Machine learning Fisher Information Metric from bitstrings

First Last[1,*]

[1] *USC affiliation*

(Dated: August 6, 2022)

We present a machine-learning based method "Bitstring-ChiFc" which, given a dataset corresponding to a family of distributions of bitstrings parameterized by a manifold, can produce a rough approximation for the corresponding Fisher Information Metric. We observe that for multiple toy models there are often enough simple patterns in the data that this approach achieves satisfactory approximation even for dataset sizes small compared to the number of possible bitstrings.

I. PRESENTATION

A. Challenge: finding phase transitions

Let's talk about phase transitions. Consider the Hamiltonian on a $2 \times L$ lattice given by the following equations.

$$H(s, K, U) = (1 - s)H_0 + sH_1, \tag{1}$$

$$H_0 = -\sum_{i=0}^{L-1} (X_{T_i} + X_{B_i}), \tag{2}$$

$$H_1 = \sum_{i=0}^{L-1} \Big(KZ_{T_i}Z_{T_{i+1}} - KZ_{T_i}Z_{B_i} - KZ_{B_i}Z_{B_{i+1}} - KZ_{T_i} + \frac{U}{2} Z_{B_i} \Big). \tag{3}$$

Here qubits T_L and B_L are identified with T_0 and B_0 respectively.

It is called "Frustrated Ladder Hamiltonian" and is schematically represented by the following diagram:

$$\cdots T_0 \cdots T_1 \cdots T_2 \cdots T_3 \cdots T_4 \cdots T_5 \cdots T_6 \cdots T_7 \cdots T_8 \cdots T_9 \cdots$$
$$- B_0 - B_1 - B_2 - B_3 - B_4 - B_5 - B_6 - B_7 - B_8 - B_9 -$$

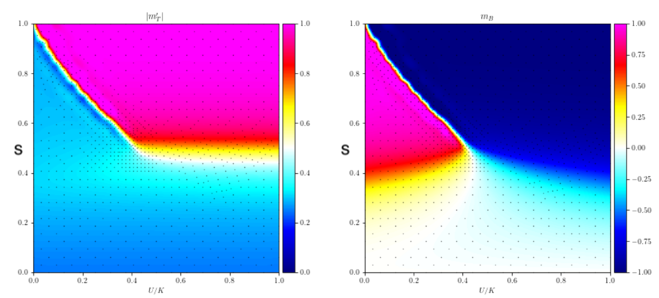
On this diagram of $L = 10$ Frustrated Ladder Hamiltonian the solid lines represent ferromagnetic couplings and dotted lines — antiferromagnetic couplings. For a fixed L the Frustrated Ladder Hamiltonian depends on 3 parameters, s, K, U. We set $K = 1$ and consider the values $s \in [0, 1]$, $U \in [0, 1]$.

How would one find phase transitions of that Hamiltonian? For that particular Hamiltonian people already know a couple of order parameters given by the following equations.

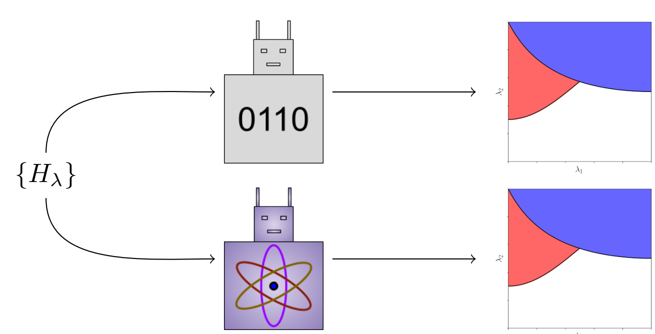
$$|m'_T| = \left| \sum_i Z_{T_i} (-1)^i \right| \tag{4}$$

$$m_B = \sum_i Z_{B_i} \tag{5}$$

These are called "staggered magnetization of the top row" and "magnetization of the bottom row" respectively. For $L = 10$ we can compute these for various values of the parameters of the Hamiltonian and produce the following diagrams:



