

Optical Fe II and near-infrared Ca II emission in active galaxies

Swayamtrupta Panda^{1,2,3}

1. Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

2. Laboratório Nacional de Astrofísica - MCTIC, R. dos Estados Unidos, 154 - Nações, Itajubá - MG, 37504-364, Brazil

3. Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, 00-716 Warsaw, Poland

The CaFe Project involves the study of the properties of the low ionization emission lines (LILs) pertaining to the broad-line region (BLR) in active galaxies. These emission lines, especially the singly-ionized iron (Fe II) in the optical and the corresponding singly-ionized calcium (Ca II) in the near-infrared (NIR) are found to show a strong correlation in their emission strengths, i.e., with respect to the broad $H\beta$ emission line, the latter also belonging to the same category of LILs. We outline the progress made in the past years that has developed our understanding of the location and the efficient production of these emission lines. We have yet to realize the full potential of Ca II emission and its connection to the black hole and the BLR parameters which can be useful in - (1) the classification of Type-1 active galactic nuclei (AGNs) in the context of the main sequence of quasars, (2) to realize an updated radius-luminosity relation wherein the inclusion of the strength of this emission line with respect to $H\beta$ can be an effective tracer of the accretion rate of the AGN, and, (3) the close connection of Ca II to Fe II can allow us to use the ratio of the two species to quantify the chemical evolution in these active galaxies across cosmic time. In this paper, we use our current sample and utilize a non-linear dimensionality reduction technique - t-distributed Stochastic Neighbour Embedding (tSNE), to understand the clustering in our dataset based on direct observables.

1 The CaFe Project - past, present and future

The CaFe Project involves the study of the properties of the low ionization emission lines (LILs) pertaining to the broad-line region (BLR) in active galaxies. These emission lines, especially the singly-ionized iron (Fe II) in the optical (4434-4684 Å blend) and the corresponding singly-ionized calcium (Ca II) in the near-infrared (NIR, the Ca II is seen as a triplet $\lambda 8498$, $\lambda 8542$ and $\lambda 8662$ Å and collectively called CaT) are found to show a strong correlation in their emission strengths, i.e., with respect to the broad $H\beta$ emission line, the latter also belonging to the same category of LILs (see Figure 1). The origin of this correlation is attributed to the similarity in the physical conditions necessary to emit these lines - especially in terms of the strength of the ionization from the central continuum source and the local number density of available matter in these regions. We refer the readers to Panda et al. (2020); Panda (2021); Martínez-Aldama et al. (2021b) for a detailed overview on the issue wherein we discuss the robustness of the correlation recovered both from the observational and modeling standpoints. The combined importance of the two species has also been recognized in addition to our findings that the CaT being an effective proxy serves to be a better alternative to Fe II-based $R_{H\beta} - L_{5100}$

relation (Martínez-Aldama et al., 2021a). This is crucial to address the scatter seen due to the inclusion of newer measurements and sources in the $R_{H\beta} - L_{5100}$ relation specifically showing a deviation from the classical two-parameter $R_{H\beta} - L_{5100}$ relation (Bentz et al., 2013). A large subset of these sources are noted to belong to the class of Narrow-line Seyfert Type-1 galaxies (NLS1s) that typically show shorter time delays, and hence a smaller radial distance of the onset of the BLR from the central continuum source (R_{BLR}) is deduced for these sources. These smaller R_{BLR} values show a marked deviation from the expected $R_{H\beta} - L_{5100}$ relation, and addressing this problem is key to our understanding of how these systems evolve and if viable corrections to the classical relation can be made to utilize AGNs as “standardizable” cosmological candles (Czerny et al., 2021). In addition to this issue, we find in Martínez-Aldama et al. (2021b) that the ratio of the CaT to Fe II (justifying the project name - CaFe) is an effective tracer of the chemical evolution of AGNs and can help us probe the co-evolution of the AGN and its host galaxy in more detail. Also, our analysis in Martínez-Aldama et al. (2021b) has shown the potential of recovering the main sequence for quasars using spectral properties of NIR emission lines, i.e. O I $\lambda 8446$ and CaT. The main sequence of quasars (Boroson & Green, 1992; Sulentic et al., 2000; Marziani et al., 2018) is an outcome of a classical linear dimensionality reduction technique - principal component analysis¹ (PCA), and the primary eigenvector relates to the correlation seen between the full-width at half maximum (FWHM) of broad H β and the optical Fe II strength (i.e., the R_{FeII}^2). As the correlation is drawn using broad emission lines seen in a quasar spectrum, this applies to only Type-1 AGNs - those where the distant observer has an almost uninterrupted view of the central engine (i.e. the accreting supermassive black hole) and its immediate surroundings, e.g. BLR that gives rise to these broad emission lines. In Martínez-Aldama et al. (2021b), we find that the FWHM(O I) recovered for our sample shows an almost 1-to-1 relation with FWHM(H β). This result compounded with the strong $R_{\text{FeII}} - R_{\text{CaT}}$ relation suggests the presence of a similar correlation between the FWHM(O I) and R_{CaT} . This is shown in Figure 1 of Martínez-Aldama et al. (2021b).

