Machine learning Fisher Information Metric from bitstrings

First Last[1, *]

[1] *USC affiliation*

(Dated: September 9, 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. INTRODUCTION

A. Problem we wish to solve

In this work we wish to solve a problem of estimating Fisher Information Metric (FIM) of a distribution of bitstrings depending on a parameter $\lambda \in \mathscr{M}$, given an access to the oracle producing the bitstrings from such distribution given the value of the parameter λ. We describe a method "Bitstring-ChiFc" which provides a rough estimate of FIM. We also consider a toy model called Frustrated Ising Ladder, and demonstrate that on a system with 20 qubits (small enough system size where we can implement the oracle using Lanczos method). We compare [VK: TODO:where] estimated FIM using our method applied to a dataset of 140140 bitstings with much more precise values obtained from Lanczos method and find that the estimated curve has similar features to the exact curve. In particular, our method reproduces the approximate value of the critical parameter [VK: TODO:precision?].

While our method estimates classical fidelity susceptibility (equal, up to a constant factor, to the FIM), in subsections II C and II D we discuss the differences between the classical and quantum fidelity susceptibility and conclude that there are cases of interest where looking at the classical fidelity susceptibility is sufficient.

We speculate, that under the following assumptions the method could have practical applications in determining phase transitions in systems where they are previously not known.

- For the system to be investigated, the oracle can be efficiently implemented in practice, e.g. using a quantum computer.
- Only a rough estimate of the location of phase transitions is needed.
- It is sufficient to obtain such estimate from a few finite system sizes.
- No efficient alternative exists. In particular:
  - The are known order parameters do not detect phase transitions or there is no known efficient method to estimate their values.
  - Oracle implementation above could not be adjusted to efficiently return the estimates of fidelity susceptibility instead of the bitstrings.
  - There is no efficient alternative method for estimating the fidelity susceptibility.

B. Background

[VK: I am not an expert in the history of that question. The following *seems* to be roughly correct, but needs to be confirmed with somebody.] Fidelity is a concept from quantum information science, which measures how how similar the two states are, and can be applied two pure states, mixed states, and probability distributions. As observed in [1], the leading term in the Taylor expansion of pure state fidelity with respect to parameters of the Hamiltonian can be used to study quantum phase transitions. The coefficient in that leading term is called fidelity susceptibility and is a Riemannian tensor. In [2], the approach was extended to the mixed state fidelity, which allows one to study finite temperature phase transitions.

Later, Gu et al. [3] discovered that the fidelity susceptibility approach can be used to characterise the universality class in quantum critical phenomena in 1D asymmetric Hubbard model.

C. Why people care about phase transitions

[VK: I'm not an expert in that, wrote to Vitaly Kresin to discuss; maybe he knows the people who do care about computational method for identifying the locations of phase transitions.]

Adiabatic theorems provide guarantees on the amount of time needed to solve a computational problem using quantum annealer. That guarantee diverges when the gap between the ground state and the first excited state becomes small. Such places are usually associated with quantum phase transitions.

[VK: If this is indeed the reason, should we instead focus on predicting the gap? What is a better predictor of the amount of time needed for QA, the gap or χ_F? These are related but not identical. Anyway, both the gap and χ_F are rigorously defined for finite systems as opposite to QPTs, hence if that is the reason, why would we talk about QPTs?]

According to the abstract of [4], QPTs can influence the behavior of electronic systems over a wide range of the phase

* email

diagram.

Computational methods to identify (quantum or classical) phase transitions could potentially be useful in the search of superconductors with given properties. In particular, according to the Wikipedia article on unconventional superconductors [5], the origins of high-temperature superconductivity are still unclear, and is one of the major unsolved problems of theoretical condensed matter physics.

D. Computational complexity of identifying phase transitions

[DL: Give a general introduction to the problem we wish to solve and include pertinent background with citations. In particular, we need to address why people care about QPTS, what is the computational complexity of identifying QPTs, and why we think that having a QC can be helpful in this regard.] [VK: I'm not sure how to discuss computational complexity of identifying QPTs in the introduction. E.g. a natural way to state the problem would be to restrict our attention to d-dimensional grid with Hamiltonian H, which has terms which are periodic with respect to the lattice structure. Then the goal would be to identify the value of s_{critical} up to n digits of precision. I have no idea what the computational complexity of that problem is. Note, that there is no parameter L here describing the lattice size, since, by definition, TODO:0]

II. RELATION TO PHASE TRANSITIONS

A. Challenge: finding phase transitions

[DL: Presentation of this Hamiltonian should come later, when we discuss a specific example. Or, you can use it in the introduction to motivate some of the specific questions we're attempting to answer, but then refer to it explicitly in terms of some specific aspects of the phase diagram.]

Consider the Hamiltonian on a $2 \times L$ lattice given by the following equations.

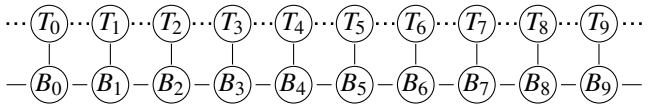
$$H(s,K,U) = (1-s)H_0 + sH_1, \tag{1}$$

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

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

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:



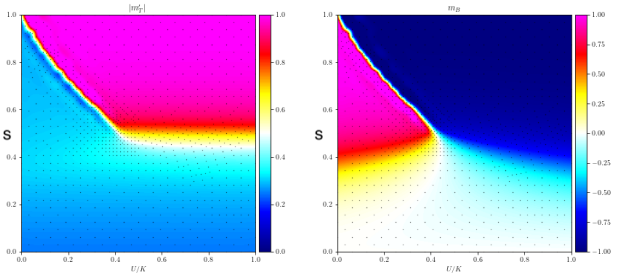
On this diagram of $L = 10$ Frustrated Ladder Hamiltonian the solid lines represent ferromagnetic couplings and dotted lines — antiferromagnetic couplings. For a fixed L the Frustrated Ladder Hamiltonian depends on 3 parameters, s, K, U. We set $K = 1$ and consider the values $s \in [0,1]$, $U \in [0,1]$.

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.

