Microlensing Light-curve Anomaly Detection

Antonio Herrera-Martín
School of Physical and Chemical Science
University of Canterbury
Private Bag 4800
Christchurch, New Zealand
antonio.herreramartin@canterbury.ac.nz

M. D. Albrow
School of Physical and Chemical Science
University of Canterbury
Private Bag 4800
Christchurch, New Zealand
michael.albrow@canterbury.ac.nz

Abstract

In recent years, there has been an increase of surveys searching for extra solar planets, giving a consequent dramatic increase in the number of detected microlensing events and the amount of data observed for each event. Although it is still possible to search and classify a small number of events by human interaction, such a process is always prone to error, and is slow and labor-intensive. In this work, we address an initial stage for identifying features that will be useful in the task of distinguishing between single-lens and non-single-lens microlensing events to increase the efficiency of exoplanet discovery.

1 Introduction

One of the first effects to be described by General Relativity was the bending of light around massive bodies, as a consequence, the light is deviated from its straight path, and the massive object acts as a lens. Although Newtonian gravity also predicts this effect, the angle is half the predicted by General Relativity. This phenomenon is called gravitational lensing, and one of its consequences is the production of multiple images of a background luminous source; when the angular separation of the images is smaller than the resolution capabilities of telescopes, it is called microlensing [4].

In microlensing, unfortunately the only observable is the amplification of the light-curve of the background source. The most basic example is the case of a single lens with a single source; the model which describes the expected amplification has a closed form and is straightforward to compute, but if more lenses are present its description is more complex. Moreover, there are other additional effects which could produce a deviation from the basic model, for example, a binary source, an extended source, etc [3].

The most common cause for observed microlensing events is due to stellar-mass objects, and these are mainly observed in the direction of the galactic bulge. Likewise, the majority are produced by single and binary lenses, although the basic model give some information about the population and distribution of stellar objects on the galaxy, there is a stronger interest in the study of binary events because the possibility of one of the lenses being an exoplanet [e.g. 12, 10]. This is main focus in the modern microlensing surveys.

Given the possibility of exoplanetary discoveries, there is a preference for the study of binary lenses. As a result, the observation of microlensing events has increased in recent years, but at the same time this means that it is necessary to study and model a greater number of events [see 6, 4]. Events due to a single stellar source and single point-mass lens are relatively straighforward to model. However, binary lens events, are far from trivial, having a discontinuous non-linear parameter probability space of at least 7 dimensions. Typically, a binary (or planetary) microlensing model takes hours of computation time on a CPU cluster or GPU [8].

Despite the increase of the number of data, the different models are best-fitted at different stages, as is it not easy beforehand to identify what model is adequate. Furthermore, once a light-curve is observed, the identification of the possible type of event is most commonly done by human selection, which with the increase of observations makes the process slow or could leave some events without being classified at all.

Here we present the first stage of an ongoing work were we try to develop an identification system which would enter the pipeline to separate the low-priority single lenses from binary, possibly planetary, or more complex models, that need a more detailed analysis.

2 Data Mining

Over the years, the application of machine learning algorithms in microlensing has been quite limited, and it has been mainly focused in detection and identification over other astronomical processes [e.g. 13, 1]. There have been a small number of works that have implemented neural networks or evolutionary algorithms to model the parameters of binary models [11], but the preferred technique has been to do a parameter estimation through a χ^2 minimization [5]. These works have focused on identifying microlensing events from ensembles of light curves with other type of signals. Our task is rather to identify anomalous (non point-source point-mass lens) light curves from an ensemble of known microlensing light curves. The task is complicated by the presence of random and systematic noise, which is difficult to remove. There is currently no standard set of features to use for the classification task.

Likewise, although there is an increasing amount of data found since the 90's, there is a lack of a collected data-sets to use for classification so we need to selected such a set from modeled events across surveys.

2.1 Outlier Clustering

Once an signal is identified as a microlensing event, it is straightforward to fit a single-lens model, which has only 3 non-linear parameters, the time of the event t_E, the time of maximum amplification t_0 and the normalized impact parameter μ_0 [see 4]. Generating this model is computationally inexpensive. Taking this as an advantage, we can use it to find the sections of the light-curve which do not agree with it. But given the existence of noise in the data it is not easy to set a simple rule to discriminate what should be considered as bad points. Instead, we used different unsupervised algorithms to find data points which are adequate for the model.

