Sample-Optimal Quantum Process Tomography with non-adaptive Incoherent Measurements

Aadil Oufkir

Sample-Optimal Quantum Process Tomography with non-adaptive Incoherent Measurements

Aadil Oufkir
Univ Lyon, Inria, ENS Lyon, UCBL, LIP, Lyon, France
aadil.oufkir@ens-lyon.fr

***Abstract*—How many copies of a quantum process are necessary and sufficient to construct an approximate classical description of it? We extend the result of Surawy-Stepney, Kahn, Kueng, and Guta (2022) to show that $\tilde{\mathcal{O}}(d^6/\varepsilon^2)$ copies are sufficient to learn a d-dimensional quantum channel to within ε in the diamond norm. Moreover, we show that $\Omega(d^6/\varepsilon^2)$ copies are necessary for any strategy using incoherent non-adaptive measurements. This lower bound applies even for ancilla-assisted strategies.**

I. Introduction

We consider the problem of quantum process tomography which consists of approximating an arbitrary quantum channel–any linear map that preserves the axioms of quantum mechanics. This task is an important tool in quantum information processing and quantum control which has been performed in actual experiments (see e.g. [1], [2], [3]). Given a quantum channel $\mathcal{N} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ as a black box, a learner could choose the input state and send it through the unknown quantum channel. Then, it can only extract classical information by performing a measurement on the output state. It repeats this procedure at different steps. After collecting a sufficient amount of classical data, the goal is to return a quantum channel $\tilde{\mathcal{N}}$ satisfying:

$$\forall \rho \in \mathbb{C}^{d\times d} \otimes \mathbb{C}^{d\times d} : \|\mathrm{id} \otimes (\mathcal{N} - \tilde{\mathcal{N}})(\rho)\|_1 \leq \varepsilon \|\rho\|_1 \qquad (1)$$

with high probability. In this work, we investigate the optimal complexity of non-adaptive strategies using incoherent measurements. These strategies can only use one copy of the unknown channel each step and must specify the input states and measurement devices before starting the learning procedure.

Contribution First, we show that the process tomography algorithm of [4] can be used to approximate an unknown quantum channel to within ε in the diamond norm (1) using a number of incoherent measurements $\tilde{\mathcal{O}}(d^6/\varepsilon^2)$. For this, we only need to relate the diamond norm between two quantum channels and the operator norm between their corresponding Choi states which improves on the usual inequality with the 1-norm: $\|\mathcal{M}\|_\diamond \leq d\|\mathcal{J}_\mathcal{M}\|_1$ (see e.g. [5]). Next, we prove that this bound is almost optimal for non-adaptive algorithms. Precisely, we prove a general lower bound of $\Omega(d^6/\varepsilon^2)$ on the number of incoherent measurements for every non-adaptive process learning algorithm. For this, we construct an $\Omega(\varepsilon)$-separated family of quantum channels close to the completely depolarizing channel of cardinal $M = \exp(\Omega(d^4))$ by choosing random Choi states of a specific form. This family is used to encode a message from $[M]$. A process tomography algorithm can be used to decode this message with the same error probability. Hence, the encoder and decoder should share at least $\Omega(d^4)$ nats of information. On the other hand, we show that the correlation between the encoder and decoder can only increase by at most $\mathcal{O}(\varepsilon^2/d^2)$ nats after each measurement. Note that the naive upper bound on this correlation is $\mathcal{O}(\varepsilon^2)$, we obtain an improvement by a factor d^2 by exploiting the randomness in the construction of the quantum channel. This result is stated in Theorem IV.1.

Related work The first works on process tomography including [6] and [7] follow the strategy of learning the quantum states images of a complete set of basis states then obtaining the quantum channel by an inversion. The problem of state tomography using incoherent measurements is fully understood even for adaptive strategies [8]–[11]: the optimal complexity is $\Theta(d^3/\varepsilon^2)$. So, learning a quantum channel can be done using $\mathcal{O}(d^5)$ measurements, but this complexity doesn't take into account the accumulation of errors. The same drawback can be seen in the resource analysis of different strategies by [12]. Another reductive approach is to use the Choi–Jamiołkowski isomorphism [13], [14] to reduce the process tomography to state tomography with a higher dimension [15], [16]. However, this requires an ancilla and only implies a sub-optimal upper bound $\mathcal{O}(d^8/\varepsilon^2)$ for learning in the diamond norm. [4] propose an algorithm for estimating the Choi state in the two norm that requires only $\tilde{\mathcal{O}}(d^4/\varepsilon^2)$ ancilla-free incoherent measurements. This article extends this result to the diamond norm and shows that this algorithm is optimal up to a logarithmic factor. Finally, ancilla-assisted quantum process tomography (which can be related to quantum imaging [17], [18]) was proven to provide an advantage over ancilla-free strategies for some problems [19], [20]. However, in this work, we show that ancilla-assisted strategies cannot overcome ancilla-free strategies for process tomography.

II. Preliminaries

We consider quantum channels acting on a system of dimension d. We use the notation $[d] := \{1, \ldots, d\}$. We adopt the bra-ket notation: a column vector is denoted $|\phi\rangle$ and its adjoint is denoted $\langle\phi| = |\phi\rangle^\dagger$. With this notation, $\langle\phi|\psi\rangle$ is the dot product of the vectors ϕ and ψ and, for a unit vector $|\phi\rangle \in \mathrm{S}^d$, $|\phi\rangle\langle\phi|$ is the rank-1 projector on the

space spanned by the vector ϕ. The canonical basis $\{e_i\}_{i\in[d]}$ is denoted $\{|i\rangle\}_{i\in[d]} := \{|e_i\rangle\}_{i\in[d]}$. A quantum state is a positive semi-definite Hermitian matrix of trace 1. A d-dimensional quantum channel is a map $\mathcal{N} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ of the form $\mathcal{N}(\rho) = \sum_k A_k \rho A_k^\dagger$ where the Kraus operators $\{A_k\}_k$ satisfy $\sum_k A_k^\dagger A_k = \mathbb{I}$. For instance, the identity map $\mathrm{id}(\rho) = \rho$ admits the Kraus operator $\{\mathbb{I}\}$ and the completely depolarizing channel $\mathcal{D}(\rho) = \mathrm{tr}(\rho)\frac{\mathbb{I}}{d}$ admits the Kraus operators $\left\{\frac{1}{\sqrt{d}}|i\rangle\langle j|\right\}_{i,j\in[d]}$. A map $\mathcal{N}$ is a quantum channel if, and only if, it is:

- **completely positive:** for all $\rho \succcurlyeq 0$, $\mathrm{id}\otimes\mathcal{N}(\rho) \succcurlyeq 0$ and
- **trace preserving:** for all ρ, $\mathrm{tr}(\mathcal{N}(\rho)) = \mathrm{tr}(\rho)$.

We define the diamond distance between two quantum channels $\mathcal{N}$ and $\mathcal{M}$ as the diamond norm of their difference:

$$d_\diamond(\mathcal{N},\mathcal{M}) := \max_\rho \|\mathrm{id}\otimes(\mathcal{N}-\mathcal{M})(\rho)\|_1$$

where the maximization is over quantum states and the Schatten p-norm of a matrix M is defined as $\|M\|_p^p = \mathrm{tr}\left(\sqrt{M^\dagger M}^p\right)$. The diamond distance can thought of as a worst case distance, while the average case distance is given by the Hilbert-Schmidt or Schatten 2-norm between the corresponding Choi states. We define the Choi state of the channel $\mathcal{N}$ as $\mathcal{J}_\mathcal{N} = \mathrm{id}\otimes\mathcal{N}(|\Psi\rangle\langle\Psi|)$ where $|\Psi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |ii\rangle$ is the maximally entangled state. However, to have comparable distances, we will normalize the 2-norm which is equivalent to unnormalize the maximally entangled state and we define the 2-distance as follows:

$$d_2(\mathcal{N},\mathcal{M}) := d\|\mathcal{J}_\mathcal{N} - \mathcal{J}_\mathcal{M}\|_2 = \|\mathrm{id}\otimes(\mathcal{N}-\mathcal{M})(d|\Psi\rangle\langle\Psi|)\|_2.$$

This is a valid distance since the map $\mathcal{J} : \mathcal{N} \mapsto \mathrm{id}\otimes\mathcal{N}(|\Psi\rangle\langle\Psi|)$ is an isomorphism called the Choi–Jamiołkowski isomorphism [13], [14]. Note that $\mathcal{J}$ should be positive semi-definite and satisfy $\mathrm{tr}_2(\mathcal{J}) = \frac{\mathbb{I}}{d}$ to be a valid Choi state (corresponding to a quantum channel).

We consider the channel tomography problem which consists of learning a quantum channel $\mathcal{N}$ in the diamond distance. Given a precision parameter $\varepsilon > 0$, the goal is to construct a quantum channel $\tilde{\mathcal{N}}$ satisfying with at least a probability $2/3$:

$$d_\diamond(\mathcal{N},\tilde{\mathcal{N}}) \le \varepsilon.$$

An algorithm $\mathcal{A}$ is $1/3$-correct for this problem if it outputs a quantum channel ε-close to $\mathcal{N}$ with a probability of error at most $1/3$. We choose to learn in the diamond distance because it characterizes the minimal error probability to distinguish between two quantum channels when auxiliary systems are allowed [21].

The learner can only extract classical information from the unknown quantum channel $\mathcal{N}$ by performing a measurement on the output state. Throughout the paper, we only consider unentangled or incoherent measurements. That is, the learner can only measure with a d (or $d\times d$)-dimensional measurement device. Precisely, a d-dimensional measurement is defined by a POVM (positive operator-valued measure) with a finite number of elements: this is a set of positive semi-definite matrices $\mathcal{M} = \{M_x\}_{x\in\mathcal{X}}$ acting on the Hilbert space $\mathbb{C}^d$ and satisfying $\sum_{x\in\mathcal{X}} M_x = \mathbb{I}$. Each element M_x in the POVM $\mathcal{M}$ is associated with the outcome $x \in \mathcal{X}$. The tuple $\{\mathrm{tr}(\rho M_x)\}_{x\in\mathcal{X}}$ is non-negative and sums to 1: it thus defines a probability. Born's rule [22] says that the probability that the measurement on a quantum state ρ using the POVM $\mathcal{M}$ will output x is exactly $\mathrm{tr}(\rho M_x)$. Depending on whether an auxiliary system is allowed to be used, we distinguish two types of strategies.

a) Ancilla-free strategies: At each step t, the learner would choose an input d-dimensional state $\rho_t \in \mathbb{C}^{d\times d}$ and a d-dimensional measurement device $\mathcal{M}_t = \{M_x^t\}_{x\in\mathcal{X}_t} \in (\mathbb{C}^{d\times d})^{\mathcal{X}_t}$. It thus sees the outcome $x_t \in \mathcal{X}_t$ with a probability $\mathrm{tr}(\mathcal{N}(\rho_t)M_{x_t}^t)$.

b) Ancilla-assisted strategies: At each step t, the learner would choose an input $d\times d$-dimensional state $\rho_t \in \mathbb{C}^{d\times d}\otimes\mathbb{C}^{d\times d}$ and a $d\times d$-dimensional measurement device $\mathcal{M}_t = \{M_x^t\}_{x\in\mathcal{X}_t} \in (\mathbb{C}^{d\times d}\otimes\mathbb{C}^{d\times d})^{\mathcal{X}_t}$. It thus sees the outcome $x_t \in \mathcal{X}_t$ with a probability $\mathrm{tr}(\mathrm{id}\otimes\mathcal{N}(\rho_t)M_{x_t}^t)$.