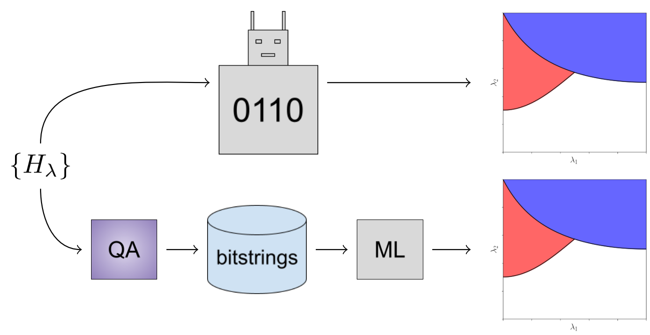
[VK: The following statement is in the direct contradiction with the statement that the main goal of this paper is to learn Fisher Information Metric given a dataset of bitstrings.] In this work we focus on the task of identifying phase transitions in a family of Hamiltonians $\{H_\lambda\}_{\lambda \in \Lambda}$ on a finite set of qubits given an access to an oracle capable, given $\lambda \in \Lambda$, of producing bitstrings measured in a computational basis from a state sampled from a low temperature distribution corresponding to the Hamiltonian H_λ. Specifically, the main goal of this paper is to make some progress towards attempting to understand whether algorithms using quantum computers can have an advantage over algorithms using the same amount of resources but running on purely classical hardware as illustrated on the following diagram.



[VK: TODO: the idea of classical and quantum "robots" was taken from some paper (probably Preskill). Find and cite that paper]

* email

More specifically, throughout this work we consider algorithms attempting to take an advantage of quantum computer having a specific structure. First, we use quantum annealer to generate a dataset of bitstrings measured in the computational basis corresponding to various values of the parameters. Then we use a classical algorithm involving machine learning to process those bitstrings into an estimates of where phase transitions are located. We also allow for an interactive version of this structure where the classical part of the algorithm can produce additional requests (values of the parameters and counts of samples requested) for the quantum annealer generating bitstrings.



There could be other algorithms for this task taking an advantage of quantum computers, but investigation of those is beyond the scope of this paper.

There are 3 main challenges which need to be discussed before we can approach specifying and solving this task.

Issue 1: fixed finite size. One may observe, that we presented diagrams for fixed $L = 10$ but wanted to discuss phase transitions which are formally only defined in the thermodynamic limit $L \to \infty$. That means, that one cannot see the actual phase transitions on these diagrams, although one can see something which looks very close to phase transitions: these are places where the color on these diagrams changes quickly. Issue 1 is how to define the task of identifying phase transitions for finite size Hamiltonians, where, strictly speaking, there are no phase transitions due to finite fixed size.

Issue 2: unknown order parameters. We want a method capable of identifying phase transitions in systems for which these are not known yet. For those systems we may not know what are the relevant order parameters. Issue 2 is how to define and determine the phase transition in the absence of relevant known order parameters.

Issue 3: loss of information at the time of measurement. ML algorithm only has access to bitstrings but, generally, bitstrings measured in the computational basis do not contain full information about the underlying quantum state.

There is a well-known approach [VK: TODO: cite] called fidelity susceptibility, which we will use to address these issues. This is a quantity which intuitively measures a squared rate of change of the underlying state. By definition fidelity between two pure states $|\phi\rangle$ and $|\psi\rangle$ is

$$F(|\phi\rangle, |\psi\rangle) = |\langle\phi \mid \psi\rangle|. \tag{6}$$

By definition fidelity between two mixed states given by density matrices ρ and σ is

$$F(\rho, \sigma) = \mathrm{Tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right). \tag{7}$$

By definition classical fidelity between two discrete probability distributions p adn q

$$F_c(p, q) = \sum_z \sqrt{p_z q_z}. \tag{8}$$

In this paper we apply classical fidelity mainly to distributions over bitstrings z obtained from measurement in the computational basis of some quantum states.