As mentioned before, there is not a simple set of features to select, for this reason, we constructed and tried several features to be derived for each data point. As a result we were able to reduce it to only two features. The simplest is just the arithmetic difference between the data point and the model, $data_i - model_i$, which for a perfect fit should be zero. The second is

$$\chi_i^2 = (data_i - model_i)^2 / \sigma_i^2 \,, \tag{1}$$

where σ_i is the measurement error of the data point. This feature considers how well each individual point is fitted. Although, this could be used by itself, we found that it tends to underestimate points with large errors.

We combined and used Robust covariance, One-Class SVM, Isolation Forest and Local Outlier Factor(LOF) to find the outliers [9, 7, 2], and we set the rule that at least three of the algorithms must find the same point to be considered as a real outlier. Examples of single and binary lenses and how it can distinguish which sections of a light-curve belong adequately to the model is shown in Figure 1. This is important, as the capacity to detect which points can be considered as bad points for the fitted model will be used later.

2.2 Time Feature

Once we have found points which can be considered outliers and do not agree with the model, we need to decide how to obtain information from them. We took into consideration expected signal from random noise and from an actual anomaly. Given that the amplification could not be cleanly used, as it is possible to have a high valued set of points due to other unknown effects that are not part of the lens or source properties, we decided that it was clear that:
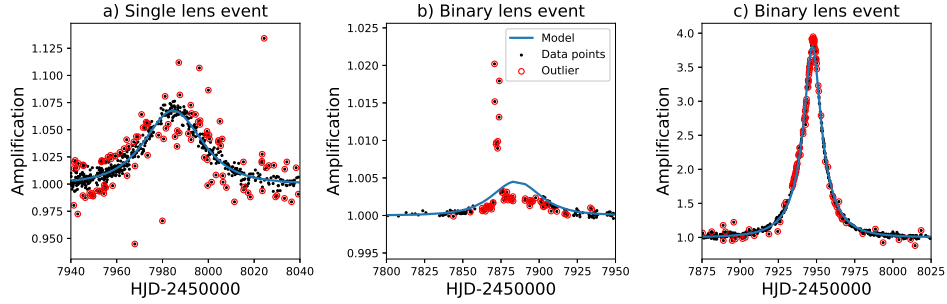
Figure 1: Comparison of the the outlier detection for microlensing events. a) Single lens event with noisy data. b) Binary lens event with a clear difference from a single lens. c) Binary lens event which very subtle difference.
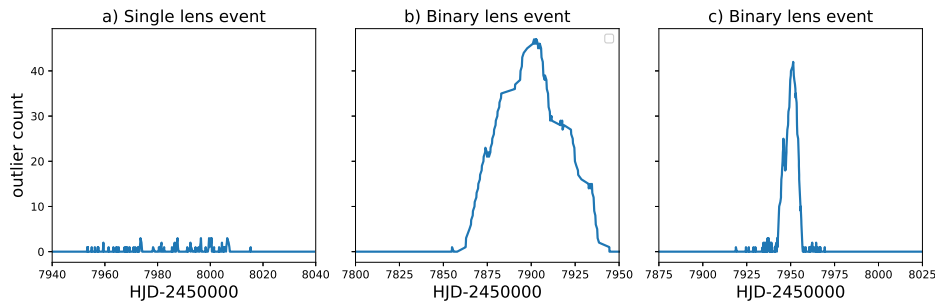

Figure 2: Comparison of the continuous number of outliers and their temporal distribution. a) A single event, the counts remains evenly across time of the event. b) Binary lens event where there is a great difference, and it consumes almost all the time of he event. c) Binary lens event with close apparent similarity with a single lens, but it is possible to see increase of outliers during a section of the event.

- An anomaly should have a consecutive and continuous set of outliers which should not happen for random noise.

- It should take a sensible proportion of time from the event, as indeed complex effects on the lensing system take some time to disappear from the light-curve.

