Machine learning Fisher Information Metric from bitstrings

First Last[1,*]
[1] *USC affiliation*
(Dated: August 15, 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:

$\cdots T_0 \cdots T_1 \cdots T_2 \cdots T_3 \cdots T_4 \cdots T_5 \cdots T_6 \cdots T_7 \cdots T_8 \cdots T_9 \cdots$

$- B_0 - B_1 - B_2 - B_3 - B_4 - B_5 - B_6 - B_7 - B_8 - B_9 -$

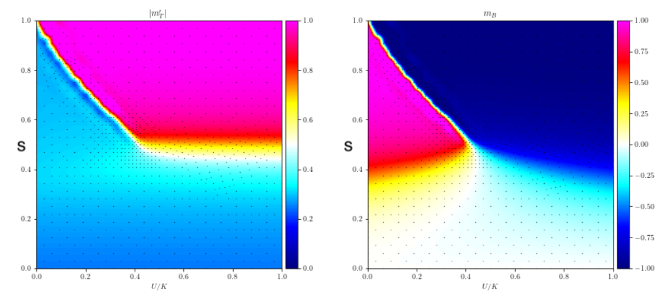
On this diagram of $L = 10$ Frustrated Ladder Hamiltonian the solid lines represent ferromagnetic couplings and dotted lines — antiferromagnetic couplings. For a fixed L the Frustrated Ladder Hamiltonian depends on 3 parameters, s, K, U. We set $K = 1$ and consider the values $s \in [0, 1]$, $U \in [0,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

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.

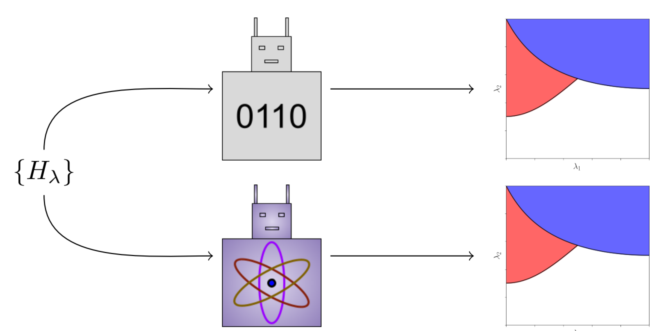
$$|m'_T| = \left| \sum_i Z_{T_i} (-1)^i \right| \tag{4}$$

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

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:



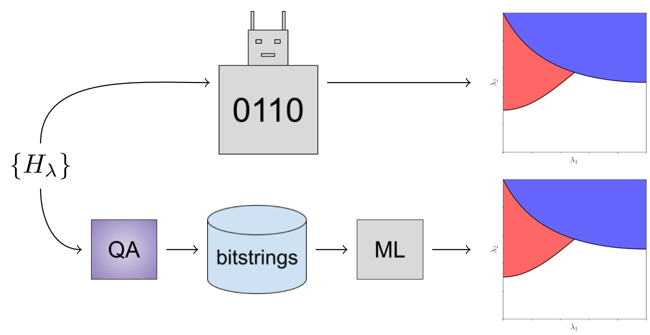
[VK: The following statement is in the direct contradiction with the statement that the main goal of this paper is to learn Fisher Information Metric given a dataset of bitstrings.] In this work we focus on the task of identifying phase transitions in a family of Hamiltonians $\{H_\lambda\}_{\lambda \in \Lambda}$ 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:

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

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.

C. 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.

- $0 \leq \chi_{F_c}(s) \leq \chi_F(s)$.
- $\mathbb{E}_{\text{measurements}} \chi_{F_c}(s) = \chi_F(s)/2$.
- For non-degenerate ground states of real-valued Hamiltonians $\chi_{F_c}(s) = \chi_F(s)$ almost everywhere.

See the below theorems for the exact statements.

The following theorem states the properties of the fidelity relevant for this paper. For proof of these, as well as many other, properties of the fidelity the reader is referred to [1, §9].

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

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

In other words, (6) *is consistent with* (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, (7) *is consistent with* (8).

