Joint machine learning analysis of muon spectroscopy data from different materials

T Tula[1], G Möller[1], J Quintanilla[1], S R Giblin[2], A D Hillier[3], E E McCabe[1,4], S Ramos[1], D S Barker[1,5] and S Gibson[1]

[1] School of Physical Sciences, University of Kent, Park Wood Rd, Canterbury CT2 7NH, United Kingdom
[2] School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, United Kingdom
[3] ISIS Facility, STFC Rutherford Appleton Laboratory, Chilton, Didcot Oxon, OX11 0QX, United Kingdom
[4] Department of Physics, Durham University, Durham, DH1 3LE, United Kingdom
[5] School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, United Kingdom

E-mail: `tt343@kent.ac.uk`

Abstract. Machine learning (ML) methods have proved to be a very successful tool in physical sciences, especially when applied to experimental data analysis. Artificial intelligence is particularly good at recognizing patterns in high dimensional data, where it usually outperforms humans. Here we applied a simple ML tool called principal component analysis (PCA) to study data from muon spectroscopy. The measured quantity from this experiment is an asymmetry function, which holds the information about the average intrinsic magnetic field of the sample. A change in the asymmetry function might indicate a phase transition; however, these changes can be very subtle, and existing methods of analyzing the data require knowledge about the specific physics of the material. PCA is an unsupervised ML tool, which means that no assumption about the input data is required, yet we found that it still can be successfully applied to asymmetry curves, and the indications of phase transitions can be recovered. The method was applied to a range of magnetic materials with different underlying physics. We discovered that performing PCA on all those materials simultaneously can have a positive effect on the clarity of phase transition indicators and can also improve the detection of the most important variations of asymmetry functions. For this joint PCA we introduce a simple way to track the contributions from different materials for a more meaningful analysis.

1. Introduction

In recent years, machine learning tools have become increasingly familiar to the scientific community, but they are not yet regarded as standard practice for analysing small-scale experimental measurements in condensed matter physics. The strength of ML techniques lies in their ability to recognise patterns and correlations in data [1, 2, 3], which is often the goal of scientific enquiries (such as the search for phase transitions). Usually, for the ML methods to work, a large data set is required and since data acquisition can be costly for many experiments (in both time and resources), it is understandable that ML tools are not always thought to be a suitable tool. In this work, we want to test whether experimental muon spectroscopy measurements can be analysed with ML even when using smaller sets of data. Specifically we

apply a linear principal component analysis algorithm to muon spectroscopy measurements [4] in order to find signs of phase transitions for a range of different materials.

In this article we report on a new development from our previous research [5], in which we demonstrated that PCA can be successfully applied to a range of different materials simultaneously, thus increasing the effectiveness of this ML method, even when the underlying physics of the materials are vastly different. To have more insight into the PCA results in this case, we present a cumulative principal component score metric, which tells us how important each PC output is for a given material. That allows us to look at phase transition indicators in correct principal component scores, and this improved our previous analysis of magnetic materials.

2. Methods
2.1. Muon spectroscopy experiments
The main quantity to analyse from muon spectroscopy experiments is the asymmetry function, which tells us the distribution of spin directions of muons implanted into the sample. This can be directly linked to the internal magnetic field of the material studied [6, 7, 8, 9]. We are interested in studying changes in the asymmetry functions at different temperatures, which can indicate phase transitions. A more comprehensive introduction to muon spectroscopy can be found in [4, 5].

2.2. PCA
Principal component analysis is an unsupervised machine learning technique [3, 10], which means that we are not required to label the input data, and given that there are some correlations between asymmetry functions, PCA will be able to find those. Applying PCA is equal to performing singular value decomposition on a matrix containing all asymmetry functions $\mathbf{A}$. Principal components (PC) are given by left-singular vectors of $\mathbf{A}$, with singular values signifying the importance of each PC. By looking at projections of asymmetry functions into the most important PCs, we can see changes in the behaviour that could indicate phase transitions. More on the PCA method as applied to μSR data can be found in our previous article [5]