Moreover, we only consider non-adaptive strategies: the input states and measurement devices should be chosen before starting the learning procedure and thus cannot depend on the observations.

Given two random variables X and Y with values over the sets $[d]$ and $[d']$ respectively. The mutual information between X and Y is the Kullback Leibler divergence between the joint distribution $P_{(X,Y)}$ and the product distribution $P_X \times P_Y$:

$$\mathcal{I}(X:Y) = \sum_{i=1}^{d}\sum_{j=1}^{d'} \mathbb{P}(X=i,Y=j)\log\left(\frac{\mathbb{P}(X=i,Y=j)}{\mathbb{P}(X=i)\mathbb{P}(Y=j)}\right).$$

All the logs of this paper are taken in base e and the information is measured in "nats".

III. Upper bound

In this section, we investigate the upper bounds on the complexity of quantum process tomography problem. We focus mainly on the algorithm proposed by [4] which is ancilla-free.

Theorem III.1. *[4] There is an ancilla-free process tomography algorithm that learns a quantum channel in the distance d_2 using only a number of measurements $N = \mathcal{O}\left(\frac{d^6\log(d)}{\varepsilon^2}\right)$.*

Actually, this algorithm proceeds by providing an unbiased estimator for the Choi state $\mathcal{J}_\mathcal{N}$, then projecting this matrix to the space of Choi states (PSD and partial trace $\mathbb{I}/d$) and finally by invoking the Choi–Jamiołkowski isomorphism we obtain an approximation of the channel. This reduction from learning the Choi state in the operator norm to learning the quantum channel in the d_2 distance uses mainly the inequality $d_2(\mathcal{N},\mathcal{M}) = d\|\mathcal{J}_\mathcal{N} - \mathcal{J}_\mathcal{M}\|_2 \le d^2\|\mathcal{J}_\mathcal{N} - \mathcal{J}_\mathcal{M}\|_\infty$. We argue that this algorithm has a similar complexity for the diamond distance. For this we show the following inequality that permits the required reduction:

Lemma III.2. *Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be two quantum channels. We have:*

$$d_\diamond(\mathcal{N}_1,\mathcal{N}_2) \leq d^2 \|\mathcal{J}_{\mathcal{N}_1} - \mathcal{J}_{\mathcal{N}_2}\|_\infty.$$

Proof. Denote by $\mathcal{M} = \mathcal{N}_1 - \mathcal{N}_2$. Let $|\phi\rangle$ be a maximizing unit vector of the diamond norm, i.e., $\|\mathrm{id} \otimes \mathcal{M}(|\phi\rangle\langle\phi|)\|_1 = d_\diamond(\mathcal{N}_1,\mathcal{N}_2)$. We can write $|\phi\rangle = A \otimes \mathbb{I}|\Psi\rangle$ where $|\Psi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |ii\rangle$ is the maximally entangled state. $|\phi\rangle$ has norm 1 so $\frac{1}{d}\mathrm{tr}(A^\dagger A) = \langle\Psi|A^\dagger A \otimes \mathbb{I}|\Psi\rangle = \langle\phi|\phi\rangle = 1$. Hence,

$$\begin{aligned} d_\diamond(\mathcal{N}_1,\mathcal{N}_2) &= \|\mathrm{id} \otimes \mathcal{M}(|\phi\rangle\langle\phi|)\|_1 \\ &= \|\mathbb{I} \otimes \mathcal{M}(A \otimes \mathbb{I}|\Psi\rangle\langle\Psi|A^\dagger \otimes \mathbb{I})\|_1 \\ &= \|(A \otimes \mathbb{I})\mathbb{I} \otimes \mathcal{M}(|\Psi\rangle\langle\Psi|)(A^\dagger \otimes \mathbb{I})\|_1 \\ &= \|(A \otimes \mathbb{I})\mathcal{J}_{\mathcal{M}}(A^\dagger \otimes \mathbb{I})\|_1. \end{aligned}$$

$J_{\mathcal{M}}$ is Hermitian so can be written as : $J_{\mathcal{M}} = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$. Using the triangle inequality, we obtain:

$$\begin{aligned} &\|(A \otimes \mathbb{I})\mathcal{J}_{\mathcal{M}}(A^\dagger \otimes \mathbb{I})\|_1 \\ &= \left\|(A \otimes \mathbb{I})\sum_i \lambda_i |\psi_i\rangle\langle\psi_i|(A^\dagger \otimes \mathbb{I})\right\|_1 \\ &\leq \sum_i |\lambda_i| \|(A \otimes \mathbb{I})|\psi_i\rangle\langle\psi_i|(A^\dagger \otimes \mathbb{I})\|_1 \\ &\leq \max_i |\lambda_i| \sum_i \|(A \otimes \mathbb{I})|\psi_i\rangle\langle\psi_i|(A^\dagger \otimes \mathbb{I})\|_1 \\ &= \|J\|_\infty \sum_i \mathrm{tr}((A \otimes \mathbb{I})|\psi_i\rangle\langle\psi_i|(A^\dagger \otimes \mathbb{I})) \\ &= \|J\|_\infty \mathrm{tr}(AA^\dagger \otimes \mathbb{I}) = d^2\|J\|_\infty. \end{aligned}$$

□

This Lemma shows that the diamond distance and 2 distance satisfy the same inequality with respect to the infinity norm between the Choi states. Since the algorithm of [4] approximates first the Choi state in the infinity norm, we obtain the same upper bound for the diamond distance:

Corollary III.3. *There is an ancilla-free process tomography algorithm that learns a quantum channel in the distance $d_\diamond$ using only a number of measurements:*

$$N = \mathcal{O}\left(\frac{d^6 \log(d)}{\varepsilon^2}\right).$$

This complexity was expected for process tomography with incoherent measurements since the complexity of state tomography with incoherent measurements is $\Theta\left(\frac{d^3}{\varepsilon^2}\right)$ [8] and learning channels can be thought as learning states of dimension d^2. We believe that the $\log(d)$-factor can be removed from the upper bound in Corollary III.3 using the same algorithm of [4] and techniques of [9]. We refer to [23] for a generalization of this result (with a complete analysis) for arbitrary input and output dimensions. The natural questions that arise here: is this complexity optimal for ancilla-free strategies? Can ancilla-assisted algorithms outperform their ancilla-free counterparts for process tomography? We answer both questions in the following Section.