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

$$F(\rho, \sigma) = F(\sigma, \rho), \tag{13}$$

$$0 \leq F(\rho, \sigma) \leq 1, \tag{14}$$

$$F(\rho, \sigma) = 1 \text{ iff } \rho = \sigma, \tag{15}$$

$$F(\rho, \sigma) = 0 \text{ iff } \rho\sigma = 0. \tag{16}$$

Before moving to fidelity susceptibility, we want to describe a generalization of the fidelity susceptibility covered in [2]. Formulas (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 (1) depends on 3 real parameters: s, K, U. In general, we can say that a state of interest depend on a parameter λ from a parameter manifold $\mathcal{M}$. In general, these parameters can include the parameters of the Hamiltonian and the parameters impacting how the state is derived from that Hamiltonian (e.g. temperature). Then, similarly to equations (9) and (10) we can expand the fidelity to the second order.

$$F(\lambda, \lambda + \delta\lambda) = 1 - \frac{1}{2}\sum_{\mu\nu} g_{\mu\nu}\delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2). \tag{17}$$

The resulting second term represents a metric

$$g = \sum_{\mu\nu} g_{\mu\nu} d\lambda^\mu d\lambda^\nu. \tag{18}$$

The metric g described by (17) and (18) is invariant with respect to the choice of the coordinates λ on $\mathcal{M}$. As explained in [2], phase transitions are expected to correspond to singularities in the metric (18) in the thermodynamic limit. We, however, are interested in the finite systems of a fixed size. Therefore, we are interested in identifying the vectors φ on $\mathcal{M}$ with high values of $g(\varphi, \varphi)/g_0(\varphi, \varphi)$, where g_0 is a metric considered to be non-singular.

Theorem 2. *1. Suppose $|\psi(\lambda)\rangle$ is a state defined in the neighbourhood of $\lambda = \lambda_0 \in \mathcal{M}$ and differentiable at $\lambda = \lambda_0$. Then the fidelity susceptibility metric is well-defined at $\lambda = \lambda_0$ and is given by*

$$\begin{aligned} g_{\mu\nu}(\lambda_0) = \mathrm{Re}\big(&\langle\partial_\mu\psi(\lambda_0)|\partial_\nu\psi(\lambda_0)\rangle\big) \\ &- \langle\partial_\mu\psi(\lambda_0)|\psi(\lambda_0)\rangle\langle\psi(\lambda_0)|\partial_\nu\psi(\lambda_0)\rangle, \end{aligned} \tag{19}$$

where $\partial_\mu\psi(\lambda_0)$ is a compact notation for $\left.\frac{\partial\psi(\lambda)}{\partial\lambda^\mu}\right|_{\lambda=\lambda_0}$.

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

$$\begin{aligned} g_{\mu\nu}(\lambda_0) = \sum_{j,k} &\frac{\mathrm{Re}\left((\partial_\mu\rho_+(\lambda_0))_{jk}(\partial_\nu\rho_+(\lambda_0))_{kj}\right)}{2(\xi_j + \xi_k)} \\ &+ \frac{1}{2}\,\partial_\mu\partial_\nu\,\mathrm{Tr}(P_0\rho(\lambda)P_0^*)\big|_{\lambda=\lambda_0}. \end{aligned} \tag{20}$$

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

$$\mathrm{Re}\left(P_0\left(\partial_\mu\rho^*(\lambda_0)\right)P_+^*\left(\rho_+(\lambda_0)\right)^{-1}P_+\left(\partial_\nu\rho(\lambda_0)\right)P_0^*\right). \tag{21}$$

If the bound is not an equality along some vector ϕ tangent to $\mathcal{M}$ at $\lambda_0 \in \mathcal{M}$ 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 $p(\lambda) = \{p_z(\lambda)\}_{z\in\mathcal{S}}$ is a discrete probability distribution on a finite set $\mathcal{S}$ defined in a neighbourhood of $\lambda = \lambda_0 \in \mathcal{M}$. Let $\mathcal{S} = \mathcal{S}_+ \cup \mathcal{S}_0$ be the split of $\mathcal{S}$ 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 $\sum_{z\in\mathcal{S}_0} p_z(\lambda)$ has the second derivative at $\lambda = \lambda_0$. Then

$$\begin{aligned} g_{\mu\nu}(\lambda_0) = \sum_{z\in\mathcal{S}_+} &\frac{\left(\partial_\mu p_z(\lambda_0)\right)\left(\partial_\nu p_z(\lambda_0)\right)}{4p_z(\lambda_0)} \\ &+ \frac{1}{2}\,\partial_\mu\partial_\nu \sum_{z\in\mathcal{S}_0} p_z(\lambda)\Bigg|_{\lambda=\lambda_0}. \end{aligned} \tag{22}$$

Part 1 of this theorem is essentially the formula (3) in [2] with the exact conditions needed from $|\varphi(\lambda)\rangle$ specified.

