}_{CP1} - \Phi\|_\infty + 2\delta)\|\hat{\Phi}_{CP1} - \Phi\|_\infty$$
$$\leq 2r(\|\hat{\Phi}_{CP1} - \Phi\|_\infty + \delta)^2.$$

$\square$

By Property 14 we may transfer this result to $\hat{\Phi}_{PLS}$:

$$\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \leq 2r(\|\hat{\Phi}_{CP1} - \Phi\|_\infty + \delta)^2. \tag{99}$$

We can now pull these results together to prove Theorem 4.

If $\|\hat{\Phi}_{LS} - \Phi_\infty\| \leq \tau$ with some probability, then $\|\hat{\Phi}_{CP1} - \Phi\|_\infty \leq 2\tau$ with at least that probability by Lemma 2. Then, by Lemma 4

$$\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \leq 2r(2\tau + \delta)^2 \tag{100}$$

with yet greater probability.

Using Equation (84), and setting the right-hand-side equal to η, we get $\tau = \sqrt{\frac{8\ln(d^2/\eta)}{3Ng(k)}}$. We thus obtain, with probability at least $1 - \eta$:

$$\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \leq 2r\left(\delta + 2\sqrt{\frac{8\ln(d^2/\eta)}{3Ng(k)}}\right)^2, \tag{101}$$

where, to remind the reader:

$$g(k) = \begin{cases} \frac{1}{3^{2k}} & \text{Scenarios 3.1, 3.2} \\ \frac{1}{2}\frac{1}{2^{2k}} & \text{Scenario 3.3} \\ \frac{1}{4}\frac{1}{2^{2k}} & \text{Scenario 3.4} \end{cases} \tag{102}$$

Recall that Equation (84) is valid if $\tau \in [0, 1]$. But if τ is more than one, then the right-hand side of Equation 90 is more than two, hence always true. The expression is thus always valid.

For the norm-one bound we use inequality (103) below. Let ρ and $\rho_{(r)}$ be states, and assume that ρ is rank r. We write $\varphi = \rho_{(r)} - \rho$ and consider its decomposition into positive and negative parts $\varphi = \varphi_+ - \varphi_-$ Since $\rho_{(r)}$ has rank r, the positive part φ_+ has at most rank r. Moreover, since $\text{Tr}(\varphi) = 0$, we have that

$$\|\rho_{(r)} - \rho\|_1 = 2\text{Tr}(\varphi_+) \leq 2\sqrt{r}\|\varphi_+\|_2 \tag{103}$$
$$\leq 2\sqrt{r}\|\rho_{(r)} - \rho\|_2. \tag{104}$$

With this inequality, we may write:

$$\|\hat{\Phi}_{PLS} - \Phi\|_1$$
$$\leq \|\hat{\Phi}_{PLS} - \Phi^r\|_1 + \|\Phi - \Phi^r\|_1$$
$$\leq 2\sqrt{r}\left(\|\hat{\Phi}_{PLS} - \Phi^r\|_2\right) + \|\Phi - \Phi^r\|_1$$
$$\leq 2\sqrt{r}\left(\|\hat{\Phi}_{PLS} - \Phi\|_2 + \|\Phi - \Phi^r\|_2\right) + \|\Phi - \Phi^r\|_1.$$

Now, if Φ^r was the approximation to the δ-almost rank r true state Φ, then $\|\Phi - \Phi^r\|_1 \leq 2r\delta$ and $\|\Phi - \Phi^r\|_2 \leq \sqrt{2r}\delta$. If Φ^r was the approximation to the δ-almost rank r CP projection $\hat{\Phi}_{CP1}$, then $\|\Phi - \Phi^r\|_1 \leq 2r(\delta + \|\Phi - \hat{\Phi}_{CP1}\|_\infty)$ and $\|\Phi - \Phi^r\|_2 \leq \sqrt{2r}(\delta + \|\Phi - \hat{\Phi}_{CP1}\|_\infty)$.

Substituting with the upper bound τ and using Inequality (103) ends the proof. $\square$

12 Error bounds and sampling complexities for the direct projection method

Theorem 5. *The squared Frobenius norm errors of the estimators $\hat{\Phi}_{PLS}$ of a Choi matrix Φ representing a k-qubit channel, found by projecting the LS estimators scenarios 3.1, 3.2, 3.3 and 3.4, directly onto $\mathcal{CPTP}$ satisfy the following bound:*

$$\Pr[\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \geq \epsilon] \leq d^2 \exp\left(-\frac{3N\epsilon}{8}f(k)\right). \tag{105}$$

where:

$$f(k) = \begin{cases} \frac{1}{3^{2k} \times 2^{2k}} & \text{Scenarios 3.1, 3.2} \\ \frac{1}{2}\frac{1}{2^{4k}} & \text{Scenario 3.3} \\ \frac{1}{4}\frac{1}{2^{4k}} & \text{Scenario 3.4} \end{cases} \tag{106}$$

for $\epsilon \in [0, 1]$.

