SciPost manuscript

Unsupervised and supervised learning of interacting topological phases from single-particle correlation functions

Response to the Referee reports

Referee 2

We thank the Referee for reading carefully our paper. We appreciate that they recognised the difficulty of this task and appreciated the novelty of using the correlation functions. We address the Referee's comments in the following:

> **Weaknesses**
>
> 1) The unsupervised learning approach based on PCA is not very convincing
> 2) The choice of the NN architecture for the supervised approach seems to be quite uncommon.
>
> **Requested changes**
>
> 1) The authors should explain more in detail how the quantities c(k) and f(k) are related to the winding number for the non-interacting model. In case there is no direct relation, they should explain why they believe such correlators should give sufficient information to determine the winding number.
> 2) The authors should comment on the difference between c(k) and f(k) for the non-interacting case, which has been computed for periodic boundary conditions, and for the interacting case which has been computed for open boundary conditions. In particular, I would expect differences coming from the boundary conditions. It would be actually fairer to provide a training set where c(k) and f(k) have been computed for open boundary conditions also for the non-interacting case.
> 3) Following the indications of the text, one should expect to have a nice clustering In the plane p1-p4. I, therefore, recommend adding this plot with a coloring that corresponds to the different phases (similar to Fig. 2 of [Phys. Rev. B 94, 195105 (2016)]). I expect to see in such a plot a nice clusterization between trivial and topological and a continuous transition between $\nu = 1$ and $\nu = -1$. I recommend doing such a figure for the non-interacting and interacting case.
> 4) It would be interesting to see whether more powerful dimensionality reduction techniques such as t-SNE or UMAP would allow for a better clustering of the data.
> 5) In the K-means section, it would actually be instructive to add a similar plot to Fig 4b with the labels of the clusters found by K-means. This could be done for one single run of K-means or with the help of a majority vote.
> 6) Could the authors also comment on the lines of low values of S in the TRI phases? It seems to me that such lines also appear in Fig. 2b of the PCA analysis.
> 7) Same comment as 5) for the interacting case. It would be interesting to see the clusterization performed by the algorithm as an additional plot in Fig. 5.
> 8) The authors should comment on the choice of the NN architecture for the supervised learning scheme. The choice of a 2D convolutional network for 1D-like data is not usual. I would have expected them to use a one-dimensional CNN with two input channels (one for c(k) and one for f(k)).
> 9) Could the authors comment on the reason for the high standard deviation on the training set after training? I would have expected a much smaller standard deviation if the networks were trained properly. Is this high error coming from the points close to the phase transitions?
> 10) Can the authors confirm that the network is also able to predict a negative winding number in the interacting case?

About the changes:

1) The winding number is obtained calculating the integral:

$$\omega = \int_0^{2\pi} \frac{h_y(k)\partial h_z(k) - h_z(k)\partial h_y(k)}{E(k)} \tag{1}$$

with $E(k) = \sqrt{h_y^2(k) + h_z^2(k)}$, $h_z(k) = J\cos k + \mu/2$ and $h_y(k) = \Delta \sin k$. Since, the correlation functions can be written as:

$$c(k) = \frac{1}{2} + \frac{\mu/2 + J\cos k}{2E(k)} = \frac{1}{2} + \frac{h_z(k)}{2E(k)}, \tag{2}$$

$$f(k) = \frac{\Delta \sin k}{2E(k)} = \frac{h_y(k)}{2E(k)}. \tag{3}$$

we can notice that the winding number can be extracted from $c(k)$ and $f(k)$.

2) In general, bulk and global properties of correlators should not depend on the choice of boundary conditions. In order to check this in the model we considered, we show this results in Fig. 1 for a system of $L = 100$ and different values of the chemical potential.

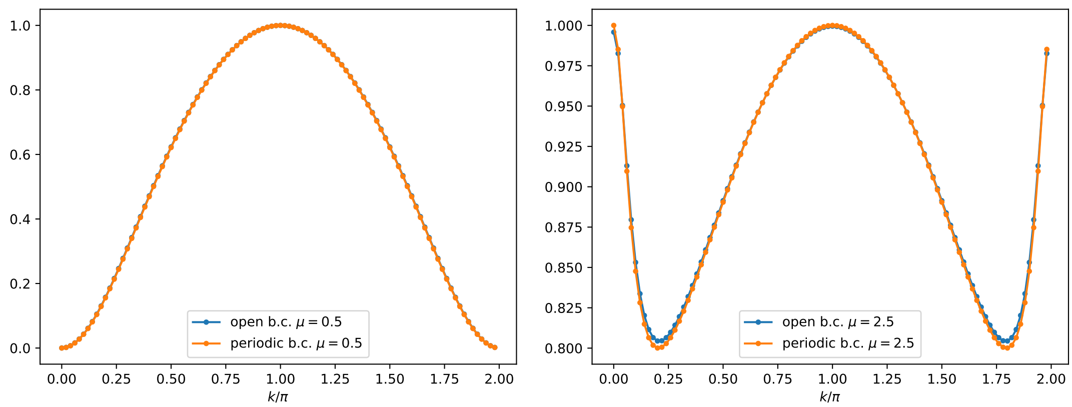

Figure 1: Correlators obtain with open and periodic boundary conditions for different values of μ.

