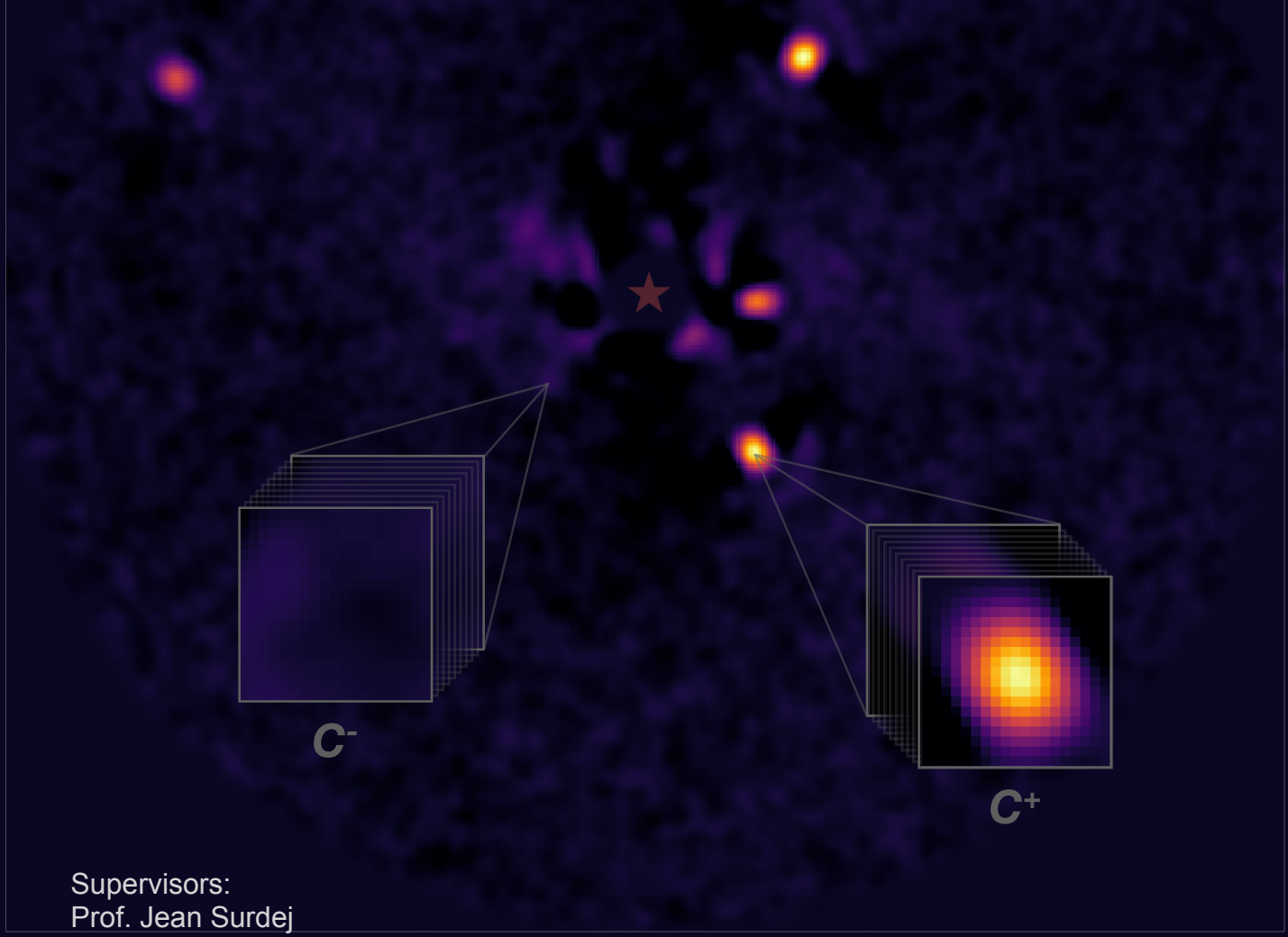
ADVANCED DATA PROCESSING FOR HIGH-CONTRAST IMAGING

PUSHING EXOPLANET DIRECT DETECTION LIMITS WITH MACHINE LEARNING

Carlos Alberto GÓMEZ GONZALEZ



Supervisors:
Prof. Jean Surdej
Prof. Marc Van Droogenbroeck
Dr. Olivier Absil

September 2017

ADVANCED DATA PROCESSING FOR HIGH-CONTRAST IMAGING

PUSHING EXOPLANET DIRECT DETECTION LIMITS WITH MACHINE LEARNING

by CARLOS ALBERTO GOMEZ GONZALEZ

Dissertation submitted in partial fullfilment of the requirements
for the degree of Doctor of Science

AGO/STAR
Faculty of Sciences
Université de Liège

Advisors:
Jean Surdej
Marc Van Droogenbroeck
Olivier Absil

August 2017

JURY MEMBERS

Louis Wehenkel, Professor at the Université de Liège (President);

Jean Surdej, Professor at the Université de Liège (Advisor);

Marc Van Droogenbroeck, Professor at the Université de Liège (Co-advisor);

Olivier Absil, Research Associate at the Université de Liège (Co-advisor);

Michaël Gillon, Research Associate at the Université de Liège;

Christian Marois, Adjunct Associate Professor at the University of Victoria (External: Canada);

Dimitri Mawet, Associate Professor at Caltech (External: US);

Sascha Quanz, Researcher at ETH Zurich (External: Switzerland).

ACKNOWLEDGMENTS

After these four years of intense and continuous learning, I have finally the pleasure of writing this thesis dedicatory. First, I would like to thank my family. This dissertation would not have been possible without their warm love and endless support. A few words to them in Spanish.

El camino que me trajo aquí, en realidad comenzó hace muchos años en las costas del Caribe Colombiano. Recuerdo cuando mi papá me invitaba con su pequeño telescopio Zeiss a ver la luna llena, y también de nuestras caminatas al Ateneo, cuando me contaba sobre filosofía, religión y ciencia. Sin duda esas experiencias me llevaron elegir dejar mi carrera en Colombia para ir a estudiar a astronomía a Rusia. Todos estos años viajando por el mundo no hubiesen sido posibles sin el apoyo incondicional de mi padre Gustavo y el amor infinito de mi madre Emilce. A ustedes les dedico esta tesis doctoral. De igual manera, toda mi gratitud para mi abue Cecilia y mis hermanas Paola y Cristina.

My utmost gratitude to my advisors Jean, Marc and Olivier for believing in me and for having granted me the opportunity of pursuing this thesis. I do not forget the kind words Jean told me when I interviewed for the PhD position, that my age was not a downside at all. I hope I have met expectations as a pupil. Thank you for giving me the perfect balance between guidance and freedom, for having your door always open and for allowing me to find my own path and interests.

I also extend my gratitude to the members of the jury for accepting to evaluate my work: Professor Louis Wehenkel, Michaël Gillon, Christian Marois, Dimitri Mawet and Sascha Quanz.

I would also like to thank the members of the VORTEX team, specially Dimitri who has always found words to encourage me and for continuously sharing his expertize. Thanks to the rest of the VORTEX team and also colleagues from the ULg: Aissa, Bikram, Brunella, Denis, Elsa, Gary, Gilles, Julien M., Maddalena, Olivier W. and Valentin. Thanks to those of you who have shared all these meals with me and have granted me their friendship. I am also grateful to the secretaries of our department, Sylvia and Angela, for their help with all the administrative matters.

Last but not least, I would like to thank my wife Veronica and my son Orión. You are my lighthouse when I am all stressed, lost in my code. Thank you for your patience, for being my home and for making me a better person. I love you both! Orión, I have realized that training artificial intelligence models does not remotely compare to the joy of teaching you new words and skills.

This research was supported by the European Research Council under the European Union's Seventh Framework Programme (ERC Grant Agreement n. 337569), and by the French Community of Belgium through an ARC grant for Concerted Research Action.

ABSTRACT

Since ancient times, mankind has wondered whether other solar systems exist around other stars somewhere in the Universe. It took many centuries to finally prove the existence of extra-solar planetary systems. Nowadays, more than 3500 exoplanets have been discovered, mostly thanks to indirect detection methods. Indeed, the task of directly detecting exoplanets through high-contrast imaging (HCI) is a formidable challenge, and has only been enabled in the last decade thanks to advances in instrumentation and dedicated image processing algorithms. This last component of the exoplanet direct imaging pipeline is what ultimately pushes the detection limits and sensitivity of HCI instruments and survey campaigns. Unfortunately, the HCI community has been slow in adopting the latest developments in data management and machine learning for analyzing the increasing amount of available data. This dissertation is an attempt to fill in this very gap, and develops at the interface of computer science, machine learning, statistics, and astrophysics. This work contributes to the field of data processing for HCI in two main ways. On one hand, I have developed an open source `Python` library for taking HCI data from the raw state up to the characterization of companions. It implements state-of-the-art approaches and is positioning itself as one of the de facto software solutions for building HCI pipelines. I have also participated to the critical analysis of data from different first and second generation HCI instruments. On the other hand, I have approached the task of exoplanet detection in angular differential imaging sequences from a computer vision and machine learning perspective. This interdisciplinary work has led to novel algorithmic solutions, extending unsupervised learning techniques widely used in HCI and proposing advanced supervised learning approaches based on cutting-edge deep learning models. My novel algorithms have been presented using a robust performance assessment framework to produce large comparative performance studies. These studies show the improved sensitivity vs specificity trade-off of the proposed supervised detection approach. The proposed algorithms bring the possibility of re-processing existing HCI databases to maximize their scientific return and potentially improve the demographics of directly imaged exoplanets.

RÉSUMÉ

Depuis l'Antiquité, l'Homme s'est demandé si d'autres systèmes solaires existent autour d'autres étoiles quelque part dans l'Univers. Il a fallu plusieurs siècles pour enfin prouver leur existence. Plus de 3500 exoplanètes ont été découvertes à ce jour, principalement grâce à des méthodes de détection indirecte. En effet, la tâche consistant à détecter directement des exoplanètes au moyen de l'imagerie à haut contraste (IHC) représente un défi considérable et n'a été rendue possible qu'au cours de la dernière décennie grâce aux progrès réalisés dans l'instrumentation astronomique et dans les algorithmes dédiés au traitement d'image. Cette dernière composante du pipeline d'imagerie directe d'exoplanètes est ce qui a permis de repousser les limites de détection des instruments IHC. Malheureusement, la communauté IHC a été lente à adopter les derniers développements pour l'exploitation des données et l'apprentissage automatique afin d'analyser la quantité croissante de données disponibles. Ma thèse représente une tentative de combler ce fossé et se situe à l'interface de l'informatique, de l'apprentissage automatique, des statistiques et de l'astrophysique. Ce travail contribue au domaine du traitement des données pour IHC principalement de deux manières. D'une part, j'ai développé une bibliothèque `Python` "open source" pour traiter des données d'observation depuis leur état brut jusqu'à la caractérisation des compagnons détectés. Cette bibliothèque met en œuvre des approches situées à la pointe des technologies informatiques et se positionne comme l'une des solutions logicielles de facto pour la construction de pipelines IHC. J'ai également participé à l'analyse critique de données en provenance de différents instruments IHC. D'autre part, j'ai développé des algorithmes de détection d'exoplanètes via imagerie différentielle angulaire à partir d'une perspective de vision par ordinateur et d'apprentissage automatique. Ce travail interdisciplinaire a conduit à de nouvelles solutions algorithmiques, en étendant les techniques d'apprentissage non supervisées au domaine IHC et en proposant des approches avancées d'apprentissage supervisé basées sur des modèles d'apprentissage automatique profonds. Mes nouveaux algorithmes ont été présentés dans un cadre rigoureux d'évaluation des performances visant à mener des études comparatives de performance. Ces études montrent l'amélioration du compromis entre sensibilité et spécificité de l'approche de détection supervisée proposée. Les nouveaux algorithmes offrent la possibilité de réanalyser les bases de données IHC existantes afin de maximiser leur exploitation scientifique et d'améliorer potentiellement la démographie des exoplanètes directement imagées.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

ADI	Angular Differential Imaging
AGPM	Annular Groove Phase Mask
ANDROMEDA	ANgular Differential OptiMal Exoplanet Detection Algorithm
AU	Astronomical Unit
CEVR	Cumulative Explained Variance Ratio
CNN	Convolutional Neural Network
ED	Eigen Decomposition
ESO	European Southern Observatory
FITS	Flexibile Image Transport System
FPR	False Positive Rate
FWHM	Full Width at Half Maximum
GPI	The Gemini Planet Imager
HCI	High-Contrast Imaging
IFS	Integral Field Spectrograph
IRDIS	InfraRed Dual-band Imager and Spectrograph
LBT	Large Binocular Telescope
LMIRCam	LBTI's L and M Infrared Camera
LLSG	Local Low-rank plus Sparse plus Gaussian-noise decomposition
LOCI	Locally Optimized Combination of Images
LSTM	Long-Short Term Memory
MCMC	Markov Chain Monte Carlo
MLAR	Multi-level Low-rank Approximation Residuals
NACO	Nasmyth Adaptive Optics System & COude Near Infrared CAmera
NEGFC	NEGative Fake Companion technique
NIR	Near Infra-Red
NMF	Non-negative Matrix Factorization
PCA	Principal Component Analysis
PC	Principal Component
PSF	Point Spread Function

RDI	Reference star Differential Imaging
ROC	Receiver Operating Characteristic curve
SDI	Spectral Differential Imaging
S/N	Signal-to-Noise ratio
SODINN	Supervised exOplanet detection via Direct Imaging with deep Neural Networks
SODIRF	Supervised exOplanet detection via Direct Imaging with Random Forests
SPHERE	Spectro Polarimetric High contrast Exoplanet REsearch
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TPR	True Positive Rate
`VIP`	Vortex Image Processing library
VLT	Very Large Telescope

COMMUNICATIONS

REFEREED JOURNALS

13. C. A. Gomez Gonzalez, O. Absil and M. Van Droogenbroeck, *Supervised detection of exoplanets through high-contrast imaging*. Submitted to A&A

12. R. Jensen-Clem, D. Mawet, C. A. Gomez Gonzalez, O. Absil, R. Belikov, T. Currie, M. Kenworthy, C. Marois, J. Mazoyer, G. Ruane and A. Tanner, *A new standard for assessing the performance of high-contrast imaging systems*. Submitted to AJ

11. V. Christiaens, S. Casassus, O. Absil, C. A. Gomez Gonzalez, J. Girard, S. Kimeswenger, R. Ramirez, O. Wertz, A. Zurlo, Z. Wahhaj, V. Salinas, A. Jordan, and D. Mawet, *Characterization of the low-mass companion HD 142527 B*, Submitted to A&A

10. G. Ruane, D. Mawet, J. Kastner, T. Meshkat, M. Bottom, B. Femenia Castella, O. Absil, C. A. Gomez Gonzalez, E. Huby, Z. Zhu, R. Jensen-Clem, E. Choquet and E. Serabyn, 2017, *Deep Imaging Search for Planets Forming in the TW Hya Protoplanetary Disk with the Keck/NIRC2 Vortex Coronagraph*. AJ, 154, 73

9. C. A. Gomez Gonzalez, O. Wertz, O. Absil, V. Christiaens, D. Defrere, D. Mawet, J. Milli, P.-A. Absil, M. Van Droogenbroeck, F. Cantalloube, P. Hinz, A. Skemer, M. Karlsson, and J. Surdej, 2017, *VIP: Vortex Image Processing Package for high-contrast direct imaging*, AJ, 154, 7

8. J. Milli, P. Hibon, V. Christiaens, E. Choquet, M. Bonnefoy, G. M. Kennedy, M. C. Wyatt, O. Absil, C. A. Gomez Gonzalez, C. del Burgo, L. Matra, J.-C. Augereau, A. Boccaletti, C. Delacroix, S. Ertel, W. R. F. Dent, P. Forsberg, T. Fusco, J. H. Girard, S. Habraken, E. Huby, M. Karlsson, A.-M. Lagrange, D. Mawet, D. Mouillet, M. Perrin, C. Pinte, L. Pueyo, C. Reyes, R. Soummer, J. Surdej, Y. Tarricq, and Z. Wahhaj, 2016, *Discovery of a low-mass companion inside the debris ring surrounding the F5V star HD 206893*. A&A, 597, L2