2 How to assess high-dimensional data properly

Visualization of high-dimensional data is crucial to improve our understanding of how a system behaves, especially factoring in the importance of observables that drive the dataset. Datasets in astronomy are no different and with the huge influx of data that is already upon us, more intelligent classification and clustering schemes are needed, in addition to making them human-readable. We thus rely upon dimensionality reduction techniques - where an n-dimensional dataset can be transformed and represented in a hyper-plane consisting of only a few *hyper*-dimensions. Dimensionality reduction aims to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map.

In the case of principal component analysis (PCA, see Jolliffe, 2011), a classical *linear* dimensionality reduction technique³, these hyper-plane dimensions are con-

¹wherein the dataset is transformed to be represented in sets of eigenvectors and their corresponding eigenvalues

² R_{FeII} is the ratio of the optical Fe II emission within 4434-4684 Å normalized to the H β emission, and, R_{CaT} is the ratio of the NIR CaT emission normalized also to the same H β emission.

³Over the last few decades, a variety of techniques for the visualization of such high-dimensional

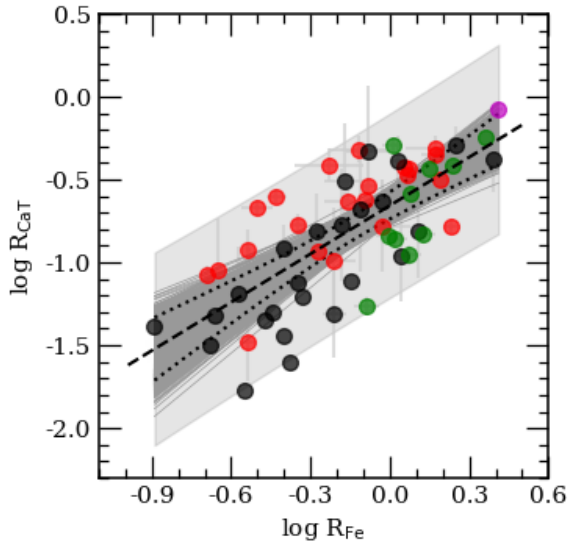


Fig. 1: $R_{\text{FeII}}-R_{\text{CaT}}$ relation shown for our sample (in log-scale). Black, red, green and magenta dots correspond to Persson (1988), Martínez-Aldama et al. (2015a,b), Marinello et al. (2016) and Marinello et al. (2020) samples, respectively. Black dotted lines mark the confidence intervals at 95% for a random 1000 realizations (dark gray lines) of a bootstrap analysis. Light gray shaded region marks the corresponding 95% prediction intervals bands. Figure courtesy: Martínez-Aldama et al. (2021b).

structured using singular value decomposition (SVD), such that the axes generated are orthogonal to each other. The original dataset is then represented in this hyper-plane where they are organised based on their respective contributions to these new dimensions⁴. But, there is a major drawback with PCA or PCA-based methods, i.e. the new axes are a linear combination of original observables. Thus, no cross-terms are involved when constructing the new dimensions. Therefore users of PCA need to be aware of this assumption and refrain from injecting datasets where prior inter-dependencies are noticed. Yet, PCA remains one of the most sought after methodologies, primarily due to its easy-to-understand algorithm.

For datasets with inter-dependent observables, one proceeds to look for other methods, for example, t-distributed Stochastic Neighbour Embedding (tSNE, van der Maaten & Hinton, 2008) which is a non-linear dimensionality reduction technique that aims to preserve the local structure of data. The foundation of tSNE is based on

data have been proposed, many of which are reviewed by Ferreira de Oliveira & Levkowitz (2003) and Baron (2019).

⁴In Martínez-Aldama et al. (2021b), we explored the observational properties of our sample and performed a principal component analysis, where 81.2% of the variance can be explained by the first three principal components drawn from the full width at half maximum (FWHMs), continuum luminosity of the sources, and the equivalent widths (EWs) of the emission lines considered. The first principal component (PC1) is primarily driven by the combination of black hole mass and luminosity with a significance over 99.9%, which in turn is reflected in the strong correlation of the PC1 with the Eddington ratio. We also discuss the biases introduced in PCA due to the use of input observables that are either derived from existing observables, i.e. black hole mass is derived from two basic observables, FWHM of the broad emission line observed in the AGN spectrum and distance of the BLR from the continuum source.

minimizing a cost function that is dependent on the nearness of a data point from a defined cluster center. The algorithm orders the dataset by assigning a rank to each data point and then organizing them on the hyper-plane. In this way, the dataset is represented based on similarity which is estimated based on the input observables for each data point. tSNE employs a heavy-tailed distribution in the low-dimensional space to alleviate both the crowding problem and the optimization problem (see van der Maaten & Hinton, 2008, for a detailed overview of the technique).