As mentioned in [2], 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 [2], as spelled out in the proof below.

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

$$|\delta\psi\rangle = \sum_\mu |\partial_\mu \psi(0)\rangle \, \delta\lambda^\mu + |r\rangle \,, \tag{23}$$

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

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

Thus,

$$\mathrm{Re}\,\langle\psi_0|\partial_\mu\psi(0)\rangle = 0 \tag{25}$$

and

$$\begin{aligned}&\mathrm{Re}\,\langle\psi_0|r\rangle \\ &\quad = -\frac{1}{2}\sum_{\mu,\nu} \langle\partial_\mu\psi(0)|\partial_\nu\psi(0)\rangle\, \delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2).\end{aligned} \tag{26}$$

We compute

$$\begin{aligned}F(0,\delta\lambda) &= |\langle\psi(0)|\psi(\delta\lambda)\rangle| = |1 + \langle\psi_0|\delta\psi\rangle| \\ &= \left|1 + \sum_\mu \langle\psi_0|\partial_\mu\psi(0)\rangle\, \delta\lambda^\mu + \langle\psi_0|r\rangle\right| \\ &= 1 + \mathrm{Re}\,\langle\psi_0|r\rangle - \frac{1}{2}\left(\sum_\mu \langle\psi_0|\partial_\mu\psi(0)\rangle\, \delta\lambda^\mu\right)^2 \\ &\qquad + o(|\delta\lambda|^2) \\ &= 1 - \frac{1}{2}\sum_{\mu\nu} g_{\mu\nu}\delta\lambda^\mu\delta\lambda^\nu + o(|\delta\lambda|^2), \end{aligned} \tag{27}$$

where $g_{\mu\nu}$ is given by (19). Note that when expanding the absolute value we used the fact that for real x, y around $x = y = 0$ we have

$$|1 + x + iy| = 1 + x + \frac{y^2}{2} + O(x^2 + y^4). \tag{28}$$

□

Lemma 3. *Let*

$$M = \begin{pmatrix} A & B \\ B^* & B^*A^{-1}B + C \end{pmatrix} \tag{29}$$

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

$$M = \begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}^* \begin{pmatrix} A^{-1} & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}. \tag{30}$$

□

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

$$\delta\rho = \sum_\mu \delta\lambda^\mu \partial_\mu \rho(0) + r, \tag{31}$$

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

$$\mathrm{Tr}\,(\partial_\mu \rho(0)) = 0, \tag{32}$$

$$\mathrm{Tr}(r) = 0. \tag{33}$$

From definition,

$$F(\rho_0, \rho_0 + \delta\rho) = \mathrm{Tr}\,\sqrt{\rho_0^2 + \sqrt{\rho_0}\delta\rho\sqrt{\rho_0}}. \tag{34}$$

Let $\rho_{0+} = P_+\rho_0 P_+^*$, $\delta\rho_+ = P_+\delta\rho P_+^*$, [VK: TODO: should adjoint be $\bullet^*$ or $\bullet^\dagger$?] $r_+ = P_+ r P_+^*$. One can see that the expression under the square root acts nontrivially only on $\rho_0(\mathcal{H})$, hence the trace can be computed in that subspace:

$$F(\rho_0, \rho_0 + \delta\rho) = \mathrm{Tr}\,\sqrt{\rho_{0+}^2 + \sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}}. \tag{35}$$

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 $0 < c_1 \leq c_2 < \infty$. 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 $[c_1, c_2]$:

$$\begin{aligned}&\sqrt{\rho_{0+}^2 + \sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}} \\ &\quad = \frac{-1}{2\pi i}\oint \sqrt{z}\left(\rho_{0+}^2 + \sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}} - z\right)^{-1} dz \\ &\quad = I_0 + I_1 + I_2 + o(|\delta\lambda|^2),\end{aligned} \tag{36}$$

where

$$\begin{aligned}I_k = \frac{(-1)^{k+1}}{2\pi i}\oint \sqrt{z}\left(\rho_{0+}^2 - z\right)^{-1} \\ \left(\sqrt{\rho_{0+}}\delta\rho_+\sqrt{\rho_{0+}}\left(\rho_{0+}^2 - z\right)^{-1}\right)^k dz.\end{aligned} \tag{37}$$

