Note on POVM and Control

Vincent Văn Dương
vv2102@nyu.edu
New York University

1 Introduction

In this note, we provide numerical evidence that quantum sensors are only efficient in a certain window of time. In particular, we sensed an external magnetic field $\boldsymbol{B}$ coupled to the quantum Ising model. Our sensor achieves a quantum advantage within a certain window of time; outside of this window, the sensor can not extract additional information about the external field parameters.

2 Quantum Fisher Information

- Introduce the Classical Fisher Information Matrix
- Introduce the Cramer-Rao Bound
- Introduce the Quantum Fisher Information (QFI)
- Derive the Quantum Cramer-Rao bound
- Explain why QFI is impractical: it shows a bound but fails to provide a parameter estimate and fails to provide the optimal measurement.

3 Quantum Channels and the Positive Operator-Valued Measure (POVM)

3.1 Positive operator-valued measure (POVM)

We briefly review the positive operator valued measure (POVM) formalism. POVMs are operators that generalise classical events. POVMs are a set of operators $\widehat{M}_\mu$ that satisfy two conditions:

$$\begin{aligned} \sum_\mu \widehat{M}_\mu &= \mathbb{I} \\ \widehat{M}_\mu &\succcurlyeq 0. \end{aligned} \tag{1}$$

If the system's density matrix is $\hat{\rho}$, then the probability of measuring event μ is given by a trace:

$$p(\mu) = \mathrm{Tr}\left[\widehat{M}_\mu \hat{\rho}\right]. \tag{2}$$

The first condition (resolution of the identity) is necessary to satisfy the law of total probability. The second condition (semidefiniteness) ensures that all measurement probabilities are positive.

The simplest POVM is the orthogonal projectors. This reduces to undergraduate quantum mechanics. For example, $\widehat{M}_\mu = |\varphi_\mu\rangle\langle\varphi_\mu|$, where $|\varphi_\mu\rangle$ is an orthonormal basis for the Hilbert space. Both POVM conditions are satisfied. Moreover, the measurement probability reduces to the probabilities of measuring a certain state:

$$
\begin{aligned}
p(\mu) &= \mathrm{Tr}\Big[\widehat{M}_\mu \hat{\rho}\Big] = \mathrm{Tr}\big[|\varphi_\mu\rangle\langle\varphi_\mu|\hat{\rho}\big] \\
&= \sum_\nu \langle\varphi_\nu|\varphi_\mu\rangle\langle\varphi_\mu|\hat{\rho}|\varphi_\nu\rangle \\
&= \langle\varphi_\mu|\hat{\rho}\ |\varphi_\mu\rangle.
\end{aligned}
\tag{3}
$$

The critical benefit of a POVM is its ability to capture measurements that are not projectors, such as entangled measurements. Unfortunately, this means they are challenging to realise experimentally. Naimark's dilation theorem saves us from that: all POVM can be expressed as rank-1 projectors of a larger Hilbert space. All we need to do is combine the initial Hilbert space with an ancillary Hilbert space and perform rank-1 measurements on the combined space.

3.2 Completely Bounded Trace Norm

Now that we introduced POVMs, I will make contact with quantum channels. It is helpful to get a notion of distances between two channels to obtain a sense of convergence. This section will introduce the completely bounded trace norm on quantum channels.

To start, we focus on the distance measure between density matrices, and extend these ideas for quantum channels. The trace distance T measures the distance between two density matrices ρ, σ:

$$
T(\rho, \sigma) = \frac{1}{2}\mathrm{Tr}\Big[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\Big]. \tag{4}
$$

The trace distance is a quantum analogue of the total variation between two classical probability distributions. The trace distance enjoys some valuable properties, so we would like to use it for measuring the distance between quantum channels instead of density matrices.

Let's be more general and consider a quantum channel Φ induced by Kraus operators L_μ. I write the quantum channel as

$$
\Phi\big(\widehat{X}\big) = \sum_\mu L_\mu \widehat{X} L_\mu^\dagger. \tag{5}
$$

My channel is a linear operator: $\Phi : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$. The domain and codomain of Φ have bases, so I can express the quantum channel Φ as a matrix. This matrix is known as the Choi matrix. It is straightforward to write the Choi matrix:

$$
\hat{\Phi} = \sum_{i,j} \hat{E}_{ij} \otimes \Phi\big(\hat{E}_{ij}\big), \tag{6}
$$

where the matrices $\hat{E}_{ij}$ forms a basis for $\mathcal{H} \times \mathcal{H}$. (For simplicity, it is best to choose the canonical basis of matrices $\hat{E}_{ij}$ that has matrix elements $\big[\hat{E}_{ij}\big]_{ab} = \delta_{ia}\delta_{jb}$.) Choi's theorem shows $\hat{\Phi}$ is completely positive! I.e., it has non-negative eigenvalues. Therefore, I can treat $\hat{\Phi}$ analogously to a density matrix.