IV. Lower bound

In this section, we would like to investigate intrinsic limitations of learning quantum channels using incoherent measurements. To avoid redundancy, we consider ancilla-assisted strategies since they contain ancilla-free strategies: one can map every d-dimensional input state ρ to the $d \times d$-dimensional input state $\tilde{\rho} = \frac{\mathbb{I}}{d} \otimes \rho$ and every d-dimensional POVM $\mathcal{M} = \{M_x\}_{x\in\mathcal{X}}$ to the $d \times d$-dimensional POVM $\tilde{\mathcal{M}} = \{\mathbb{I} \otimes M_x\}_{x\in\mathcal{X}}$. Mainly, we prove the following theorem:

Theorem IV.1. *Let $d \geq 8$. Any non-adaptive algorithm for process tomography in diamond distance requires*

$$N = \Omega\left(\frac{d^6}{\varepsilon^2}\right)$$

incoherent measurements.

This theorem shows that the algorithm of [4] is almost optimal.

Proof. We use here the construction of the Choi state

$$\mathcal{J}_U = \frac{\mathbb{I}}{d^2} + \frac{\varepsilon}{d^2}(U + U^\dagger) - \frac{\varepsilon}{d^2}\mathrm{tr}_2(U + U^\dagger) \otimes \frac{\mathbb{I}}{d}$$

where $U \sim \mathrm{Haar}(d^2)$. $\mathcal{J}$ is Hermitian and satisfies $\mathrm{tr}_2(\mathcal{J}) = \frac{\mathbb{I}}{d}$. Moreover, $\mathcal{J} \succcurlyeq 0$ for $\varepsilon \leq 1/4$. Indeed, U is a unitary so it has an operator norm 1 thus $\|U + U^\dagger\|_\infty \leq 2$. Besides, $\|\mathrm{tr}_2(U + U^\dagger) \otimes \frac{\mathbb{I}}{d}\|_\infty \leq \frac{1}{d}\|\mathrm{tr}_2(U + U^\dagger)\|_\infty \leq \max_i \|\mathbb{I} \otimes \langle i|(U + U^\dagger)\mathbb{I} \otimes |i\rangle\|_\infty \leq 2$. we claim that:

Lemma IV.2. *We can construct an $\varepsilon/2$-separated (according to the diamond distance) family $\{\mathcal{N}_x\}_{x\in[M]}$ of cardinal $M = \exp(\Omega(d^4))$.*

Proof. It is sufficient to show that for $U, V \sim \mathrm{Haar}(d^2)$:

$$\mathbb{P}(\|\mathcal{J}_U - \mathcal{J}_V\|_1 \leq \varepsilon/2) \leq \exp\big(-\Omega(d^4)\big).$$

because, once this concentration inequality holds, we can choose our family randomly, and by the union bound, it will be $\varepsilon/2$-separated with an overwhelming probability $(1 - \exp\big(-\Omega(d^4)\big))$ using the inequality $d_\diamond(\mathcal{N}_U,\mathcal{N}_V) \geq \|\mathcal{J}_U - \mathcal{J}_V\|_1$. First, let us lower bound the expected value.

$$\begin{aligned} \mathbb{E}(\|\mathcal{J}_U - \mathcal{J}_V\|_1) \geq \frac{\varepsilon}{d^2}\mathbb{E}\big(\|U + U^\dagger - V - V^\dagger\|_1\big) \\ - \frac{\varepsilon}{d^3}\mathbb{E}\big(\|\mathrm{tr}_2(U + U^\dagger - V - V^\dagger) \otimes \mathbb{I}\|_1\big). \end{aligned}$$

On one hand, we can upper bound the second expectation using the triangle and the Cauchy Schwartz inequalities:

$$\begin{aligned} &\mathbb{E}\big(\|\mathrm{tr}_2(U + U^\dagger - V - V^\dagger) \otimes \mathbb{I}\|_1\big) \leq 4\mathbb{E}(\|\mathrm{tr}_2(U) \otimes \mathbb{I}\|_1) \\ &\leq 4d\mathbb{E}(\|\mathrm{tr}_2(U) \otimes \mathbb{I}\|_2) \leq 4d\sqrt{\mathbb{E}(\mathrm{tr}(\mathrm{tr}_2(U)\mathrm{tr}_2(U^\dagger) \otimes \mathbb{I}))} \\ &= 4d\sqrt{\mathbb{E}\left(\sum_{i,j}\sum_{k,l}\langle i| \otimes \langle k|U\mathbb{I} \otimes |k\rangle\langle l|U^\dagger|i\rangle \otimes |l\rangle\right)} \\ &= 4d\sqrt{\mathbb{E}\left(\sum_{i,j}\sum_{k,l}\frac{d\delta_{k,l}}{d^2}\right)} = 4d^2. \end{aligned}$$

On the other hand, we can lower bound the first expectation using the Hölder's inequality.

