Quantum speedup in testing causal hypotheses: Nutshell

aqasch

1 Introduction

To test hypothesis about the causal structure in a setting where we have different *prior hypothesis* of how we infer a bunch of variables are causally related to each other. **We aim to find which hypothesis is the correct one among the set of prior considerations**.

1.1 Test

To find the correct causal hypothesis one need to emphasize an allotted role of intervention and how it is important to plan carefully the interventions in order to maximize our chances to find out the right hypothesis.

1.2 Causality

For a set of variables in A and in B, we can say A **is a cause for** B if and only if **our ability to intervene** A **has a visible effect on the statistics of** B.

1.3 Caveat

Famous conclusion says **"Correlation does not imply Causation"**. So it is not enough to study the natural correlations between variables in order to establish a causal link. **Hence it is important to be able to test and probe different causal settings** as *single joint probability distribution is not enough for that*.

1.4 Recent advances in quantum

Recent advances in "causal relation" and "causal networks" to quantum theory and beyond considers the **causal variables as physical systems**. The causal relations are defined as A **is a cause for** B **if changing the parameters of the state of** A **induces a change of the state** B.

1.5 Could you please motivate me to invest my precious time in this quantum extension?

Sure! So there are foundational and practical reasons. Follow me as I say

- There are some well known quantum features such as **superposition**, **entanglement**, **nonlocality** and not-so-well known quantum features like **contextuality**. And there are some features that are involved in discovering causal relationships.

  Finding the relation between quantum and causal-relation discovering features may help us to get better insight to a theory that combine important role of causality and quantum $\rightarrow$ **QUANTUM GRAVITY**.

- It would be fascinating to **axiomatize quantum theory which has something to do with ability to distinguish between causal relationships**.

- **To identify working principles for a new quantum device**. Also, to develop a technology for quantum causality.
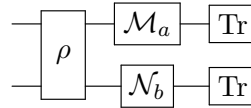
1.6 An Example

To distinguish between

- 1st hypothesis: A **causes** B
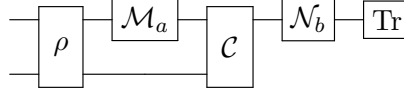
  The variables are distinguished in two parts *input* and *output*. Whatever comes from $\mathcal{M}_a$ (which is our choice of measurement on A) causes b i.e. it effects the measurement outcome of B presented through $\mathcal{N}_b$.

- 2nd hypothesis: A **and** B **have a common cause**

  It means there was a physical system prepared and the two measurements $\mathcal{M}_a$ and $\mathcal{N}_b$ are acting parallel on two different subsystems.

1.6.1 Solution

1st situation: Consider in 1st hypothesis $\mathcal{C} = \mathcal{I}$ i.e. identity. Then if $\mathcal{M}_a$ and $\mathcal{N}_b$ are the same projective measurement then we expect to observe the same outcome. And by investigating this correlation over all possible measurements we can be assured that we are in 1st hypothesis. It is not possible in 2nd hypothesis.

2nd situation: If we use the same orthogonal measurement in 2nd hypothesis then we will observe perfect anti-correlation. i.e. if $\mathcal{M}_a = 0$ then $\mathcal{N}_b = 1$. If this appears for each and every possible setting then we can conclude that A and B are of common cause.

Conclusion: For a specific ρ and $\mathcal{C}$ it is possible to distinguish between 1st and 2nd hypothesis using only *projective measurements*. In the hypothesis by scanning the two projective measurements $\mathcal{M}_a$ and $\mathcal{N}_b$ over all possible settings we can reach the above conclusion.

Classical: In projective classical measurement we really see the value of a variable and we do not change it. So in classical theory we would have no way to to distinguish two hypothesis.

1.7 Question

The type of advantage we show above is restricted to the specific kinds of allowed measurements where **classical theory is restricted to passive observational strategies** and no intervention is allowed.
Can we find the advantage in a situation where arbitrary interventions are allowed?

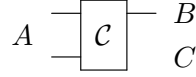## 1.8 Causal discovery vs Causal hypothesis testing

In **Causal discovery** we have variables as *input* and we get a causal relation among them as *output*.

In **Causal hypothesis testing** we have variables and a *set of hypothesis on causal relations among them* as *input* and we get correct hypothesis as *output*.
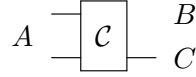
2 Causal Hypothesis

Two causal hypothesis

- A **causes** B **but not** C

$$A - \boxed{\mathcal{C}} \begin{array}{l} B \\ C \end{array}$$

- A **causes** C **but not** B

$$A - \boxed{\mathcal{C}} \begin{array}{l} B \\ C \end{array}$$

This is also called signaling vs no-signaling condition.