How can I calculate the distance d between two quantum channels Φ, Ψ? Let's use the trace distance with the matrices induced from interpreting Φ, Ψ as density matrices:

$$d(\Phi, \Psi) = \frac{1}{2}\mathrm{Tr}\left[\sqrt{\left(\hat{\Phi} - \hat{\Psi}\right)^{\dagger}\left(\hat{\Phi} - \hat{\Psi}\right)}\right]. \tag{7}$$

Finally, I can calculate the distance between two POVMs M_μ, N_ν by measuring their trace distance on their induced Choi matrix. That is,

$$d\left(M_\mu, N_\nu\right) = \frac{1}{2}\mathrm{Tr}\left[\sqrt{\left(\hat{\Phi} - \hat{\Psi}\right)^{\dagger}\left(\hat{\Phi} - \hat{\Psi}\right)}\right], \tag{8}$$

where the Choi matrices $\hat{\Phi}, \hat{\Psi}$ for each channel is induced from their linear map:

$$\begin{aligned}
\Phi(X) &= \sum_\mu M_\mu X M_\mu^\dagger \\
\Psi(X) &= \sum_\nu N_\nu X N_\nu^\dagger \\
\hat{\Phi} &= \sum_{i,j} \hat{E}_{ij} \otimes \Phi\left(\hat{E}_{ij}\right) \\
\hat{\Psi} &= \sum_{i,j} \hat{E}_{ij} \otimes \Psi\left(\hat{E}_{ij}\right).
\end{aligned} \tag{9}$$

In this section, I have shown how to measure distances between sets of POVMs, $d\left(M_\mu, N_\nu\right)$, by measuring the induced norm on their Choi matrices.

4 Quantum Parameter Estimation

- Introduce the problem of quantum parameter estimation
- Provide practical examples, such as NMR spectroscopy or qubit readout using
- Introduce the Bayesian approach to parameter estimation
- Introduce the optimisation procedure for finding the optimal parameters of a Hamiltonian
- Verify that entanglement improves parameter estimation
- Study a simple system numerically

How do we build sensors from quantum devices? The idea is to couple the quantum device to the external field that we want to sense, and then perform inference. In what ways is this different from classical physics? The measurement outcome after coupling the system to the external field is probabilistic, due to the quantum mechanics. This is commonly referred to as 'Hamiltonian Learning.'

Suppose that I have a Hamiltonian H with unknown parameters θ. How can I learn the parameters? I run an experiment that depends on the Hamiltonian to perform inference on the unknown parameters θ. Then, I measure the outcome of my experiment and update my knowledge about the parameters. This strategy is well understood and studied under the Bayesian framework.

How do I learn the parameters optimally? The optimal strategy consists of constructing an experiment that allows one to extract as much information about the parameters as possible.

I.e., One reduces the total uncertainty on the inferred parameters. This is a minimisation problem!

4.1 The Spin-1/2 system

4.2 Uncertainty and optimisation

4.3 Optimisation (Dual Problem)

4.4 Two spin-1/2 system

4.5 Spin-1/2 Tomography

5 Quantum Control

- Introduce the efficiency advantage of designing measurement devices that use the Markov Decision Process Theory
- Find a system that is exactly solvable using the theory. Maybe this is two qubits in a Magnetic Field?

test

6 Numerical strategy

The goal of numerics is to learn the POVM and estimates $\Theta = \left\{\widehat{M}_\mu, \boldsymbol{\xi}_\mu\right\}$ that minimise the cost function:

$$\begin{aligned} L(\Theta) &= \sum_\mu \int d\boldsymbol{\theta} P(\boldsymbol{\theta}) \ \mathrm{Tr}\left[\widehat{M}_\mu e^{-i\widehat{H}(\boldsymbol{\theta})t}\hat{\rho} e^{+i\widehat{H}(\boldsymbol{\theta})t}\right]\left[\boldsymbol{\xi}_\mu - \boldsymbol{\theta}\right]^2 \\ &\text{subject to} \\ &\widehat{M}_\mu \succcurlyeq 0 \\ \sum_\mu &\widehat{M}_\mu = \mathbb{I}. \end{aligned} \tag{10}$$