$$\mathbb{E}\big(\|U+U^\dagger-V-V^\dagger\|_1\big) \geq \sqrt{\frac{(\mathbb{E}(\mathrm{tr}(U+U^\dagger-V-V^\dagger)^2))^3}{\mathbb{E}(\mathrm{tr}(U+U^\dagger-V-V^\dagger)^4)}} \geq \sqrt{\frac{(4d^2)^3}{28d^2}} \geq \frac{3d^2}{2}.$$

Therefore:

$$\begin{aligned}&\mathbb{E}(\|\mathcal{J}_U-\mathcal{J}_V\|_1)\\ &\geq \frac{\varepsilon}{d^2}\mathbb{E}\big(\|U+U^\dagger-V-V^\dagger\|_1\big)-\frac{4\varepsilon}{d^3}\mathbb{E}(\|\mathrm{tr}_2 U\otimes\mathbb{I}\|_1)\\ &\geq \frac{3\varepsilon}{2}-\frac{4\varepsilon}{d}\geq\varepsilon \quad \text{for} \quad d\geq 8.\end{aligned}$$

Now, the function $(U,V)\mapsto\|\mathcal{J}_U-\mathcal{J}_V\|_1$ is $\frac{8\varepsilon}{d}$-Lipschitz. Indeed, we have $\|\mathrm{tr}_2(X)\otimes\mathbb{I}\|_1 \leq d\|\mathrm{tr}_2(X)\otimes\mathbb{I}\|_2 = d\sqrt{d}\|\mathrm{tr}_2(X)\|_2 \leq d^2\|X\|_2$ where the last inequality can be found in [24]. Therefore, by letting $X=U-U'$ and $Y=V-V'$ and using the triangle inequality we obtain:

$$\begin{aligned}&|\|\mathcal{J}_U-\mathcal{J}_V\|_1-\|\mathcal{J}_{U'}-\mathcal{J}_{V'}\|_1|\\ &\leq\frac{2\varepsilon}{d^2}\left[\|X\|_1+\|Y\|_1+\left\|\mathrm{tr}_2(X)\otimes\frac{\mathbb{I}}{d}\right\|_1+\left\|\mathrm{tr}_2(Y)\otimes\frac{\mathbb{I}}{d}\right\|_1\right]\\ &\leq\frac{4\varepsilon}{d}(\|U-U'\|_2+\|V-V'\|_2)\overset{(CS)}{\leq}\frac{8\varepsilon}{d}\|(U-U',V-V')\|_2\end{aligned}$$

so by the concentration inequality for Lipschitz functions [25]:

$$\begin{aligned}&\mathbb{P}(\|\mathcal{J}_U-\mathcal{J}_V\|_1\leq\varepsilon/2)\\ &\leq\mathbb{P}(\|\mathcal{J}_U-\mathcal{J}_V\|_1-\mathbb{E}(\|\mathcal{J}_U-\mathcal{J}_V\|_1)\leq-\varepsilon/2)\\ &\leq\exp\left(-\frac{d^2\varepsilon^2}{48\times 64\varepsilon^2/d^2}\right)=\exp\big(-\Omega(d^4)\big).\end{aligned}$$

□

Now we follow a standard strategy for proving lower bounds of learning problems (see e.g., [26], [27]). We use this $\varepsilon/2$-separated family of quantum channels $\{\mathcal{N}_x\}_{x\in[M]}$ (corresponding to the Choi states $\{\mathcal{J}_x\}_{x\in[M]}$ found in Lemma IV.2) to encode a uniformly random message $X\sim\mathrm{Uniform}([M])$ by the map $X\mapsto\mathcal{N}_X$. Using a learning algorithm for process tomography with precision $\varepsilon/4$ and an error probability at most $1/3$, a decoder Y can find X with the same error probability because the family $\{\mathcal{N}_x\}_{x\in[M]}$ is $\varepsilon/2$-separated. By Fano's inequality, the encoder and decoder should share at least $\Omega(\log(M))$ nats of information.

Lemma IV.3 (Fano, [28])**.**

$$\mathcal{I}(X:Y)\geq 2/3\log(M)-\log(2)\geq\Omega(d^4).$$

The remaining part of the proof is to upper bound this mutual information in terms of the number of measurements N, the dimension d and the precision parameter ε. Intuitively, the mutual information after few measurements is very small and then it increases when the number of measurements increases. To make this intuition formal, let N be the number of measurement sufficient for process tomography and let $(I_1,\ldots,I_N)$ be the observations of the learning algorithm, we apply first the data processing inequality to relate the mutual information between the encoder and the decoder with the mutual information between the uniform random variable X and the observations $(I_1,\ldots,I_N)$:

$$\mathcal{I}(X:Y)\leq\mathcal{I}(X:I_1,\ldots,I_N).$$

Now we can apply the chain rule for the mutual information:

$$\mathcal{I}(X:I_1,\ldots,I_N)=\sum_{t=1}^N\mathcal{I}(X:I_t|I_{\leq t-1})$$

where we use the notation $I_{\leq t}=(I_1,\ldots,I_t)$ and $\mathcal{I}(X:I_t|I_{\leq t-1})$ is the conditional mutual information between X and I_t given $I_{\leq t-1}$. A learning algorithm $\mathcal{A}$ would choose the input states $\{\rho_t\}_{t\in[N]}$ and measurement devices $\{\mathcal{M}_t\}_{t\in[N]}$ which can be chosen to have the form $\mathcal{M}_t=\{\mu_i^t|\phi_i^t\rangle\langle\phi_i^t|\}_{i\in\mathcal{I}_t}$ where $\mu_i^t\geq 0$ and $\langle\phi_i^t|\phi_i^t\rangle=1$ for all t,i. Using Jensen's inequality, we can prove the following inequality:

Lemma IV.4. *Let* $\eta=48N\varepsilon^2\sqrt{\frac{\log(10)}{M}}$. *Let* $\mathcal{M}=\mathcal{N}-\mathcal{D}$ *where* $\mathcal{D}(\rho)=\mathrm{tr}(\rho)\frac{\mathbb{I}}{d}$.*We have:*

$$\begin{aligned}&\mathcal{I}(X:I_1,\ldots,I_N)=\sum_{t=1}^N\mathcal{I}(X:I_t|I_{\leq t-1})\\ &\leq 3\sum_{t,i}\mu_i^t\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle\mathbb{E}\left(\frac{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{M}_U(\rho_t)|\phi_i^t\rangle}{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle}\right)^2+\eta\\ &\leq 3N\sup_{t,i}\mathbb{E}\left(\left(\frac{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{M}_U(\rho_t)|\phi_i^t\rangle}{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle}\right)^2\right)+48N\varepsilon^2\sqrt{\frac{\log(10)}{M}}.\end{aligned}$$

See [23] for a detailed proof. Actually, we only prove this statement with a probability at least $9/10$. The error probability can be absorbed in the construction above by asking the unitaries $\{U_x\}_{x\in[M]}$ not only to satisfy the separability condition, but also to satisfy the inequalities:

$$\begin{aligned}&\frac{1}{M}\sum_{t,i,x}\mu_i^t\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle\left(\frac{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{M}_x(\rho_t)|\phi_i^t\rangle}{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle}\right)^2\\ &\leq\sum_{t,i}\mu_i^t\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle\mathbb{E}_U\left(\frac{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{M}_U(\rho_t)|\phi_i^t\rangle}{\langle\phi_i^t|\mathrm{id}\otimes\mathcal{D}(\rho_t)|\phi_i^t\rangle}\right)^2\\ &+48N\varepsilon^2\sqrt{\frac{\log(10)}{M}}.\end{aligned}$$

Now fix $t\in[N], i_t\in\mathcal{I}_t$ and $|\phi\rangle=|\phi_{i_t}^t\rangle$. We can write $\rho_t=\sum_i\lambda_i|\psi_i\rangle\langle\psi_i|$. Recall that the maximally entangled state is denoted $|\Psi\rangle=\frac{1}{\sqrt{d}}\sum_{i=1}^d|ii\rangle$, every $|\psi_i\rangle$ can be written as $|\psi_i\rangle=A_i\otimes\mathbb{I}|\Psi\rangle$ such that $\mathrm{tr}(A_iA_i^\dagger)=d$ so:

$$\begin{aligned}\mathrm{id}\otimes\mathcal{D}(\rho_t)&=\sum_i\lambda_i(\mathrm{id}\otimes\mathcal{D})(A_i\otimes\mathbb{I}|\Psi\rangle\langle\Psi|A_i^\dagger\otimes\mathbb{I})\\ &=\sum_i\lambda_i(A_i\otimes\mathbb{I})\mathrm{id}\otimes\mathcal{D}(|\Psi\rangle\langle\Psi|)(A_i^\dagger\otimes\mathbb{I})\\ &=\sum_i\lambda_i(A_i\otimes\mathbb{I})\frac{\mathbb{I}}{d^2}(A_i^\dagger\otimes\mathbb{I})\\ &=\frac{\sum_i\lambda_iA_iA_i^\dagger}{d}\otimes\frac{\mathbb{I}}{d}.\end{aligned}\tag{1}$$

On the other hand, using the notation $V = U - \mathrm{tr}_2(U) \otimes \frac{\mathbb{I}}{d}$, we can write:

$$\begin{aligned}
&\mathrm{id} \otimes \mathcal{M}(\rho_t) = \sum_i \lambda_i \mathrm{id} \otimes \mathcal{M}(A_i \otimes \mathbb{I}|\Psi\rangle\langle\Psi|A_i^\dagger \otimes \mathbb{I}) \\
&= \sum_i \lambda_i (A_i \otimes \mathbb{I}) \mathrm{id} \otimes (\mathcal{N} - \mathcal{D})(|\Psi\rangle\langle\Psi|)(A_i^\dagger \otimes \mathbb{I}) \\
&= \sum_i \lambda_i (A_i \otimes \mathbb{I}) \left(\mathcal{J}_{\mathcal{N}} - \frac{\mathbb{I}}{d^2} \right) (A_i^\dagger \otimes \mathbb{I}) \\
&= \frac{\varepsilon}{d^2} \sum_i \lambda_i A_i \otimes \mathbb{I} \left(U + U^\dagger - \mathrm{tr}_2(U + U^\dagger) \otimes \frac{\mathbb{I}}{d} \right) A_i^\dagger \otimes \mathbb{I}) \\
&= \frac{\varepsilon}{d^2} \sum_i \lambda_i \Big[(A_i \otimes \mathbb{I}) V (A_i^\dagger \otimes \mathbb{I}) + (A_i \otimes \mathbb{I}) V^\dagger (A_i^\dagger \otimes \mathbb{I}) \Big].
\end{aligned}$$

Now, by Lemma IV.4, we need to control the expectation $\mathbb{E}_U \langle\phi|\mathrm{id} \otimes \mathcal{M}_U(\rho_t)|\phi\rangle^2$. First, we replace $\mathrm{id} \otimes \mathcal{M}(\rho_t)$ with the latter expression, then we apply the inequality $(x+y)^2 \leq 2x^2 + 2y^2$ to separate the terms involving U and the terms involving $\mathrm{tr}_2(U)$. The first term can be computed and bounded as follows.