Fidelity susceptibility $\chi_F(s)$ is defined when there is a state $\rho(s)$ depending on some parameter s. In this case we write $F(s_1, s_2)$ instead of $F(\rho(s_1), \rho(s_2))$. Then term in the Taylor expansion of the fidelity:

$$F(s, s + \delta s) = 1 - \frac{\delta s^2}{2}\chi_F(s) + O(\delta s^3). \tag{9}$$

Similarly, the classical fidelity susceptibility $\chi_{F_c}(s)$ is defined by

$$F_c(s, s + \delta s) = 1 - \frac{\delta s^2}{2}\chi_{F_c}(s) + O(\delta s^3). \tag{10}$$

Issues 1 and 2 are then solved by defining the task we are trying to solve as the task of identifying the local maxima of fidelity susceptibility. To address issue 3 we look at the properties of classical fidelity susceptibility, fidelity susceptibility, adn relations between them.

B. Properties of fidelity susceptibility

Roughly speaking, we plan to prove the following properties.

- Formula (7) is consistent with (6) for pure states and with (8) for probability distributions.
- Usually, for fidelity susceptibility (or classical fidelity susceptibility) to be defined, only one derivative of wave function (or probabilities) needs to exist.
- $0 \le \chi_{F_c}(s) \le \chi_F(s)$.
- $\mathbb{E}_{\text{measurements}}\chi_{F_c}(s) = \chi_F(s)/2$.
- For non-degenerate ground states of real-valued Hamiltonians $\chi_{F_c}(s) = \chi_F(s)$ almost everywhere.

See the below theorems for the exact statements.

Theorem 1. *1. If $|\phi\rangle$ and $|\psi\rangle$ are pure states, then*

$$F(|\phi\rangle, |\psi\rangle) = F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|). \tag{11}$$

2. *If ρ and σ are diagonal matrices with diagonal entries equal to $\rho_{zz} = p_z$ and $\sigma_{zz} = q_z$ respectively, then*

$$F(\rho, \sigma) = F(p, q). \tag{12}$$

Proof. Suppose $|\phi\rangle$ and $|\psi\rangle$ are pure states. Then

$$\begin{aligned} F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) \\ = \mathrm{Tr}\left(\sqrt{\sqrt{|\psi\rangle\langle\psi|}\,|\phi\rangle\langle\phi|\sqrt{|\psi\rangle\langle\psi|}}\right) \\ = \mathrm{Tr}\left(\sqrt{|\psi\rangle\langle\psi|\phi\rangle\langle\phi|\psi\rangle\langle\psi|}\right) \\ = |\langle\psi|\phi\rangle|\,\mathrm{Tr}\left(|\psi\rangle\langle\psi|\right) = |\langle\psi|\phi\rangle| = F(|\phi\rangle, |\psi\rangle). \end{aligned} \tag{13}$$

Suppose now ρ and σ are diagonal matrices with diagonal entries equal to $\rho_{zz} = p_z$ and $\sigma_{zz} = q_z$ respectively. Then

$$F(\rho, \sigma) = \mathrm{Tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right) = \sum_z \sqrt{\sqrt{q_z}p_z\sqrt{q_z}} = \sum_z \sqrt{p_z q_z} = F(p, q). \tag{14}$$

□

Theorem 2. 1. *Suppose $|\varphi(s)\rangle$ is a state defined in the neighbourhood of $s = s_0 \in \mathbb{R}$ and differentiable at $s = s_0$. Then the fidelity susceptibility is well-defined at $s = s_0$ and is given by*

$$\chi_F(s_0) = \|\dot{\varphi}_\perp(s_0)\|^2, \tag{15}$$

where $\varphi_\perp(s)$ is the component of $|\varphi(s)\rangle$ orthogonal to $|\varphi(s_0)\rangle$ and $\dot{\bullet}$ denotes the derivative with respect to s.