7. E. Choquet, J. Milli, Z. Wahhaj, R. Soummer, A. Roberge, J.-C. Augureau, M. Booth, O. Absil, A. Boccaletti, C. H. Chen, J. H. Debes, C. del Burgo, W. R.F. Dent, S. Ertel, J. H. Girard, E. Gofas-Salas, D. A. Golimowski, C. A. Gomez Gonzalez, J. B. Hagan, D. C. Hines, G. M. Kennedy, A.-M. Lagrange, L. Matra, D. Mawet,

D. Mouillet, M. N'Diaye, M. D. Perrin, C. Pinte, L. Pueyo, A. Rajan, G Schneider, S. Wolff, and M. Wyatt, 2016, *First scattered-light images of the gas-rich debris disk around 49 Ceti*. ApJ, 834, L12

6. Z. Wahhaj, J. Milli, G. Kennedy, S. Ertel, L. Matra, A. Boccaletti, C. del Burgo, M. Wyatt, C. Pinte, A. M. Lagrange, O. Absil, E. Choquet, C. Gomez Gonzalez, H. Kobayashii, D. Mawet, D. Mouillet, L. Pueyo, W. R.F. Dent, J.-C. Augereau, and J. Girard, 2016, *The SHARDDS survey: First resolved image of the HD 114082 debris disk in the Lower Centaurus Crux with SPHERE*. A&A 596, L4

5. E. Serabyn, E. Huby, K. Matthews, D. Mawet, O. Absil, B. Femenia, P. Wizinowich, M. Karlsson, M. Bottom, R. Campbell, B. Carlomagno, D. Defrere, C. Delacroix, P. Forsberg, S. Habraken, C. A. Gomez Gonzalez, A. Jolivet, K. Liewer, S. Lilley, P. Piron, M. Reggiani, J. Surdej, H. Tran, E. Vargas Catalan, and O. Wertz, 2016, *The W. M. Keck Observatory infrared vortex coronagraph and a first image of HIP79124B*, AJ, 153, 43

4. D. Mawet, E. Choquet, O. Absil, E. Huby, M. Bottom, E. Serabyn, B. Femenia, J. Lebreton, K. Matthews, C. A. Gomez Gonzalez, O. Wertz, B. Carlomagno, V. Christiaens, D. Defrere, C. Delacroix, P. Forsberg, S. Habraken, A. Jolivet, M. Karlsson, J. Milli, C. Pinte, P. Piron, M. Reggiani, J. Surdej, E. Vargas Catalan, 2016, *Characterization of the Inner Disk around HD 141569 A from Keck/NIRC2 L-Band Vortex Coronagraphy*. AJ, 153, 44

3. O. Wertz, O. Absil, C. A. Gomez Gonzalez, J. Milli, J. H. Girard, D. Mawet, and L. Pueyo, 2016, *VLT/SPHERE robust astrometry of the HR8799 planets at milliarcsecond-level accuracy*, A&A 598, A83

2. C. A. Gomez Gonzalez, O. Absil, P.-A. Absil, M. Van Droogenbroeck, D. Mawet, and J. Surdej, 2016, *Low-rank plus sparse decomposition for exoplanet detection in direct-imaging ADI sequences. The LLSG algorithm*. A&A, 589, A54

1. F. Cantalloube, D. Mouillet, L. M. Mugnier, J. Milli, O. Absil, C. A. Gomez Gonzalez, G. Chauvin, J.-L. Beuzit, and A. Cornia, 2015, *Direct exoplanet detection and characterization using the ANDROMEDA method*. A&A, 582, A89

NON REFEREED JOURNALS

- B. Femenia Castella et al. 2016. *Commissioning and first light results of an L'-band vortex coronagraph with the Keck II adaptive optics NIRC2 science instrument*. SPIE proceedings vol. 9909

- O. Absil et al. 2016, *Three years of harvest with the vector vortex coronagraph in the thermal infrared*. SPIE proceedings vol. 9908
- B. Pairet et al. 2016, *Low Rank and Group-Average Sparsity Driven Convex Optimization for Direct Exoplanets Imaging*. Proceedings of the international Traveling Workshop on Interactions between Sparse models and Technology
- D. Defrere et al. 2014, *L'-band AGPM vector vortex coronagraph's first light on LBTI/LMIRCam*. Search for Life Beyond the Solar System. Exoplanets, Biosignatures and Instruments, ed. D. Apai & P. Gabor, P4.75

TALKS AND POSTERS

- Poster: *VIP - Vortex Image Processing package*. Astroinformatics IAU symposium, Sorrento, Italy (October 2016)
- Talk: *Post-processing for high-contrast imaging of exoplanets/disks*. Invitation-only KISS workshop on Exoplanet Imaging, Caltech, Pasadena, US (August 2016)
- Talk: *Image processing with VIP and RDI post-processing*. First Vortex international workshop, Caltech, Pasadena, US (August 2016)
- Poster: *VIP - Vortex Image Processing package*. Resolving Planet Formation in the Era of ALMA and Extreme AO, ESO, Chile (May 2016)
- Talk: *Applications of PCA and low-rank plus sparse decompositions in high-contrast Exoplanet imaging*. Seminar, ICTEAM, Catholic University of Louvain, Belgium (February 2016)
- Talk: *Beyond PCA, Low-rank plus Sparse decomposition of ADI image sequences for exoplanet detection*. Conference, In the Spirit of Lyot, Montreal, Canada (June 2015)
- Talk: *VIP - Vortex Image Processing package tutorial*. AO data processing workshop, LAM, Marseille, France (May 2015)
- Poster: *Python based pipeline for post-processing in astronomical high-contrast imaging*. SciPy Conference 2014, Austin, US (July 2014)

1 INTRODUCTION

Contents

1.1 DETECTION OF EXTRA-SOLAR PLANETARY SYSTEMS

The idea of other worlds or planets outside our Solar System was proposed in the Ancient Greece by Epicurus in the IV century b.C. Since these ancient times, we have wondered whether other solar systems exist around other stars somewhere in the universe. It took us many centuries to finally prove the existence of extra-solar planetary systems empowered by recent technological developments. In only two decades of exoplanetary science, we count about 3500 confirmed discoveries[1], most of which have been made possible thanks to indirect detection methods (Pepe et al., 2014; Fischer et al., 2014). During the same period of time, we have enriched our understanding of the formation mechanisms of these planetary systems and the immense diversity of their architectures. However, exoplanetary science is still in its infancy, as we are still grasping the full complexity of multi-planet systems configurations, developing a general framework of planet formation, and ultimately hunting for Earth analogues that could host life as we know it. The task of finding exoplanets around their host

1 http://exoplanetarchive.ipac.caltech.edu/

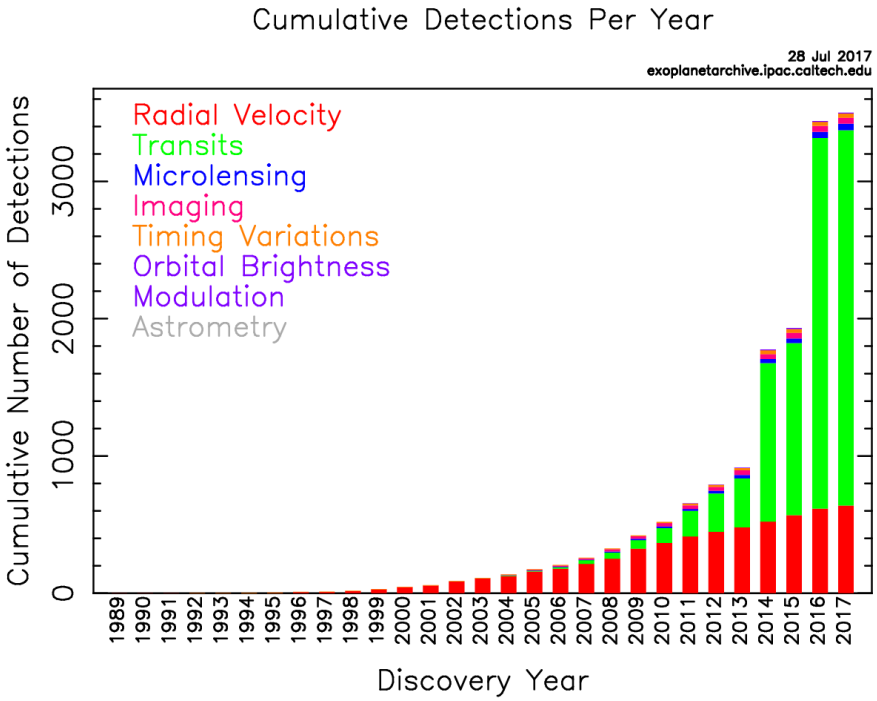


Figure 1.1: Bar plot showing the cumulative number of discoveries per year for each detection method. The contribution of the direct imaging detection method is hardly noticeable. Indirect methods like radial velocity and transit method are the most prolific methods. Credits: NASA exoplanet archive.

stars with direct observations is very challenging. Until recently, indirect detection methods were the only way to study exoplanets.

1.2 INDIRECT DETECTION METHODS

Indirect detection methods account for the majority of the exoplanet detections as clearly seen in Fig. 1.1. The distinction between direct and indirect methods is based on the ability of obtaining photons from the planets. Each detection method has an observational bias, meaning that it is suited for a different region of the physical parameter space. Different techniques are complementary but unfortunately our coverage of the parameter space is still incomplete. In Fig. 1.2 we see, for instance, how direct imaging is a successful method for detecting exoplanets at large distances, larger than 10 AU, while radial velocity and transit method are well suited for close-in companions.

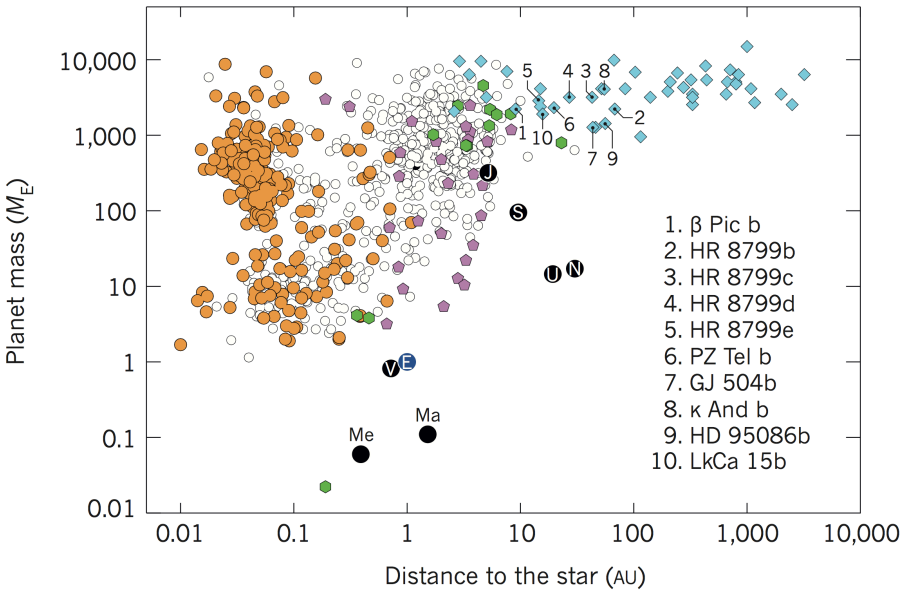


Figure 1.2: Diagram showing the mass as a function of semi-major axis of known planets. The colors correspond to the detection technique: light blue for direct imaging, orange for transits, white for radial velocity, pink for microlensing and green for pulsation timing. Black filled circles show solar system planets. Taken from Milli et al. (2016)

1.2.1 Pulsar timing

This method deserves to be listed first as it enabled the discovery of the first ever exoplanet. Pulsars are neutron stars: ultradense star remnants of intermediate mass stars, roughly $\geqslant 8M_{\odot}$[2] on the main sequence, which exploded as supernovae. They emit radiowaves regularly as they rotate. Anomalies in the regularity of these pulses were used, back in 1992, by Wolszczan and Frail (1992) to discover the first know exoplanets orbiting the 6.2-ms radio pulsar PSRB1257 + 12. Unfortunately, millisecond pulsars are rare and so are the pulsar planets. To date we count a couple of dozens of exoplanet detections via pulsar timing.

1.2.2 Radial velocities

The first exoplanet orbiting a Sun-like star, 51 Peg, was discovered by Mayor and Queloz (1995) in 1995 using the radial velocity technique. This Jupiter-mass planet in a short 4.23-day orbit, what we call today a Hot-Jupiter, is the first exoplanet detected around a normal star, other than the Sun. The radial velocity technique measures the re-

2 $M_{\odot}$ - Solar Mass.

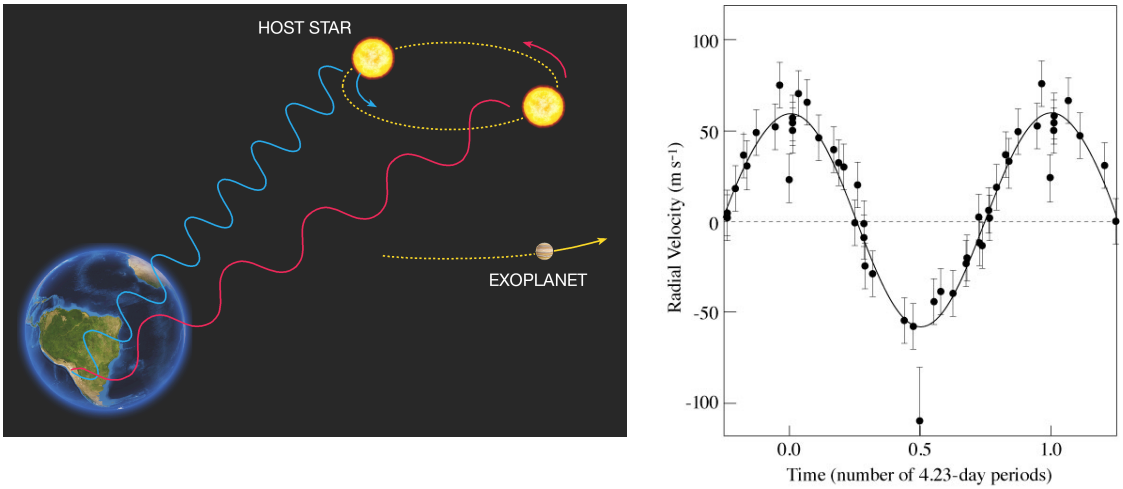


Figure 1.3: Left: Radial velocity method representation. Credits: ESO. Right: Radial velocity curve of 51 Peg, phased to a period of 4.23 days. The solid line is the fit to the observational data (black dots). Taken from Mayor and Queloz (1995).

flex velocity that an orbiting planet induces on a star via Doppler spectroscopy. This imposes a preference for close-in planets, since the gravitational force decreases with the square of the distance to their parent stars. Also, these short orbital periods are easier to detect, as they require shorter observation timescales. Besides this detection bias, we also note the preference for massive planets and planetary systems in edge-on configurations.

The Doppler precision improved from 10 ms^{-1} to 1 ms^{-1} in the ten years after the discovery of 51 Peg (see Fig. 1.3), enabling the discovery of hundreds of extra-solar planetary systems. This technique enables the estimation of the minimum mass of the companion and several orbital parameters, such as the orbital period, eccentricity and inclination. The ultimate limitation of this technique, besides the instrumental precision, is given by the stellar activity or jitter, which can produce stellar line profile variations that can be confused as a velocity change in the star (Fischer et al., 2014).