Proof. Using Lemma 1, one can take the bound defined in (85) and simply apply it to $\hat{\Phi}_{PLS}$ to arrive at equation (105). $\qquad\square$

Inverting Theorem 5 we prove following:

Corollary 3. *To achieve the following accuracy:*

$$\Pr[\|\hat{\Phi}_{PLS} - \Phi\|_2^2 \geq \epsilon] \leq \eta \tag{107}$$

one requires a sample size of

$$N \geq \frac{1}{f(k)}\frac{8}{3\epsilon}\log(\frac{d^2}{\eta}). \tag{108}$$

The sampling complexities for each method of data collection, with direct projection, are then:

$$N(k) = \begin{cases} \mathcal{O}(\frac{1}{\epsilon}3^{2k} \times 2^{2k}) \approx \mathcal{O}(\frac{1}{\epsilon}d^{5.17}) & \text{Scenarios 1/2} \\ \mathcal{O}(\frac{1}{\epsilon}d^4) & \text{Scenarios 3/4} \end{cases} \tag{109}$$

up to logarithmic factors.

Equation (16) is to be contrasted with (105). Clearly, the bound (16) is stricter than that of equation (105) when $r < \frac{d^2}{32}$. Therefore, the first method of projecting $\hat{\Phi}_{CP1}$ onto $\mathcal{CPTP}$ improves upon the error bound when the channel acts on more than 5 qubits, and for certain ranks in states of 3 to 5 qubits, while for 1 and 2 qubits the direct projection bound is tighter. Most importantly, the bound for the two steps projection method captures the qualitative dependence on the rank of the channels, which is confirmed by the numerical simulations described in Section 6.5.

13 Expressions for Direct Projections

13.1 Solution to the TP Projection

The projection, with respect to the Frobenius norm, of a matrix X onto the set of matrices representing trace-preserving maps is the solution to the following optimisation problem:

$$\text{Proj}_{\mathcal{TP}}[X] = \underset{X'}{\text{argmin}}\|X - X'\|_2^2$$
$$\text{s.t. } \text{Tr}_s(X') = \frac{1}{d}\mathbb{1}_d \tag{110}$$

.

Proposition 11 (Closed-form expression for TP-projection)**.** *The unique solution to the optimization problem (110) is*

$$Proj_{\mathcal{TP}}[X] = X + \frac{1}{d}\mathbb{1} \otimes \left(\frac{1}{d}\mathbb{1} - \text{tr}_s(X)\right).$$

Proof. Let us start by defining the projection onto the linear subspace of trace annihilating maps $\mathcal{TA} = \{X : \operatorname{tr}_s(X) = 0\}$:

$$\operatorname{Proj}_{\mathcal{TA}}[X] = X - \tfrac{1}{d}\mathbb{1} \otimes \operatorname{tr}_s(X).$$

It is easy to check that this operator defines a projection with respect to the Frobenius norm $(\operatorname{Proj}_{\mathcal{TA}}[\operatorname{Proj}_{\mathcal{TA}}[X]] = \operatorname{Proj}_{\mathcal{TA}}[X]$ and $\operatorname{tr}(\operatorname{Proj}_{\mathcal{TA}}[X]Y) = \operatorname{tr}(X\operatorname{Proj}_{\mathcal{TA}}[Y])$ for all matrices X, Y with compatible dimension) and we denote its ortho-complement by

$$\operatorname{Proj}_{\mathcal{TA}}^{\perp}[X] = X - \operatorname{Proj}_{\mathcal{TA}}[X] = \tfrac{1}{d}\mathbb{1} \otimes \operatorname{tr}_s(X).$$

For the matrix X we introduce the short-hand notation $X_{\mathcal{TA}} = \operatorname{Proj}_{\mathcal{TA}}[X]$ and $X_{\mathcal{TA}}^{\perp} = \operatorname{Proj}_{\mathcal{TA}}^{\perp}[X]$. The following identity is essentially the Pythagorean theorem:

$$\|X\|_2^2 = \|X_{\mathcal{TA}} + X_{\mathcal{TA}}^{\perp}\|_2^2 = \|X_{\mathcal{TA}}\|_2^2 + \|X_{\mathcal{TA}}^{\perp}\|_2^2. \tag{111}$$