Proof. Due to equivariance of the definitions with respect to translations of s, without loss of generality we can prove the statements in the theorem for $s_0 = 0$. Due to the absolute value in (6) we can multiply $|\varphi(s)\rangle$ by $e^{-i\theta}$, where $\theta = \arg(\langle\varphi(0)|\varphi(s)\rangle)$, without changing truthfullness of the conditions of the theorem or the value of the fidelity susceptibility. Now we can express

$$|\varphi(s)\rangle = (1 - \alpha(s))\,|\varphi(0)\rangle + \varphi_1(s), \tag{16}$$

where $\alpha(s)$ is real and $\varphi_1(s)$ is orthogonal to $\varphi(0)$. Also note that $\varphi_1(0) = 0$ and $\varphi_1(s)$ is differentiable at $s = 0$.

$$\begin{aligned} F(0, s) = \langle\varphi(0), \varphi(s)\rangle = 1 - \alpha(s) = \sqrt{1 - \|\varphi_1(s)\|^2} \\ = 1 - \frac{s^2}{2}\|\dot{\varphi}_1(0)\|^2 + o(s^2). \end{aligned} \tag{17}$$

Comparing (16) with (9) we see that $\chi_F(0)$ is defined and is equal to

$$\chi_F(0) = \|\dot{\varphi}_\perp(0)\|^2. \tag{18}$$

□

Theorem 3. *Suppose a pure s-dependent state in a finite-dimensional Hilbert space is described by a function $(s_0 - \epsilon, s_0 + \epsilon) \to \mathcal{H} : s \mapsto |\psi(s)\rangle$, which has a 1st derivative with respect to s at $s = s_0$. Then $\chi_F(s_0)$ is well-defined, χ_{F_c} is well-defined for any orthogonal measurement basis, and*

$$\mathbb{E}_{measurements}\chi_{F_c}(s_0) = \chi_F(s_0)/2, \tag{19}$$

where the expectation $E_{measurements}$ is taken accross all orthogonal bases in $\mathcal{H}$ using Haar measure (unique measure invariant with respect to unitary rotations). [VK: TODO: should "measurements" be italic in a formula inside a theorem?]

Proof. [VK: TODO:4] □

Theorem 4. *Suppose there is a Hamiltonian $H(s)$ acting on a finite-dimensional Hilbert space $\mathcal{H}$ with a basis $\{e_j\}_{j=1}^d$, which we will call "computational basis". [VK: TODO:5]*

- *.*
- *a non-degenerage ground state $|\psi(s)\rangle$.*

Proof. [VK: TODO:4] □

II. TODO

1. Complete presentation section above.
2. Write down the proofs for the fidelity susceptibility claims below.
3. Describe the models and practical results for them.

III. INTRODUCTION

TODO:

Such a family can arise, e.g., from measurements of a low-temperature Gibbs ensemble of Hamiltonians parametrized by a parameter λ.

IV. CLASSICAL FIDELITY SUSCEPTIBILITY

Classical fidelity between 2 probability distributions p and q of bitstrings z is defined as

$$F_c(p, q) = \sum_z \sqrt{p(z)q(z)}. \tag{20}$$

We are interested in the fidelity between bitstring distributions at different s (e.g. $s = s_1$ and $s = s_2$), which we will denote as $F_c(s_1, s_2)$.

[NE: This is an example of a commonly used in-line comment which is separated by color. I could say something like: "This sentence is awkward" or "Needs citation" or very meta "Please use enquote for real quotes and not literal quotes."]

Fidelity susceptibility is defined as the term $\chi_{F_c}(s)$ in the Taylor expansion

$$F_c(s, s+\delta s) = 1 - \frac{\delta s^2}{2}\chi_{F_c}(s) + O(\delta s^3). \tag{21}$$

For such Taylor expansion to exist it is sufficient that the probabilities have a Taylor expansion up to $O(\delta s^3)$. More generally, probability distribution can depend on a point λ on a manifold Λ, in which case the Tailor expansion (20) would become

$$F_c(\lambda, \lambda+\delta\lambda) = 1 - \frac{\delta\lambda_j \delta\lambda_k}{2}\chi^{jk}_{F_c}(\lambda) + O(\delta\lambda^3). \tag{22}$$

A. Classical and quantum fidelity susceptibility

Fact 1: For pure states $\mathbb{E}\chi_{F_c}(s) = \frac{1}{2}\chi_F(s)$ where the expectation is over all orthogonal bases to perform the measurement in.