1.2.3 Astrometry

This technique also exploits the gravitational perturbation of a planetary companion on its parent star. Here we use precise astrometric measurements to study the motion of the star around the mutual center of mass. This technique has not been very fruitful so far mainly due to the difficulty in measuring the tiny changes in stellar position. Several confirmations of radial velocity detections, but only one

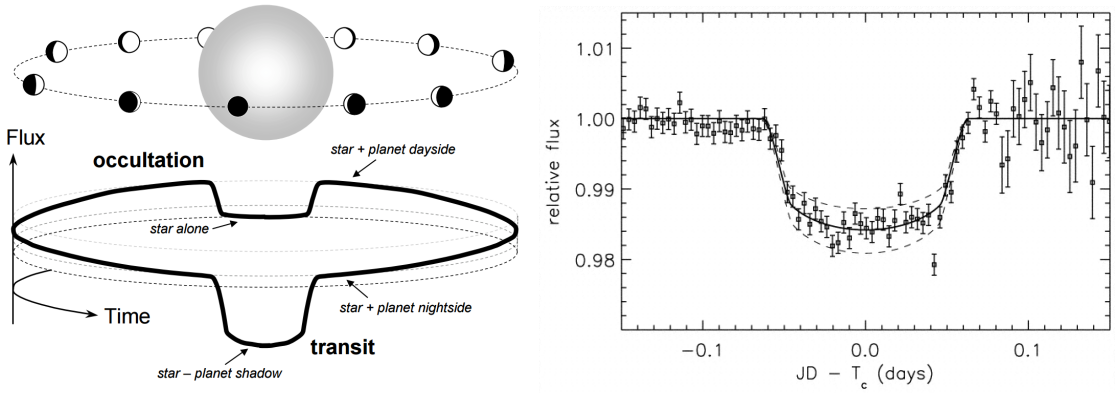


Figure 1.4: Transit detection method. Left: Illustration of the transit method showing how an orbiting planet can create variations in the brightness of its parent star. Credits: Pale Red Dot project. Right: Folded photometric time series of HD 209458 and the transit shape that would occur for the best-fit model (solid line). In this case the star dims by 1.7% due to the presence of the companion. Taken from Charbonneau et al. (2000).

detection (Muterspaugh et al., 2010), have been possible thanks to the astrometry method. This technique is most sensitive to planets with large orbits complementing the radial velocity method. With the launch of Gaia, a rise in the number of detected exoplanets by astrometric measurements is expected.

1.2.4 Transit method

The field of exoplanet detection has entered its most prolific period thanks to the transit method. In the last ten years, the transit detection technique has gone from nine to over 2,600 confirmed candidates. This method relies on tracking the periodic eclipses that a star and a planet cause to each other, in the case when the system is in a (nearly) edge-on configuration (see left panel of Fig. 1.4). The right panel of Fig. 1.4 shows the transit light curve of HD 209458, the first exoplanetary transit detected by Charbonneau et al. (2000).

The transit method gives access to information that cannot be obtained from radial velocity data, such as the orientation of the planet's orbit relative to the sky plane, and the relative dimensions of the planet and the star. Companions that have been studied by both radial velocity and transit methods are well-characterized. The access to the density (given the mass and the diameter) of the companion allows to learn about the planet's physical structure. Even though the transit method is indirect, analysis of the spectrum of the still star enables to study the atmosphere of the transiting planet. When the planet transits the star, light from the star passes through the upper

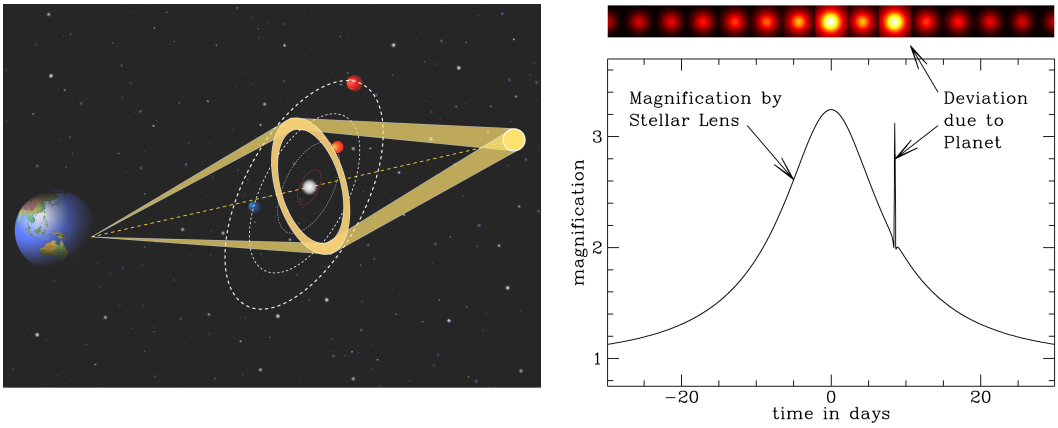


Figure 1.5: Left: Illustration of the microlensing detection method. Right: Light curve of a distant star in a microlensing scenario. Credits: NASA Exoplanet Science Institute.

atmosphere of the planet. This allows us to detect the absorption lines caused by the fraction of the star's light filtered through the planetary atmosphere. Also, secondary eclipses (when the planet is blocked by its star) allow to measure dayside emission by disentangling star atmosphere and planetary atmosphere contributions.

The NASA *Kepler* space observatory has been instrumental in the great success of this exoplanet detection technique. This mission scanned a hundred thousand stars in a close regions of our galaxy, and has been able to detect Earth-size planets (370 to date), generating the first statistics on the number of such planets around Sun-like stars. Another notable project is the Belgian TRAPPIST (TRAnsiting Planets and PlanetesImals Small Telescope) instrument, which recently announced the discovery of seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1 (Gillon et al., 2017). Four of these planets belong to the list of potentially habitable exoplanets according to the Habitable Exoplanets Catalogue[3].

The biases of this method are mainly due to the inclination requirement, which is more likely to happen for very short separations. Also, there is a preference for planets orbiting relatively small stars. A related detection technique is the timing variation method, which looks for variations in the periodicity of known exoplanet transits to identify additional exoplanets with masses comparable to Earth's.

1.2.5 Microlensing

Gravitational microlensing is another indirect method of exoplanets detection. This method exploits the situation when a foreground star, with an orbiting planet, passes very close to our line of sight to a more distant star (see left panel of Fig. 1.5). The gravitational field of the foreground star acts as a lens, splitting in two or more lensed images and magnifying the light from the background star (observed as an increase in its brightness). When a planet is present, it perturbs the lensing effect resulting in a characteristic, short-lived signature of the planet (see Fig. right panel of 1.5). This methods only accounts for about 50 detections to date.

Microlensing is a rare phenomenon and large surveys (OGLE, MOA) monitoring crowded fields of the galactic bulge are actively looking for these phenomena. Another disadvantage is the fact that the microlensing events cannot be repeated (the chance alignment never occurs again). This technique is sensitive to wide orbits just beyond the snow line (orbits comparable to Saturn and Uranus), and also to multi-planet systems. It can also detect free-floating planets and companions around very distant stars (most distant than any other detection method).

1.3 DIRECT DETECTION OF EXOPLANETS

Indirect exoplanet detection methods have dominated the field for many years due to the fact that taking a picture of a planet is extremely difficult. To date, only 44 exoplanets could be directly resolved through high-contrast imaging (HCI), which accounts for about 1% of the confirmed exoplanet candidates known to date.

In the last decade, direct imaging of exoplanets has been enabled thanks to technological advances in ground-based infrared instruments (installed on the largest telescopes on Earth), adaptive optics, and coronagraphy. Direct observations of exoplanets provide a powerful complement to indirect detection techniques and enable the exploration, thanks to its high sensitivity for wide orbits, of different regions of the parameter space (see Fig. 1.2). Direct imaging also allows us to put important constraints on planet formation models and planetary system dynamics and, since we obtain the photons from the planets themselves, we can proceed with further photometric and spectroscopic characterization.

The first direct detection of an exoplanet was carried out by Chauvin et al. (2005) at VLT in 2004 (using simple adaptive optics imaging), when they imaged the brown dwarf[4] 2M1207 and found a hot gas gi-

3 http://phl.upr.edu/projects/habitable-exoplanets-catalog

4 Brown dwarfs are objects larger than gas giant planets and smaller than the lightest stars. They are not massive enough to sustain nuclear fusion.

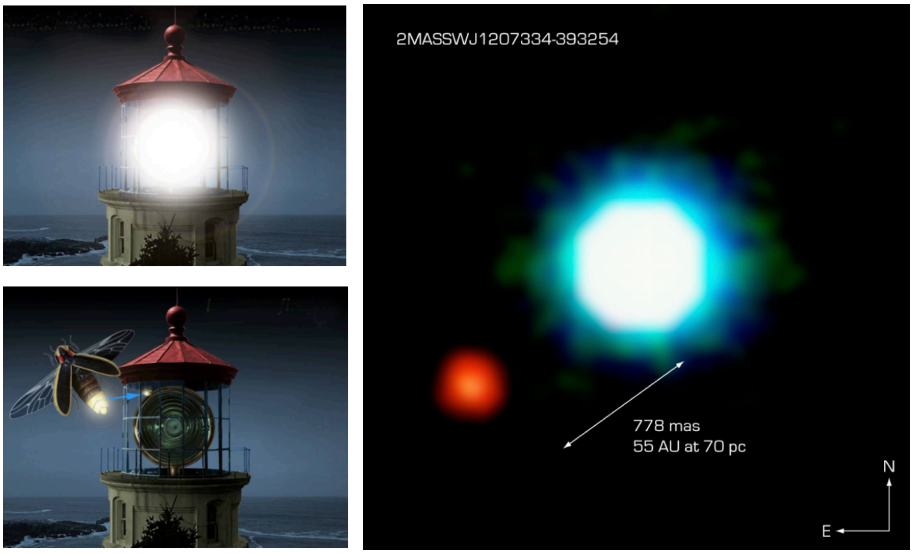


Figure 1.6: Left: Observing a firefly close to a lighthouse 1000 miles away as an analogy to directly observing exoplanets. Right: First exoplanet directly imaged by Chauvin et al. (2005). Credits: ESO.

ant, 100 times dimmer than its host (see the right panel of Fig. 1.6). This was followed by the detections of Fomalhaut b using the Hubble Space Telescope and a coronagraphic mask, four planets around the star HR8799 (Marois et al., 2008b, 2010b) using the Keck and Gemini telescopes and β Pic b at VLT (with adaptive optics and special observing techniques) (Lagrange et al., 2010). These detection were enabled by the techniques that will be introduced in the following sections: adaptive optics, coronagraphs, special observing techniques and post-processing. The two later cases are representative of high-contrast direct imaging from the ground in the infrared aided by hardware and software technologies, and of the data that will be studied throughout this dissertation.

1.3.1 Challenges of ground-based observations

High-contrast direct imaging from the ground presents three main challenges:

1. the image degradation caused by the Earth's turbulent atmosphere and optical imperfections of the telescope and instruments,

2. the small angular separation between the stars and their planets ($\sigma''.01$ to $1''$) and,

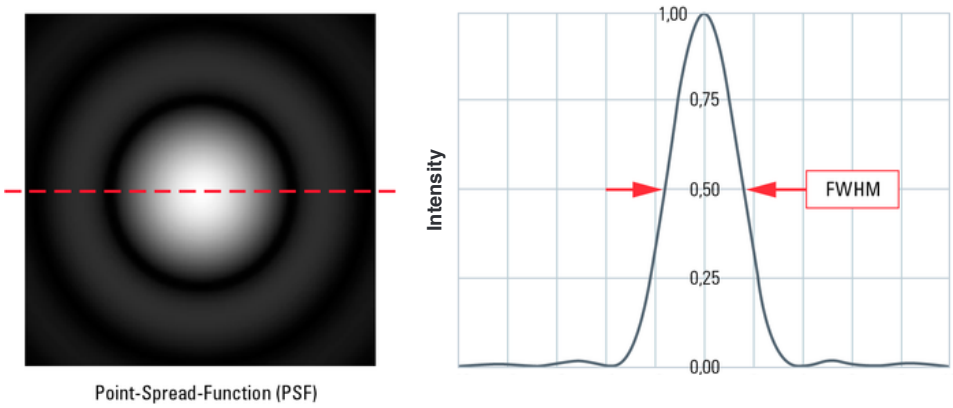


Figure 1.7: Left: Point spread function by a circular aperture showing the Airy-disk and the first Airy ring. Right: Cut on the PSF to show the full width at half maximum. Credits: Leica mycrosystems.

3. the huge difference in brightness between them, with contrasts typically ranging from 10^{-3} to 10^{-10}.

These challenges are addressed with a combined effort of state-of-the-art coronagraphy, optimized wavefront control, dedicated observing techniques, and advanced image post-processing (Guyon, 2005; Mawet et al., 2012). These are the four corner stones enabling high-contrast imaging.

1.3.1.1 *Image degradation challenge*

Even for the largest ground-based telescopes stars can generally be considered as point-like sources of light. Due to the finite size of the aperture of our telescopes, the image of a star is a diffraction pattern called the point spread function (PSF, see left panel of Fig. 1.7). An important measure of the PSF is the full width at half maximum (FWHM, see right panel of Fig. 1.7) which represents the resolving power of an instrument. For an Airy pattern, it is equal to $1.028\lambda/D$ were λ is the wavelength at which the observations are performed and D the diameter of the instrument's aperture. This means, in theory, that using a larger aperture we would directly improve angular resolution.

In practice, atmospheric turbulence and imperfections on the optical surfaces of our instrument distort the incoming wavefronts and degrade the PSF's sharpness (Marois et al., 2003). This creates a halo or fuzzy blob instead of a sharp PSF core. The halo of starlight originating from wavefront errors is particularly troublesome because it does not form a smooth background, but is broken up into a pattern of speckles (Racine et al., 1999). These speckles resemble bright

blobs on the λ/D scale. These static and quasi-static speckles exist on timescales of several minutes to several hours. On top of this, the turbulence and density fluctuations in the layers of the atmosphere produce a second speckle pattern with timescales of a few milliseconds.

The rapidly evolving atmospheric speckles easily average out unlike the quasi-static component of the speckle pattern. This noise follows a modified Rician statistics (Soummer and Aime, 2004; Fitzgerald and Graham, 2006) instead of having a Gaussian nature. Moreover, the fact that speckles resemble the diffraction-limited PSF of the telescope, and can therefore mimic the signature of planets (since they are comparable in angular size and brightness), greatly affects the detectability of such companions. This harmful effect is particularly evident in the small-angle regime, where the quasi-static speckle noise becomes dominant.

Adaptive optics (AO) was originally proposed in 1953 and was first applied in astronomy by ESO in 1990 at La Silla Observatory (Rousset et al., 1990). This technique makes use of deformable mirrors (DM) to partially correct the distorted incoming wavefront coming from astronomical objects. The DM is composed by hundreds or thousands of piezo-electric actuators, which function in a closed-loop receiving real-time instructions from a computer. These actuators reshape the mirror in such a way that it cancels out the wavefront errors, to some extent sharpening the images and restoring the energy back into the Airy disk. For a complete review of AO in high-contrast imaging refer to Milli et al. (2016).

Figures 1.8 and 1.9 show images from the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE, Beuzit et al., 2008), a second generation high-contrast imaging instrument installed at the VLT ESO telescope facility. Fig. 1.8 shows the reconstructing effect of adaptive optics at SPHERE while Fig. 1.9 shows the evolution or decorrelation of speckles with time in a controlled experiment where a realistic temperature gradient was induced (Martinez et al., 2013).

1.3.1.2 *High-angular and high-contrast challenges*

For addressing the small separation problem, we rely on small-angle coronagraphs and observations at shorter wavelengths. For the future, direct detection instruments have been proposed for the next generation of 20-40 m extremely large telescopes. These instruments will be able to achieve direct imaging at smaller inner working angles and probe the close vicinity of stars.

The left panel of Fig. 1.6 shows the best analogy for the direct imaging of exoplanets *high-contrast* challenge. Observing a firefly located near a lighthouse, miles away from us, is as difficult as observing an exoplanet orbiting a distant star. The difference with this scenario is that the light from the stars we observe never shuts down. For ad-

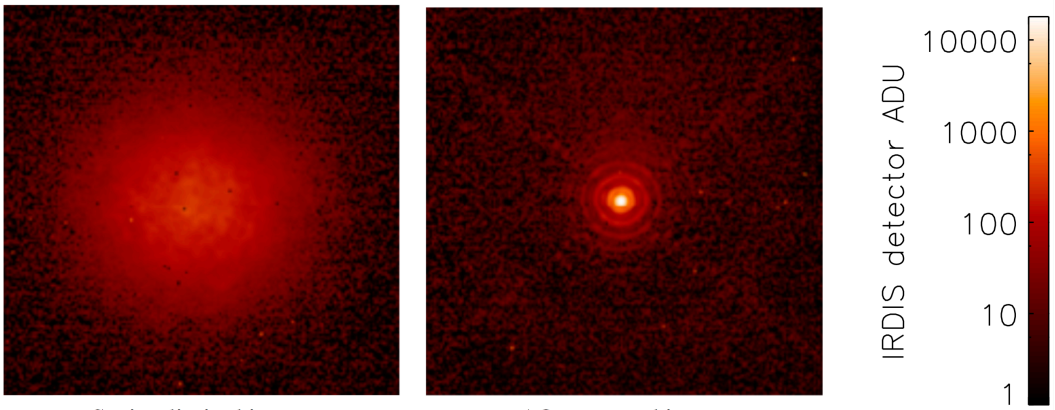


Figure 1.8: Adaptive optics corrections in H band at VLT/SPHERE. Left: Long exposure of a point source without AO correction. Right: AO corrected image in nominal seeing conditions. Taken from Sauvage et al. (2016).