3) We thank the Referee for pointing out the reference [Wang, PRB 2016] that we incorporated in the reference list of the revised manuscript as Ref. [10]. In Fig. 2 (left panel) we show the component p_4 (y-axis) vs. the component p_1 (x-axis) for the non interacting data labelled by their winding number. As predicted by the referee, we can indeed see clusterization of the topological trivial vs. the non-trivial phases. However we prefer not to include this plot in the main text of the manuscript as we believe it would provide redundant information. Regarding the interacting model, the clustering of the topological vs non-topological phases is not clearly visible when plotting the component p_4 vs. the component p_1 (Fig. 2 (right panel)).

4) For the sake of simplicity and interpretability we only considered PCA and Kmeans. For completeness in Figure 3 we show the clusterization performed by tSNE and we can see that the algorithm recognizes points in different clusters.

5) We thank the Referee for pointing this out. We added the plots as the Referee suggested.

6) We agree with the Referee that these lines are peculiar but unfortunately we do not have a clear understanding of the behaviour of the silhouette at these points. This could probably be due to the change of sign of Δ which affects the shape of the $f(k)$ correlators. Nonetheless, the values of the silhouettes along those lines are around 0.5 so still very far from 0, which excludes a possibile phase transition.

7) See item 5)

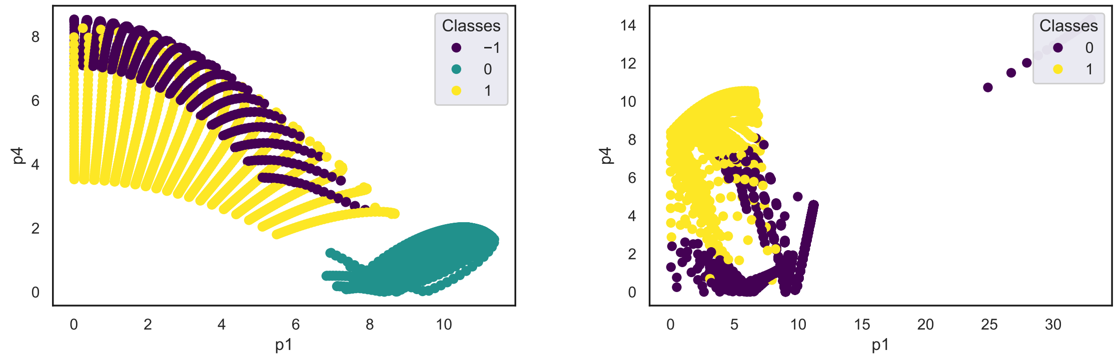

Figure 2: Plot of the two principal components for the non-interacting (left panel) and the interacting (right panel) models.

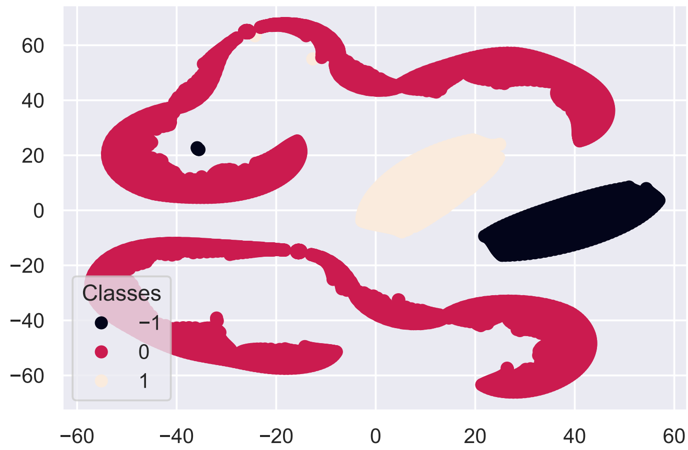

Figure 3: Clusters obtained with the tSNE algorithm.

8) The choice of a 2D CNN was motivated by the work of [Zhang et al. PRL(2018)] where the authors train a CNN with a 2D input.

9) We thank the Referee for pointing this out. The sentence is indeed referring to the predictions of the test set. Therefore we moved the sentence to the paragraph *Testing*.

10) In the model we consider, there is no phase with negative topological indicator. The correlators of the TOP phase of the interacting model, loosely speaking, resemble the correlators of the TOP+1 non-interacting phase as they are computed for $\Delta = +1$. We believe that if the interacting Hamiltonian had $\Delta = -1$ the correlators would resemble the ones of the TOP-1 phase and so the CNN could predict negative winding numbers.