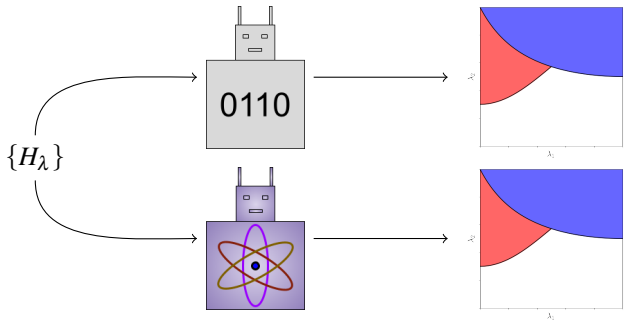
$$|m'_T| = \left| \sum_i Z_{T_i} (-1)^i \right| \tag{4}$$

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

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

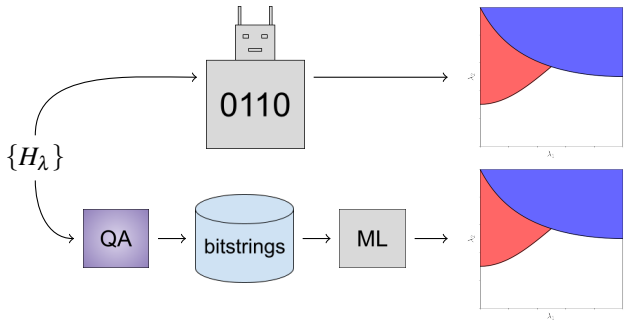


[VK: The following statement is in the direct contradiction with the statement that the main goal of this paper is to learn Fisher Information Metric given a dataset of bitstrings.] In this work we focus on the task of identifying phase transitions in a family of Hamiltonians $\{H_\lambda\}_{\lambda \in \Lambda}$ acting on a finite set of qubits, and given access to an oracle capable – given $\lambda \in \Lambda$ – of producing bitstrings measured in the computational basis from a state sampled from a low temperature distribution corresponding to the Hamiltonian H_λ. [DL: The following sentence is very generic; it describes almost all attempts to demonstrate a quantum advantage. It can be moved to the very beginning of the introduction, or even deleted altogether.] More generally, 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 in the following diagram.



[VK: TODO: the idea of classical and quantum "robots" was taken from some paper (probably Preskill). Find and cite that paper] [DL: yes: Fig.1 of https://www.science.org/doi/10.1126/science.abn7293]

More specifically, throughout this work we consider algorithms attempting to take advantage of quantum computers having a specific structure. First, we use a 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 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 queries (values of the parameters and counts of samples requested) for the quantum annealer generating bitstrings.



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

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

Issue 1: fixed finite size. [DL: This is a well known problem for physicists and goes under the name "finite size scaling". It's not at the same level as issues 2 and 3 and can probably just be mentioned without making it a full-fledged "issue".] [VK: We can discuss. I do not yet understand your comment for the following reason. The issues listed here are the issues making the description of the challenge above incomplete. In this work we are not doing finite size scaling, and, instead, compare the performance of the algorithms on systems of finite size. Hence, strictly speaking, the above challenge needs to be altered in order to match what we actually do.] 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

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

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 and 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 $F(\rho(s_1),\rho(s_2))$. The fidelity susceptibility is the lowest order non-zero term in the Taylor expansion of the fidelity:

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

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

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

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 the fidelity susceptibility, its classical counterpart, and relations between them.

B. Motivation for the definition of the task

Finite size systems in the sequence of systems experiencing a phase transition in the thermodynamic limit are known to often experience maxima of fidelity susceptibility at or around the location of the phase transition. Intuitively, that makes sense because the fidelity susceptibility measures the square rate of change of the underlying state and it is known that phase transitions correspond to rapid change of the underlying state. The study of the properties of infinite systems, including phase transitions, based on the properties of systems of finite size is called finite size scaling and described, e.g., in [6]. In this work we pursue a humble goal of studying the properties of the finite size system and do not attempt to investigate the finite size scaling.

C. Properties of fidelity susceptibility

In this section we prove the following properties:

- Usually, for the fidelity susceptibility (or classical fidelity susceptibility) to be defined, only the first derivative of the wave function (or probabilities) needs to exist.
- $0 \leq \chi_{F_c}(s) \leq \chi_F(s)$.
- $\mathbb{E}_{\text{measurements}} \chi_{F_c}(s) = \chi_F(s)/2$.
- For non-degenerate ground states of real-valued Hamiltonians $\chi_{F_c}(s) = \chi_F(s)$ almost everywhere.

The theorems below provide the exact statements. [VK: TODO: references to Zanardi papers: https://www.worldscientific.com/doi/abs/10.1142/S0217979210056335]

We first collect some well known properties of the fidelity relevant for this paper. For proofs of these, as well as many other properties of the fidelity, see [7, §9].

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

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

   In other words, Eq. (6) is consistent with Eq. (7).

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

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

   In other words, Eq. (7) is consistent with Eq. (8).

3. If ρ and σ are two density matrices, we have

$$F(\rho, \sigma) = F(\sigma, \rho) \tag{13a}$$
$$0 \leq F(\rho, \sigma) \leq 1 \tag{13b}$$
$$F(\rho, \sigma) = 1 \text{ iff } \rho = \sigma \tag{13c}$$
$$F(\rho, \sigma) = 0 \text{ iff } \rho\sigma = 0. \tag{13d}$$

Note that Eqs. (9) and (10) assume dependence on a single parameter s. In practice, the Hamiltonian of interest depends on multiple parameters; e.g., the frustrated ladder Hamiltonian in Eq. (1) depends on 3 real parameters: s, K, U. In general, a state of interest depends on a parameter set λ from a parameter manifold $\mathscr{M}$. This set can include the internal parameters of the Hamiltonian and external parameters impacting how the state is derived from that Hamiltonian (e.g., temperature). Then, similarly to Eqs. (9) and (10) we can expand the fidelity to the second order:

$$F(\lambda, \lambda + \delta\lambda) = 1 - \frac{1}{2} \sum_{\mu\nu} g_{\mu\nu} \delta\lambda^\mu \delta\lambda^\nu + o(|\delta\lambda|^2). \tag{14}$$