The causal hypothesis can be formulated independently of the underlying theory. The theory can be either classical or quantum or some fictional theory and still the question of distinguishing hypothesis makes sense.

2.1 Testing Causal Hypothesis

As an experiment one can probe the same process for a finite number of times in between performing arbitrary interventions. By considering

$$\mathcal{HP} = \quad A - \boxed{\mathcal{C}} \begin{array}{l} B \\ C \end{array}$$

the most general intervention is given by **Description of the figure below:**

Figure 1: General strategy
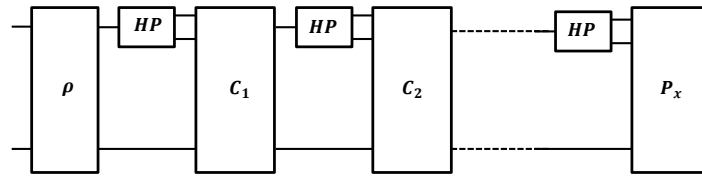
- **Till $\mathcal{C}_1$:** We prepare an input A of the first use of the process i.e. $\mathcal{C}$ and the other variables are directly fed to $\mathcal{C}_1$ then we process again the output of the $\mathcal{C}_1$ with some variables in the lab (untouched ones) at $\mathcal{C}_1$.

- **From $\mathcal{C}_1$ to $\mathcal{C}_2$:** The processed variables at $\mathcal{C}_1$ are then partially fed to the $\mathcal{C}_2$ and the remaining variables are kept in the lab (untouched).

3

- We keep continuing this till some time and at the end we do observation which is basically **a guess on the correct hypothesis** i.e. $\mathcal{P}_x$, where x defines the **correct hypothesis**. The guess x can be either 1 or 2 or any number and is decided by the number of available prior hypothesis.

This is a very general kind of intervention which has **special cases** in *process tomography*, *parallel strategies* etc.

NOTE: The only thing that would not fit in the above framework is when we take into account the ordering of actions i.e. the ordering of $\mathcal{C}$s. Under experimentalist point of view it is *ill definite*.

2.2 Discrimination Rate

An important notion when we do an experiment with *finite number of data* is **how the probability of error or success depends on the number of experiments we do** which is equivalent to **how many times we query the black box or the oracle**.

- At the end of the cascaded experiment (check the circuit [1]) we are going to get a guess on the correct hypothesis and we would like to minimize the *probability of error* i.e. probability of choosing the wrong hypothesis as much as possible.

- **As we do not know what is the causal structure i.e the relation between different variables under consideration so we are going to take a wrost case approach**. Under this consdieration we choose **worst-case error probability**

$$\mathbf{P}_{\text{err}}(\mathbf{N}), \tag{1}$$

as a function of number of times we probe the black box i.e. $\mathbf{N}$.

- $\mathbf{N}$ is very large then the rate at which the $\mathbf{P}_{\text{err}}(\mathbf{N})$ goes to zero is the most important factor. It is easy to show that the $\mathbf{P}_{\text{err}}(\mathbf{N})$ will indeed go to zero. **This will reach to zero *exponetially* fast, as this is what happens when we do statistics over N times**. Hence

- We want to see **what is the exponential rate of the decay of the error** which in turns defined as follows

$$\frac{-\log \mathbf{P}_{\text{err}}(\mathbf{N})}{\mathbf{N}},$$

as per probability theory. In the limit $\mathbf{N} \to \infty$ we get the quantifier of the distinguishibality between the available hypothesis as follows

$$\mathbf{R} = \lim_{\mathbf{N} \to \infty} \frac{-\log \mathbf{P}_{\text{err}}(\mathbf{N})}{\mathbf{N}}. \tag{2}$$

3 Causal Intermediary Identification

To be noted here we are basically considering **complete causal intermediary**. Which means B is a causal intermediary of A if all the influences on A must propagate through B. In the spacetime picture we can explain the Fig.[2] as follows
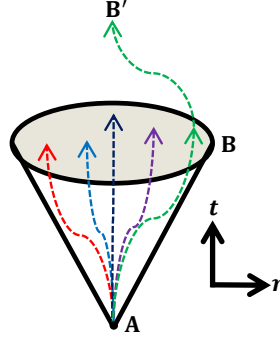
Figure 2: Causal intermediary

- **A** is the point which is localized in *spacetime*

- The *causal intermediary* **B** is *localized in time* but *captures a section in space* i.e. **A section on the *lightcone* going out from variable A**.