To minimise the cost function, we will use the typical Newton-Raphson Method, treating Θ as a one-dimensional vector parameterising all $\widehat{M}_\mu$ and $\boldsymbol{\xi}_\mu$. Expanding the loss function simplifies the problem:

$$L(\Theta) = \mathrm{Tr}\left[\sum_\mu \widehat{M}_\mu \left(\boldsymbol{\xi}_\mu^2 \hat{F}[1] - 2\boldsymbol{\xi}_\mu \cdot \hat{F}[\boldsymbol{\theta}] + \hat{F}[\boldsymbol{\theta}^2]\right)\right], \tag{11}$$

where the matrix-valued function $\hat{F}[...]$ is defined via integration

$$\hat{F}[g(\boldsymbol{\theta})] = \int d\boldsymbol{\theta} P(\boldsymbol{\theta}) e^{-i\widehat{H}(\boldsymbol{\theta})t}\hat{\rho} e^{+i\widehat{H}(\boldsymbol{\theta})t} g(\boldsymbol{\theta}). \tag{12}$$

The three matrices $\hat{F}[1], \hat{F}[\boldsymbol{\theta}], \hat{F}[\boldsymbol{\theta}^2]$ are evaluated to arbitrary accuracy via Monte Carlo integration. This is straightforward from sampling the prior:

$$
\begin{aligned}
\widehat{F}[g(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}\sim P(\boldsymbol{\theta})}\left[e^{-i\widehat{H}(\boldsymbol{\theta})t}\widehat{\rho}e^{+i\widehat{H}(\boldsymbol{\theta})t}g(\boldsymbol{\theta})\right] \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\left[e^{-i\widehat{H}(\boldsymbol{\theta}_i)t}\widehat{\rho}e^{+i\widehat{H}(\boldsymbol{\theta}_i)t}g(\boldsymbol{\theta}_i)\right], \boldsymbol{\theta}\sim P(\boldsymbol{\theta}).
\end{aligned} \tag{13}
$$

The time-evolved density matrices

$$
e^{-i\widehat{H}(\boldsymbol{\theta}_i)t}\widehat{\rho}e^{+i\widehat{H}(\boldsymbol{\theta}_i)t} \tag{14}
$$

can be calculated using a variety of techniques. The one used in this paper makes use of the singular value decomposition of $\widehat{\rho}$:

$$
\widehat{\rho} = \widehat{U}\widehat{S}\widehat{U}^{\dagger}. \tag{15}
$$

Therefore,

$$
e^{-i\widehat{H}(\boldsymbol{\theta}_i)t}\widehat{\rho}e^{+i\widehat{H}(\boldsymbol{\theta}_i)t} = \left[e^{-i\widehat{H}(\boldsymbol{\theta}_i)t}\widehat{U}\sqrt{\widehat{S}}\right]\left[e^{-i\widehat{H}(\boldsymbol{\theta}_i)t}\widehat{U}\sqrt{\widehat{S}}\right]^{\dagger}. \tag{16}
$$

In this form, the density matrix is the square of one matrix.

The loss function from Equation 10 is constrained: POVM must satisfy the two conditions. Positive semidefiniteness is simple to enforce since every positive semidefinite matrix admits a unique Cholesky decomposition:

$$
\widehat{M}_{\mu} = \widehat{L}_{\mu}\widehat{L}_{\mu}^{\dagger}, \tag{17}
$$

where $\widehat{L}_{\mu}$ is a lower triangular complex matrix. That is, we re-parameterise the problem with these lower triangular matrices. The remaining constraint is non-trivial to enforce. To enforce the constraint, we relax the loss function with a term that penalises deviations from the constraint:

$$
L(\Theta) \to L(\Theta) + \lambda\left\|\mathbb{I} - \sum_{\mu} M_{\mu}\right\|_2^2. \tag{18}
$$

The upshot is that we must increase λ during the optimisation to enforce the constraint.

