Contents lists available at ScienceDirect

Astronomy and Computing

Full length article

PynPoint code for exoplanet imaging

A. Amara *, S.P. Quanz, J. Akeret

ETH Zurich, Department of Physics, Wolfgang Pauli Strasse 27, 8093 Zurich, Switzerland

ARTICLE INFO

ABSTRACT

We announce the public release of PynPoint, a Python package that we have developed for analysing exoplanet data taken with the angular differential imaging observing technique. In particular, PynPoint is designed to model the point spread function of the central star and to subtract its flux contribution to reveal nearby faint companion planets. The current version of the package does this correction by using a principal component analysis method to build a basis set for modelling the point spread function of the observations. We demonstrate the performance of the package by reanalysing publicly available data on the exoplanet β Pictoris b, which consists of close to 24,000 individual image frames. We show that PynPoint is able to analyse this typical data in roughly 1.5 min on a Mac Pro, when the number of images is reduced by co-adding in sets of 5. The main computational work, the calculation of the Singular-Value-Decomposition, parallelises well as a result of a reliance on the SciPy and NumPy packages. For this calculation the peak memory load is 6 GB, which can be run comfortably on most workstations. A simpler calculation, by co-adding over 50, takes 3 s with a peak memory usage of 600 MB. This can be performed easily on a laptop. In developing the package we have modularised the code so that we will be able to extend functionality in future releases, through the inclusion of more modules, without it affecting the users application programming interface. We distribute the PynPoint package under GPLv3 licence through the central PyPI server, and the documentation is available online (http://pynpoint.ethz.ch).

1. Introduction

The field of exoplanet research has grown enormously in recent decades, with the current census of confirmed exoplanets and exoplanet candidates exceeding 5000 objects (see, www.exoplanet.eu, Schneider et al. (2011), www.exoplanets.org, Wright et al. (2011)). The vast majority of the exoplanets known today have been detected through radial velocity (RV) and transit measurements (e.g., Batalha et al., 2013; Borucki et al., 2011; Mayor et al., 2014; Howard et al., 2010; Cumming et al., 2008). These studies, and in particular the *Kepler* space mission, allow us to derive robust statistics for the occurrence rate of planets on close-in orbits as a function of planet size and mass (e.g., Marcy et al., 2014; Fressin et al., 2013; Dressing and Charbonneau, 2013; Mayor et al., 2014).

In addition to RV and transit searches for exoplanets, a considerable and increasing effort has been devoted to direct imaging searches for exoplanets in recent years. These activities are currently culminating in the commissioning of dedicated high-contrast planet imaging instruments at 8-m class telescopes

(SPHERE at the VLT (Beuzit et al., 2008), and GPI at Gemini (Macintosh et al., 2008)). The scientific motivation for direct imaging searches is at least twofold: (1) determining the occurrence rate of exoplanets at larger orbital separations not accessible by RV and transit searches; (2) obtaining spectra and spectral energy distributions (SEDs) of extrasolar planets to constrain their atmospheric parameters. In principle, exoplanet atmospheres can also be studied using transit and secondary eclipse observations (e.g., Seager and Deming, 2010), but only a small fraction of the exoplanet population is transiting in front of or behind their host star. Examples for directly imaged exoplanets that likely formed from a circumstellar disk around their host star include β Pictoris b (Lagrange et al., 2010), the 4-planet system around HR8799 (Marois et al., 2008, 2010), HD95086 b (Rameau et al., 2013) and GJ504 b (Kuzuhara et al., 2013). In addition, in a few cases, there is observational evidence from direct imaging campaigns for very young planets that are embedded in or interact with the gas rich disk of their host stars (e.g. HD100546 Quanz et al., 2013).

The biggest challenge for the direct detection of an exoplanet is the stark brightness contrast between the planet and its host star in combination with the small on-sky separation. Hence, direct planet searches are typically done with ground-based, adaptive optics assisted, very large telescopes at near-infrared (NIR) wavelengths (1–5 μm), where one aims at the detection of the planet's thermal

---

* Corresponding author.
E-mail address: adam.amara@phys.ethz.ch (A. Amara).

emission rather than reflected starlight. To enhance the contrast capabilities of the instruments, coronagraphs with different designs have been introduced in the past years, in particular for observations between 3 and 5 μm (Kenworthy et al., 2010; Quanz et al., 2010; Mawet et al., 2011). Furthermore, a big improvement in contrast performance was achieved with the introduction of the angular differential imaging (ADI) technique (Marois et al., 2006), which allows for a much more accurate subtraction of the central star's point spread function (PSF).