- **A** can influence another variable **B'** which should be mediated by **B**.

- **More formally:**

  - **B** is an effect of **A**.
  - The whole process from **A** to **B'** can be decomposed into as process as follows

$$\mathbf{A} \to \mathbf{B} \to \mathbf{B'} \tag{3}$$

3.1 Exampler Hypothesis on Intermediary

For variables A, B and C

- 1st hypothesis B is a *causal intermediary* of A, where C **fluctuates uniform randomly**

- 2nd hypothesis C is a **fluctuates uniform randomly** of A, where B *fluctuates uniform randomly*.

Find out the correct hypothesis.

3.1.1 CLASSICAL SOLUTION

Assume that **the random variables A, B and C have the same dimension d**. This assumption makes the two hypothesis very simple as follows

- 1st hypothesis b (value of variable B) is a **permutation** of a (value of variable A), where c (value of variable C) *fluctuates uniform randomly*

- 2nd hypothesis c is a **permutation** of a, where b *fluctuates uniform randomly*.

Naive strategy is to initialize the known variable lets say $A = (000 \ldots \vee 111 \ldots)$. The advantage of this initialization is that the value that is permutation of A will always have the same value as A and the value that is random at some point will deviate from the A with *high-probability*. Then we repeat the experiment **N** times. In the the exampler table [1] we see that B suppose to be the *bit-flip* of A and C is the *identity* but in the **4th row** it deviates. C **is the fake**

	1	2	3	4
A	0	0	1	1
B	1	1	0	0
C	0	0	0	0

Table 1: An example for $\mathbf{N} = 5$, $d = 2$.

effect and B **is the true one**. So if **we are unlucky** we get $\mathbf{P}_{\text{err}}(\mathbf{N}) = 1/2$. If we try v different values of A then the probability of being **unlucky**

$$\mathbf{P}_{\text{unlucky}} = \frac{\text{Number of injective functions from set of } v \text{ values to a set of } d \text{ values}}{d^{\mathcal{N}}} \tag{4}$$

$$= \frac{d(d-1)(d-2)...(d-v-1)}{d^{\mathbf{N}}}. \tag{5}$$

In the above case the best thing we can do is to test just one value i.e. $v = 1$ then we will have

$$\mathbf{P}_{\text{unlucky}} = \frac{d}{d^{\mathbf{N}}} \tag{6}$$

Discrimination rate for $v = 1$ is given by

$$\mathbf{P}_{\text{err}}(\mathbf{N}) = \frac{\mathbf{P}_{\text{unlucky}}}{2} = \frac{1}{2d^{\mathbf{N}-1}}$$
$$\mathbf{R} = \lim_{\mathbf{N} \to \infty} \frac{-\log \mathbf{P}_{\text{err}}(\mathbf{N})}{\mathbf{N}}$$
$$\mathbf{R} = \log d. \tag{7}$$

Conclusion We do not need to consider the most general strategies but **the best** we can do is to put the channels in parallel and we can prove that **asymptotically there is no classical strategy that is better than the naive classical strategy**.

3.1.2 QUANTUM SOLUTION

Once again we start with the assumption that A, B and C are of same dimension d. By considering this the two becomes

- 1st hypothesis $\mathcal{C}_{A \to BC}(\rho_A) = (U\rho U^\dagger)_B \otimes \left(\frac{I}{d}\right)_C$, U is some unknown unitary. Which can be explained as follows. The channel from A to BC should be *Unitary channel* from A to B and some *maximally mixed state on* C.

- 2nd hypothesis $\mathcal{C}_{A \to BC}(\rho_A) = \left(\frac{I}{d}\right)_B \otimes (V\rho V^\dagger)_C s$, V is some unknown unitary. Which can be explained as follows. The channel from A to BC should be *Unitary channel* from A to C and some *maximally mixed state on* B.