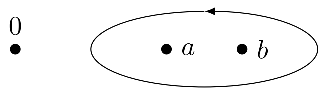
In order to evaluate the metric, we only need to compute the diagonal elements of I_0, I_1, I_2 discarding any terms with order $o(|\delta\lambda|^2)$. We pick the basis where ρ_{0+} is diagonal with diagonal elements $\xi_0 \geq \xi_1 \geq \cdots \geq \xi_{n_+-1} > 0$.

$$I_0 = \rho_{0+}, \tag{38}$$

$$(I_1)_{jj} = \frac{1}{2}\delta\rho_{+,jj} = \frac{1}{2}\sum_\mu \delta\lambda^\mu \partial_\mu \rho_{+,jj}(0) + \frac{1}{2}r_{+,jj}, \tag{39}$$

[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

$$\frac{1}{2\pi i}\oint \frac{\sqrt{z}}{(z-a)^2(z-b)}dz = \frac{-1}{2\sqrt{a}(\sqrt{a}+\sqrt{b})^2}. \tag{40}$$

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

$$(I_2)_{jj} = -\sum_k \frac{\xi_j \delta\rho_{+,jk}\xi_k\delta\rho_{+,kj}}{2\xi_j(\xi_j+\xi_k)^2} = -\sum_k \frac{|\delta\rho_{+,jk}|^2 \xi_k}{2(\xi_j+\xi_k)^2}$$
$$= -\sum_{k,\mu,\nu} \frac{\mathrm{Re}\left((\partial_\mu\rho_+(0))_{jk}(\partial_\nu\rho_+(0))_{kj}\right)\xi_k}{2(\xi_j+\xi_k)^2}\delta\lambda^\mu\delta\lambda^\nu$$
$$+ o(|\delta\lambda|^2). \tag{41}$$

Now we are ready to evaluate the trace

$$F(\rho_0, \rho_0+\delta\rho) = \mathrm{Tr}(I_0+I_1+I_2) + o(|\delta\lambda|^2). \tag{42}$$

In order to evaluate (17) we will include the terms up to the order $o(|\delta\lambda|^2)$:

$$\mathrm{Tr}(I_0) = \mathrm{Tr}(\rho_0) = 1. \tag{43}$$

The first term in $(I_1)_{jj}$ cannot have a non-zero contribution to $\mathrm{Tr}(I_1)$ due to the fact that $\rho(\lambda)$ is non-negative and has $\mathrm{Tr}(\rho(\lambda)) = 1$. For the second term, notice that $\mathrm{Tr}(r) = 0$, hence $\mathrm{Tr}(r_+) + \mathrm{Tr}(P_0 r P_0) = 0$, giving

$$\mathrm{Tr}(I_1) = -\mathrm{Tr}(P_0 r P_0^*)/2. \tag{44}$$

Note that according to lemma 3

$$P_0 r P_0^* \geq (P_+\delta\rho P_0^*)^* \rho_{0+}^{-1} P_+ \delta\rho P_0^*$$
$$= \sum_{\mu,\nu} \mathrm{Re}\left(P_0\left(\partial_\mu\rho^*(0)\right)P_+^*\rho_{0+}^{-1}P_+\left(\partial_\nu\rho(0)\right)P_0^*\right)\delta\lambda^\mu\delta\lambda^\nu$$
$$+ o(|\delta\lambda|^2). \tag{45}$$

Here $\mathrm{Re}(a) = (a+a^*)/2$ for a matrix or an operator a. Combining (41) and (44) we get

$$g_{\mu\nu}(0) = \sum_{j,k}\frac{\mathrm{Re}\left((\partial_\mu\rho_+(0))_{jk}(\partial_\nu\rho_+(0))_{kj}\right)\xi_k}{(\xi_j+\xi_k)^2}$$
$$+ \frac{1}{2}\,\partial_\mu\partial_\nu \,\mathrm{Tr}(P_0\rho(\lambda)P_0^*)|_{\lambda=0}. \tag{46}$$