To extract this, we simply count the consecutive outliers and calculated the time it takes on the light curve while compensating for small separation of good data points. This is presented in Figure 2, where the count is in the y axis while the x axis is the light-curve time.

From the different events, it is possible to see that the single lens, as expected, has a more or less constant number of outliers which is expected if these are indeed coming from random noise. For the binary cases, the increase of the count gives the idea of a clear deviation from the model which does not come from noisy data. From this, we can extract the fractional time and number of consecutive outliers which will become a feature that can give information about the existence of an anomaly. This anomaly would give the information we require to be able to classify the different type of events.

3   Future work

This is an ongoing work, and we are continuing to evaluate potential features that can be used to extract our required information from the different events. From this we aim to build an appropriate data-set which could be used, as a first step, to separate between single and non-single lenses. In the future, we hope to be able to no only select anomalous light curves from an ensemble, but also to classify the type of binary that a selected light-curve belongs to.

References

[1] Vasily Belokurov, N. Wyn Evans, and Yann Le Du. Light-curve classification in massive variability surveys – I. Microlensing. *Monthly Notices of the Royal Astronomical Society*, 341(4):1373–1384, jun 2003.

[2] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF: identifying density-based local outliers. In *ACM sigmod record*, pages 93–104. ACM, 2000.

[3] Martin Dominik. Studying planet populations by gravitational microlensing. *General Relativity and Gravitation*, 42(9):2075–2100, sep 2010.

[4] B. Scott Gaudi. Microlensing Surveys for Exoplanets. *Annual Review of Astronomy and Astrophysics*, 50(1):411–453, sep 2012.

[5] Kyu-Ha Hwang, Chung-Uk Lee, Tianshu Wang, Matthew T. Penny, Richard W. Pogge, Weicheng Zang, Pascal Fouqué, Sang-Mok Cha, Andrew Gould, Michael D. Albrow, Yoon-Hyun Ryu, In-Gu Shin, Jennifer C. Yee, Yossi Shvartzvald, Hyoun-Woo Kim, Sun-Ju Chung, Cheongho Han, Dong-Joo Lee, Shude Mao, Yongseok Lee, Dong-Jin Kim, Seung-Lee Kim, Youn Kil Jung, Wei Zhu, and Byeong-Gon Park. KMT-2017-BLG-0165Lb: A Super-Neptune-mass Planet Orbiting a Sun-like Host Star. *The Astronomical Journal*, 157(2):72, 2019.

[6] Seung Lee Kim, Chung Uk Lee, Byeong Gon Park, Dong Jin Kim, Sang Mok Cha, Yongseok Lee, Cheongho Han, Moo Young Chun, and Insoo Yuk. KMTNet: A network of 1.6 m wide -field optical telescopes installed at three southern observatories. *Journal of the Korean Astronomical Society*, 49(1):37–44, 2016.

[7] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, dec 2008.

[8] Alistair McDougall and Michael D. Albrow. Microlensing observations rapid search for exoplanets: MORSE code for GPUs. *Monthly Notices of the Royal Astronomical Society*, 456(1):565–570, Feb 2016.

[9] Tri-Dzung Nguyen and Roy E. Welsch. Outlier detection and robust covariance estimation using mathematical programming. *Advances in Data Analysis and Classification*, 4(4):301–334, dec 2010.

[10] M Ollivier, T Encrenaz, F Roques, F Selsis, and F Casoli. *Planetary Systems: Detection, Formation and Habitability of Extrasolar Planets*. Astronomy and Astrophysics Library. Springer Berlin Heidelberg, 2008.

[11] V. Rajpaul. Evolutionary-algorithm-based analysis of gravitational microlensing light curves. *Monthly Notices of the Royal Astronomical Society*, 427(2):1755–1768, 2012.

[12] Yiannis Tsapras. Microlensing Searches for Exoplanets. *Geosciences*, 8(10):365, 2018.

[13] Georgios Vernardos and Grigorios Tsagkatakis. Quasar microlensing light-curve analysis using deep machine learning. *Monthly Notices of the Royal Astronomical Society*, 486(2):1944–1952, jun 2019.