The naive quantum strategy If we use naive quantum strategy as shown in Fig.[3] then we get an error probability

$$\mathbf{P}_{\text{err}} = \frac{\binom{\mathbf{N}+d-1}{d-1}}{2d^{\mathbf{N}}} \tag{8}$$
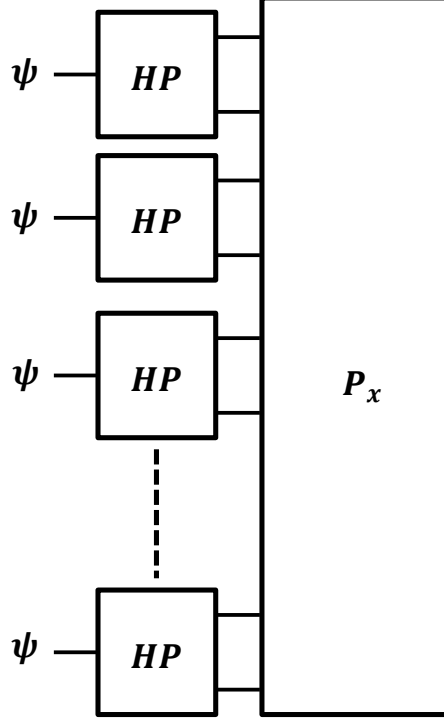
6

Figure 3: Parallel strategy.

which is **worst than the classical** but at least the **discrimination rate** is

$$\log d \tag{9}$$

Full Tomography of the Channel which is illustrated in the **Figure [4]**. We prepare a correlated system Ψ of the input variable A together with some reference system R in a fixed state, repeating it **N** times, and then measuring the output state. this also give us the same discrimination rate of

$$\log d \tag{10}$$

Parallel Strategies Previously we had the **initialization** in such a way that the input state was of the following form $\Psi \otimes \Psi \otimes \Psi \otimes \Psi \cdots \otimes \Psi$ so there was correlation inbetween the inputs. In the current strategy we will allow correlation in the initialization without a reference state as depicted in **Figure [5]** and with a reference system as depicted in **Figure [6]** The main motivation behind analysing the *without reference* and *with reference* scenerio is that **the without reference case is more general hence the solution of this setting will give us a better understanding of the outcomes in with reference settings**

Parallel Strategy without reference the **optimal** strategy is to assume $d = 2$ and **N** is even say $\mathbf{N} = 2p$ so that **we can divide the N input variables in the groups of two and then to initialize each pair in a singlet state**

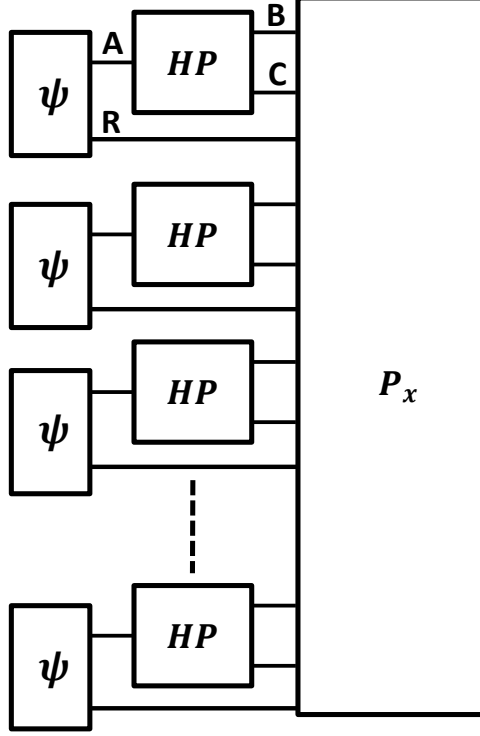$$|\Psi\rangle = \frac{|0\rangle \otimes 1\rangle - 1\rangle \otimes 0\rangle}{\sqrt{2}} \tag{11}$$

7

Figure 4: Complete tmography of the quantum channel.

The reason behind initializing in singlet is the fact that **the singlet state is invarient under Unitary changes of basis** i.e.

$$U \otimes U |\Psi\rangle = |\Psi\rangle \tag{12}$$

hence **we get rid of that we do not know the functional dependency between the cause and effect by choosing a quantum state that is entangled and by removing that dependency i.e we can test the causal structure by not knowing the fucntional dependency between cause and effect.** For a general dimention d, we divide the $\mathcal{N}$ input variables in groups of d and prepare each group in the following sinlget

$$|S_d\rangle = \frac{1}{\sqrt{d!}} \sum_{a_1,a_2,\cdots,k_d} \epsilon_{a_1,a_2,\cdots,a_d} |a_1\rangle |a_2\rangle \cdots |a_d\rangle \tag{13}$$