3 Methods and Analysis

Our analysis is based on the observational properties of H β , optical Fe II and NIR CaT triplet collected from Persson (1988), Martínez-Aldama et al. (2015a,b), Marinello et al. (2016) and Marinello et al. (2020). The full sample includes 58 objects with $42.5 < \log L_{\text{opt}} (5100 \text{ \AA}) < 47.7$ at $0.01 < z < 1.68$. Table A1 in Martínez-Aldama et al. (2021b) reports the properties of the each source in the sample such as redshift, optical (at 5100 \AA) and NIR (at 8542 \AA) continuum luminosity, the flux ratios R_{FeII} and R_{CaT} , as well as the equivalent width (EW) and Full-Width at Half Maximum (FWHM) of Fe II, H β , CaT and O I (see Table A1 in Martínez-Aldama et al. 2021b).

In this work, we utilize the PYTHON implementation of the tSNE⁵ to analyze the clustering in our sample accounting for the aforementioned spectral properties for each source, specifically, the L_{opt} , FWHMs of Fe II, H β , CaT and O I, and the EWs for the same set of emission lines. This prescription of choosing only the direct observables was shown to provide clear evidence of the primary driver of our sample reducing the biases arising from inter-dependencies due to the inclusion of derived parameters, like BH mass or Eddington ratio ($L_{\text{bol}}/L_{\text{Edd}}$). The 2D hyper-plane generated using the tSNE is shown in Figure 2. The sources are categorized based on the papers that reported them.

4 Results & Conclusions

In this work, we utilize a non-linear dimensionality reduction technique - tSNE, to understand better the organization of our current sample from our CaFe Project. The preliminary testing has revealed a dichotomy in our sample when visualized in a 2D tSNE-based hyper-plane - the overall sample is composed of two clusters driven by their observed optical luminosities and black hole masses (and hence, their Eddington ratios). The analysis has also revealed the movement of sources in this hyper-plane which we observed more than once. As was found earlier from our direct correlation in Panda et al. (2020); Martínez-Aldama et al. (2021b), the source PHL1092 stands out as an outlier. Here, in addition to that, we notice that this source doesn't belong to either of the clusters. This can be due to the prominence of strong soft X-ray excess seen in its spectrum. This may explain the recovery of high values for $R_{\text{FeII}} (= 2.576 \pm 0.108)$ and $R_{\text{CaT}} (= 0.839 \pm 0.038)$ for this source. Another source, Mrk335, is also of potential interest, as it seems to have jumped to the high-luminosity cluster (red and blue dots) albeit belonging to a sub-sample whose remaining sources are located in the low-luminosity cluster (green and black

⁵<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

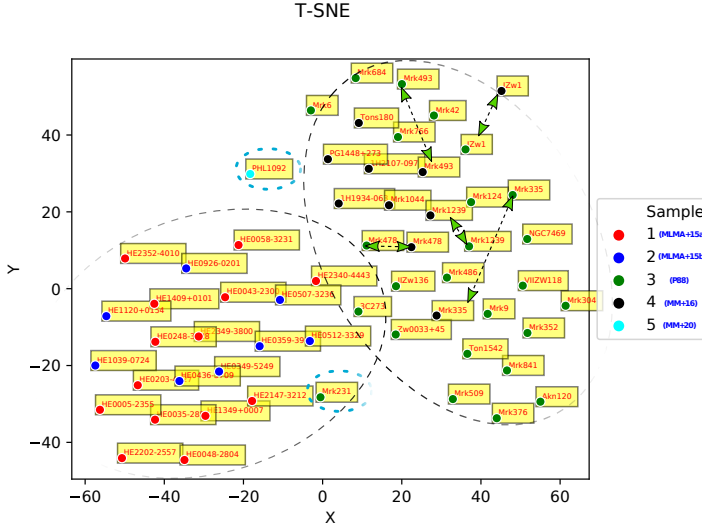


Fig. 2: Two-dimensional hyper-plane generated using the tSNE. Green, red, blue, black and cyan dots correspond to P88 - Persson (1988), MLMA+15a, MLMA+15b - Martínez-Aldama et al. (2015a,b), MM+16 - Marinello et al. (2016), and, MM+20 - Marinello et al. (2020) samples, respectively. The dashed ellipses are shown for illustrative purposes to highlight the dichotomy in the sample and mark sources of interest. The green arrows mark the movement of the five sources which were observed twice at different epochs.