Together with these technical improvements, it has also been realised that there are significant potential gains from developing better and more sophisticated data analysis methods. A particularly crucial step in the data processing chain is the modelling and subtraction of the star's PSF, as typically the star outshines a planetary companion by orders of magnitude. Hence, only an accurate subtraction of the stellar flux contribution might reveal the existence of a nearby companion. A nice example here was the introduction of the LOCI algorithm (Lafrenière et al., 2007) that was developed for the Gemini Deep Planet Search (Lafrenière et al., 2007) and is still commonly used today. More recently, principal component analysis (PCA) based algorithms were introduced by our group and a group in the US (Amara and Quanz, 2012; Soummer et al., 2012). As the quality and volume of exoplanet direct imaging data will continue to increase in the coming years, in particular with the advent of the dedicated instruments mentioned above, the analysis methods used to process the data will also have to grow in sophistication.

[We therefore announce the public release of PynPoint, a Python packaged for processing ADI data. The package is distributed through the central PyPI server. The main engine is based on PCA using Singular-Value-Decomposition (SVD). One of the advantages of this approach is that we are able to rely on optimised and tested Python libraries. As an example, when compiled against Intel's math kernel library (MKL), the SciPy `linalg.svd` function naturally parallelises without further input from the user. This does not suffer from the Python Global Interpreter Lock problem since MKL builds on its own native parallelism. Such features of these libraries make Python a very attractive development language. As part of the release we also provide example data. The data set was taken on 2009 December 26 at the Very Large Telescope with the high-resolution, adaptive optics assisted, near-infrared camera NACO in the L′ filter (central wavelengths 3.8 μm) in ADI mode. The raw data are publicly available from the European Southern Observatory (ESO) archive (Program ID: 084.C-0739(A))].

In this paper, we give a brief overview of the PynPoint package. We begin with a brief description of the data and processing steps that currently need to be performed before, the data is passed to PynPoint (Section 2). In Section 3, we give an overview of the PynPoint package; and we describe the main user interfaces in Section 4. We show performance results in Section 5, and in Sections 6 and 7 we present a discussion of quality assurance and a summary.

2. The data and initial data reduction steps

Before describing in detail the functionalities of the PynPoint package, it is worth laying out the typical input data that PynPoint uses and the basic data reduction and analysis steps that are carried out before the PSF fitting and subtraction are performed.

As described above, the standard high-contrast observing technique to image exoplanets is ADI on adaptive optics-assisted NIR cameras. Typically during this process, the pupil of the telescope is fixed throughout the observing sequence, but the stellar field rotates on the detector due to the continuously changing parallactic angle caused by Earths' rotation. Depending on the observing

wavelength, the brightness of the star and whether or not a coronagraph is used, the typical exposure time of a single image frame varies between a fraction of a second to several seconds. A typical size of an individual raw image is 1024 × 1024 pixels. However, in some cases smaller detector sub-arrays are read out and stored (e.g., 512 × 512 or 256 × 256 pixels). A typical observing sequence lasts for up to a few hours, resulting in thousands if not tens of thousands of individual exposures that are stored individually. To be able to correct for bad pixels and sample the PSF more accurately, while properly subtracting background emission, the star is made to cycle through different pre-defined positions on the detector during the observing sequence ('dithering'). The number of exposures taken at a given dither position before the star is moved to the next point depends on the observing wavelengths, the individual exposure time and the observing conditions. While observing at 1–2.5 μm, one can normally spend up to several minutes at the same dither position. At longer wavelengths (3–5 μm) and with more strongly varying background (see below), one switches to another dither position every 30–60 s.

The raw images undergo several steps before they are processed by the PynPoint package. These processes include:

Parallactic angle determination. Depending on the instrument, not all individual raw images will have a parallactic angle assigned to them. In that case, the parallactic angle for a given raw image has to be computed and stored in the header of the image file. PynPoint will search for a header keyword called `NEW_PARA` in each individual input image to access the parallactic angle.

Dark current subtraction. The dark current (and, where applicable, also the bias level) for a given exposure time and detector readout scheme is subtracted from all raw images. This step can be skipped if in the background subtraction step (see below) two individual images are subtracted from each other, as this will also subtract any dark current signal.

Bad pixel cleaning. Dead or bad detector pixels, as well as pixels possibly hit by a cosmic ray, have to be identified. These pixels are either masked out and disregarded from further analyses or they are replaced with the mean value of surrounding pixels.

Flatfielding. A flat-field correction is applied to the individual images to correct for large scale sensitivity gradients across the detector. In some cases, depending on the required calibration accuracy, this step can be skipped if the detector is reasonably 'flat'.

Background subtraction. Depending on the wavelength, the background emission has to be subtracted from all individual images. This step becomes more important with increasing wavelength and is absolutely crucial for data taken between 3 and 5 μm. At these wavelengths, thermal emissions from the sky-background, the telescope and the instrument contribute to the total background flux. Without a nearby bright star, observations at 3–5 μm are typically limited by the Poisson noise of the background flux and temporal changes therein. To subtract the background emission from a given raw image, typically another image observed close in time but at a different dither position is subtracted. Alternatively, a sequence of several images from a different dither position is mean combined, to reduce the Poisson noise in the background, and the resulting average image is removed. It has been suggested, that additional sensitivity can be gained if the LOCI algorithm is used to construct and subtract the background emission in the immediate vicinity of a star for a given input image (Galicher et al., 2011).