

Identification of black hole states using matrix based methods: Time series analysis of RXTE satellite data

Abstract—Black hole is one of the fascinating, however mysterious, astro-physical objects. In order to identify it one has to look at its environment, often forming a disc-like structure. This disc, called accretion disc, evolves with time transiting from one state to another. For example, in one extreme regime it shows temperature dependent radiations making the disc geometrically thin, and in other extreme regime of time span however radiation turns out to be temperature independent making the disc hot and geometrically thick. Nevertheless in general accretion disc lies in a state intermediate between the two extremes. The present mission is to capture black hole states explicitly using PCA and SVD based decompositions. In order to do that we rely on time series data of black hole *GRS 1915 + 105* obtained from RXTE satellite. As a black hole cannot be seen directly, identifying its states accurately could help in characterizing its properties. Earlier time series analysis based on correlation dimension methods, supplemented by theory, argued for four specific states. However there are caveats when data themselves are not free from noise and the appropriate method for such an analysis itself is exploratory. Present interdisciplinary study aims at, on one hand to cross-verify the previous inference, on the other hand to identify, if any, novel characteristics of black holes. This is expected to have long standing implications in astrophysics and otherwise.

I. INTRODUCTION

A challenging problem in astrophysics is the understanding of black holes. In this work, we focus on the black hole source *GRS 1915+105*, which presents several intriguing facets. One of the fundamental aspects of the understanding is to determine if the black hole source is a stochastic system or a non-stochastic one. The latter one is related to the well turbulent nature of the system. There are several studies that utilize the Correlation Integral (CI) approach to determine the characterization of the black hole data [1], [2]. However, there can also be other approaches to understanding the same data by applying, for e.g., matrix-based methods such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). It is useful to compare the inferences obtained using these two distinct approaches; the implications of the (dis)similarities in inferences, if any, could lead to questions about understanding the temporal dynamics of the system.

Following are the major contributions of this paper:

- We use PCA, an approach widely used for decorrelating features and dimensionality reduction, for characterizing a time series as stochastic vs non-stochastic. We propose a novel approach by iteratively computing eigen value ratios of covariance matrix for different subintervals of

the time series. We derive multiple features from the eigen value ratios and use them to characterize the time series. Utility of the proposed approach is illustrated by comparing the results with previously established methods.

- We use SVD decomposition of the data matrix for identifying the temporal dynamics of the time series. A plot involving top two right singular vectors of the data matrix shows a clear distinction between stochastic and non-stochastic time series.

II. RELATED WORK

Several groups have worked on distinguishing between stochastic and non-stochastic time series. In the work reported in [4], a Neural Network based approach is used. A Neural Network is trained with noise to learn the parametrization of stochastic signals. The paper explores the idea of utilizing permutation entropy (PE) of a time series to determine if it is strongly correlated with stochastic signal (noise). The idea of utilizing Permutation Entropy to determine the complexity measure of a time series was previously explored in [3]. The claim is that for Non-stochastic signals the deviation of the parameter is relatively large as compared to the parameter of a stochastic signal.

Yet another approach has been to utilize graph theoretical tools. In the work reported in [5], the authors have utilized the horizontal visibility algorithm in order to distinguish between stochastic and non-stochastic processes.

In the approach outlined in [7], the authors combine the idea of sparsity and machine learning with non-linear dynamical systems, in order to determine the governing dynamics. Sparse regression is used to determine the fewest terms in the equations that govern the dynamics of the phenomenon. The user-defined basis consists of well-known functions such as polynomials, trigonometric functions and exponentials. The coefficients corresponding to very few of these basis functions will be non-zero for a non-stochastic system. However, the optimal choice of dictionary remains a challenge.

In this work we propose to utilize matrix based methods which do not require any assumptions about the underlying phenomenon.