We may perform Newton-Raphson minimisation since the problem is amenable to auto-differentiation. Updating the parameters allow us to find a minima:

$$
{\Theta_a}' = \Theta_a - \sum_b \left[H^{-1}\right]_{ab} D_b, \tag{19}
$$

where

$$
\begin{aligned}
D_a &= \frac{\partial L}{\partial \Theta_a} \\
H_{ab} &= \frac{\partial^2 L}{\partial \Theta_a \partial \Theta_b}.
\end{aligned} \tag{20}
$$

7 Results

In this section, we provide evidence that optimal sensing can only be performed within a certain time interval. Inside the time interval, we can reduce the total uncertainty on the external magnetic field by measuring the system with a POVM. I.e., any POVM measurement outside this window yields no added benefit.

7.1 Set-up

The first step is to provide a Hamiltonian $\widehat{H}$ that describes our sensor. The Hamiltonian $\widehat{H}$ that we considered is the Ising model coupled to an unknown magnetic field $\boldsymbol{B}$:

$$\widehat{H} = J \sum_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z + \boldsymbol{B} \cdot \hat{\boldsymbol{L}}, \tag{21}$$

where $\boldsymbol{B}$ is the unknown external field and $\hat{\boldsymbol{L}}$ is the total angular momentum operator. In our analyses, the Ising coupling J is known. (We studied whether large or small J increase or reduce sensing abailities.)

The second step is to include our prior knowledge of the field that we want to sense. The only information available about the external field $\boldsymbol{B}$ is its prior $P(\boldsymbol{B})$, which we take to be a spherically symmetric Gaussian probability distribution:

$$P(\boldsymbol{B}) = \left(2\pi\sigma^2\right)^{-\frac{3}{2}} e^{-\frac{1}{2\sigma^2}\boldsymbol{B}^2}. \tag{22}$$

In the following analyses, we fix the width of the distribution: $\sigma = 0.1$.

The third step is to intialise the state of the system $\hat{\rho}$. For simplicity, we initialise each spin in the $|-z\rangle$ state. Though this corresponds to a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$, we can (and should) consider mixed states.

The fourth step is to let the system evolve for a duration t:

$$\hat{\rho}(t) = e^{-i\widehat{H}t} \hat{\rho} e^{+i\widehat{H}t}. \tag{23}$$

In our analyses, the evolution time t is fixed. (We will find that there is an interval where sensing can be made optimal.)

The fifth, and final step is to make a judicious choice of POVM $\left\{\widehat{M}_\mu\right\}$ that provides the most information about the magnetic field $\boldsymbol{B}$. Put another way, we want to minimise the total uncertainty on the external field $\boldsymbol{B}$. This is the most challenging step because finding the optimal POVM is non-trivial. The technique used in this paper is outlined in the Numerics section.

In the following sub-sections, we chart the regions where there is a quantum advantage. These are regions where the cost (or loss) function can be reduced via a judicious choice of POVM and estimates. The regions where there is a quantum advantage take place when the evolution time t and Ising strength J take particular values – one should tune t and J to guarantee optimal performance.

7.2 $N = 2$

This is a two-body system. The system is initiated with all spins in $|-z\rangle$. The number of POVM is