The resulting second term represents a metric

$$g = \sum_{\mu\nu} g_{\mu\nu} d\lambda^\mu d\lambda^\nu. \tag{15}$$

The metric $g(\lambda)$ described by Eqs. (14) and (15) is invariant with respect to the choice of the coordinates λ on $\mathscr{M}$. As explained in [1], phase transitions are expected to correspond to singularities in g in the thermodynamic limit. We, however, are interested in finite-size systems. Therefore, we are interested in identifying the vectors φ on $\mathscr{M}$ with high values of $g(\varphi, \varphi)/g_0(\varphi, \varphi)$, where g_0 is a metric considered to be non-singular. We also note that Eq. (15) is $1/4$ of Fisher information metric for the case of classical fidelity when for all z we have $p_z > 0$ [DL: need a ref for this]. Let $\partial_\mu \psi(\lambda_0) \equiv \frac{\partial \psi(\lambda)}{\partial \lambda^\mu}\Big|_{\lambda=\lambda_0}$.

Theorem 1. *1. Suppose $|\psi(\lambda)\rangle$ is a state defined in the neighbourhood of $\lambda = \lambda_0 \in \mathscr{M}$ and differentiable at $\lambda = \lambda_0$. Then the fidelity susceptibility metric is well-defined at $\lambda = \lambda_0$ and is given by*

$$g_{\mu\nu}(\lambda_0) = \operatorname{Re}\Big(\langle \partial_\mu \psi(\lambda_0) | \partial_\nu \psi(\lambda_0) \rangle\Big) - \langle \partial_\mu \psi(\lambda_0) | \psi(\lambda_0) \rangle \langle \psi(\lambda_0) | \partial_\nu \psi(\lambda_0) \rangle. \tag{16}$$

2. Suppose $\rho(\lambda)$ is a density matrix defined in the neighbourhood of $\lambda = \lambda_0 \in \mathscr{M}$, is differentiable at $\lambda = \lambda_0$, and $\operatorname{Tr}(P_0 \rho(\lambda) P_0^\dagger)$ is twice differentiable at $\lambda = \lambda_0$, where P_0 is the orthogonal projector $\mathscr{H} \to \ker \rho(\lambda_0)$. Let P_+ be the orthogonal projector $\mathscr{H} \to \rho(\lambda_0)(\mathscr{H})$, $\rho_+(\lambda) = P_+ \rho(\lambda) P_+^\dagger$. Then, in the basis where $\rho_+(\lambda_0) = \operatorname{diag}(\xi_0, \ldots, \xi_{n_+})$, we have

$$g_{\mu\nu}(\lambda_0) = \sum_{j,k} \frac{\operatorname{Re}\left((\partial_\mu \rho_+(\lambda_0))_{jk} (\partial_\nu \rho_+(\lambda_0))_{kj} \right)}{2(\xi_j + \xi_k)} + \frac{1}{2} \partial_\mu \partial_\nu \operatorname{Tr}(P_0 \rho(\lambda) P_0^\dagger)\Big|_{\lambda=\lambda_0}. \tag{17}$$

In that expression the second term can be bounded from below by

$$\operatorname{Re}\left(P_0 \left(\partial_\mu \rho(\lambda_0)\right) P_+^\dagger \left(\rho_+(\lambda_0)\right)^{-1} P_+ \left(\partial_\nu \rho(\lambda_0)\right) P_0^\dagger \right). \tag{18}$$

If the bound is not an equality along some vector ϕ tangent to $\mathscr{M}$ at $\lambda_0 \in \mathscr{M}$ then $\mathrm{rank}(\rho(\lambda))$ *is larger than* $\mathrm{rank}(\rho(\lambda_0))$ *in some punctured neighbourhood of λ_0 (i.e. the neighbourhood excluding the point λ_0 itself) along the direction ϕ.*

3. *Suppose $p(\lambda) = \{p_z(\lambda)\}_{z\in\mathscr{S}}$ is a discrete probability distribution on a finite set $\mathscr{S}$ defined in a neighbourhood of $\lambda = \lambda_0 \in \mathscr{M}$. Let $\mathscr{S} = \mathscr{S}_+ \cup \mathscr{S}_0$ be the split of $\mathscr{S}$ into subsets where $p_z(\lambda_0)$ is positive or zero respectively. Assume that p_z has the first derivative at $\lambda = \lambda_0$ and $\sum_{z\in\mathscr{S}_0} p_z(\lambda)$ has the second derivative at $\lambda = \lambda_0$. Then*

$$g_{\mu\nu}(\lambda_0) = \sum_{z\in\mathscr{S}_+} \frac{\left(\partial_\mu p_z(\lambda_0)\right)\left(\partial_\nu p_z(\lambda_0)\right)}{4p_z(\lambda_0)} + \frac{1}{2}\,\partial_\mu\partial_\nu \sum_{z\in\mathscr{S}_0} p_z(\lambda)\Bigg|_{\lambda=\lambda_0}. \quad (19)$$

Part 1 of this theorem is essentially the formula (3) in [1] with the exact conditions needed from $|\varphi(\lambda)\rangle$ specified. As mentioned in [1], the proof is essentially done by the Taylor expansion of (6) and the usage of the fact that the Hilbert space elements representing the states have the norm of 1. Since we do not require the second derivative to exist, we have to do that expansion with a bit more care than [1], as spelled out in the proof below. Note also that part 2 is a generalization of part 1, and the proof for part 2 won't use part 1, hence we could just leave the proof for part 2. However, the proof for part 1 is much simpler, so we leave it here.