III. PROPOSED METHOD

In this work, we use two different matrix based approaches, one using PCA and the second using SVD, to characterize time

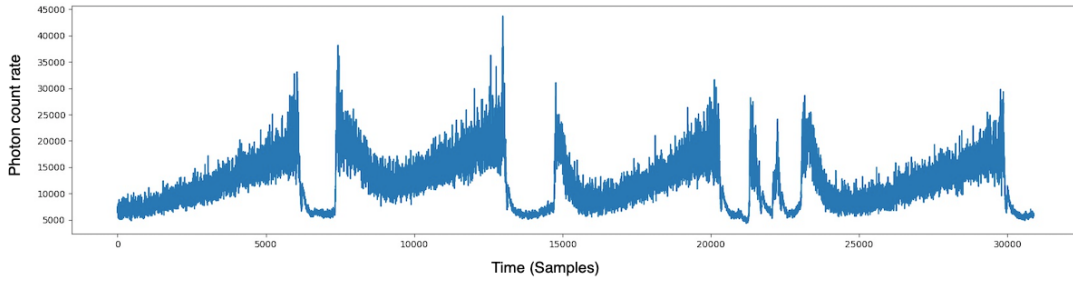


Fig. 1: A representative of stochastic time series of class θ

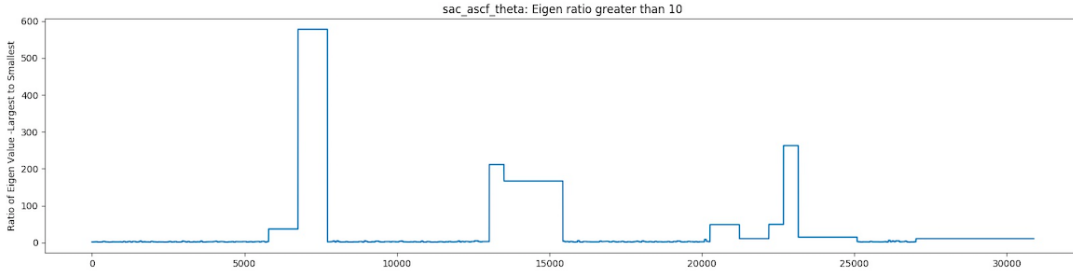


Fig. 2: Plot of eigen ratio of the stochastic time series shown in Figure 1

series as stochastic vs non-stochastic.

A. PCA Based approach

In the first approach, we utilize PCA to understand if the available data occupy a preferred orientation. This can be computed by splitting the time series into two halves, and computing the covariance matrix of these observations. The eigen values of this 2×2 covariance matrix will show one of the signatures : If the data indeed shows any preferred direction (as in Non-stochastic time series), then the larger eigen value will be significantly greater than the other. This will lead to a large ratio of the eigen values. On the other hand, if the data does not show any preferred direction (as in Stochastic time series), then the two eigen values of the covariance matrix will be comparable. This will lead to small values of eigen value ratio.

Consider a time series consisting of n values $z_1, z_2 \dots z_n$. We begin by computing the eigen ratio for the entire series using the following steps:

- Split the series into two halves $(z_1, z_2 \dots z_{\frac{n}{2}})$ and $(z_{\frac{n}{2}+1}, \dots z_n)$.
- Compute covariance matrix, C , by treating the samples in two halves as $\frac{n}{2}$ observations of two dimensional vectors.
- Compute eigen values, λ_1 and λ_2 of C , and the eigen ratio is computed as $\frac{\lambda_1}{\lambda_2}$ where $\lambda_1 > \lambda_2$ (Eigen values of a covariance matrix are real).

If eigen ratio for an interval is less than a predefined threshold (empirically determined as 10), the interval is split into two subintervals of equal size and eigen ratio for each sub-interval is computed. The process is repeated as long as the length

of the sub-interval is greater than a predefined number of samples.

Using PCA analysis we derived the following three features for each time series:

- Maximum eigen ratio (MER): This is the maximum value obtained as the ratio of the two eigen values of the covariance matrix of any sub-interval of the time series.
- Variance of eigen ratio: This is the variance of the eigen ratios of covariance matrices across sub-interval in the entire time series.
- Area under the eigen ratio curve: This measure captures the area under the curve of the eigen ratio for the entire time series.