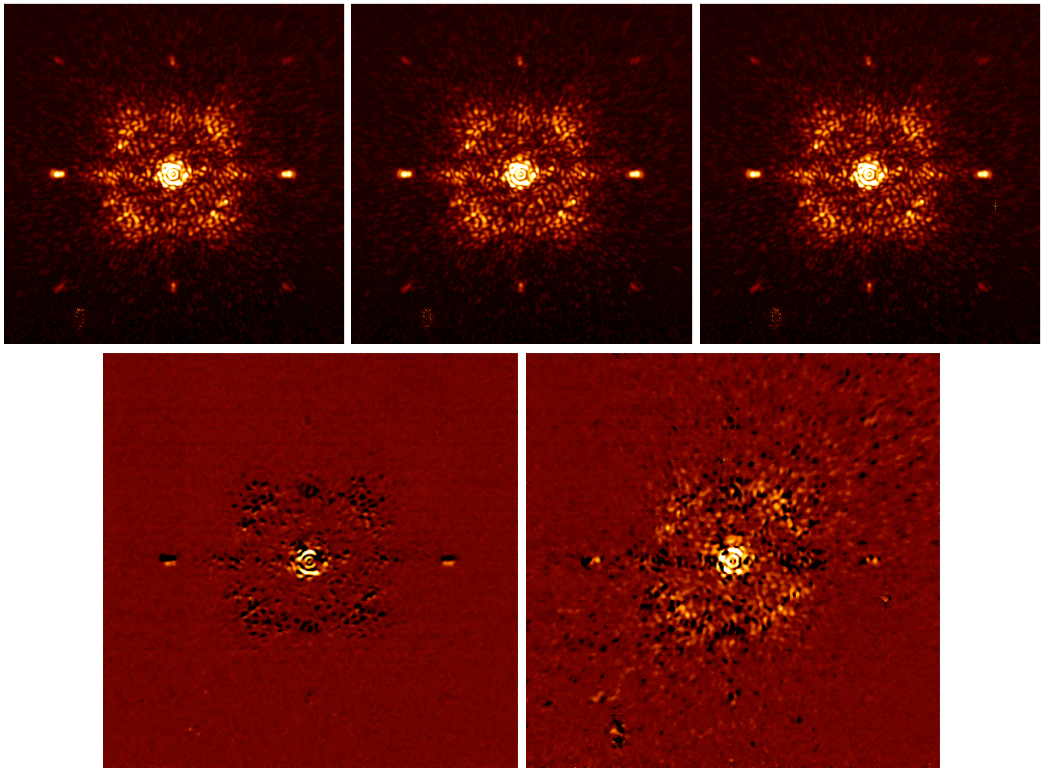


Figure 1.9: Top: Coronagraphic images recorded with VLT/SPHERE/IRDIS in closed-loop at three consecutive times t_0, t_0 + 10 mn, and t_0 + 100 mn. Bottom: Pairwise subtraction of the image at t_0 and the images taken 10 minutes later (bottom left) and 100 minutes later (bottom right). Taken from Martinez et al. (2013).

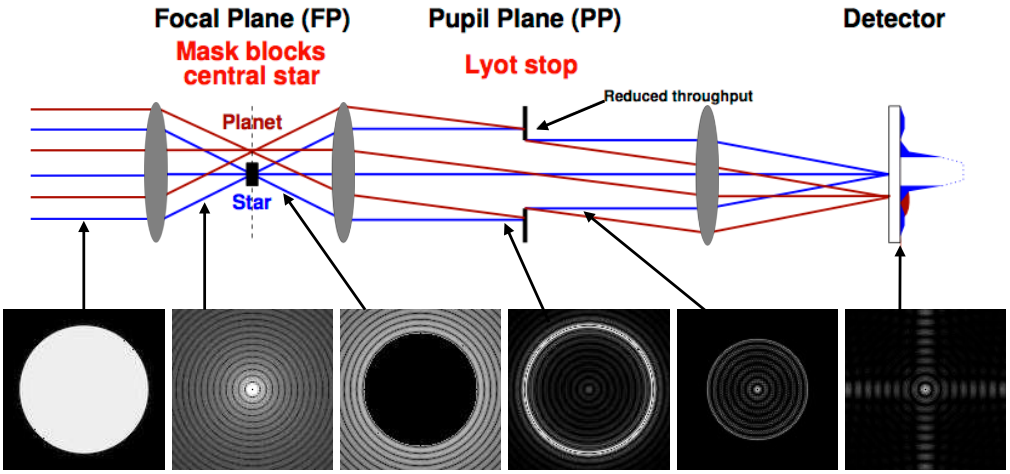


Figure 1.10: Optical layout of the Lyot coronagraph (first coronagraph used to observe the Sun's corona) and images at different stages of the optical path. Based on figure by M. Kenworthy.

dressing this problem of detecting companions several orders of magnitude fainter than their parent stars, we rely on coronagraphy. The role of a coronagraph is basically to reduce the contrast between the star and its potential companions by canceling the diffracted starlight from the on-axis star while allowing to pass that of the off-axis planets. Fig. 1.10 shows a schematic view of the Lyot coronagraph. In this layout a circular, opaque disk, the Lyot mask, is placed at the primary focus to remove from the optical path the core of the stellar PSF. The diffracted light outside of the mask is propagated to the pupil plane where a second opaque mask, the Lyot stop, is used to cancel most of the remaining diffracted starlight. The final image only contains a small contribution from the initial starlight. This simple setup has several disadvantages, such as the non-perfect cancellation of the starlight even when the Lyot mask and the star are aligned, the large inner working angle, and the reduced throughput due to the smaller size of the Lyot stop.

Several different coronagraphic technologies have emerged to improve the Lyot setup with the goal of improving the contrast at very short separations (small inner working angle) while preserving the signal from the potential companions (high throughput). Also, important characteristics of these different coronagraphic solutions are the sensitivity to instrumental aberrations and the spectral bandwidth at which they operate. Among the masks that are also placed on the focal plane we count the four quadrant phase mask (4QPM, Rouan

et al., 2000) and the vortex phase mask, such as the annular groove phase mask (AGPM, Mawet et al., 2005). Unlike the Lyot mask, the 4QPM and the vortex phase mask do not block the light (act on the amplitude) but introduce a phase-shift to the on-axis starlight, rejecting it outside the geometric image of the pupil. The AGPM brings great improvements in terms of inner working angle (better contrast in the close vicinity of the star) and in throughput (Mawet et al., 2013), at the cost of a high sensitivity to centering and tilt errors. It has been installed in several of the largest ground-based telescopes, such as the LBT, VLT, Subaru and Keck. Other approaches consist in the use of amplitude or phase pupil apodization (Spergel and Kasdin, 2001; Kenworthy et al., 2007), or combined solutions, such as the apodized pupil Lyot coronagraph (Soummer, 2005). See Mawet et al. (2012) for a thorough review of the coronagraphic solutions used in high-contrast imaging.

1.3.2 Observing strategies

Unfortunately, even with optimized wavefront control and state-of-the-art coronagraphy combined, speckle noise hinders the detectability of faint close-in companions. Therefore, special observing strategies are used to introduce various diversities with the goal of disentangling the signal of potential companions from the speckle noise field (Mawet et al., 2012; Fischer et al., 2014). These techniques differ on how the data is acquired at the instrumental level, which in turn changes the final distribution of the pixel signals with respect to the starlight. These techniques are differential by nature, as a key step is the subtraction of a model of the remaining starlight and speckle pattern from the science images. In this section we will focus on the observing strategies and their motivations. The next section will describe the state-of-the-art image processing techniques that are built on top of these observing techniques to exploit the particular structures of the data.

1.3.2.1 *Angular differential imaging*

The angular differential imaging observing strategy (ADI, Marois et al., 2006) aims to decouple, on the image plane, the planetary signal from the speckle noise field. In the case of ADI, the images are acquired in pupil-stabilized mode, so the image rotates with time while the aberrations and speckle noise field remains at the same orientation. On the image sequence, this results in the rotation (this is a fake movement caused by the pupil-stabilized observation) of the planet around the star in a circular trajectory. Keeping the pupil orientation constant, we obtain a more stable speckle halo around the star. This is thanks to the fact that the main sources of quasi-static speckles are locked to the pupil (e.g. spider structure). Owing to the field rotation,

the residual noise (after post-processing) averages incoherently after rotating the images to a common north. From the central limit theorem, the noise in the final combined image mostly becomes Gaussian (Marois et al., 2008a; Mawet et al., 2014).

ADI is the most commonly used technique for high-contrast imaging in spite of its limitations, such as being a time-demanding procedure (using significant telescope time, usually from one to a few hours) and the self-subtraction of close-in companions. See Subsections 1.3.3.1 and 2.5.1 for details on why signal self-subtraction occurs, when differential imaging post-processing is applied, and what are the approaches exist for mitigating this undesirable effect.

1.3.2.2 *Spectral differential imaging*

Another observing strategy that seeks to decouple the planetary signal from the speckles is multiple-channel spectral differential imaging (SDI, Sparks and Ford, 2002). The basic case, with only two images or wavelengths, was proposed by Racine et al. (1999) aiming to exploit the presence/absence of an exoplanet's signal in two or more adjacent bands.

Multiple-channel SDI employs integral field spectrographs (after pre-processing the images so as to have a frame per spectral channel) to exploit further the wavelength diversity and the different wavelength dependence of speckles and companions. In a multiple-channel SDI sequence, the speckles stretch (they appear to move radially from the center of the image) while the companion signal remains fixed as the wavelength increases (the PSF stretches as well but remains at the same position).

Multiple-channel SDI is usually combined with ADI on modern high-contrast imaging instruments to exploit both rotation and wavelength diversities. The resulting data cube has 4 dimensions in this case: time, wavelength and spatial dimensions.

1.3.2.3 *Reference differential imaging*

Reference star differential imaging (RDI, Mawet et al., 2009, 2010; Rameau et al., 2012; Mawet et al., 2012) uses observations of a different star with the goal of using pairwise subtraction to get rid of the speckle pattern. The motivation of this technique is to alleviate the weakness of ADI and SDI at small angles.

Unfortunately, this technique comes with important constraints, besides high stability and correlation that we expect from the next generation systems, such as short duty cycles, efficient low-order aberrations correction on both stars, matching magnitudes and parallactic angles[5]. The observations of the reference star can serve as a reference PSF but the complexity of this method is the determination of the flux

5 The angles associated to each frame in an ADI sequence.

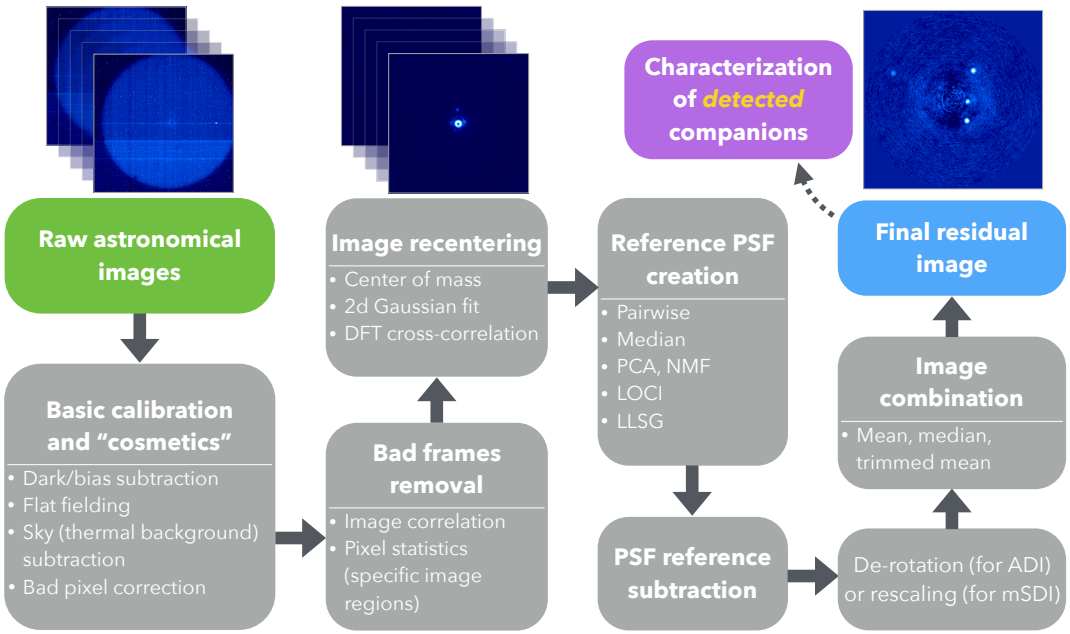


Figure 1.11: Differential imaging typical pipeline, from raw images coming from the telescope until the generation of a residual flux image.

scaling factor to avoid the creation of artifacts. A preferred solution is to build a reference PSF from the reference images in a similar way as how we proceed with ADI and multiple-channel SDI.

A special case of RDI, binary differential imaging (Rodigas et al., 2015), relies on the advantage of observing two stars simultaneously at the same wavelength for obtaining a highly correlated reference PSF. This technique comes with some limitations, like the fact that it is limited to visual binary in which only one star must have companions and that it cannot be used with coronagraphs. Observing techniques such as polarimetric and coherent differential imaging also seek to overcome the limitations of ADI and SDI at short separations.

1.3.3 Image processing techniques for high-contrast imaging

Data processing constitutes a critical component of high-contrast exoplanet imaging. Its role is almost as important as the choice of a coronagraph or a wavefront control system, and it is intertwined with the chosen observing strategy. Data-processing for HCI is a multi-step process, represented in Fig. 1.11. Raw images coming from the telescope are stored in FITS files (Wells et al., 1981). FITS stands for Flexibile Image Transport System, and is the International Astronomical Union recognized standard for storing astronomical data. A FITS file consists of a header, stored in human-readable ASCII, and a data section as a sequence of pixel values. These pixels can be integers (8-

bit unsigned, 16-bit, 32-bit or 64-bit signed integers) or floating point numbers (32-bit or 64-bit). In raw form, the images are usually in 32-bit integer precision. As the images progress down the pipeline, the data type changes to 32-bit (single) or 64-bit (double) float precision. As a first step, the images need to be calibrated. These procedures are common to all astronomical images: dark current (and bias offset) subtraction, flat field correction to mitigate the nonuniformity of the detector response, and correction of bad pixels (dead and hot detector pixels). The dark current and bias offset are negligible for most of modern detectors. For details in astronomical image calibration, see Berry and Burnell (2000).

Since most of the ground-based HCI observations are done in the near- to mid-infrared (where we have better contrast for young exoplanets), we often have to subtract the background radiation due to the sky and warm surfaces in the telescope/instrument. The purpose of this subtraction is to go from a frame such as the ones on the top-left of Fig. 1.11 to the ones on the top-middle. These images will have a smoother background with near zero values (without outlying pixels) revealing the central star and the speckle pattern. From the calibrated observing sequence we discard the outlying frames, those where the AO loop opened, the conditions worsened, or the star was misaligned with the coronagraph. These are detected by measuring the correlation of the images or based on the pixels statistics of specific regions of the frame, and confirmed visually. The last preprocessing procedure, frames alignment or re-centering, is carried out to correct for small shifts of the star center and to put the star at the center of each frame (which is critical for ADI). For instruments with real-time coronagraphic centering capabilities like VLT/SPHERE we usually skip the frame-to-frame re-centering. At this point, we have a calibrated data cube or image sequence we can work with.

The ultimate goal of post-processing procedures is to combat the speckle noise (increasing the ability to detect faint and close-in companions) by improving the contrast and reducing the image dynamic range. Fine-tuned post-processing is then the last piece in the landscape of HCI for exoplanet (and disk) detection. As mentioned before, there is a close connection between the observing technique and the image processing algorithm. Throughout this dissertation, a distinction will be made between the observing technique and the post-processing algorithm (ADI will refer only to the way the data is acquired).