Proof of part 1. Due to equivariance of the definition of $g_{\mu\nu}$ with respect to the change of the coordinates $\lambda \mapsto \lambda - \lambda_0$, without loss of generality we can prove the statements in the theorem for $\lambda_0 = 0$. Let $|\psi_0\rangle = |\psi(0)\rangle$, $|\delta\psi\rangle = |\psi(\delta\lambda)\rangle - |\psi(0)\rangle$. We know that

$$|\delta\psi\rangle = \sum_\mu |\partial_\mu\psi(0)\rangle\,\delta\lambda^\mu + |r\rangle, \quad (20)$$

where $|r\rangle = o(|\delta\lambda|)$. We also know that $\langle\psi(\delta\lambda)|\psi(\delta\lambda)\rangle = 1$. On the other hand,

$$\begin{aligned}\langle\psi(\delta\lambda)|\psi(\delta\lambda)\rangle &= 1 + 2\,\mathrm{Re}\,\langle\psi_0|\delta\psi\rangle + \langle\delta\psi|\delta\psi\rangle \\ &= 1 + 2\sum_\mu \mathrm{Re}\,\langle\psi_0|\partial_\mu\psi(0)\rangle\lambda^\mu + 2\,\mathrm{Re}\,\langle\psi_0|r\rangle \\ &\quad + \sum_{\mu,\nu}\langle\partial_\mu\psi(0)|\partial_\nu\psi(0)\rangle\,\delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2). \quad (21)\end{aligned}$$

Thus,

$$\mathrm{Re}\,\langle\psi_0|\partial_\mu\psi(0)\rangle = 0 \quad (22)$$

and

$$\begin{aligned}&\mathrm{Re}\,\langle\psi_0|r\rangle \\ &\quad = -\frac{1}{2}\sum_{\mu,\nu}\langle\partial_\mu\psi(0)|\partial_\nu\psi(0)\rangle\,\delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2). \quad (23)\end{aligned}$$

We compute

$$\begin{aligned}F(0,\delta\lambda) &= \left|\langle\psi(0)|\psi(\delta\lambda)\rangle\right| = \left|1 + \langle\psi_0|\delta\psi\rangle\right| \\ &= \left|1 + \sum_\mu \langle\psi_0|\partial_\mu\psi(0)\rangle\,\delta\lambda^\mu + \langle\psi_0|r\rangle\right| \\ &= 1 + \mathrm{Re}\,\langle\psi_0|r\rangle - \frac{1}{2}\left(\sum_\mu\langle\psi_0|\partial_\mu\psi(0)\rangle\,\delta\lambda^\mu\right)^2 \\ &\quad + o(|\delta\lambda|^2) \\ &= 1 - \frac{1}{2}\sum_{\mu\nu} g_{\mu\nu}\delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2), \quad (24)\end{aligned}$$

where $g_{\mu\nu}$ is given by (16). Note that when expanding the absolute value we used the fact that for real x, y around $x = y = 0$ we have

$$|1 + x + iy| = 1 + x + \frac{y^2}{2} + O(x^2 + y^4). \quad (25)$$

□

Lemma 2. *Let*

$$M = \begin{pmatrix} A & B \\ B^\dagger & B^\dagger A^{-1} B + C \end{pmatrix} \quad (26)$$

be a finite-dimensional block diagonal matrix over $\mathbb{C}$ with positive definite A. Then $M \geq 0$ iff $C \geq 0$. In that case $\mathrm{rank}(M) = \mathrm{rank}(A) + \mathrm{rank}(C)$.

Proof. The lemma follows from the decomposition

$$M = \begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}^\dagger \begin{pmatrix} A^{-1} & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}. \quad (27)$$

□

For part 2 see, e.g., [8, section 15.1].

Proof of part 2. As in the previous proof, without loss of generality set $\lambda_0 = 0$. Denote $\rho_0 = \rho(0)$, $\delta\rho = \rho(\delta\lambda) - \rho_0$. We know that

$$\delta\rho = \sum_\mu \delta\lambda^\mu\partial_\mu\rho(0) + r, \quad (28)$$

where $r = o(|\delta\lambda|)$. We also know that

$$\mathrm{Tr}\left(\partial_\mu\rho(0)\right) = 0, \quad (29)$$

$$\mathrm{Tr}(r) = 0. \quad (30)$$

From definition,

$$F(\rho_0, \rho_0 + \delta\rho) = \mathrm{Tr}\sqrt{\rho_0^2 + \sqrt{\rho_0}\delta\rho\sqrt{\rho_0}}. \quad (31)$$

Let $\rho_{0+} = P_+\rho_0 P_+^\dagger$, $\delta\rho_+ = P_+\delta\rho P_+^\dagger$, $r_+ = P_+ r P_+^\dagger$. One can see that the expression under the square root acts nontrivially

only on $\rho_0(\mathscr{H})$, hence the trace can be computed in that subspace:

$$F(\rho_0,\rho_0+\delta\rho)=\operatorname{Tr}\sqrt{\rho_{0+}^2+\sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}}. \tag{32}$$

For $\delta\lambda=0$ the expression under the square root is equal to ρ_{0+}^2 and has only positive eigenvalues. Thus, for $\delta\lambda$ in some neighbourhood of 0 the spectrum of the expression under the square root lies in (c_1,c_2) for some c_1,c_2 satisfying $0<c_1\le c_2<\infty$. Thus, in that neighbourhood the square root is an analytic function and can be expressed as an integral with the corresponding resolvent over a contour surrounding $[c_1,c_2]$:

$$\begin{aligned}&\sqrt{\rho_{0+}^2+\sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}}\\&\quad=\frac{-1}{2\pi i}\oint\sqrt{z}\left(\rho_{0+}^2+\sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}-z\right)^{-1}dz\\&\quad=I_0+I_1+I_2+o(|\delta\lambda|^2),\end{aligned} \tag{33}$$

where

$$\begin{aligned}I_k=\frac{(-1)^{k+1}}{2\pi i}\oint\sqrt{z}\left(\rho_{0+}^2-z\right)^{-1}\\\left(\sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}\left(\rho_{0+}^2-z\right)^{-1}\right)^k dz.\end{aligned} \tag{34}$$

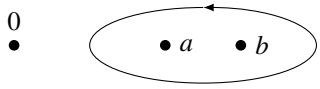
In order to evaluate the metric, we only need to compute the diagonal elements of I_0,I_1,I_2 discarding any terms with order $o(|\delta\lambda|^2)$. We pick the basis where ρ_{0+} is diagonal with diagonal elements $\xi_0\ge\xi_1\ge\cdots\ge\xi_{n_+-1}>0$.

$$I_0=\rho_{0+}, \tag{35}$$

$$(I_1)_{jj}=\frac{1}{2}\delta\rho_{+,jj}=\frac{1}{2}\sum_\mu\delta\lambda^\mu\partial_\mu\rho_{+,jj}(0)+\frac{1}{2}r_{+,jj}, \tag{36}$$

[VK: TODO: what is the proper way to separate $+$ and jj in $r_{+,jj}$? $(r_+)_{jj}$, r_{+jj}, $r_{+\,jj}$?]