B. SVD based approach

In this approach, we form uncorrelated observation vectors from the raw time series data by utilizing the optimal value of embedding dimension [2]. A data matrix, D , is formed with each row as the time shifted version of the original time series. The time shift is chosen to be large enough so that each row can be viewed as a different observation vector of the same phenomenon. The temporal dynamics is understood by utilizing the right singular vectors of the SVD decomposition of the data matrix as given in equation (1). Columns of U and V form the left and right singular vectors respectively and Σ is a block diagonal matrix with diagonal elements as the singular values.

$$D = U\Sigma V^T \quad (1)$$

The plot of pair-wise observation vectors is compared against the plot of the top two right singular vectors.

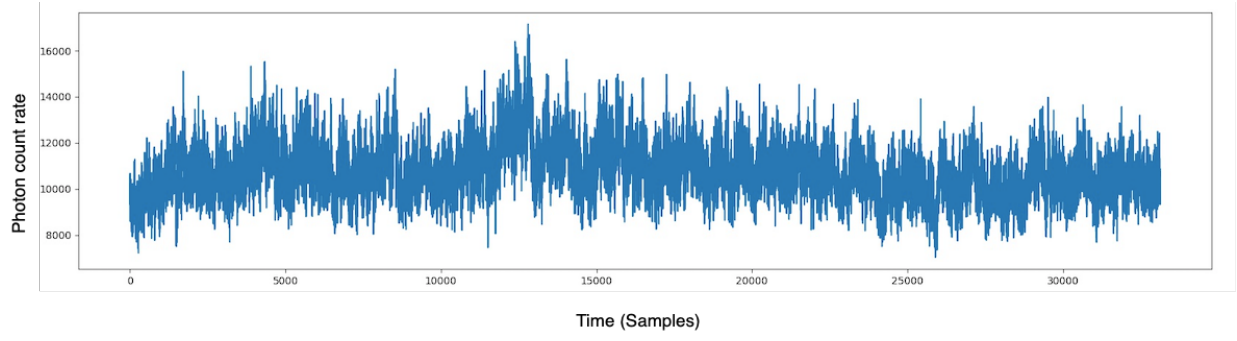


Fig. 3: A representative of non-stochastic time series of class ϕ

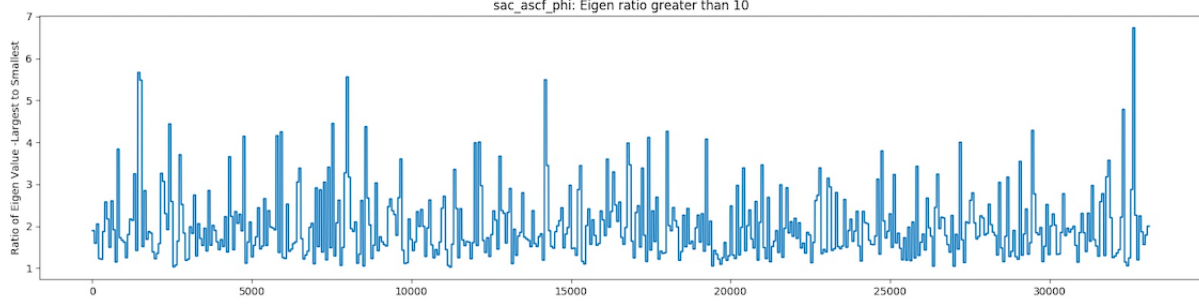


Fig. 4: Plot of eigen ratio of the non-stochastic signal shown in Figure 3

IV. RESULTS AND DISCUSSIONS

A. PCA Based Analysis

Figure 2 and 4 shows the maximum eigen ratio plot for stochastic and non-stochastic time series respectively. In our experiments using time series of 12 different classes, we noticed the following behavior for each of the PCA features:

- Max eigen ratio: For stochastic time series, eigen ratio values are small across the entire time series, typically lying in the range 1-20. This implies the maximum eigen ratio value will also be small. On the other hand, for the non-stochastic time series the eigen ratio values are significantly high, typically reaching a few thousands in certain sub-intervals. Hence the maximum eigen ratio value for a non-stochastic time series is typically large.
- Variance of eigen ratios: For a stochastic signal since the range of eigen ratio values is typically small, the variance is also small. On the other hand, for a non-stochastic signal, since the eigen ratio values occupy a large range of values, their variance is typically high.
- Area under the eigen ratio curve: For a stochastic time series, since the eigen ratio values are small the area under the curve is also small. However, for a non-stochastic signal the eigen ratio values remain high for longer time intervals. Hence the area under the curve is significantly higher.