Algorithms of different complexities are used to build optimal reference PSFs, exploiting the diversities and the data particularities of each observing technique. For some observing techniques like dual-band SDI and RDI, pairwise subtraction could be enough to subtract most of the speckle noise and leaked starlight. For ADI sequences, we exploit the fact that the planet moves in circular trajectories with

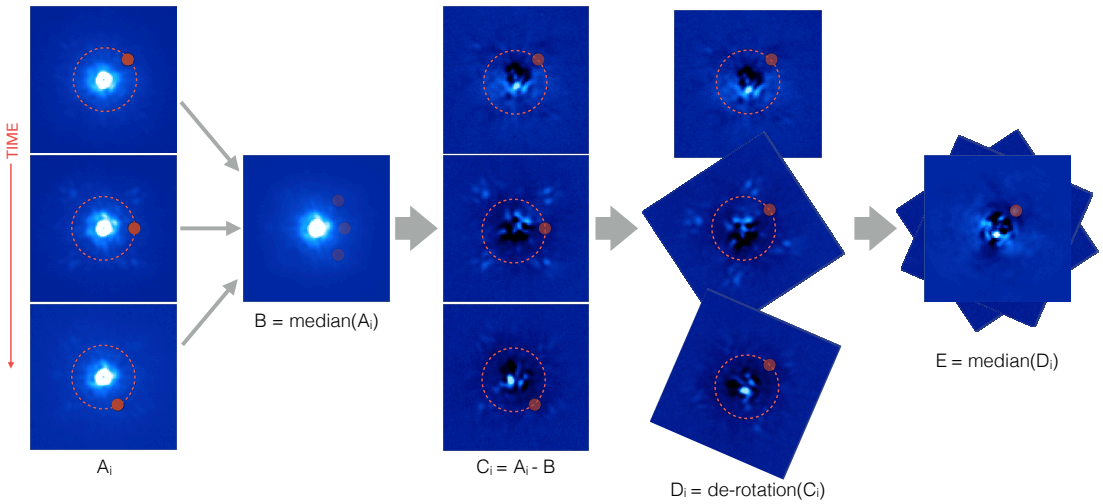


Figure 1.12: Diagram of ADI median subtraction, where the reference PSF is constructed by taking the median of the image sequence.

respect to the speckle field. In the case of multiple-channel SDI, we exploit the radial movement of the planet PSF, after re-scaling the images with respect to the longest wavelength channel. From now on, and throughout this dissertation, we will call reference PSF the algorithmically built image that we use with differential observing techniques for subtracting the scattered starlight and speckle noise pattern, to enhance the signal of disks and exoplanets.

1.3.3.1 *Median subtraction*

For ADI, modeling the speckle pattern with a simple median combination of the sequence is the simplest approach. This was originally proposed by Marois et al. (2006) for obtaining a final science-ready image with the ADI observing strategy. A schematic representation of the ADI median subtraction is shown in Fig. 1.12. The main idea behind this, is that the moving planet does not show up in the median image of the sequence, which can then be used as a first approximation of the leaked starlight and speckle noise pattern. As mentioned earlier, after subtracting the median, the residual noise averages incoherently once the images are rotated to a common north. From the central limit theorem, the noise in the final combined image mostly becomes Gaussian (Marois et al., 2008a; Mawet et al., 2014). Some improvement can be achieved by processing the frames in an annulus-wise fashion and applying a rotation criterion for selecting the references. These filtered reference frames (annuli) are used to build a PSF for each image of the sequence.

By principle, ADI is limited in the close-in vicinity of the star, where the planet rotation is more constrained (see Fig. 2.2 from Chapter

2). The basic median subtraction approach fails to clean up the innermost region of the images, where strong residual speckle noise is mixed with planetary signal, weakened by self-subtraction. More advanced techniques aim to improve this situation by constructing in more complex ways, for each image, a reference PSF from appropriately selected images contained in the same sequence. These will be introduced in the following subsections. In the case of multiple-channel SDI the radial displacement of the planet (after re-scaling the images with respect to the longest wavelength channel) is generally not large enough to keep the median free from the companion signal.

1.3.3.2 *LOCI*

The median subtraction approach can be improved significantly by employing computationally more expensive approaches. The family of LOCI (locally optimized combination of images, Lafrenière et al., 2007) algorithms aims to create a reference PSF as a linear combination of reference images (in the case of ADI, these are the rest of the frames in the sequence after applying a rotation/proximity threshold) independently inside multiple subregions, in which the residual noise is minimized in the least-squares sense. A schematic representation of the ADI-LOCI processing is shown in Fig. 1.13. The LOCI image partition can be seen in Fig. 1.14. The optimization regions (each annulus subsection) are larger than the subtraction zone (dark grey regions). The coefficients used for subtraction of the speckles in the subtraction zones are determined by a minimization of the noise in the larger optimization subsections (Lafrenière et al., 2007). This approach provides a significant gain in signal-to-noise but introduces significant flux and position biases.

Throughout this dissertation, upper-case letters are used to denote matrices. Let us consider a sequence of n images (ADI sequence with n images) and a matrix $M \in \mathbb{R}^{n\times p}$ whose rows are vectorized versions of patches from those images (it does not matter if the image vectors are in column form). As this is done in local patches, p are the number of pixels in an optimization zone. According to a proximity (parallactic angle) threshold, we filter the reference images for each image m. These are stored in the rows of matrix $R \in \mathbb{R}^{k\times p}$ where $k < n$. Therefore the sample covariance of the reference set is:

$$S = \frac{1}{p-1} RR^T, \tag{1.1}$$

where $S \in \mathbb{R}^{k\times k}$. LOCI consists in finding the k-dimensional vector of optimal coefficients, c, such that (Savransky, 2015):

$$c = \underset{c}{\operatorname{argmin}} \left\| m - R^T c \right\|, \tag{1.2}$$

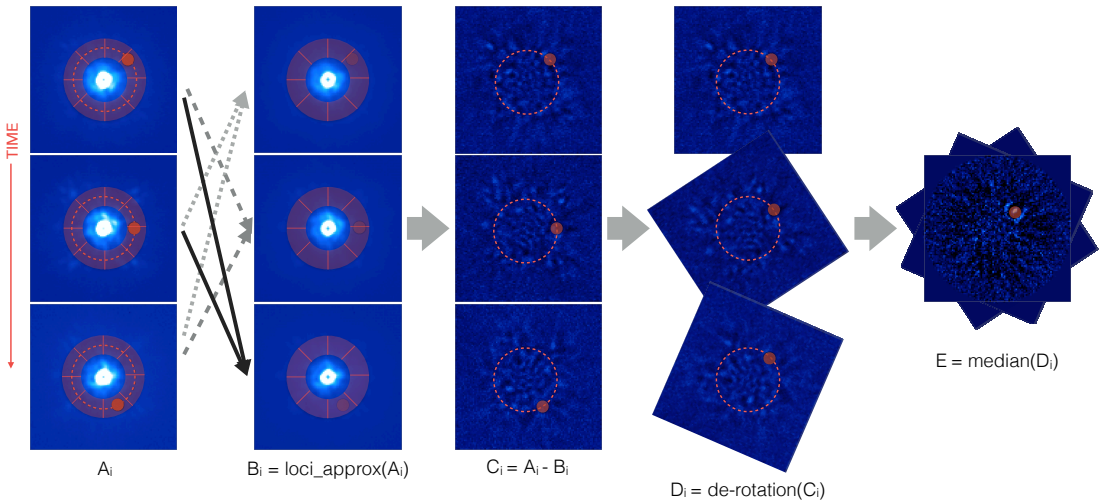


Figure 1.13: Diagram of LOCI post-processing. *Loci_approx* denotes the patch-wise least-squares approximation of the input images.

where the norm is typically the ℓ^2. This is analogous to solving the overdetermined linear system:

$$R^T c = m, \tag{1.3}$$

which usually has no unique solution, considering that $p > k$ (more pixels than independent images). However, when RR^T is full rank[6], the left pseudo-inverse of R^T gives the minimum least-squares error solution satisfying 1.2:

$$c = (RR^T)^{-1} Rm, \tag{1.4}$$

and the residual signal res is given by the subtraction of the linear combination of reference images from the image m:

$$res = m - R^T c = \left(I - R^T (RR^T)^{-1} R\right) m = \left(I - R^T \frac{S^{-1}}{p-1} R\right) m, \tag{1.5}$$

where I is the identity matrix. S is only guaranteed to be positive semi-definite[7] and it is not necessarily invertible. We then rely on the pseudoinverses of S to get the residual signal.

Modifications of this algorithm, such as adaptive LOCI (Currie et al., 2012a) and matched LOCI (Wahhaj et al., 2015) incorporate image correlation and synthetic sources injection to maximize the signal-to-noise of companions. Other versions of LOCI, such as damped

6 A matrix is full row rank when each of the rows of the matrix are linearly independent and full column rank when each of the columns of the matrix are linearly independent. For a square matrix these two concepts are equivalent and we say the matrix is full rank if all rows and columns are linearly independent.

7 Rows of R are not linearly independent.

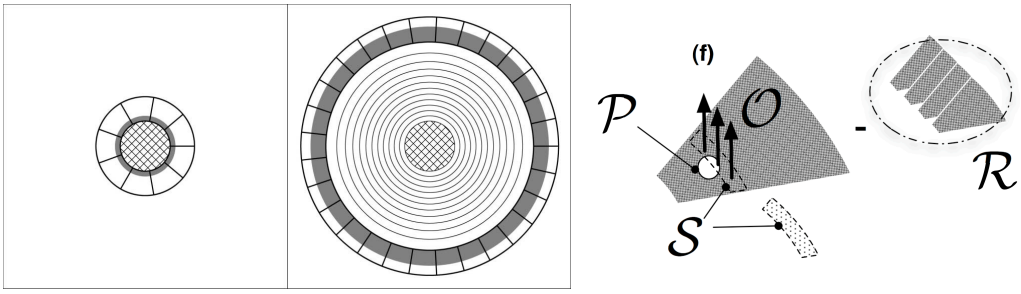


Figure 1.14: Left: Diagram of LOCI image partition. The optimization regions (each annulus subsection) are larger than the subtraction zone (dark grey regions). The coefficients used for subtraction of the speckles (in the subtraction zones) are determined by a minimization of the noise in the larger optimization subsections. Taken from Lafrenière et al. (2007). Right: Damped LOCI with companion masking. *O* is the optimization zone, *P* denotes the position of the companion, *S* is the subtraction zone and *R* is the set of reference images. Taken from Pueyo et al. (2012).

LOCI (Pueyo et al., 2012) and template LOCI (Marois et al., 2014), are adapted to multiple-channel SDI (using integral field spectrographs). Damped LOCI modifies the original cost function of LOCI that determines the approximation linear coefficients (minimization of the residuals over the optimization zone) to enforce the maximization of the residuals in the subtraction zone. Further improvement on the retrieved spectra can be achieved by masking the pixels of the candidate companion (Pueyo et al., 2012). Template LOCI uses an input spectrum and template PSFs to optimize the reference image least-squares coefficients to minimize the planet self-subtraction (maximizing its throughput per wavelength) while simultaneously providing a maximum suppression of the speckle noise (Marois et al., 2014).

1.3.3.3 *Frame differencing*

Given an ADI sequence, one simple solution to getting a reference PSF is to set pairs of frames I_i and I_j based on the minimum distance between them, and make subtractions in both directions $I_i - I_j$ and $I_j - I_i$. The distance criterion used is the sum of absolute differences (SAD) of the pixels intensities in the quadrants of an annulus inside the frames. These pairs are unique, and both frames are chosen taking into account a parallactic angle threshold to ensure enough field rotation (of a given FWHM displacement) and avoid self-subtraction. The SAD can be expressed as: $SAD = \sum_{i=1}^{p} |x_i - y_i|$, where p is the number of pixels in a portion of the images, x_i and y_i are the values of those pixels for I_i and I_j respectively. The image is partitioned

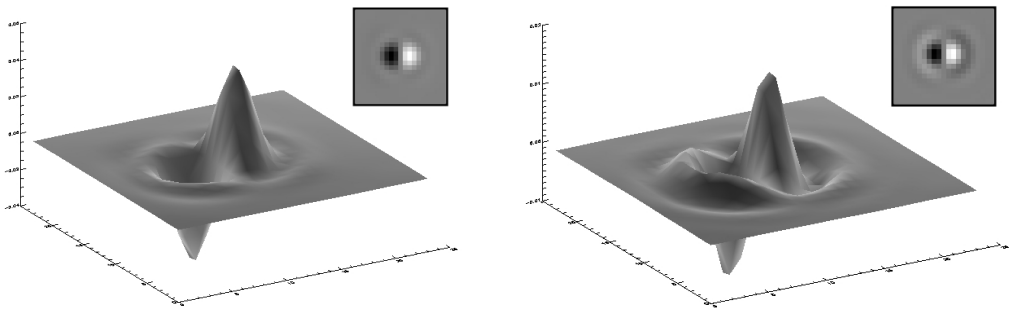


Figure 1.15: Planet signatures generated after performing the ADI pairwise frame differencing. Left: Signature obtained without high-pass filtering. Right: Signature after high-pass filtering has been applied to the frame sequence. Taken from Cantalloube et al. (2015).

in quadrants of annuli of variable size, typically 1xFWHM width. A similar frame-differencing approach is proposed by Hagelberg et al. (2016) with their GRAPHIC ADI reduction pipeline.

1.3.3.4 *ANDROMEDA*

ANDROMEDA (ANgular Differential OptiMal Exoplanet Detection Algorithm, Mugnier et al., 2009; Cantalloube et al., 2015) treats the ADI sequence in a pairwise way and ensures that the images are chosen close enough in time to guarantee the stability of the speckle noise and thereby allow its suppression (the second image from every pair is used as a reference PSF for the first). This enables the creation of a particular signature, as shown in Fig. 1.15, that can be modeled. ANDROMEDA uses a maximum likelihood approach with the aim of estimating the position and the flux of any point source present in the field of view (Cantalloube et al., 2015). This approach differs from the rest of the differential imaging algorithms in that it characterizes the potential companions.

1.3.3.5 *PCA*

The most recent ADI post-processing algorithmic approach is the family of principal component analysis (PCA) based algorithms (Soummer et al., 2012; Amara and Quanz, 2012). PCA, or Karhunen-Loeve (KL) transform, is a widely used statistical tool developed during the first half of the past century. PCA serves, in this case, as a subspace projection technique for constructing a reference PSF. This reference PSF is constructed for each image as the projection of the image onto a lower-dimensional orthogonal basis (the principal components) extracted from the reference images via PCA. A schematic representation of an ADI-PCA workflow is shown in Fig. 1.16.

The principal components of our image sequence can be obtained using eigen decomposition or singular value decomposition (SVD). Let us consider again a matrix $M \in \mathbb{R}^{n \times p}$ whose rows are vectorized versions of the ADI sequence frames. p is the number of pixels in an image. Here we focus on the full-frame PCA case, but the approach can be extended to patch processing (for details on improvements over the full-frame version of PCA, see Chapter 2). The frame-to-frame covariance S, defined as before, is Hermitian and also diagonizable as $S\Phi = \Phi\Lambda$ where $\Phi \in \mathbb{R}^{n \times n}$ is the unitary matrix whose columns are the eigenvectors of S and $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix of the eigenvalues. Instead of using a pseudoinverse, we can project the target signal onto a subset of an optimally energy-compacting basis given by (Savransky, 2015):

$$Z = \Phi^T M, \tag{1.6}$$

where Z is the $n \times p$ matrix of KL transform vectors. Z is truncated, keeping only k rows to reconstruct our images:

$$\hat{m} = Z_k^T Z_k m, \tag{1.7}$$

and obtain the residual signal:

$$res = (I - Z_k^T Z_k) m. \tag{1.8}$$

Optionally, one can use the SVD to compute a projection matrix that preserves the desired amount of variance. SVD is a matrix factorization such that:

$$M = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T, \tag{1.9}$$

where the vectors u_i and v_i are the left and right singular vectors, and σ_i the singular values of M. SVD is involved in several least-squares problems, such as finding the best low-rank approximation of M in the least-squares sense, i.e.,

$$\underset{X}{\operatorname{argmin}} \|M - X\|_F^2, \tag{1.10}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm, i.e. the sum of the squared elements of its argument. If we consider that σ_i are ordered by decreasing value (see Absil et al., 2008, chap. 2), by keeping the first k right singular vectors, we form the orthonormal basis for the low-dimensional subspace capturing most of the variance of M. Notice that the right singular vectors of M are the eigen vectors of S, the covariance of M.