$$\begin{aligned}
&\frac{4\varepsilon^2}{d^4} \mathbb{E} \left(\left| \langle\phi| \left(\sum_i \lambda_i (A_i \otimes \mathbb{I}) U (A_i^\dagger \otimes \mathbb{I}) \right) |\phi\rangle \right|^2 \right) \\
&= \frac{4\varepsilon^2}{d^4} \sum_{i,j} \frac{\lambda_i \lambda_j}{d^2} \left| \mathrm{tr} \Big(A_i^\dagger \otimes \mathbb{I}|\phi\rangle\langle\phi|A_j \otimes \mathbb{I} \Big) \right|^2 \\
&\overset{\text{(CS)}}{\leq} \frac{4\varepsilon^2}{d^4} \sum_{i,j} \frac{\lambda_i \lambda_j}{d^2} \langle\phi|A_i A_i^\dagger \otimes \mathbb{I}|\phi\rangle\langle\phi|A_j A_j^\dagger \otimes \mathbb{I}|\phi\rangle \\
&= \frac{4\varepsilon^2}{d^6} \left(\langle\phi| \sum_i \lambda_i A_i A_i^\dagger \otimes \mathbb{I}|\phi\rangle \right)^2. \qquad (2)
\end{aligned}$$

Let's move to the second term which involves the partial trace. Let $M_{ij} = (A_i^\dagger \otimes \mathbb{I})|\phi\rangle\langle\phi|(A_j \otimes \mathbb{I})$.

$$\begin{aligned}
&\frac{4\varepsilon^2}{d^4} \mathbb{E} \left(\left| \langle\phi| \sum_i \lambda_i (A_i \otimes \mathbb{I}) \left(\mathrm{tr}_2(U) \otimes \frac{\mathbb{I}}{d} \right) (A_i^\dagger \otimes \mathbb{I}) |\phi\rangle \right|^2 \right) \\
&= \frac{4\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathbb{E} \Big(\mathrm{tr} \big[(\mathrm{tr}_2(U) \otimes \mathbb{I}) M_{i,j} \big(\mathrm{tr}_2(U^\dagger) \otimes \mathbb{I} \big) M_{ij}^\dagger) \big] \Big) \\
&\overset{\text{(CS)}}{\leq} \frac{2\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathbb{E} \Big(\mathrm{tr} \big[\big(\mathrm{tr}_2(U^\dagger) \mathrm{tr}_2(U) \otimes \mathbb{I} \big) M_{ij} M_{ij}^\dagger) \big] \Big) \\
&+ \frac{2\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathbb{E} \Big(\mathrm{tr} \big[\big(\mathrm{tr}_2(U) \mathrm{tr}_2(U^\dagger) \otimes \mathbb{I} \big) M_{ij}^\dagger M_{ij}) \big] \Big) \\
&= \frac{4\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathbb{E} \Big(\mathrm{tr} \big[\big(\mathrm{tr}_2(U^\dagger) \mathrm{tr}_2(U) \big) \mathrm{tr}_2(M_{ij} M_{ij}^\dagger) \big] \Big) \\
&= \frac{4\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathrm{tr} \big[\mathrm{tr}_2(M_{ij} M_{ij}^\dagger) \big] \\
&= \frac{4\varepsilon^2}{d^6} \sum_{i,j} \lambda_i \lambda_j \mathrm{tr}(M_{ij} M_{ij}^\dagger) = \frac{4\varepsilon^2}{d^6} \left(\langle\phi| \sum_i \lambda_i A_i A_i^\dagger \otimes \mathbb{I}|\phi\rangle \right)^2 \qquad (3)
\end{aligned}$$

since $\mathbb{E}\big(\mathrm{tr}_2(U^\dagger)\mathrm{tr}_2(U)\big) = \mathbb{I}_d$. Indeed for $k, l \in [d]$:

$$\begin{aligned}
\mathbb{E}\big(\langle k|\mathrm{tr}_2(U^\dagger)\mathrm{tr}_2(U)|l\rangle\big) &= \sum_{i,x,y} \mathbb{E}\big(\langle k|\langle x|U^\dagger|i\rangle|x\rangle\langle i|\langle y|U|l\rangle|y\rangle\big) \\
&= \frac{1}{d^2} \sum_{i,x,y} \delta_{x,y}\delta_{k,l} = \delta_{k,l}.
\end{aligned}$$

Using the equality (1) and the two inequalities (2) and (3), we deduce:

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{\langle\phi|\mathrm{id} \otimes \mathcal{M}_U(\rho_t)|\phi\rangle}{\langle\phi|\mathrm{id} \otimes \mathcal{D}(\rho_t)|\phi\rangle} \right)^2 \right) \\
&\leq \frac{\frac{8\varepsilon^2}{d^6} \left(\langle\phi| \sum_i \lambda_i A_i A_i^\dagger \otimes \mathbb{I}|\phi\rangle \right)^2}{\langle\phi| \frac{\sum_i \lambda_i A_i A_i^\dagger}{d} \otimes \frac{\mathbb{I}}{d} |\phi\rangle^2} = \frac{8\varepsilon^2}{d^2}.
\end{aligned}$$

Hence by Lemma IV.4:

$$\begin{aligned}
&\mathcal{I}(X : I_1, \ldots, I_N) = \sum_{t=1}^N \mathcal{I}(X : I_t | I_{\leq t-1}) \\
&\leq 3N \sup_{t, i_t} \mathbb{E} \left(\left(\frac{\langle\phi_{i_t}^t|\mathrm{id} \otimes \mathcal{M}_U(\rho_t)|\phi_{i_t}^t\rangle}{\langle\phi_{i_t}^t|\mathrm{id} \otimes \mathcal{D}(\rho_t)|\phi_{i_t}^t\rangle} \right)^2 \right) + \eta \\
&\leq 24N \frac{\varepsilon^2}{d^2} + 48N\varepsilon^2 \sqrt{\frac{\log(10)}{M}} \leq \mathcal{O}\left(N \frac{\varepsilon^2}{d^2} \right)
\end{aligned}$$

because $M = \exp(\Omega(d^4))$. But from the data processing inequality and Lemma IV.3, $\mathcal{I}(X : I_1, \ldots, I_N) \geq \mathcal{I}(X : Y) \geq \Omega(d^4)$, we deduce that:

$$\mathcal{O}\left(N \frac{\varepsilon^2}{d^2} \right) \geq \mathcal{I}(X : I_1, \ldots, I_N) \geq \Omega(d^4).$$