The above characteristics have been exploited in the present study to distinguish between stochastic vs non-stochastic signals. Visualization of the time series in the space of these

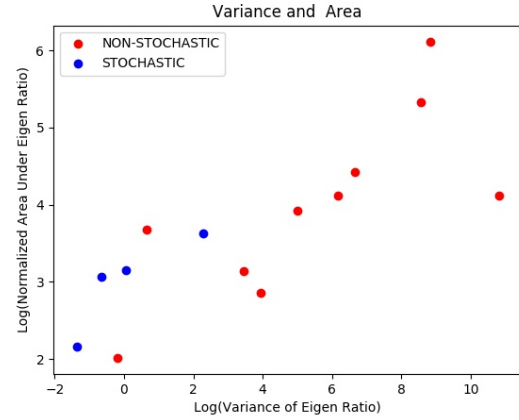


Fig. 5: Feature space (variance and area) shows that the two classes are well separated

derived features as shown in Figure 5 shows the clear separation between stochastic and non-stochastic time series for most of the cases. These two computed measures, area under the ER curve and variance take into account the entire span of time series and hence form robust feature space. Our inference for each of the time series using above guidelines is shown in Table I. It is noticed that our inference matches with analysis result based on CI in all the cases except for class δ . According to the CI analysis δ turns out to be in between class Slim disc and General advective accretion flow (GAFF) [6]. However, the present analysis shows that δ falls in between Advection

dominated accretion flow (ADAF) and Keplerian disc.

B. SVD based approach

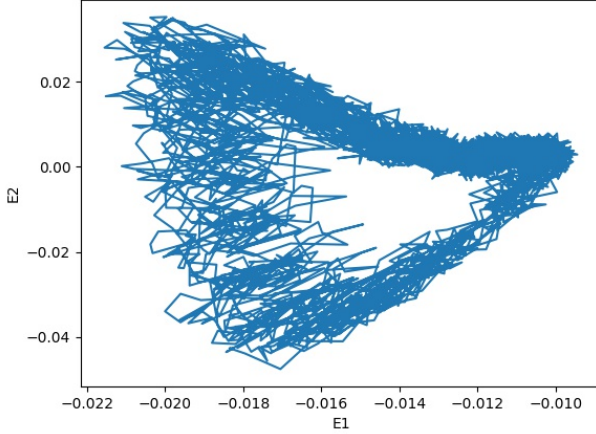


Fig. 6: Plot of E1 Vs E2 (top two right singular vectors) of data matrix for time series ρ (non-stochastic)

From SVD decomposition of the data matrix, we pick up the top 2 right singular vectors (E1, E2) corresponding to the temporal dynamics and plot E1 vs E2 for each time series. Figure 6 shows the plot for time series ρ which is of type Limit Cycle [2]. Figures 7 and 8 show the plots for class χ and λ which are stochastic and non-stochastic respectively. It is observed that the plot of right singular vectors follows certain pattern (attractor behaviour) in case of non-stochastic signal whereas no such pattern is visible in stochastic case.

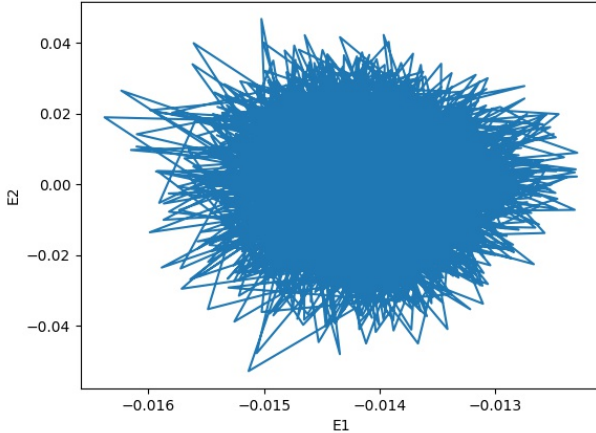


Fig. 7: Plot of E1 Vs E2 (top two right singular vectors) of data matrix for time series χ (stochastic)

V. CONCLUSION

Exploring different techniques in order to have a conclusive inference for black hole systems turns out to be indispensable.