3. Results and Discussion
We applied PCA to a range of different materials. For each of them we had around 10-30 asymmetry functions, measured at different temperatures and at vanishing external magnetic field throughout. Among the systems studied are time-reversal symmetry breaking superconductors ($LaNiGa_2$, $LaNiC_2$, $LaNi_{0.9}Cu_{0.1}C_2$, $LaNi_{0.75}Cu_{0.25}C_2$), a proposed spin liquid ($LuCuGaO_4$) and a layered antiferromagnet ($BaFe_2Se_2O$). We initially tried performing the analysis for each material individually, but found out that for some materials the small number of measurements resulted in most PCs having a significant contribution to covariance. To overcome that, we performed PCA on the joint dataset of measurements from all materials simultaneously. This improved the visibility of phase transitions for those measurements that were previously hard to analyse, and did not significantly alter the other results. This is remarkable given that the materials in the data set have very different underlying physics. We can interpret this result as follows: by feeding all the data to the PCA simultaneously, the algorithm is not learning anything new about any particular material, but it is learning more about the experimental technique being employed in all cases (μSR). A comparison of the shapes of the ensuing principal components – with PCA performed on one material at a time or simultaneously on all of them – can be found in our previous work [5].

We now demonstrate that more insight can be gained from the results of a PCA analysis performed on multiple samples by calculating the total score of a given PC for each individual material. Unlike analysis for a single material, where PC scores are naturally ordered in
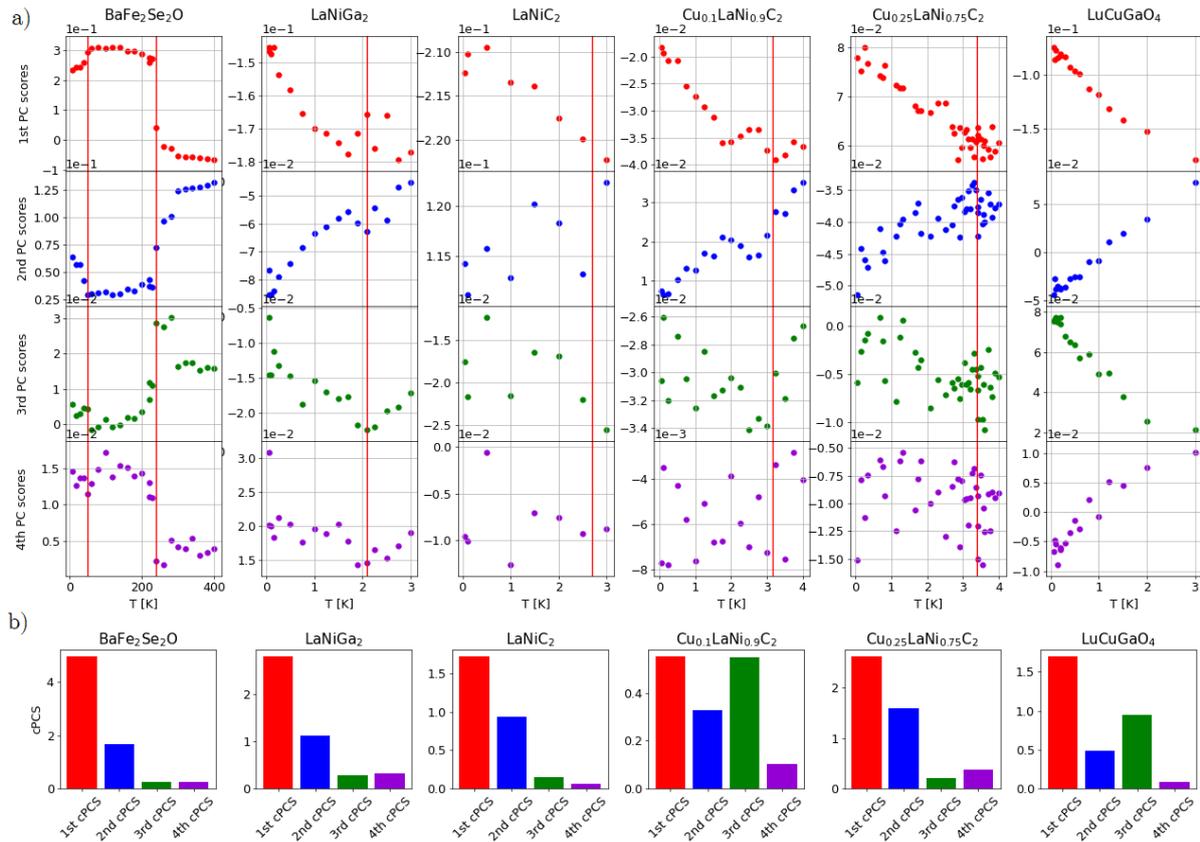
Figure 1. a) Temperature dependence for principal component scores when applying PCA to multiple materials simultaneously. Red vertical lines correspond to expected phase transitions, taken from the literature; b) Cumulative principal component scores for all materials.