Applying it to $X - X'$ with $X' \in \mathcal{TA}$ ($\operatorname{tr}_s(X') = \tfrac{1}{d}\mathbb{1}$) yields

$$\begin{aligned}
\|X - X'\|_2^2 &= \|(X - X')_{\mathcal{TA}}\|_2^2 + \|(A - Y)_{\mathcal{TA}}^{\perp}\|_2^2 \\
&= \left\|X + \tfrac{1}{d}\mathbb{1} \otimes \left(\tfrac{1}{d}\mathbb{1} - \operatorname{tr}_s(X)\right) - X'\right\|_2^2 \\
&\quad + \left\|\tfrac{1}{d}\mathbb{1} \otimes \left(\tfrac{1}{d}\mathbb{1} - \operatorname{tr}_s(X)\right)\right\|_2^2.
\end{aligned}$$

The second term is independent from the actual choice of X'. The first term is minimal for

$$X'_{\sharp} = X + \tfrac{1}{d}\mathbb{1} \otimes \left(\tfrac{1}{d}\mathbb{1} - \operatorname{tr}_1(X)\right) \in \mathcal{TP}.$$

$\square$

The argument above conveys another interesting piece of information. Projecting onto the set of trace-preserving maps decreases the Frobenius distance. Suppose that $Y \in \mathcal{TP}$ and X is arbitrary. Then,

$$\|\operatorname{Proj}_{\mathcal{TP}}[X] - Y\|_2^2 = \|X - Y\|_2^2 - \tfrac{1}{d}\left\|\tfrac{1}{d}\mathbb{1} - \operatorname{tr}_s(X)\right\|_2^2.$$

The right hand side is equal to $\|X - Y\|_2^2$ if and only if X is also trace preserving. Otherwise it is strictly smaller.

13.2 Solution to the CP Projection

The projection of a matrix X onto the set of positive-semidefinite matrices is the solution to the following optimisation problem:

$$\begin{aligned}
\operatorname{Proj}_{\mathcal{CP}}[X] = \operatorname*{argmin}_{X'} &\|X - X'\|_2^2 \\
\text{s.t. } &X' \text{ positive semi-definite}
\end{aligned} \tag{112}$$

The condition of positive semidefiniteness is that all eigenvalues be greater than or equal to zero. An obvious method, therefore, for enforcing the positive semidefiniteness of a matrix is to set all negative eigenvalues to zero. This turns out to be the unique solution to equation (112) [16]. We provide a short proof for the sake of being self-contained.

Proposition 12 (Closed-form expression for PSD-projection). *Fix a self-adjoint matrix X with eigenvalue decomposition $X = \sum_i \xi_i |x_i\rangle\langle x_i|$. Then, the unique solution to optimization problem (112) is*

$$Proj_{\mathcal{CP}}[X] = X_+ = \sum_i \max\{0, \xi_i\} |x_i\rangle\langle x_i| \tag{113}$$

Proof. Decompose $X = X_+ - X_-$ such that X_+, X_- are both psd and orthogonal with respect to the Frobenius inner product $(\text{tr}(X_+ X_-) = 0)$. For X' positive semidefinite, we can expand the squared Frobenius norm difference as

$$
\begin{aligned}
\|X - X'\|_2^2 =& \|(X_+ - X') - X_-\|_2^2 \\
=& \|X_+ - X'\|_2^2 - 2\text{tr}((X_+ - X')X_-) + \|X_-\|_2^2 \\
=& \|X_+ - X'\|_2^2 + 2\text{tr}(X_- X') + \|X_-\|_2^2,
\end{aligned}
$$

where we have used orthogonality between X_+ and X_-. Note that $\text{tr}(X_- X') \geq 0$, because both X_- and X' are psd and $\|X_+ - X'\|_2^2 \geq 0$. Hence,

$$
\|X - X'\|_2^2 \geq \|X_-\|_2^2 \quad \text{for all } X' \text{ positive semidefinite.}
$$

Equality holds if and only if $X = X_+$. $\qquad\square$

This argument also conveys another interesting piece of information: Projecting onto the set of positive-semidefinite matrices decreases the squared Frobenius distance. Suppose that Y is positive semidefinite and X is arbitrary. Then,

$$
\|\text{Proj}_{\mathcal{CP}}[X] - Y\|_2^2 \leq \|X - Y\|_2^2 - \|X - \text{Proj}_{\mathcal{CP}}[X]\|_2^2
$$

and we can always compute the minimal shrinkage $\|X - \text{Proj}_{\mathcal{CP}}[X]\|_2^2$ explictly. With the projection defined as in (113), we note that the computational cost of projecting onto the the CP cone comes almost entirely from that of computing the eigen-decomposition of the estimator.