Finally the lower bound yields:

$$N \geq \Omega\left(\frac{d^6}{\varepsilon^2} \right).$$

□

V. Conclusion and open questions

In this work, we find the optimal complexity of quantum process tomography using non-adaptive incoherent measurements. Furthermore, we show that ancilla-assisted strategies cannot outperform their ancilla-free counterparts contrary to Pauli channel tomography [19]. Still, many questions remain open. First, it is known that adaptive strategies have the same complexity of non-adaptive ones for state tomography [11], could adaptive strategies overcome non-adaptive ones for quantum process tomography? Secondly, can entangled strategies exploit the symmetry and show a polynomial (in d) speedup as they do for state tomography [8]? Lastly, what would be the potential improvements for simpler problems such as testing identity to a fixed quantum channel or learning the expectations of given input states and observables?

Acknowledgments

We acknowledge support from the European Research Council (ERC Grant AlgoQIP, Agreement No. 851716).

REFERENCES

[1] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. James, N. K. Langford, T. C. Ralph, and A. G. White, "Quantum process tomography of a controlled-not gate," *Physical review letters*, vol. 93, no. 8, p. 080502, 2004.

[2] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, M. Steffen *et al.*, "Quantum process tomography of a universal entangling gate implemented with josephson phase qubits," *Nature Physics*, vol. 6, no. 6, pp. 409–413, 2010.

[3] T. Yamamoto, M. Neeley, E. Lucero, R. Bialczak, J. Kelly, M. Lenander, M. Mariantoni, A. O'Connell, D. Sank, H. Wang *et al.*, "Quantum process tomography of two-qubit controlled-z and controlled-not gates using superconducting phase qubits," *Physical Review B*, vol. 82, no. 18, p. 184515, 2010.

[4] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, "Projected least-squares quantum process tomography," *Quantum*, vol. 6, p. 844, 2022.

[5] A. Jenčová and M. Plávala, "Conditions for optimal input states for discrimination of quantum channels," *Journal of Mathematical Physics*, vol. 57, no. 12, p. 122203, 2016.

[6] I. L. Chuang and M. A. Nielsen, "Prescription for experimental determination of the dynamics of a quantum black box," *Journal of Modern Optics*, vol. 44, no. 11-12, pp. 2455–2467, 1997.

[7] J. Poyatos, J. I. Cirac, and P. Zoller, "Complete characterization of a quantum process: the two-bit quantum gate," *Physical Review Letters*, vol. 78, no. 2, p. 390, 1997.

[8] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016, pp. 913–925.

[9] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp, "Fast state tomography with optimal error bounds," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 20, p. 204001, 2020.

[10] A. Lowe and A. Nayak, "Lower bounds for learning quantum states with single-copy measurements," *arXiv preprint arXiv:2207.14438*, 2022.

[11] S. Chen, B. Huang, J. Li, A. Liu, and M. Sellke, "Tight bounds for state tomography with incoherent measurements," *arXiv preprint arXiv:2206.05265*, 2022.

[12] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, "Quantum-process tomography: Resource analysis of different strategies," *Physical Review A*, vol. 77, no. 3, p. 032322, 2008.

[13] M.-D. Choi, "Completely positive linear maps on complex matrices," *Linear algebra and its applications*, vol. 10, no. 3, pp. 285–290, 1975.

[14] A. Jamiołkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators," *Reports on Mathematical Physics*, vol. 3, no. 4, pp. 275–278, 1972.

[15] D. W. C. Leung, *Towards robust quantum computation*. stanford university, 2000.

[16] G. D'Ariano and P. L. Presti, "Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation," *Physical review letters*, vol. 86, no. 19, p. 4195, 2001.

[17] M. Ghalaii, M. Afsary, S. Alipour, and A. Rezakhani, "Quantum imaging as an ancilla-assisted process tomography," *Physical Review A*, vol. 94, no. 4, p. 042102, 2016.

[18] M. Ghalaii and A. Rezakhani, "Scheme for coherent-state quantum process tomography via normally-ordered moments," *Physical Review A*, vol. 95, no. 3, p. 032336, 2017.

[19] S. Chen, S. Zhou, A. Seif, and L. Jiang, "Quantum advantages for pauli channel estimation," *Physical Review A*, vol. 105, no. 3, p. 032435, 2022.

[20] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, "Exponential separations between learning with and without quantum memory," in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2022, pp. 574–585.

[21] J. Watrous, *The theory of quantum information*. Cambridge university press, 2018.

[22] M. Born, "Zur Quantenmechanik der Stoßvorgänge," *Zeitschrift fur Physik*, vol. 37, no. 12, pp. 863–867, Dec. 1926.

[23] A. Oufkir, "Sample-optimal quantum process tomography with non-adaptive incoherent measurements," *arXiv preprint arXiv:2301.12925*, 2023.

[24] D. A. Lidar, P. Zanardi, and K. Khodjasteh, "Distance bounds on quantum dynamics," *Physical Review A*, vol. 78, no. 1, p. 012308, 2008.

[25] E. Meckes, M. Meckes *et al.*, "Spectral measures of powers of random matrices," *Electronic communications in probability*, vol. 18, 2013.

[26] 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*, vol. 14, no. 9, p. 095022, 2012.

[27] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," in *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2016, pp. 913–925.

[28] R. M. Fano, "Transmission of information: A statistical theory of communications," *American Journal of Physics*, vol. 29, no. 11, pp. 793–794, 1961.