terms of decreasing significance, in a sample of several materials considered jointly, the order of significance can depend on the material: Some PC shapes might be dominant in most of materials, meaning that they will have large contribution to the covariance, but for other materials another PC might be more prominent. To interpret the analysis, we should always examine the most important PC(s) for each sample. To identify the relevant scores, we define a cumulative principal component score (cPCS) as

$$\mathrm{cPCS}(i;n) = \sum_{j=1}^{N_i} |\mathrm{PCS}_n^i(j)|, \tag{1}$$

where i iterates over different materials, n stands for n-th principal component, $\mathrm{PCS}_n^i(j)$ corresponds to n-th principal component score for a j-th measurement of a given material and N_i stands for number of measurements for i-th material. We then use this measure to identify the principal component scores with the largest cPCS for each material.

Our findings are presented in figure 1, which is an extension of figure 5 from [5]. We present the temperature dependence of the first four principal component scores for all the materials in a), and their cPCS in b). From the cPCS we can see that although the 1st PC has the largest contribution to all of measurements, the 2nd PC is less important than the 3rd one for $Cu_{0.1}LaNi_{0.9}C_2$ and $LuCuGaO_4$. If we compare the principal component scores for $Cu_{0.1}LaNi_{0.9}C_2$, we can clearly see that the 3rd PC has a much more visible indication of the

expected phase transition (red vertical line) than the 2nd or even 1st one. For $LuCuGaO_4$ material, we expected improvement from our previous joint analysis, and although we find that 3rd PC is more important than 2nd, there seems to be no feature indication at 0.6 K, which we have found in principal component analysis of just this material ([5], figure 4, last column, third row). This tells us that joint analysis is most helpful whenever the separate PCA is not successful, which we can always tell by looking at scree plots. In the case of $LuCuGaO_4$, the cPCS metric looks worse than scree plot of its separate analysis ([5], figure 4, last column, last row), and we should take that into consideration when analysing the material. One important note here is that adding $LuCuGaO_4$ did not affect negatively the results for other materials, which we have showed previously [5].

4. Conclusions

We have suggested a new analysis method to extend our recently propsed PCA of μSR data [5]. We are more confident that performing a joint PCA of measurements taken for different materials is a viable method of analysis, given that for each set of measurements we are looking at the right PC scores, which can be ensured by calculating a simple metric – cumulative principal component scores. We should also compare the cPCS to scree plots from separate analyses conducted for each material, in case the simultaneous PCA might not improve the clarity of each individual study. The ability to thus enhance our understanding gained from a new experiment based on existing data goes beyond the possibilities of preexisting approaches, where data for each material is necessarily analyzed and fitted in isolation, and overarching commonalities are anticipated in advance by the formulation of a suitable fitting function. We anticipate this could offer great advantage when deployed in large-throughput user facilities.

References

[1] Zdeborová L 2017 *Nat Phys* **13** 420–421
[2] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 2773
[3] Géron A 2019 *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (O'Reilly Media)
[4] Blundell S J 1999 Muon-spin rotation : a brief introduction [Online; accessed 11. Sep. 2020] URL http://users.ox.ac.uk/ sjb/musr/musr.html
[5] Tula T, Möller G, Quintanilla J, Giblin S R, Hillier A D, McCabe E E, Ramos S, Barker D S and Gibson S 2021 *Journal of Physics: Condensed Matter* **33** 194002
[6] Schenck A G E 1985 *Muon Spin Rotation Spectroscopy Principles and Applications in Solid State Physics* (CRC Press) ISBN 978-085274551-9
[7] S F J Cox 1987 *Journal of Physics C: Solid State Physics* **20** 3187
[8] Blundell S J 1999 *Contemporary Physics* **40** 175–192
[9] Lee S L, Cywinski R and Kilcoyne S H (eds) 1999 *Muon science: Muons in physics, chemistry and materials* (CRC press)
[10] Jolliffe I T and Cadima J 2016 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374** 20150202