To evaluate the diagonal entries of I_2 we note that for the contour



where a,b are positive real numbers (a could be equal to b), we have

$$\frac{1}{2\pi i}\oint\frac{\sqrt{z}}{(z-a)^2(z-b)}dz=\frac{-1}{2\sqrt{a}(\sqrt{a}+\sqrt{b})^2}. \tag{37}$$

We then evaluate $(I_2)_{jj}$:

$$\begin{aligned}(I_2)_{jj}&=-\sum_k\frac{\xi_j\delta\rho_{+,jk}\xi_k\delta\rho_{+,kj}}{2\xi_j(\xi_j+\xi_k)^2}=-\sum_k\frac{|\delta\rho_{+,jk}|^2\xi_k}{2(\xi_j+\xi_k)^2}\\&=-\sum_{k,\mu,\nu}\frac{\operatorname{Re}\left((\partial_\mu\rho_+(0))_{jk}(\partial_\nu\rho_+(0))_{kj}\right)\xi_k}{2(\xi_j+\xi_k)^2}\delta\lambda^\mu\delta\lambda^\nu\\&\quad+o(|\delta\lambda|^2).\end{aligned} \tag{38}$$

Now we are ready to evaluate the trace

$$F(\rho_0,\rho_0+\delta\rho)=\operatorname{Tr}(I_0+I_1+I_2)+o(|\delta\lambda|^2). \tag{39}$$

In order to evaluate (14) we will include the terms up to the order $o(|\delta\lambda|^2)$:

$$\operatorname{Tr}(I_0)=\operatorname{Tr}(\rho_0)=1. \tag{40}$$

The first term in $(I_1)_{jj}$ cannot have a non-zero contribution to $\operatorname{Tr}(I_1)$ due to the fact that $\rho(\lambda)$ is non-negative and has $\operatorname{Tr}(\rho(\lambda))=1$. For the second term, notice that $\operatorname{Tr}(r)=0$, hence $\operatorname{Tr}(r_+)+\operatorname{Tr}(P_0rP_0)=0$, giving

$$\operatorname{Tr}(I_1)=-\operatorname{Tr}(P_0rP_0^\dagger)/2. \tag{41}$$

Note that according to lemma 2

$$\begin{aligned}&P_0rP_0^\dagger\ge\left(P_+\delta\rho P_0^\dagger\right)^\dagger\rho_{0+}^{-1}P_+\delta\rho P_0^\dagger\\&=\sum_{\mu,\nu}\operatorname{Re}\left(P_0\left(\partial_\mu\rho(0)\right)P_+^\dagger\rho_{0+}^{-1}P_+\left(\partial_\nu\rho(0)\right)P_0^\dagger\right)\delta\lambda^\mu\delta\lambda^\nu\\&\quad+o(|\delta\lambda|^2).\end{aligned} \tag{42}$$

Here $\operatorname{Re}(a)=(a+a^\dagger)/2$ for a matrix or an operator a. Combining (38) and (41) we get

$$\begin{aligned}g_{\mu\nu}(0)=\sum_{j,k}\frac{\operatorname{Re}\left((\partial_\mu\rho_+(0))_{jk}(\partial_\nu\rho_+(0))_{kj}\right)\xi_k}{(\xi_j+\xi_k)^2}\\+\frac{1}{2}\,\partial_\mu\partial_\nu\operatorname{Tr}(P_0\rho(\lambda)P_0^\dagger)\Big|_{\lambda=0}.\end{aligned} \tag{43}$$

The remaining statements of the part 2 of the theorem follow from the lemma 2. □

Note that part 3 trivially follows from part 2 when applied to diagonal ρ.

[VK: The third term is weird. Consider an example where $\mathscr{S}=\{0,1\}$, $\mathscr{M}=\mathbb{R}$, $p_0=\sin^2(\theta)$, $p_1=\cos^2(\theta)$. Then $g=(1+\delta_{\sin(4\theta),0})ds^2$.]

Theorem 3. *Suppose $|\psi(\lambda)\rangle$ is a state defined in the neighbourhood of $\lambda=\lambda_0\in\mathscr{M}$ and differentiable at $\lambda=\lambda_0$. Let g be the corresponding fidelity metric and g_c be its classical counterpart dependent on a projective measurement. then*

$$\mathbb{E}_{measurements}g_c(\lambda_0)=g(\lambda_0)/2, \tag{44}$$

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

Proof. The second term in (19) is only relevant when one of the vectors in the measurement basis is orthogonal to $|\psi(\lambda_0)\rangle$ — a subset of measure 0 in the space of all measurements. Thus, we can safely discard it in (44). To simplify the proof we notice that a quadratic form can be recovered from its values of the form $g(\lambda_0)(v,v)$, and both sides of the formula (44)

are invariant with respect to the choice of coordinates. For a fixed v we can always choose coordinates λ such that $v = \partial_0$, $\lambda_0 = 0$. It remains to prove that

$$\mathbb{E}_{\text{measurements}} \sum_z \frac{\left(\partial_0 p_z(0)\right)^2}{4 p_z(0)} = \frac{1}{2}\left\| |\partial_0 \psi(0)\rangle \right\|^2 - \frac{1}{2}\left|\langle \partial_0 \psi(0) | \psi(0)\rangle\right|^2. \quad (45)$$

One can simplify the l.h.s. by noting that the expectation of each term in the sum is the same. Thus, it remains to average over $|\varphi\rangle$ s.t. $\left\| |\varphi\rangle \right\| = 1$:

$$\text{l.h.s. of (45)} = n\mathbb{E}_{|\varphi\rangle} \frac{\left(\text{Re}\left(\langle\varphi|\partial_0\psi(0)\rangle \langle\psi(0)|\varphi\rangle\right)\right)^2}{\left|\langle\varphi|\psi(0)\rangle\right|^2}. \quad (46)$$