then we get an error probability

$$\mathbf{P}_{\mathrm{err}}(\mathbf{N}) = \frac{1}{2d^{\mathbf{N}}} \tag{14}$$

which is better than the classical $\frac{1}{2d^{\mathbf{N}-1}}$ dependence but the discrimination rate is still

$$\log d, \tag{15}$$

Parallel Strategy with reference in this case we choose $d = 2$, just like the case above, **N** is once again even. And there are many ways to partition the inputs into pairs as depicted in **Figure 7 We can associate different reference states to different partition configutation and suporpose them** which practically means we are creating an **entangled**
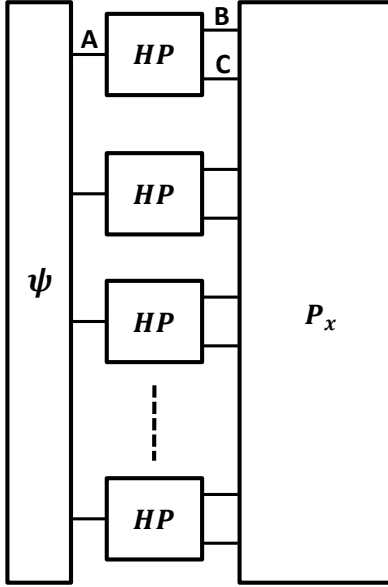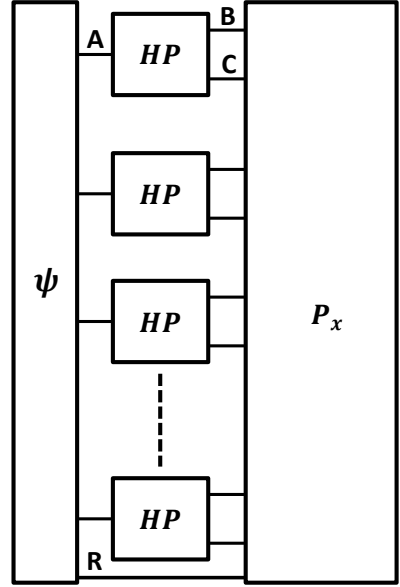
Figure 5: Without reference.
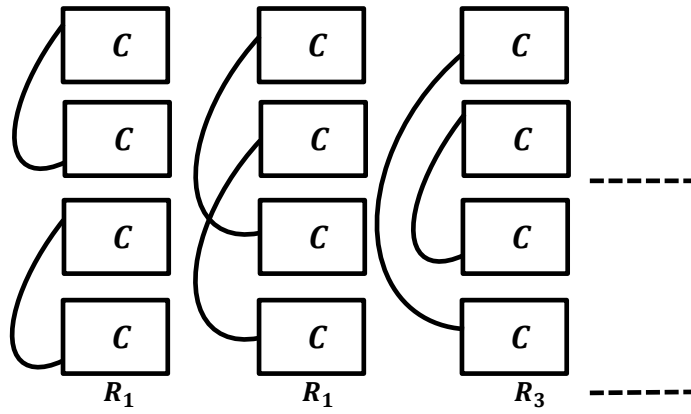

Figure 6: With reference.

Figure 7: Many ways to partition with corresponding reference system.

state where we **entangle the reference system with the partition configuration of the system** as follows

$$|\Psi\rangle = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} \left(|S_d\rangle^{\otimes \mathbf{N}/d} \right)_i \otimes |i\rangle_R. \tag{16}$$

i defines the possible ways of grouping the systems and L is the number of groupings in superposition. If there are r linearly independent groupings then the error probability becomes

$$\mathbf{P}_{\text{err}}(r) = \frac{r}{2d^{\mathcal{N}}} \left(1 - \sqrt{1 - r^{-2}} \right) \rightarrow \left[\frac{1}{4rd^{\mathbf{N}}} \right]_{r \gg 1}, \tag{17}$$

picking the maximum r we obtain the **discrimination rate**

$$\mathbf{R_Q} = -\lim_{\mathbf{N} \to \infty} \frac{\log \mathbf{P}_{\text{err}}}{\mathbf{N}} = 2\log d \tag{18}$$

4   Conclusion

The **general quantum strategy** uses the **fidelity divergence** method to get the divergence in quantum states from each other when two quantum channels representing the two hypothesis are considered. Then the **general strategy** actually upperbounds the discrimination rate that is

$$\mathbf{P}_Q^{\text{general}} \leq 2\log d \tag{19}$$

Asymptotically the best quantum discrimination rate is $2\log d$.