Fig. 1.17 gives an example of the combined power of hardware and software (in this case a PCA processing) HCI solutions. PCA post-processing has become the standard in the high-contrast community.

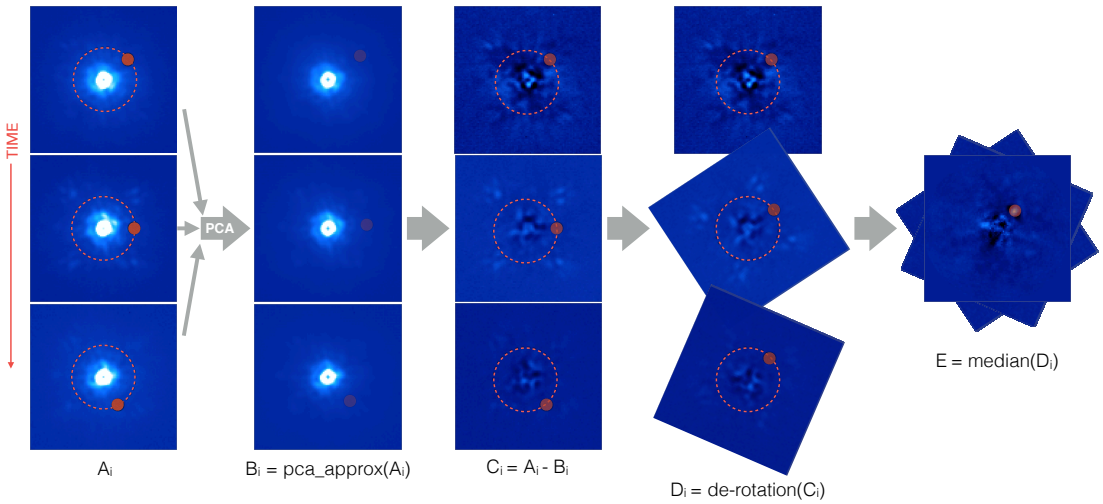


Figure 1.16: Diagram of PCA post-processing. *Pca_approx* denotes the projection of the images onto the first k principal components to obtain the reference PSFs.

The main advantages of this approach are that it can be applied to the full images very efficiently using SVD, the reduced number of free parameters (basically the size of the basis) and that PCA enables forward modeling of astrophysical sources by fitting an astrophysical model directly to the reduced images without introducing degeneracies (Soummer et al., 2012; Pueyo, 2016). Unfortunately, when building this reference PSF from the science data itself, PCA comes with companion self-subtraction biases similar to the rest of differential imaging approaches.

Amara and Quanz (2012) propose a comparison of LOCI and full-frame PCA using synthetic companions. The improvements of PCA come in both signal-to-noise and in the detectability of fake companions, especially those close to the star. Recently, Pueyo (2016) has further developed a forward modeling technique for PCA (called KLIP-FM) that retrieves the unbiased companion photometry at the expense of a higher computational cost. Another PCA improvement was presented by Ruffio et al. (2017), who coupled the KLIP-FM algorithm with a matched filtering approach. The KLIP-induced distortion of the astrophysical signal is included in the matched filter template by computing a forward model of the PSF at every position in the image.

PCA based post-processing is very versatile and can be applied in a similar way to RDI, multiple-channel SDI, and IFS data with wavelength and rotational diversities (mSDI+ADI). In the case of RDI the principal components are learned from the reference star observations (different than the target) and used for modeling the reference

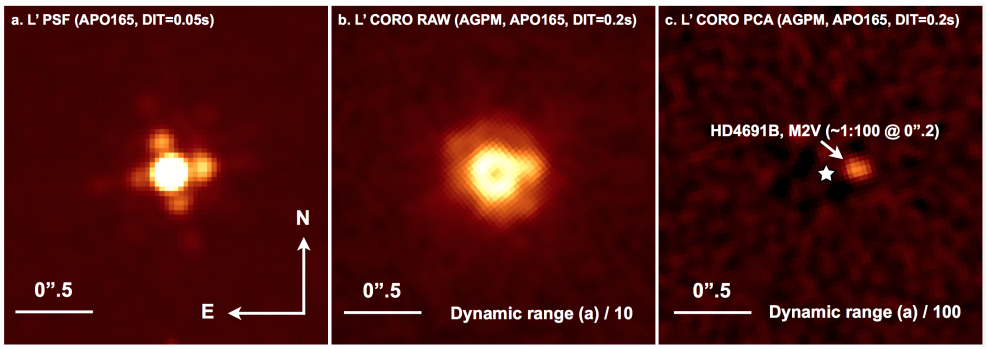


Figure 1.17: Illustration of the combined effect of wavefront-sensing, coronagraphy and post-processing on the dynamic range of the images (and therefore on the reduced contrast close to the star). Left: L'-band VLT/NACO adaptive optics observations without a coronagraph. Middle: Image (from the ADI sequence) where the star has been centered on the vortex coronagraph of NACO, featuring a dynamic range 10 times smaller than on the previous image (the dark hole at the center is caused by the coronagraph). Right: Image after ADI-PCA post-processing was applied, revealing a companion. The dynamic range is a factor 100 smaller than in the left-most image. Image taken from Mawet et al. (2013).

PSF of the target sequence. For multiple-channel SDI, the frames are re-scaled to match the speckles and create an apparent radial movement of a potential companion. Then PCA is applied model the reference PSF as a low-rank approximation of the dataset. In the case of mSDI+ADI, several options are available. Perhaps the simplest one is to perform a two-stages PCA, the first for each SDI cube (same as the multiple-channel SDI) and one on the cube of residuals in an ADI fashion.

1.4 SCOPE AND OUTLINE OF THIS DISSERTATION

This dissertation describes the results obtained during my PhD research program, focused on novel approaches to data processing of HCI sequences, with a special emphasis on the detection of exoplanets through angular differential imaging. My work as a graduate student was developed at the interface of several disciplines, as illustrated in Fig. 1.18.

In the first part of this manuscript, I present my contribution to the field of HCI in terms of scientific software development. In Chapter 2, I describe the Vortex Image Processing library, an open-source `Python` package, where I implemented state-of-the-art and novel pre- and post-processing algorithms for the reduction and analysis of HCI

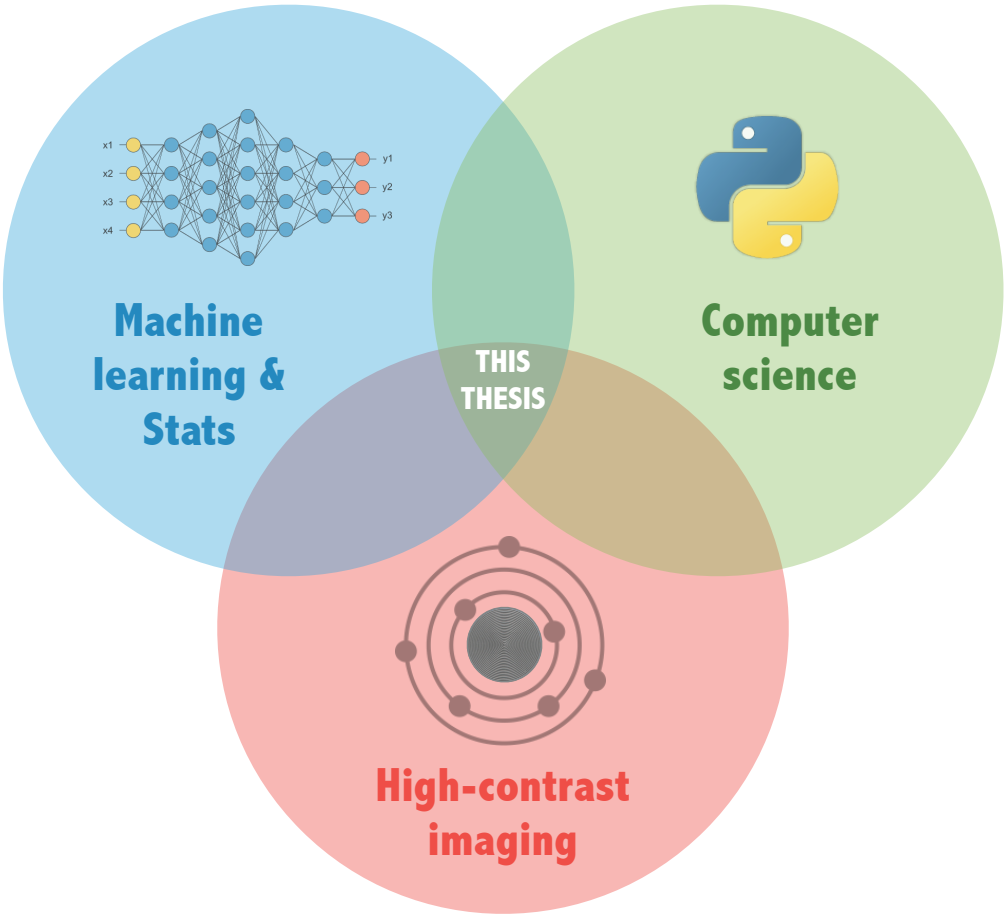


Figure 1.18: Venn diagram showing the three main components of my PhD thesis project.

data. This work started from scratch, with the goal of replacing an existing data reduction pipeline written in IDL[8] by Dimitri Mawet and Olivier Absil. VIP was inspired by other `Python` open source libraries, and now serves the whole HCI community beyond my core team. In Chapter 3, I showcase the versatility and capabilities of my library, presenting examples of its use on real on-sky HR8799 datasets.

In the second part of this dissertation, I discuss novel algorithmic approaches focused on the ADI observing technique. The HCI community consensus is that ADI and SDI are very efficient at large angular separations, but these techniques cannot be applied efficiently at small angles (Mawet et al., 2012). My work shows that there is a lot of room for improvement in the data processing department, and that the limited sensibility of ADI at small separations can be mitigated with advanced post-processing and discriminative models. In particular, I present two different approaches to ADI post-processing with the goal of detecting exoplanets, both based on machine learning recent developments: the LLSG algorithm and the SODIRF/SODINN supervised detection framework. The first is an evolution of PCA-based post-processing approaches for differential imaging, and the latter is a totally new approach where the detection of companions is formulated as a classification process. SODINN employs state-of-the-art neural network architectures to exploit the ADI datasets in a supervised learning framework. Finally, I compare both approaches to their predecessors (e.g. PCA) in a robust signal detection theory framework and comment on the use of metrics for assessing algorithms performance.

8 IDL is a commercial interactive programming language widely used in Astronomy.

Part I

STATE-OF-THE-ART DIFFERENTIAL IMAGING POST-PROCESSING TECHNIQUES

2 VORTEX IMAGE PROCESSING PACKAGE FOR HCI

Contents

ABSTRACT

In this chapter I present the Vortex Image Processing (`VIP`) library, a `Python` package dedicated to astronomical high-contrast imaging. This package relies on the extensive `Python` stack of scientific libraries and aims to provide a flexible framework for high-contrast data and image processing. I describe the capabilities of `VIP` related to processing image sequences acquired using the angular differential imaging (ADI) observing technique. `VIP` implements functionalities for building high-contrast data processing pipelines, encompassing pre- and post-processing algorithms, potential sources position and flux estimation, and sensitivity curves generation. Among the reference point-spread function subtraction techniques for ADI post-processing, `VIP` includes several flavors of principal component analysis (PCA) based algorithms, such as annular PCA and incremental PCA algorithm capable of processing big datacubes (of several gigabytes) on a computer with limited memory. Also, I present a novel ADI algorithm based on non-negative matrix factorization

(NMF), which comes from the same family of low-rank matrix approximations as PCA and provides fairly similar results. `VIP` is available at `http://github.com/vortex-exoplanet/VIP` and is accompanied with Jupyter notebook tutorials illustrating the main functionalities of the library. *This chapter is based on previous work published in Gomez Gonzalez et al. (2017).*

2.1 INTRODUCTION

In this chapter we present a `Python` library for image processing of high-contrast astronomical data: the Vortex Image Processing (`VIP`, Gomez Gonzalez et al., 2016b, 2015) package. `VIP` provides a wide collection of pre- and post-processing algorithms and currently supports three high-contrast imaging observing techniques: angular, reference-star, and multi-spectral differential imaging. The code encompasses not only well-tested and efficient implementations of known algorithms but also state-of-the-art new approaches to high-contrast imaging tasks. Our library has been designed as an instrument-agnostic toolbox featuring a flexible framework where functionalities can be plugged in according to the needs of each particular dataset or pipeline. This is accomplished while keeping `VIP` easy-to-use and maintaining an extensive documentation. Finally, our package is released as open-source, hoping that it will be useful to the whole high-contrast imaging community.

This chapter is organized as follows. Section 2.2 gives a general overview of the design and structure of `VIP`. Section 2.3 introduces the signal-to-noise (S/N) definition used in `VIP`. In section 2.4 we briefly describe the pre-processing and cosmetic functionalities implemented in our package. Section 2.5 goes into the details of reference PSF subtraction for ADI data, exploring the available post-processing algorithmic approaches in `VIP`. Section 2.7 describes the photometric and astrometric extraction procedures and finally Section 2.8 describes the sensitivity limits estimation implemented in our package.

2.2 PACKAGE OVERVIEW

The design and development of `VIP` follow modern practices for scientific software development such as code modularity, the active use of a version control system (`git`) and extensive documentation (Wilson et al., 2014). The code is being developed in `Python`, and relies on its vast ecosystem of scientific open-source libraries/packages including `numpy` (van der Walt et al., 2011), `scipy` (Jones et al., 2001), `matplotlib` (Hunter, 2007), `astropy` (Astropy Collaboration

Table 2.1: `VIP` subpackages.

Subpackage	General description
`conf`	Timing, configuration and internal utilities
`exlib`	Code borrowed from external sources
`fits`	Fits input/output functionality
`llsg`	Local low-rank + sparse + Gaussian-noise decomposition for ADI data
`madi`	Standard ADI recipe (median PSF reference)
`negfc`	Negative fake companion technique
`nmf`	Non-negative matrix factorization for ADI data
`pca`	PCA-based algorithms for ADI, RDI and mSDI data
`phot`	Signal-to-noise and detection of point-like sources. Contrast curve generation
`preproc`	Low-level image operations. Pre-processing and cosmetic procedures
`stats`	Statistics from frames and cubes, correlation and sigma clipping procedures
`var`	Filtering, 2d-fitting, shapes extraction and other utilities

et al., 2013), `scikit-learn` (Pedregosa et al., 2011), `pandas` (McKinney, 2010) and `scikit-image` (van der Walt et al., 2014). For low-level image processing operations, `VIP` can optionally use, through its `Python` bindings, `OpenCV` (Bradski, 2000), a fast and robust C/C++ library for computer vision and image processing. The latest development version of `VIP` is available on GitHub[1], which is also the platform where users and/or collaborators can report bugs and make change requests. Every function and class in `VIP` has its own internal documentation attached describing the aim, arguments (inputs), and outputs. The internal documentation is part of the `VIP`'s web documentation[2], which also provides help in installation and troubleshooting. A tutorial dedicated to ADI, in the form of a Jupyter notebook, is shipped in a separate repository[3].

The structure of `VIP`, shown in Table 2.1, is modular and allows easy extension and re-utilization of functionalities. The code is organized, as any other `Python` library, in subpackages (directories) encapsulating modules (`Python` files), which in turn contain the functions

1 http://github.com/vortex-exoplanet/VIP
2 http://vip.readthedocs.io/en/latest/
3 http://github.com/carlgogo/vip-tutorial

and classes. It is important noting that `VIP` is not a pipeline per se but a library, inspired in well established projects such as `astropy` or `scikit-learn`, and does not provide a predefined linear workflow. Instead, the user must choose which procedures to use and in which order. The results of `VIP`'s calculations are kept in memory or displayed, e.g. in plots or figures, and can be later on saved to disk in the form of fits files. In the following paragraphs, we briefly describe the most relevant functionalities of each subpackage of `VIP`.

The subpackage `fits` includes functions for handling files in the FITS format, through `Astropy` functionalities. It also includes a `Python` class which allows controlling `SAOImage DS9` windows (based on the interface to `SAOImage DS9` through `XPA` from the `RO Python` package) and displaying `numpy` arrays. Thanks to these functions, `VIP` can be fed from disk with any FITS file containing a high-contrast imaging datacube.