Let's denote $|\psi_0\rangle = |\psi(0)\rangle$, $\psi_\parallel = \langle\psi_0|\partial_0\psi(0)\rangle$, $|\psi_\perp\rangle = |\partial_0\psi(0)\rangle - \psi_\parallel |\psi_0\rangle$, $\varphi_\parallel = \langle\psi_0|\varphi\rangle$, $|\varphi_\perp\rangle = |\varphi\rangle - \varphi_\parallel |\psi_0\rangle$. Note that $\text{Re}(\psi_\parallel) = 0$. With this notation

$$\text{r.h.s. of (45)} = \frac{1}{2}\left\| |\psi_\perp\rangle \right\|^2, \quad (47)$$

$$\langle\varphi|\partial_0\psi(0)\rangle \langle\psi(0)|\varphi\rangle = \left|\varphi_\parallel\right|^2 \psi_\parallel + \varphi_\parallel \langle\varphi_\perp|\psi_\perp\rangle. \quad (48)$$

Substituting this and averaging over arguments of $\varphi_\parallel$ we get

$$\text{l.h.s. of (45)} = \frac{n}{2}\mathbb{E}_{|\varphi\rangle}\left|\langle\varphi_\perp|\psi_\perp\rangle\right|^2 = \frac{1}{2}\|\psi_\perp\|^2. \quad (49)$$

□

Given theorem 3 is for pure states, one could wonder whether a similar statement is true for mixed states. In general the answer is negative. Moreover, it is possible to find a special case where the average of the classical fidelity is 0 while the quantum fidelity is non-zero:

$$\rho(\lambda) = \begin{pmatrix} 1-|\lambda|^2 & 0 \\ 0 & |\lambda|^2 \end{pmatrix} \quad (50)$$

gives $g_{\mu\nu}(0) = 1$ while apart for one specific measurement basis where one measures ρ along the coordinates in which it is given the classical fidelity is 0 at $\lambda = 0$ due to vanishing first derivatives. Such extreme examples are rare. Indeed, the example relies on the bound (18) not being an equality, hence the rank of ρ being higher in the neighbourhood of a point where similar example applies along the direction for which it applies. In one dimension that a set of such values of λ has to be discrete, in any number of dimensions it has to have measure 0.

In today's quantum annealer, though, the measurement basis is not random: instead, there is a fixed basis called "computational basis" in which all measurements are performed. At the same time, today's quantum annealers have a limited set of biases and couplings allowed to the user. Below theorem shows, that in this case too for pure states classical fidelity almost everywhere is a good approximator of quantum fidelity. Even better, in this case they are equal! The main restriction on the Hamiltonian is that its matrix elements are real, something which is true e.g. for X and Z biases, and XX, YY, and ZZ couplings.

Theorem 4. • *Suppose there is a continuously differentiable Hamiltonian $H(\lambda)$ acting on a finite-dimensional Hilbert space $\mathscr{H}$ with a basis which we will call "computational basis". Assume that matrix elements of the Hamiltonian are real in that basis. Also assume that the ground state of that Hamiltonian is non-degenerate. Then the ground state could be chosen to have real elements and continuously differentiable w.r.t. λ. elements.*

• *For a state differentiable w.r.t. λ with real matrix elements, fidelity susceptibility metric coincides with its classical counterpart.*

Proof. First, let's fix λ and show that the ground state can be picked to be real for a Hamiltonian with real matrix elements. Indeed, let $|\psi\rangle$ be the ground state with ground energy ε_0. We can represent it as

$$|\psi\rangle = \cos(\theta)|\psi_0\rangle + i\sin(\theta)|\psi_1\rangle, \quad (51)$$

where both $|\psi_j\rangle$ are normalized and have real vector elements, then we have

$$\begin{aligned}\varepsilon_0 &= \langle\psi|H(\lambda)|\psi\rangle \\ &= \cos^2(\theta)\langle\psi_0|H(\lambda)|\psi_0\rangle + \sin^2(\theta)\langle\psi_1|H(\lambda)|\psi_1\rangle \\ &\geq \varepsilon_0(\cos^2(\theta)+\sin^2(\theta)) = \varepsilon_0. \quad (52)\end{aligned}$$

Here the inequality is due to the fact that ε_0 is the lowest energy level of $H(\lambda)$. Since it has to be an equality, we have

$$\langle\psi_0|H(\lambda)|\psi_0\rangle = \varepsilon_0 \text{ or } \cos(\theta) = 0, \quad (53)$$

$$\langle\psi_1|H(\lambda)|\psi_1\rangle = \varepsilon_0 \text{ or } \sin(\theta) = 0. \quad (54)$$

Thus, at least one of ψ_0 or ψ_1 is the ground state with real vector elements. We can now finish the proof of part 1 by first applying implicit function theorem to $\det(H(\lambda)-\varepsilon_0(\lambda))$ to show $\varepsilon_0(\lambda)$ is differentiable, and then applying implicit function theorem to the system of equations obtained from $(H(\lambda)-\varepsilon_0)|\psi(\lambda)\rangle$ by replacing one of the redundant rows with the equation

$$\left\| |\psi(\lambda)\rangle \right\|^2 = 1. \quad (55)$$

To prove the second part, notice that for the same reason as in the proof of theorem 3, it is sufficient to prove it for g_{00}. Then expanding both (16) and (19) for a state with real vector elements, we get the same expression

$$g_{00}(\lambda) = \left\| |\partial_0\psi(\lambda)\rangle \right\|^2. \quad (56)$$

□

On the other hand, Hamiltonians with non-real matrix elements can have $\chi_{F_c}(s) = 0$ with non-zero $\chi_F(s)$. For example, consider

$$H(s) = -(1-s)\sum_j X_j - s\sum_j Y_j. \tag{57}$$

The eigenstate is $|\psi(s)\rangle = |\psi_0(s)\rangle^{\otimes n}$, where

$$|\psi_0(s)\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + \frac{(1-s)+is}{\sqrt{1-2s(1-s)}}|1\rangle\right). \tag{58}$$