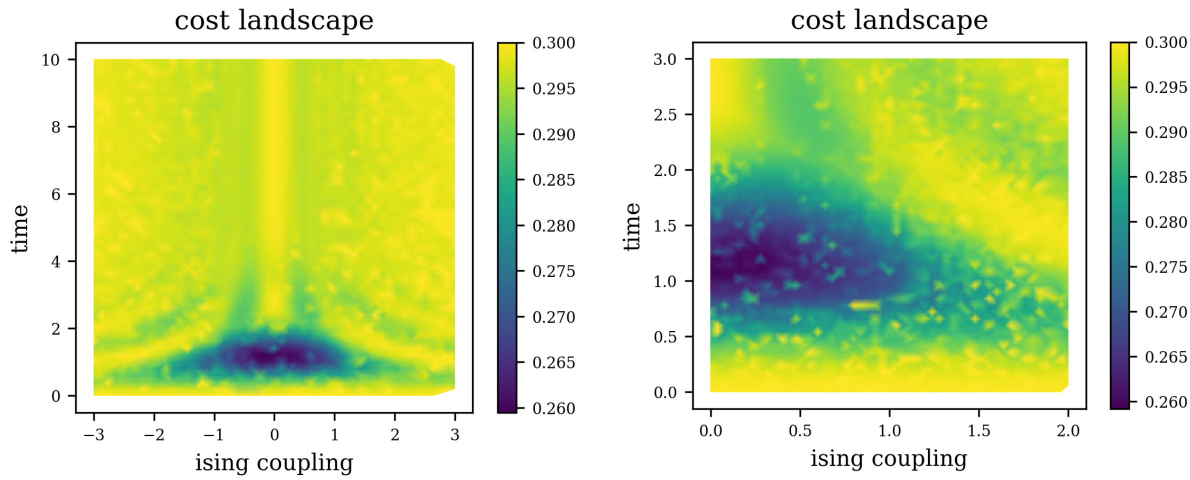

Figure 1: some caption

7.3 Three spins

In the

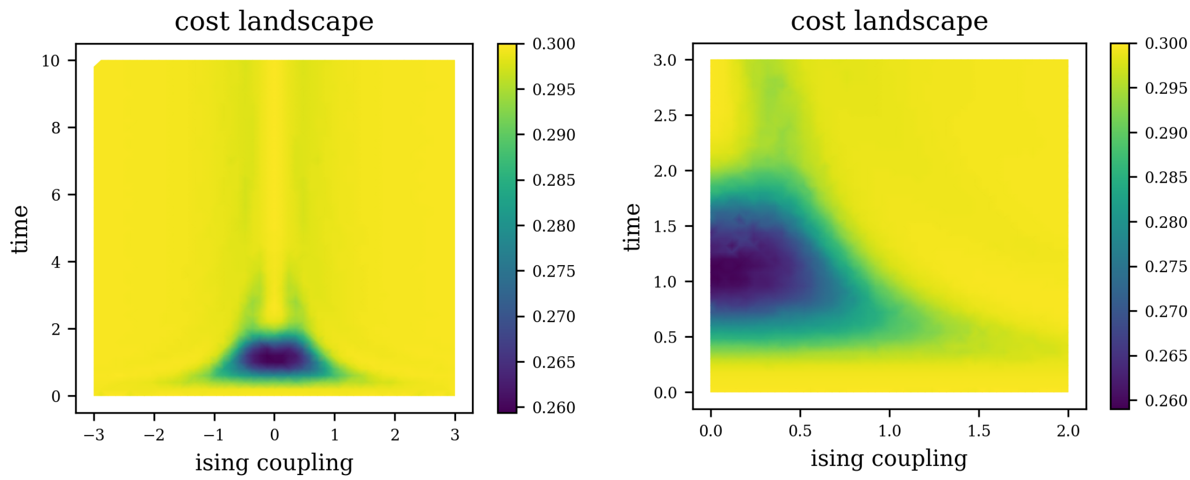

Figure 2: some caption

7.4 Four spins

In the

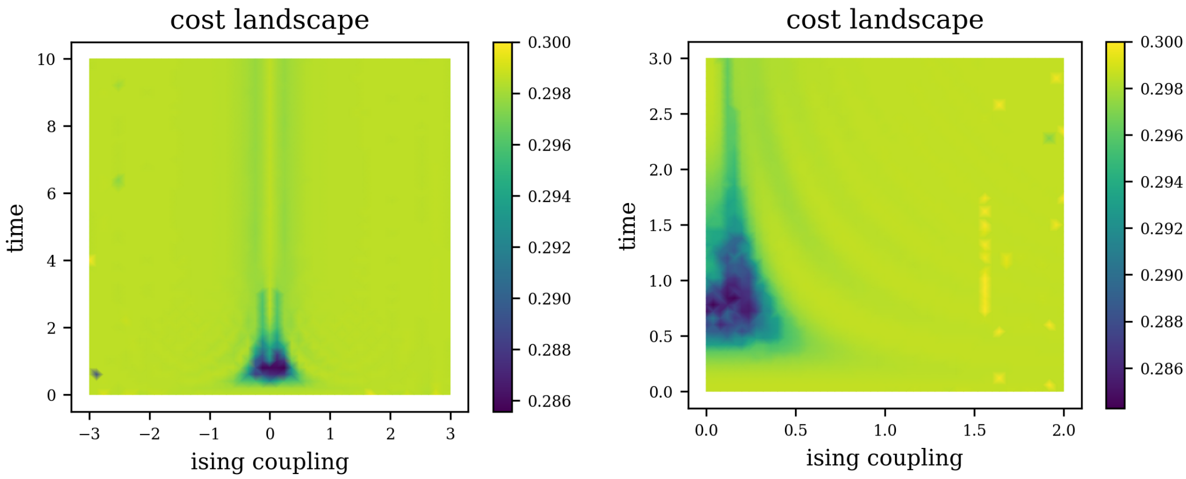

Figure 3: some caption

Bibliography