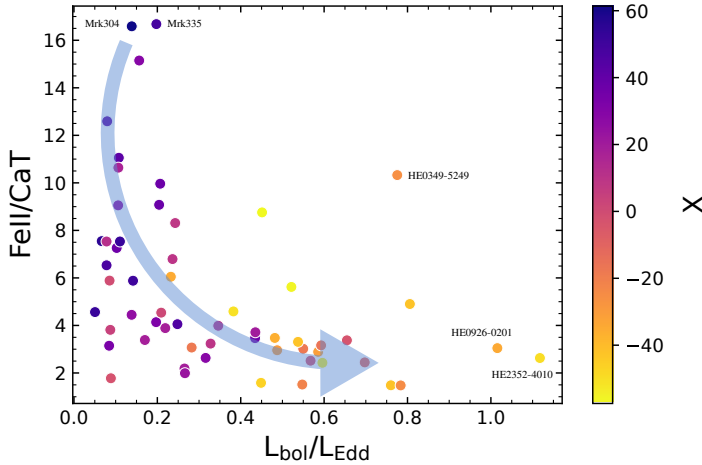


Fig. 3: $\text{Fe II}/\text{CaT}$ versus Eddington ratio ($L_{\text{bol}}/L_{\text{Edd}}$) for the sources in our sample. The color-scale represents the primary tSNE hyper-plane axis, X. The arrow marks the increase in the $L_{\text{bol}}/L_{\text{Edd}}$. Sources at extremes are marked.

dots).

We explore further the results from the tSNE analysis which reveals a strong and clear connection between the ratio $\text{Fe II}/\text{CaT}$ (this ratio has been studied in

Martínez-Aldama et al. (2021b) as a tracer of the chemical evolution in galaxies hosting accreting super-massive black holes) and the Eddington ratio ($L_{\text{bol}}/L_{\text{Edd}}$) for the sources in our sample - sources with lower Fe II/CaT belong to the high-luminosity class of objects that typical have relatively large $L_{\text{bol}}/L_{\text{Edd}}$ (see Figure 3). This result needs to be tested more carefully.

The advent of newer, deeper surveys including the James Webb Space Telescope (JWST, Gardner et al. 2006) and Maunakea Spectroscopic Explorer (MSE, Marshall et al. 2019), will lead to populating our sample by orders of magnitudes and firmly confirming these aforementioned avenues.

Acknowledgements. The project was partially supported by the Polish National Science Centre, project 2017/26/A/ST9/00756 (MAESTRO 9), MNiSW grant DIR/WK/2018/12 and acknowledges partial support from CNPq Fellowship (164753/2020-6).

References

- Baron, D., *arXiv e-prints* arXiv:1904.07248 (2019)
- Bentz, M. C., et al., *ApJ* **767**, 149 (2013)
- Boroson, T. A., Green, R. F., *ApJS* **80**, 109 (1992)
- Czerny, B., et al., *Acta Physica Polonica A* **139**, 4, 389 (2021)
- Gardner, J. P., et al., *Space Sci. Rev.* **123**, 4, 485 (2006)
- Jolliffe, I., Principal Component Analysis, 1094–1096, Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
- van der Maaten, L., Hinton, G., *Journal of Machine Learning Research* **9**, 86, 2579 (2008)
- Marinello, M., et al., *ApJ* **820**, 2, 116 (2016)
- Marinello, M., et al., *MNRAS* **494**, 3, 4187 (2020)
- Marshall, J., et al., in BAAS, volume 51, 126 (2019)
- Martínez-Aldama, M. L., Panda, S., Czerny, B., in XIX Serbian Astronomical Conference, volume 100, 287–293 (2021a)
- Martínez-Aldama, M. L., et al., *ApJS* **217**, 3 (2015a)
- Martínez-Aldama, M. L., et al., *JApA* **36**, 4, 457 (2015b)
- Martínez-Aldama, M. L., et al., *ApJ* **918**, 1, 29 (2021b)
- Marziani, P., et al., *FrASS* **5**, 6 (2018)
- Ferreira de Oliveira, M., Levkowitz, H., *IEEE Transactions on Visualization and Computer Graphics* **9**, 3, 378 (2003)
- Panda, S., *A&A* **650**, A154 (2021)
- Panda, S., et al., *ApJ* **902**, 1, 76 (2020)
- Persson, S. E., *ApJ* **330**, 751 (1988)
- Sulentic, J. W., Zwitter, T., Marziani, P., Dultzin-Hacyan, D., *ApJL* **536**, L5 (2000)