These states generate uniform probability distribution independent of s when measured in the computational basis. However, the state does change with s, hence it has non-zero

$$g(s) = \frac{n}{4}\left(d\arctan\left(\frac{s}{1-s}\right)\right)^2 = \frac{n\,ds^2}{4(1-2s(1-s))^2}. \tag{59}$$

D. Resolution of issue 3

As we have seen in theorems 3 and 4, classical fidelity susceptibility is expected to be between $\frac{1}{2}\chi_F(s)$ and $\chi_F(s)$ in many cases of interest. Thus, there is a hope that the maximum of classical fidelity susceptibility would be close to the maximum of fidelity susceptibility in cases of interest. In particular, the Hamiltonians which could be implemented on the current quantum annealers are stoquastic and, in particular, have real matrix elements in the computational basis. Thus if the ground state is non-degenerate $\chi_{F_c}(s) = \chi_F(s)$ for the ground state and no information is lost when replacing $\chi_F(s)$ with $\chi_{F_c}(s)$ if we are looking for phase transitions at zero temperature. [VK: TODO: should we use here $g_c(s) = g(s)$? Should we then explain above that we plan to use g_c for classical fidelity susceptibility metric whenever we want to avoid the ambiguity with the fidelity susceptibility?]

Thus, we can reformulate the setting of the paper as the challenge between two approaches for identifying the maximums of $\chi_F(s)$. The first approach is to use classical state of the art algorithms. The second approach involves the following steps.

- Use a quantum annealer to produce a dataset of pairs (s,z), where s is a value of the parameter and z is a bitstring measured in the computational basis.
- Use an ML-based method, which we call Bitstring-ChiFc, to produce an estimate $\widehat{\chi}_{F_c}(s)$ of $\chi_{F_c}(s)$. See [VK: TODO] for the description of the method.
- Find local maxima $\hat{s}_{\text{critical},i}$ of the estimate.
- If necessary, refine the estimates $\hat{s}_{\text{critical}}$ by repeating the algorithm with a smaller range of values of s around $\hat{s}_{\text{critical}}$.

Now we move to describe classical algorithms and their performance in this task.

E. Classical algorithm: Lanczos

If we are interested in the fidelity suceptibility associated to the ground states of the Hamiltonian, and the system size is small, one of the powerful classical algorithms is based on Lanczos algorithm. In the context of this algorithm we think about $H(\lambda)$ as a sparse 2^n by 2^n matrix together with an algorithm to compute $H(\lambda)|\psi\rangle$ for a vector $|\psi\rangle$ of the length 2^n. Typically, Hamiltonians we consider have $\Theta(n)$ terms and the computation $H(\lambda)|\psi\rangle$ takes $\Theta(n2^n)$ time. Lanczos algorithm starts from a random vector $|\psi\rangle$ and an integer k (typically, k is of the order of 40). In exact arithmetic the algorithm would return a unit vector $|\varphi\rangle$ from $\text{span}(\{H(\lambda)^j|\psi\rangle : j = 0,\ldots,k-1\})$ s.t. $\big\||\varphi\rangle\big\| = 1$ and $\langle\varphi|H(\lambda)|\varphi\rangle$ is the smallest possible. In practice, one performs the algorithm using floating point numbers which causes an additional source of errors. In fact, the Lanczos algorithm currently used in practice is different from the one originally proposed by Lanczos, with most of the differences intended to make the algorithm more resilient to numerical errors of finite precision arithmetics while minimizing the additional computational cost spent on these efforts.

One then would apply the algorithm for values of parameters on a grid and estimate the fidelity susceptibility directly. E.g. if H depends on a single parameter $s \in [0,1]$ one could choose the number of points N and compute $H(s)$ for $s \in \{j/N : j = 0,\ldots,N\}$, then estimate

$$\hat{\chi}_F(s+\Delta s/2) \simeq 2(1-\big|\langle\varphi(s)|\varphi(s+\Delta s)\rangle\big|)/\Delta s^2, \tag{60}$$

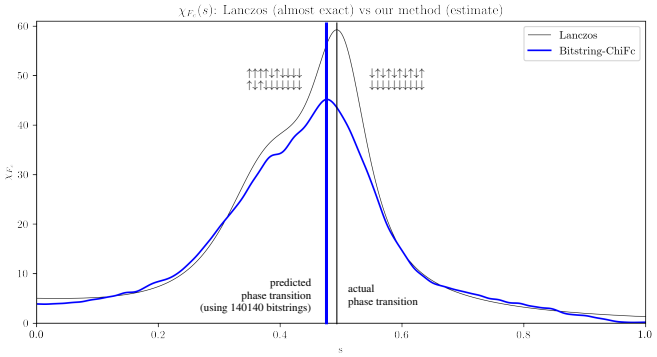
where $\Delta s = 1/N$, $s = j/N$, $j = 0,\ldots,N-1$.

The main downside of this approach is the need to perform k multiplications of $H(\lambda)$ by a vector of size 2^n, something which typically takes time $\Theta(kn2^n)$, in each of N executions of the Lanczos algorithm. In practice, that means that this method is the go to method for $n \leq 20$, being able to estimate the fidelity to a good enough accuracy for $n = 20$ in under 2 hours on a single machine but impractical for $n > 40$.

There are cases where it fails due to either algorithmic or numerical issues, especially if the gap between the ground state and first excited state is small, or there is a large number of energy levels close to ground state energy. However, in these cases other algorithms we reviewed would typically fail as well and such failures could be easily detected by running the algorithm with a different initial random state.

F. Results

Here is the comparison based on Lanczos diagonalization for L=10 (20-qubit) frustrated ladder Hamiltonian with $K = U = 1$: ground truth from Lanczos vs reconstruction from Bitstring-ChiFc method.