The subpackage `phot` includes functionalities such as S/N estimation, S/N maps generation, automatic detection of point-like sources, fake companion injection in ADI cubes, and sensitivity limits computation. The formal definition of S/N is given in Section 2.3. The subpackage `stats` contains functions for computing statistics from regions of frames or cubes, sigma filtering of pixels in frames, and for computing distance and correlation measures between frames. The subpackage `var` includes image filtering, shapes extraction from images and 2d-fitting (Gaussian, Moffat) among other functionalities.

Finally, the subpackage `preproc` contains low-level image operations and pre-processing functionalities as described in Section 2.4, while the subpackages `llsg`, `madi`, `pca` and `negfc` contain the post-processing algorithms, which are described in Section 2.5 for the case of ADI data.

2.3 SIGNAL-TO-NOISE RATIO

For calculating the S/N, we depart from the previously used definition in high-contrast imaging where the pixels were assumed to be statistically independent and the S/N was basically considered as the ratio of the flux in an aperture centered on the planet to the standard deviation of the pixels in an annulus at the same radius. We instead adopt the definition proposed by Mawet et al. (2014), which is based on a Student *t*-test (Student, 1908) and considers the problem of small sample statistics applied to small-angle high-contrast imaging. The number of resolution elements (λ/D) decreases rapidly toward small angles, thereby dramatically affecting confidence levels and false alarm probabilities. In this small sample regime, a two-sample t-test is used to see whether the intensity of a given resolution element is statistically different from the flux of similar λ/D circular

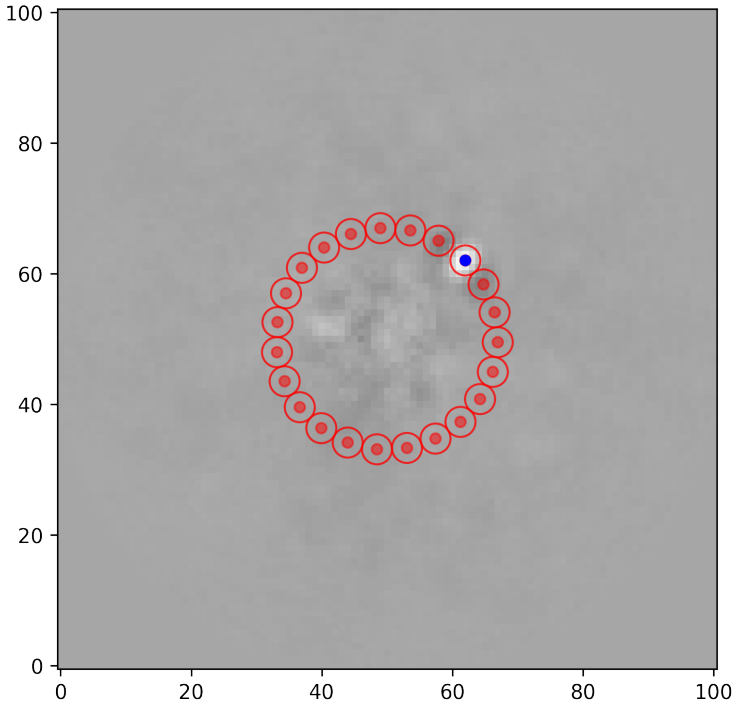


Figure 2.1: Circular apertures used for the S/N calculation. The test aperture is marked with a blue dot. The noise estimation takes into account the information at the same radial distance.

apertures at the same radius r from the star. When one of the samples contains the single test resolution element, the two-sample t-test brings a better definition of S/N and is given by:

$$S/N = \frac{\overline{x}_1 - \overline{x}_2}{s_2\sqrt{1+\frac{1}{n_2}}}, \tag{2.1}$$

where $\overline{x}_1$ is the intensity of the single test resolution element, n_2 the number of background resolution elements at the same radius ($n_2 = \mathrm{round}(2\pi r) - 1$, where r is given in λ/D units), and $\overline{x}_2$ and s_2 are the mean intensity and the empirical standard deviation computed over the n_2 resolution elements. In Fig. 2.1 the background apertures are marked with red dots while the test aperture is marked with a blue dot.

2.4 PRE-PROCESSING

`VIP` accepts datacubes, or sequence of images stacked in a 3d FITS file, that have undergone basic astronomical calibration procedures. These procedures, such as flat fielding and dark subtraction, in spite of their simplicity, are not included in `VIP` due to the heterogeneity

of data coming from different observatories. This is a sacrifice that we made in order to maintain `VIP` as an instrument-agnostic library. We let the users perform these procedures with their own tools or with dedicated instrument pipelines. `VIP` requires frames that have been at least successfully flat fielded and provides algorithms for any subsequent pre-processing task.

Subpackage `preproc` contains the functions related to low-level image operations, such as resizing, upscaling/pixel binning, shifting, rotating and cropping frames. All these functions have a counterpart for handling cubes or images sequences. Also, it is possible to temporally sub-sample and collapse/combine sequences in a single frame. Combining the images can be done via a pixel-wise mean, median or trimmed mean operation (Brandt et al., 2013).

Pre-processing steps are an important in the high-contrast imaging processing pipeline. In the case of ADI sequences, it is critical to have the star at the very center of the frames and have them all well aligned. `VIP` (subpackage `preproc`), makes it possible to register the frames by using 2d-Gaussian or Moffat fits to the data, applying Fourier cross-correlation (DFT upsampling method, Guizar-Sicairos et al., 2008), computing the Radon transform (Pueyo et al., 2015) for broadband images, or by fitting the position of satellite spots specifically created by ripples on the deformable mirror (Wertz et al., 2016). `VIP` includes procedures for detecting bad pixels from images and bad frames from datacubes. Bad pixels are replaced with the median of the nearest neighbor pixels inside a square window of variable size. Bad frame detection is implemented using pixel statistics (i.e. using the pixels in a centered annulus at a given radius), frame correlation, or ellipticities of point-like sources for detecting and discarding outliers. We suggest to discard the bad frames from a sequence before proceeding to the post-processing stage.

In certain scenarios, sky subtraction might be a desirable step. We implemented in `VIP` an algorithm for computing optimal sky background frames, learned from the sky frames taking during the same observing run. We apply a mask to the central core (and optionally to the spider arms) in both the science frames and the eigenvectors learned from the sky frames. Then we obtain the coefficients to approximate each science frame as a combination of the masked sky eigenvectors in the least-square sense. These are the optimal sky backgrounds for each science frame, which generally outperform the median sky subtraction common procedure. Recently, a similar approach was independently proposed by Hunziker et al. (2017).

2.5 ADI POST-PROCESSING

2.5.1 Median reference PSF subtraction

Subpackage `madi` contains the implementation of the most basic reference PSF subtraction for ADI data (Marois et al., 2006), usually called classical ADI in the literature. In this procedure a single reference PSF is modeled as the median of the stack, trying to capture the static and quasi-static structures of the sequence. This algorithm can also work in annular mode, where an optimized PSF reference can be built for each annulus, taking into account a parallactic angle threshold ω for enforcing a given amount of field rotation in the reference frames. The threshold ω is defined as:

$$\omega = 2 \arctan \frac{\delta \cdot \mathrm{FWHM}}{2r}, \tag{2.2}$$

where FWHM is the Gaussian full width at half maximum in pixels, δ a user-defined threshold parameter, and r the angular separation for which the parallactic angle threshold ω is defined (see Fig. 2.2). The enhanced reference PSF is built for each annulus by median combining the m closest in time frames, after discarding neighboring frames according to the threshold ω. Median reference PSF subtraction has limited performance in the small-angle regime, and it has been superseded by more advanced post-processing techniques.

2.5.2 PCA-based algorithms for reference PSF subtraction

PCA is an ubiquitous method in statistics and data mining for computing the directions of maximal variance from data matrices. It can also be understood as a low-rank matrix approximation (Absil et al., 2008). PCA-based algorithms for reference PSF subtraction on ADI data can be found in the `VIP` subpackage `pca`. For ADI-PCA, the reference PSF is constructed for each image by projecting the image onto a lower-dimensional orthogonal basis extracted from the data itself via PCA. Subtracting from each frame its reference PSF produces residual frames where the signal of the moving planets is enhanced. The most basic implementation of ADI-PCA uses the whole images by building a matrix $M \in \mathbb{R}^{n \times p}$, where n is the number of frames and p the number of pixels in a frame.

The basic structure of the full-frame ADI-PCA algorithm is the following:

1. the datacube is loaded in memory and M is built by storing on each row a vectorized version of each frame;
2. optionally M is mean-centered or standardized (mean-centering plus scaling to unit variance);

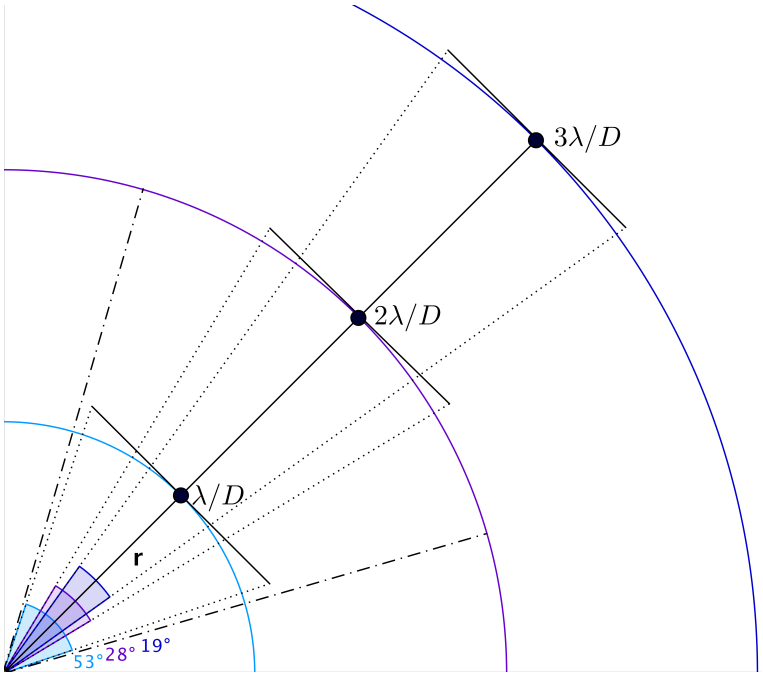


Figure 2.2: Illustration of the ADI rotation thresholds at different separations in λ/D. The dot-dashed lines show the rejection zone at 2λ/D with $\delta = 1$ that ensures a rotation by at least 1×FWHM (λ/D) of the PSF.

3. $k \leqslant \min(n, p)$ principal components (PCs) are chosen to form the new basis B;
4. the low-rank approximation of M is obtained as MB^TB, which models the reference PSF for each frame;
5. this low-rank approximation is subtracted from M and the result is reshaped into a sequence of frames;
6. all residual frames are rotated to a common north and are median combined in a final image.

The PCs can be obtained by computing the eigen decomposition (ED) and choosing the eigenvectors corresponding to the k largest eigenvalues of the covariance matrix M^TM, or equivalently by computing the SVD of M and extracting the k dominant right singular vectors. Fig. 2.3 shows intermediate resulting images corresponding to some of the steps of the frull-frame ADI-PCA algorithm.

Instead of computing the ED of M^TM (which is a large square matrix $p \times p$ that must fit in working memory) we can perform the ED of MM^T for a cheaper PCA computation. In a similar way, taking the SVD of M^T is faster and yields the same result as computing the SVD of M. Both speed tricks are implemented in `VIP`. `Python`, as well as other modern programming environments such as `Mathematica`,

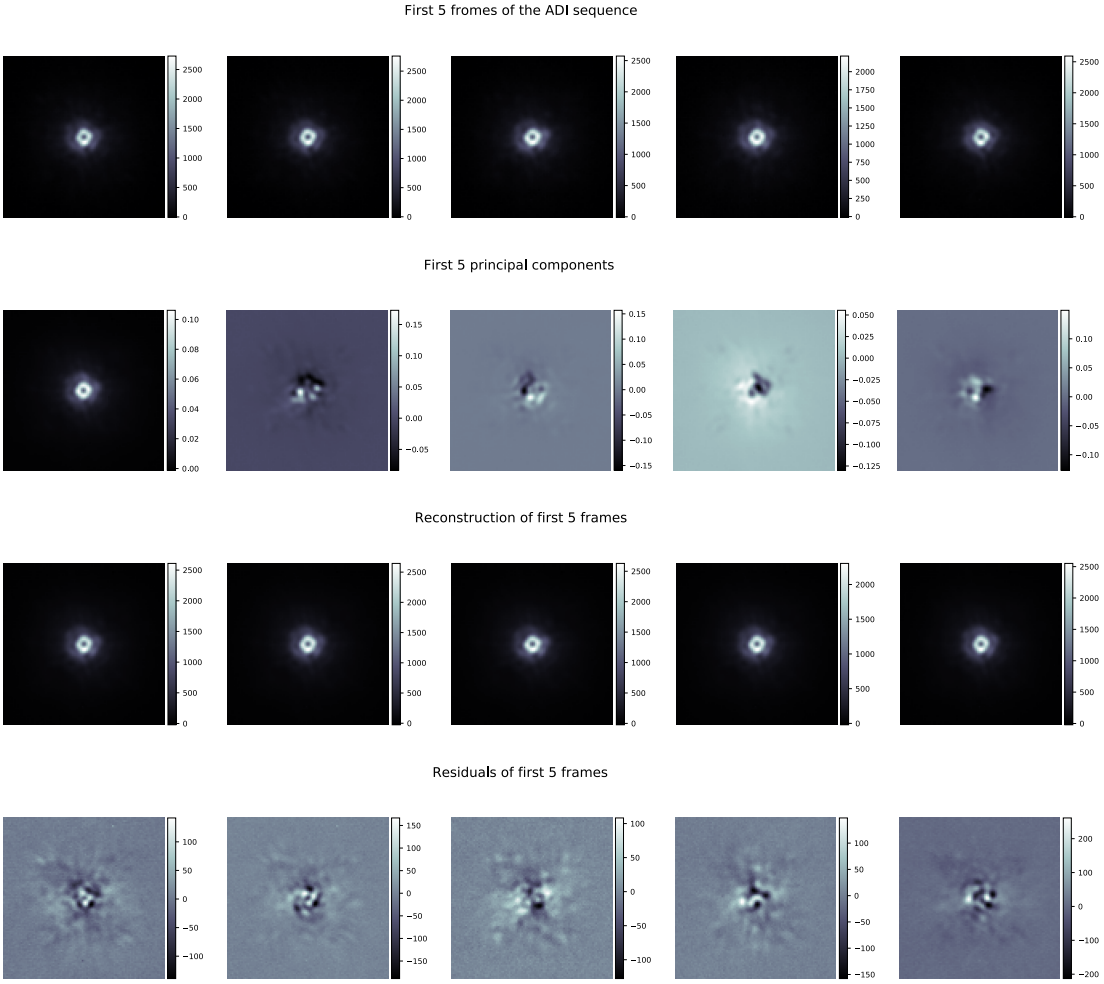


Figure 2.3: Full-frame ADI-PCA post-processing of a VLT/NACO (in AGPM mode) including intermediary frames. Top row shows the first five input frames. Second row, from top to bottom, presents the first five singular vectors (when reshaped into image space) of the matrix formed by vectorizing the images in the ADI sequence. Third row shows the frames reconstructions (low-rank approximations), when projected onto the first five principal components. These are the reference PSFs corresponding to each frame in the top row. Bottom row shows the residuals after subtracting the reference PSFs from the input frames, shown in the top row.