TODO:proof

Fact 2: For computational basis measurement of a non-degenerate ground state of a real-valued Hamiltonian H, then $\chi_{F_c}(s) = \chi_F(s)$ almost everywhere.

TODO:proof

V. PROBLEM SETUP

- In this work we consider a family of distributions of bitstrings $\{\mathcal{D}_\lambda\}_{\lambda\in\Lambda}$, each of length n.
- We are given a finite sample $\mathcal{D}_{\text{train}}$ of size N of pairs (λ, z) s.t. $P(z|\lambda) = P_{\mathcal{D}_\lambda}(z)$.
- We are also given (possibly implicitly via coordinate description of Λ) a naive metric g^0 on Λ.
- We are asked to estimate the Fisher information metric g on Λ corresponding to distributions $\mathcal{D}_\lambda$.
- Locations with high g/g^0 are then considered to be conjectured locations of possible phase transitions.

We focus on the task of identifying phase transitions in that family. Rigorously speaking, phase transitions are only defined in the limit $n \to \infty$, while we are dealing with finite size systems. A solution to that is to look at Fisher information metric: high distances according to Fisher information metric for points close according to naive metric likely correspond to phase transitions.

VI. BITSTRING-CHIFC METHOD

In this work we propose the following method:

- Collect training dataset $\mathcal{D}_{\chi_{F_c}\text{-train}}$ of the form $(\lambda_0, \delta\lambda, z, y)$, where z is sampled from $p(\bullet, \lambda = \lambda_z)$, $p_+ = p(\lambda_z = \lambda_0 + \delta\lambda/2 | \lambda_z = \lambda_0 \pm \delta\lambda/2)$, and $\mathbb{E}(y|\lambda_0, \delta\lambda, z) = p_+$. In practice $y \in \{0, 1\}$. Do it in the following way:
  - Consider $\mathcal{D}_{\text{train}}$ consisting of pairs (z, λ).
  - Sample pairs (z_{i+}, λ_{i+}), (z_{i-}, λ_{i-}) from $\mathcal{D}_{\text{train}}$.
  - Compute $\lambda_i = (\lambda_{i+} + \lambda_{i-})/2$, $\delta\lambda_i = \lambda_{i+} - \lambda_{i-}$.
  - Add tuples $(\lambda_i, \delta\lambda_i, z_{i+}, 1)$ and $(\lambda_i, \delta\lambda_i, z_{i-}, 0)$ to the dataset $\mathcal{D}_{\chi_{F_c}\text{-train}}$.
- Train a model M, which given $(\lambda_0, \delta\lambda, z)$ will predict $l = M(\lambda_0, \delta\lambda, z)$ s.t. $p_+ = (1 + e^{-l\cdot\delta\lambda})^{-1}$. Do this by minimizing cross-entropy loss on the dataset $\mathcal{D}_{\chi_{F_c}\text{-train}}$.
- Estimate

$$\chi^{jk}_{F_c}(\lambda) = \text{smoothen}\Big(\lambda_1 \mapsto \text{mean}_{(z,\lambda_1)\in\mathcal{D}_{\text{train}}}\big(M(\lambda_1, 0, z)_j M(\lambda_1, 0, z)_k\big)\Big)(\lambda). \tag{23}$$

TODO: expand the explanation.

$$\chi_{F_c}(\lambda) = \lim_{\delta\lambda\to 0}\frac{2}{\delta\lambda^2}\left(1 - \mathbb{E}_{z\sim Q(\bullet)}\frac{\sqrt{P(z|\lambda-\delta\lambda/2)P(z|\lambda+\delta\lambda/2)}}{Q(z)}\right)$$
$$\simeq \lim_{\delta\lambda\to 0}\mathbb{E}_Q\frac{2}{\delta\lambda^2}\frac{2\sinh^2(l\delta\lambda/4)}{\cosh(l\delta\lambda/2)} \simeq \frac{1}{4}\mathbb{E}_{z|\lambda}M(\lambda, 0, z)^2.$$

TODO: models
TODO: experiments