III. 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.

IV. INTRODUCTION

TODO:

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

V. CLASSICAL FIDELITY SUSCEPTIBILITY

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

$$F_c(p,q) = \sum_z \sqrt{p(z)q(z)}. \tag{61}$$

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

$$F_c(s, s+\delta s) = 1 - \frac{\delta s^2}{2}\chi_{F_c}(s) + O(\delta s^3). \tag{62}$$

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 (62) would become

$$F_c(\lambda, \lambda+\delta\lambda) = 1 - \frac{\delta\lambda_j \delta\lambda_k}{2}\chi^{jk}_{F_c}(\lambda) + O(\delta\lambda^3). \tag{63}$$

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

VI. PROBLEM SETUP

- In this work we consider a family of distributions of bitstrings $\{\mathscr{D}_\lambda\}_{\lambda\in\Lambda}$, each of length n.
- We are given a finite sample $\mathscr{D}_{\text{train}}$ of size N of pairs (λ, z) s.t. $P(z|\lambda) = P_{\mathscr{D}_\lambda}(z)$.
- 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 $\mathscr{D}_\lambda$.
- 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.

VII. BITSTRING-CHIFC METHOD

In this work we propose the following method:

- Collect a training dataset

$$\mathscr{D}_{\chi_{F_c}\text{-train}} = \{(\lambda_0, \delta\lambda, z, y), \dots\}, \tag{64}$$

  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 $\mathscr{D}_{\text{train}}$ consisting of pairs (z, λ).
  - Sample pairs (z_{i+}, λ_{i+}), (z_{i-}, λ_{i-}) from $\mathscr{D}_{\text{train}}$.
  - 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 $\mathscr{D}_{\chi_{F_c}\text{-train}}$.

- 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 $\mathscr{D}_{\chi_{F_c}\text{-train}}$.

- Estimate

$$(\hat{g}_c)_{\mu\nu}(\lambda) = \text{smoothen}\Big(\lambda_1 \mapsto \text{mean}_{(z,\lambda_1)\in\mathscr{D}_{\text{train}}}\Big(M(\lambda_1,0,z)_\mu M(\lambda_1,0,z)_\nu\Big)\Big)(\lambda). \quad (65)$$

A. Justification for the formula for Bitstring-ChiFc method.

Here we justify the formula used for Bitstring-ChiFc method.

Theorem 5. *Assume that the model M is continuous with respect to $\delta\lambda$ and has the least possible out of sample loss near $\delta\lambda = 0$. Then*

$$(g_c)_{\mu\nu}(\lambda) = \mathbb{E}_{(z,\lambda)\in\mathscr{D}}\left(M(\lambda,0,z)_\mu M(\lambda,0,z)_\nu\right). \quad (66)$$

In some sense the theorem justifies the use of formula (65): it essentially shows that if the model is perfect, and the dataset used in (65) is infinite, then that formula (65) without the smoothing gives a perfect estimate for the classical fidelity susceptibility metric g_c. Of course, in reality these assumptions are not realistic: the model is not perfect, the dataset is finite, and the smoothing is necessary to dampen high frequency noise arising from the finite dataset size. [VK: TODO: Should we discuss the imperfections in a later subsection?]

Proof. Both lhs and rhs of (66) are quadratic forms equivariant under the rotations of the coordinate system of the parameter space. Therefore, it is sufficient to show they are equal for $\mu = \nu = 0$. That is, it is sufficient to prove the theorem in the scalar case, where λ consists of a single parameter s. In this case (66) takes the form

$$\chi_{F_c}(s) = \mathbb{E}_{(z,s)\in\mathscr{D}}\left(M(s,0,z)^2\right). \quad (67)$$

By definition (10) and (8) we have

$$\chi_{F_c}(s) = \lim_{\delta s\to 0}\frac{2}{\delta s^2}\left(1 - \sum_z \sqrt{P(z|s-\delta s/2)P(z|s+\delta s/2)}\right). \quad (68)$$

Let's introduce the distribution Q on bitstrings z by $Q(z) = (P(z|s-\delta s/2) + P(z|s+\delta s/2))/2$. We can rewrite the above expression as

$$\chi_{F_c}(s) = \lim_{\delta s\to 0}\frac{2}{\delta s^2}\left(1 - \mathbb{E}_{z\sim Q}\frac{\sqrt{P(z|s-\delta s/2)P(z|s+\delta s/2)}}{Q(z)}\right). \quad (69)$$

According to the definition of model M and the assumption that it has the least possible out of sample loss, we have

$$P(z|s\pm\delta s/2)/Q(z) = \frac{e^{\pm M(s,\delta s,z)\delta\lambda/2}}{\cosh(M(s,\delta s,z)\delta\lambda/2)}. \quad (70)$$

Substituting (70) to (71) we get

$$\chi_{F_c}(s) = \lim_{\delta s\to 0}\frac{2}{\delta s^2}\left(1 - \mathbb{E}_{z\sim Q}\frac{1}{\cosh(M(s,\delta s,z)\delta\lambda/2)}\right) = . \quad (71)$$

[VK: TODO:0 == Work in progress: START] □

Then we show that under these (unrealistic) assumptions the formula (65) without the smoothing part produces the perfect estimate for χ_{F_c}. By definition, we have
[VK: TODO:2: expand the explanation for s instead of λ.]

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

TODO: models
TODO: experiments
[VK: TODO:2: 2nd part of the presentation with an image: presentation is a more low-low hanging fruit.] [VK: Work in progress: END]

[1] P. Zanardi, P. Giorda, and M. Cozzini, Physical review letters **99**, 100603 (2007).
[2] P. Zanardi, H. Quan, X. Wang, and C. Sun, Physical Review A **75**, 032109 (2007).
[3] S.-J. Gu, H.-M. Kwok, W.-Q. Ning, H.-Q. Lin, *et al.*, Physical Review B **77**, 245109 (2008).
[4] M. Vojta, Reports on Progress in Physics **66**, 2069 (2003).
[5] Wikipedia, "Unconventional superconductor — Wikipedia, the free encyclopedia," http://en.wikipedia.org/w/index.php?title=Unconventional%20superconductor&oldid=1101177511 (2022), [Online; accessed 09-September-2022].
[6] J. Cardy, *Finite-size scaling* (Elsevier, 2012).
[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[8] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, 2nd ed. (Cambridge University Press, 2017).