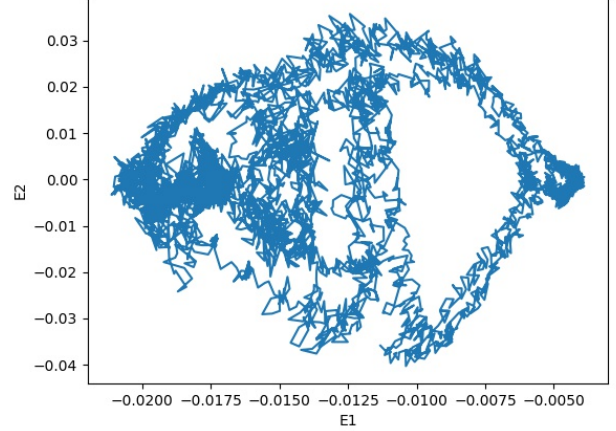


Fig. 8: Plot of E1 Vs E2 (top two right singular vectors) of data matrix for time series λ (non-stochastic)

We explore two different techniques for the first time in the literature to uncover properties of black holes from the time series obtained from satellite data. Based on our analysis, we are able to identify two extreme temporal dynamical classes of accretion around black holes. We use Principal Component Analysis, a widely used technique to identify an prominent directionality in the data, in order to characterize the time series as stochastic vs non-stochastic. We further extend the Correlation Integral based studies by performing Singular Value Decomposition on the data matrix and using the right singular vectors in order to study temporal dynamics. Our results are matching with previous Correlation Integral based studies in most of the cases.

REFERENCES

- [1] Mukhopadhyay, Banibrata. "Chaotic behavior of micro quasar GRS 1915+ 105." AIP Conference Proceedings. Vol. 714. No. 1. American Institute of Physics, 2004.
- [2] Misra, Ranjeev, et al. "The nonlinear behavior of the black hole system grs 1915+ 105." The Astrophysical Journal 643.2 (2006): 1114.
- [3] Bandt, Christoph, and Bernd Pompe. "Permutation entropy: a natural complexity measure for time series." Physical review letters 88.17 (2002): 174102.
- [4] Boaretto, B. R. R., et al. "Discriminating chaotic and stochastic time series using permutation entropy and artificial neural networks." Scientific reports 11.1 (2021): 1-10.
- [5] Lacasa, Lucas, and Raul Toral. "Description of stochastic and chaotic series using visibility graphs." Physical Review E 82.3 (2010): 036120.
- [6] Adegoke, Oluwashina, et al. "Correlating non-linear properties with spectral states of RXTE data: possible observational evidences for four different accretion modes around compact objects." Monthly Notices of the Royal Astronomical Society 476.2 (2018): 1581-1595.
- [7] Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." Proceedings of the national academy of sciences 113.15 (2016): 3932-3937.

TABLE I: Time series: comparison between correlation dimension behavior vs proposed PCA features. The mismatched time series class is shown in bold.

Class	CI-Behavior	Diskbb	PL	MER	Variance	Area	Inference	Match
β	F	46	52	214	483	43	Non-stochastic	Yes
θ	F	11	88	577	778	58	Non-stochastic	Yes
λ	F	54	46	600	6782	314	Non-stochastic	Yes
κ	F	59	51	700	5199	144	Non-stochastic	Yes
μ	F	56	41	50	51	12	Non-stochastic	Yes
ν	F	28	72	30	32	16	Non-stochastic	Yes
α	F	23	77	30	1.9	27.7	Non-stochastic	Yes
ρ	LC	28	72	60	147	35	Non-stochastic	Yes
δ	S	48	50	42	9.74	26.2	Non-stochastic	No
ϕ	S	50	34	7	0.5	15	Stochastic	Yes
γ	S	60	31	12	1	16	stochastic	Yes
χ	S	09	89	5.6	0.25	6.05	Stochastic	Yes