The remaining statements of the part 2 of the theorem follow from the lemma 3. ☐

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

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

Theorem 4. *Suppose $|\psi(\lambda)\rangle$ is a state defined in the neighbourhood of $\lambda = \lambda_0 \in \mathcal{M}$ 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*

$$\mathbb{E}_{measurements}\, g_c(\lambda_0) = g(\lambda_0)/2, \tag{47}$$

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. The second term in (22) 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 (47). 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 (47) 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

$$\mathbb{E}_{\text{measurements}}\sum_z \frac{(\partial_0 p_z(0))^2}{4p_z(0)}$$
$$= \frac{1}{2}\left\||\partial_0\psi(0)\rangle\right\|^2 - \frac{1}{2}\left|\langle\partial_0\psi(0)|\psi(0)\rangle\right|^2. \tag{48}$$

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 integrate over $|\varphi\rangle$ s.t. $\||\varphi\rangle\| = 1$:

$$\text{l.h.s. of (48)} = n\mathbb{E}_{|\varphi\rangle}\frac{\left(\mathrm{Re}\left(\langle\varphi|\partial_0\psi(0)\rangle\langle\psi(0)|\varphi\rangle\right)\right)^2}{|\langle\varphi|\psi(0)\rangle|^2}. \tag{49}$$

Let's denote $|\psi_0\rangle = |\psi(0)\rangle$, $\alpha_\| = \langle\psi_0|\partial_0\psi(0)\rangle/i \in \mathbb{R}$, $|\psi_\|\rangle = i\alpha_\|\,|\psi_0\rangle$, $|\psi_\perp\rangle = |\partial_0\psi(0)\rangle - |\psi_\|\rangle$. With this notation

$$\text{r.h.s. of (48)} = \frac{1}{2}\left\||\psi_\perp\rangle\right\|^2, \tag{50}$$

$$\langle\varphi|\partial_0\psi(0)\rangle\langle\psi(0)|\varphi\rangle = i\alpha_\|\left|\langle\varphi|\psi_0\rangle\right|^2, \tag{51}$$

[VK: TODO:0]

$$\text{l.h.s. of (48)} = n\mathbb{E}_{|\varphi\rangle}\frac{\left(\mathrm{Re}\left(\langle\varphi|\partial_0\psi(0)\rangle\langle\psi(0)|\varphi\rangle\right)\right)^2}{|\langle\varphi|\psi(0)\rangle|^2}. \tag{52}$$

☐

Theorem 5. *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 $|\psi(s)\rangle$.*

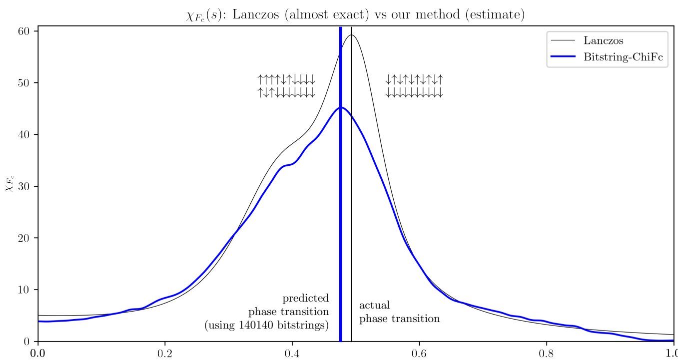
Proof. [VK: TODO:4] □

D. Resolution of issue 3

As we have seen in [VK: TODO: ref], [VK: TODO: ref], [VK: TODO: ref] 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 $\chi_{F_c}(s) = \chi_F(s)$ almost everywhere for the ground state and no information is lost when replacing $\chi_F(s)$ with $\chi_{F_c}(s)$ (for almost all values of s) if we are looking for phase transitions at zero temperature.

E. 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.



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

$$F_c(p, q) = \sum_z \sqrt{p(z)q(z)}. \quad (53)$$

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

$$F_c(s, s + \delta s) = 1 - \frac{\delta s^2}{2}\chi_{F_c}(s) + O(\delta s^3). \quad (54)$$

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 (54) would become

$$F_c(\lambda, \lambda + \delta\lambda) = 1 - \frac{\delta\lambda_j \delta\lambda_k}{2}\chi^{jk}_{F_c}(\lambda) + O(\delta\lambda^3). \quad (55)$$

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

$$\chi^{jk}_{F_c}(\lambda) = \text{smoothen}\Big(\lambda_1 \mapsto \text{mean}_{(z,\lambda_1)\in\mathcal{D}_{\text{train}}}\big(M(\lambda_1, 0, z)_j M(\lambda_1, 0, z)_k\big)\Big)(\lambda). \quad (56)$$

[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.]

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[2] P. Zanardi, P. Giorda, and M. Cozzini, Physical review letters **99**, 100603 (2007).