13.3 Solution to the CP1 Projection

The final projection we discuss is that on trace one positive matrices, which is required in the first step of the PLS method. The projection itself is similar to the CP projection, the difference being that the output obeys the additional constraint

$$
\text{Tr}(\,\text{Proj}_{CP1}(\hat{\Phi}_{LS})\,) \overset{!}{=} 1 \tag{114}
$$

Define $\boldsymbol{\lambda'} = \boldsymbol{\lambda} - \boldsymbol{x_0}$, where $\boldsymbol{x_0} = (x_0, \cdots, x_0)$ is a vector that will ensure unit trace. Then the solution is a simple modification of the CP projection [16]:

$$
\text{Proj}_{CP1}[\hat{\Phi}_{LS}] = \mathcal{V}\text{diag}(\boldsymbol{\lambda'}_+)\mathcal{V}^\dagger \tag{115}
$$

where $\boldsymbol{\lambda'}_+ = (\max(\lambda'_1, 0), \cdots, \max(\lambda'_{d^2}, 0))$. The parameter x_0 can be found by solving the equation:

$$
2 + \text{Tr}(\hat{\Phi}_{LS}) - d^2 x_0 + \sum_{i=1}^{d^2} |\lambda_i - x_0| = 0. \tag{116}
$$

Such constrained projection is well-known, see for instance [16, 37]. (The result appears to be have been known earlier in the optimization community but we were unable to find a reference.)

14 Additional Projection Algorithms

The algorithms HIPswitch, OneHIP and PureHIP explored experimentally in Section 6.3 are based on Algorithm 3. Here, we show explicitly their individual implementations. We rely on Algorithm 2.

First, the simplest implementation: OneHIP, in which we carry out no AP iterations, and only project onto the single, most recent, hyperplane (Algorithm 4).

Next, PureHIP, in which still do no AP iterations, and we keep a number of hyperplanes in memory (the list `w_active` in the algorithm below), and project onto their intersection (Algorithm 5).

Finally, we show HIPswitch with the condition that we do 6 iteration of AP, followed by 30 iterations of HIP (Algorithm 6)

Algorithm 4 oneHIP

1: **function** ONEHIP(Φ initial state, ϵ tolerance)
2:     **while** $\lambda_{\min}(\Phi) < -\epsilon$ **do**
3:         w_active $\leftarrow$ empty list
4:         $\Phi_{\mathcal{CP}} \leftarrow \mathtt{proj}_{\mathcal{CP}}(\Phi)$
5:         w $\leftarrow$ half-space containing $\mathcal{CP}$ orthogonal at $\Phi_{\mathcal{CP}}$ to $\Phi_{\mathcal{CP}} - \Phi$
6:         Add w as first element of w_active
7:         w_active, $\Phi \leftarrow \mathtt{HIP\_inner}(\mathtt{w\_active}, \Phi)$
8:     **return** Φ

Algorithm 5 pureHIP

1: **function** PUREHIP(Φ initial state, ϵ tolerance)
2:     **while** $\lambda_{\min}(\Phi) < -\epsilon$ **do**
3:         $\Phi_{\mathcal{CP}} \leftarrow \mathtt{proj}_{\mathcal{CP}}(\Phi)$
4:         w $\leftarrow$ half-space containing $\mathcal{CP}$ orthogonal at $\Phi_{\mathcal{CP}}$ to $\Phi_{\mathcal{CP}} - \Phi$
5:         Add w as first element of w_active
6:         w_active, $\Phi \leftarrow \mathtt{HIP\_inner}(\mathtt{w\_active}, \Phi)$
7:     **return** Φ

Algorithm 6 HIPswitch

1: **function** HIPSWITCH(Φ initial state, ϵ tolerance)
2:     SwitchCondition_toHIP: i is divisible by 6
3:     SwitchCondition_toAP: j is divisible by 30
4:     i = 1
5:     j = 1
6:     mode $\leftarrow$ AP
7:     **while** $\lambda_{\min}(\Phi) < -\epsilon$ **do**
8:         **if** mode = AP **then**
9:             $\Phi \leftarrow \mathtt{proj}_{\mathcal{CP}}(\Phi)$
10:            $\Phi \leftarrow \mathtt{proj}_{\mathcal{TP}}(\Phi)$
11:            i = i + 1
12:            **if** SwitchCondition_toHIP **then**
13:                mode $\leftarrow$ HIP
14:                w_active $\leftarrow$ empty list
15:        **if** mode = HIP **then**
16:            $\Phi_{\mathcal{CP}} \leftarrow \mathtt{proj}_{\mathcal{CP}}(\Phi)$
17:            w $\leftarrow$ half-space containing $\mathcal{CP}$ orthogonal at $\Phi_{\mathcal{CP}}$ to $\Phi_{\mathcal{CP}} - \Phi$
18:            Add w as first element of w_active
19:            w_active, $\Phi \leftarrow \mathtt{HIP\_inner}(\mathtt{w\_active}, \Phi)$
20:            j = j + 1
21:            **if** SwitchCondition_toAP **then**
22:                mode $\leftarrow$ AP
23:     **return** Φ