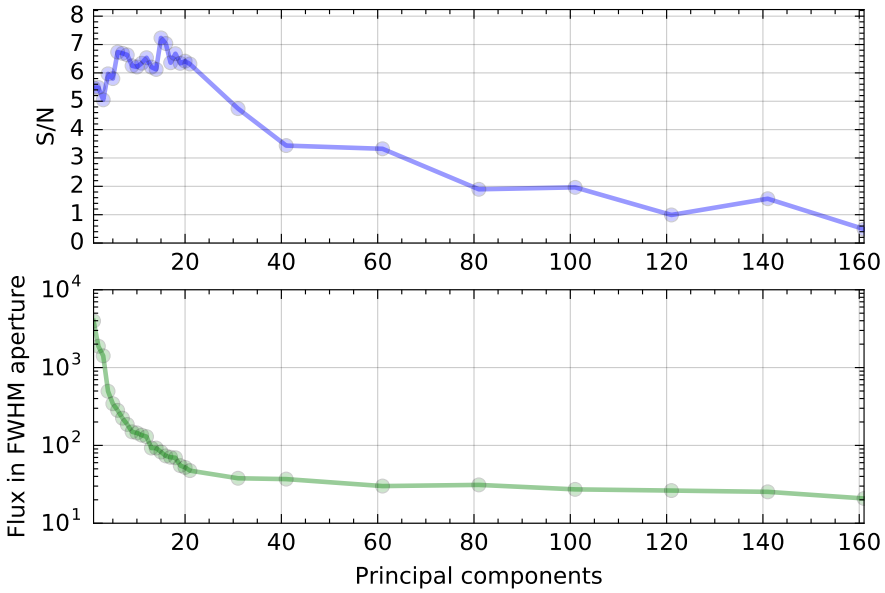


Figure 2.4: Top: grid optimization of the number of PCs for full-frame ADI-PCA at the location of a known planet. In this example, the mean S/N in a FWHM aperture was maximized with 16 PCs. Bottom: flux of the planet in a FWHM aperture in the final, post-processed residual image.

`R`, `Julia` and `Matlab`, relies on `LAPACK` (Linear Algebra PACKage)[4], which contains the state-of-the-art implementations of numerical dense linear algebra algorithms. We use the `Intel MKL` libraries, which provide multi-core optimized high performance `LAPACK` functionality consistent with the standard. For the SVD, `LAPACK` implements a "divide-and-conquer" algorithm that splits the task of a big matrix SVD decomposition into some smaller tasks, achieving good performance and high accuracy when working with big matrices (at the expense of a large memory workspace).

2.5.2.1 *Optimizing* k *for ADI-PCA*

The most critical parameter in every PCA-based algorithms is the number of principal components (PCs) k. `VIP` implements an algorithm to find the k that maximizes the S/N metric, described in section 2.3, for a given location in the image, by running a grid search varying the value of k and measuring the S/N for the given coordinates. This algorithm can also define an adaptative grid refinement to avoid computing the S/N in regions of the parameter space far from the maximum. This algorithm does not deal with the reliability of the candidate point-source located at the coordinates of interest. The computational cost remains close to that of a single full-frame ADI-PCA

4 http://www.netlib.org/lapack/

run thanks to the fact that we compute the PCA basis once with the maximum k we want to explore. Having this basis, we truncate it for each k PCs in the grid and proceed to project, subtract and produce the final frames where the S/N is computed. An example of such optimization procedure is shown in Fig. 2.4 In this case we maximized the mean S/N in a FWHM aperture. We can observe that the optimal S/N reaches a plateau near the maximum. For true planets, the S/N decreases slowly when increasing the number of PCs, as shown in the top panel of Fig. 2.4, in contrast with a more abrupt S/N decay for noise artifacts or bright speckles (which have significant S/N only for a few PCs and quickly fade away). The maximum S/N does not correspond to the maximum algorithm throughput, and in the illustrated case occurs for a throughput of about 0.1.

2.5.2.2 *Optimizing the library for ADI-PCA (annular ADI-PCA)*

Full-frame ADI-PCA suffers from companion self-subtraction when the signal of interest, especially that of a close-in companion, gets absorbed by the PCA-based low-rank approximation that models the reference PSF (Gomez Gonzalez et al., 2016a). A natural improvement of this algorithm for minimizing the signal loss is the inclusion of a parallactic angle threshold for discarding rows from M when learning the reference PSF. This frame selection for full-frame ADI-PCA is optional and can be computed for only one separation from the star. The idea is to leave in the reference library those frames where the planet has rotated by at least an angle ω, as described in Section 2.5.1. The computational cost increases when performing the selection of library frames (for each frame according to its index in the ADI sequence) since n singular value decompositions (SVD) need to be computed for learning the PCs of matrices with less rows than M. Following the same motivation of refining the PCA library, `VIP` implements an annular ADI-PCA algorithm, which splits the frames in annular regions (optionally in quadrants of annuli) and computes the reference PSF for each patch taking into account a parallactic angle rejection threshold for each annulus. This ADI-PCA algorithm processes $n \times n_{\text{annuli}}$ (or $4n \times n_{\text{annuli}}$ in case quadrants are used) smaller matrices.

The annular ADI-PCA comprises the following steps:

1. the datacube is loaded in memory, the annuli are constructed and a parallactic angle threshold is computed for each one of them;
2. for each annulus a matrix $M_{\text{annulus}} \in \mathbb{R}^{n \times p_{\text{annulus}}}$ is built;
3. optionally M_{annulus} is mean centered or standardized;
4. for each frame and according to the rotation threshold, a new M_{opt} matrix is formed by removing adjacent rows;

5. from M_{opt} the $k \leqslant \min(n_{opt}, p_{annulus})$ principal components are chosen to form the new basis optimized for this annulus and this frame;
6. the low-rank approximation of the annulus patch is computed and subtracted;
7. the residuals of this patch are stored in a datacube of residuals which is completed when all the annuli and frames are processed;
8. the residual frames are rotated to a common north and median combined in a final image.

This algorithm has been implemented with multiprocessing capabilities allowing to distribute the computations on each zone separately. According to our experience, using more than four cores for the SVD computation (through `LAPACK/MKL`) of small matrices, like the ones we produce in the annular ADI-PCA, does not lead to increased performance due to the overhead in the multi-threading parallelism. When used on a machine with a large number of cores, this algorithm can be set to process each zone in parallel, coupling both parallelisms for higher speed performance. Computing the PCA-based low-rank approximation for smaller patches accounts for different pixel statistics at different parts of the frames. This algorithm can also define automatically the parameter k for each patch by minimizing the standard deviation in the residuals, similar to the objective of the original LOCI algorithm (Lafrenière et al., 2007), at the expense of an increased computation time.

2.5.2.3 *Full-frame ADI-PCA for big ADI datasets*

Also, `VIP` implements variations of the full-frame ADI-PCA tailored to reduce the computation time and memory consumption when processing big datacubes (tens of GB in memory) without applying temporal frame sub-sampling. The size of an ADI dataset may vary from case to case and depends on the observing strategy and the preprocessing steps taken. Typically, a datacube contains several tens to several thousands of frames, each one of typically 1000×1000 pixels for modern detectors used in high-contrast imaging. In selected instruments (VLT/NACO, LBTI/LMIRCam) that are able to record high-frame rate cubes, a typical one-hour ADI sequence can contain up to ~20000 frames. After cropping down the frames to 400×400 pixels, we get a datacube in single float values occupying more than 10 GB of disk space. Loading this dataset at once in memory, for building M, would not be possible on a typical personal computer. Even if we manage to load the file, the PCA algorithm itself requires more RAM memory for SVD/ED calculations, which will eventually cause slowdowns (or system crashes) due to heavy disk swapping.

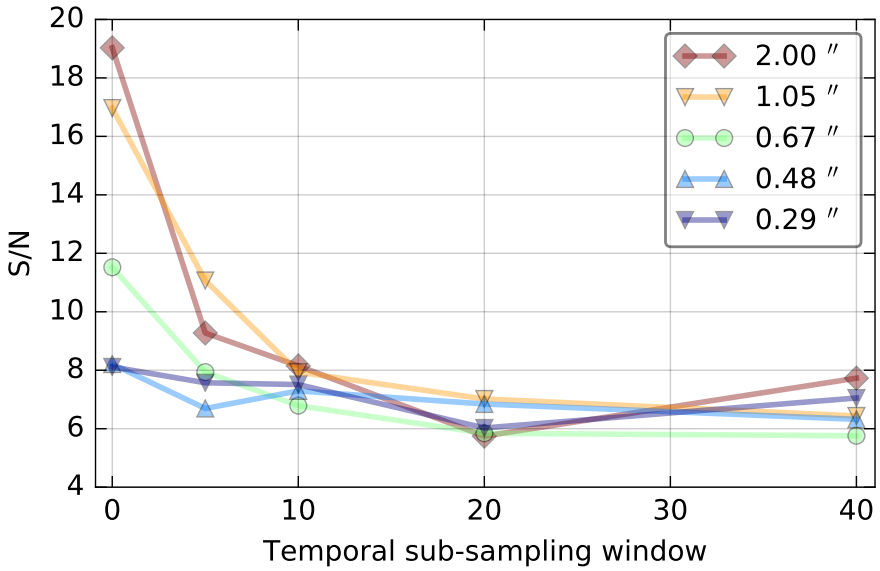


Figure 2.5: Fake companions S/N for different angular separations as a function of the temporal sub-sampling applied to the ADI sequence. The horizontal axis shows the amount of frames that were mean combined, with zero meaning that the whole ADI sequence (10k frames) is used. Full-frame ADI-PCA is applied on each datacube, with 21 PCs.

The most common approach for dealing with big datasets of this kind is to temporally and/or spatially sub-sample the frames. Reducing the size of the dataset effectively reduces the computation time of full-frame ADI-PCA to a few seconds but at the cost of smearing out the signal (depending on the amount of rotation). Also, depending on the temporal window used for co-adding the frames and on the PSF decorrelation rate, we might end up combining sections of the sequence where the PSF has a very different structure. It has been stated that there is an optimal window for temporal sub-sampling, which results in increased S/N (Meshkat et al., 2014a). After running simulations with fake companion injections at different angular separations and measuring the obtained mean S/N in a λ/D aperture, we came to the conclusion that using the whole sequence of frames (data without temporal sub-sampling) is the best choice and delivers the best results in terms of S/N. For this test, we used a datacube of ~10000 frames, each with 0.5 second of integration time. In Fig. 2.5 we show the S/N of the recovered companions in datacubes sub-sampled using different windows, for an arbitrarily fixed number of 21 PCs (even though 21 PCs do not necessarily represent the same explained variance for datacubes with different numbers of frames). There is an agreement with these results and those in Meshkat et al. (2014a) for sub-sampling windows larger than 20, but unfortunately Meshkat et al. (2014a) did not consider smaller sub-sampling win-

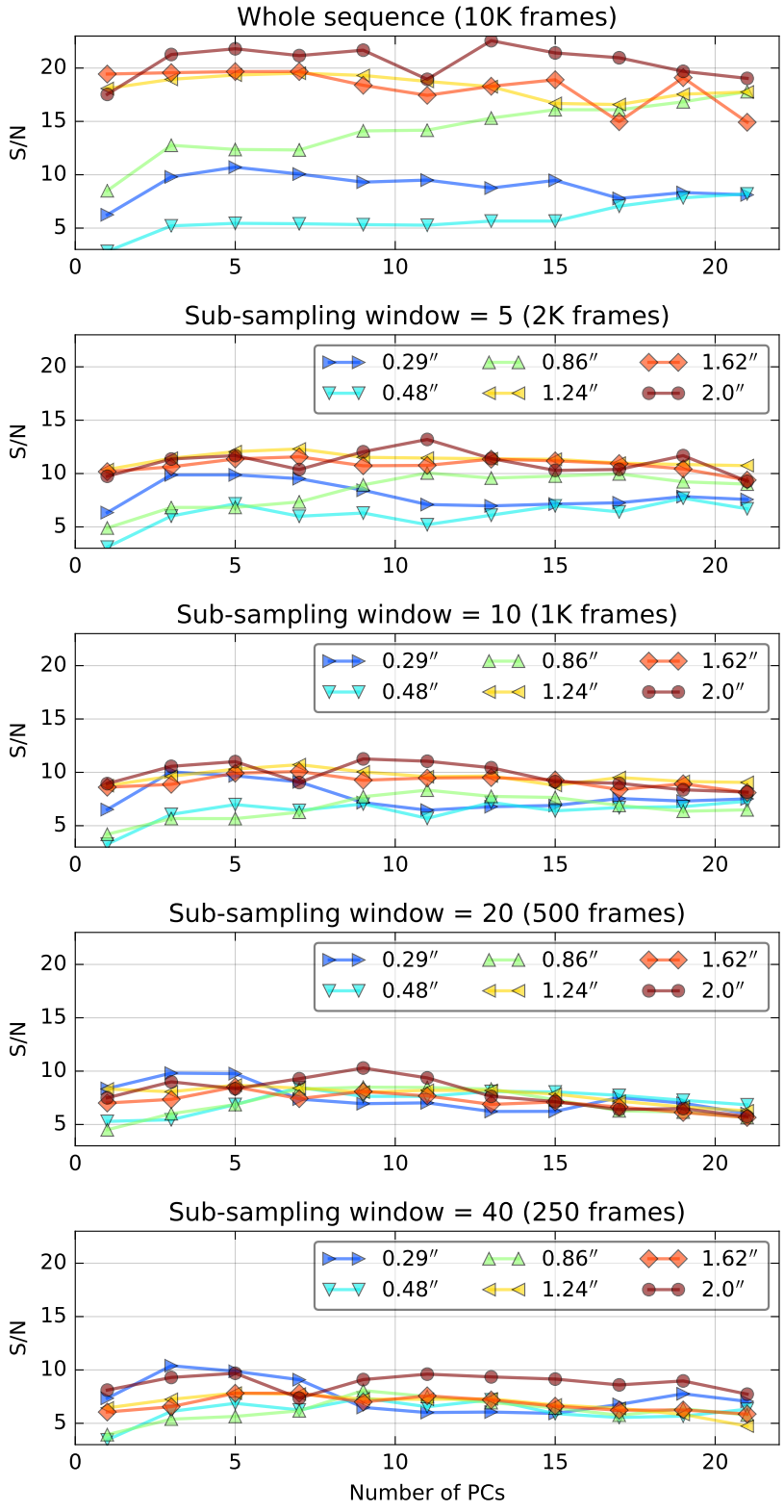


Figure 2.6: Retrieved S/N on fake companions injected at different angular separations and with a constant flux. The top panel shows the results of varying the number of principal components of the full-frame ADI-PCA algorithm when processing the full resolution ADI sequence. The rest of the panels show the same S/N curves obtained on sub-sampled versions of the sequence using different windows.

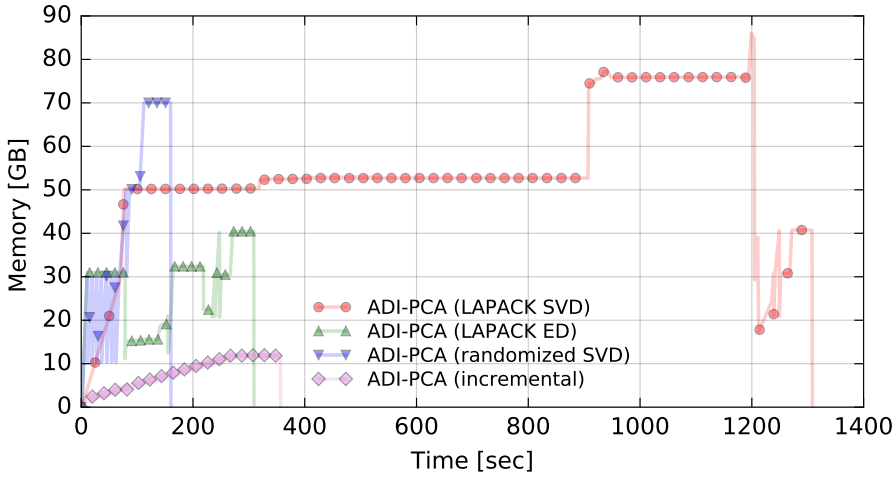


Figure 2.7: Memory usage as a function of the processing time for different variations of the full-frame ADI-PCA algorithm on a large datacube (20 PCs were requested). This is valid for datacubes occupying several GB on disk (in this particular case a 10 GB FITS file was used). It is worth noting that for short or sub-sampled ADI sequences the full-frame ADI-PCA through `LAPACK` SVD is very efficient and the difference in processing time may become negligible.

dows, nor the full ADI sequence. More information can be found in Fig. 2.6, which shows the dependence of the fake companions S/N on the number of PCs for different angular separations. Our simulations show very significant gain in S/N when temporal sub-sampling of frames is avoided, especially at large separations where smearing effects are the largest.

`VIP` offers two additional options when it comes to compute the full-frame ADI-PCA through SVD, tailored to reduce the computation time and memory consumption when data sub-sampling needs to be avoided (see Fig. 2.7). These variations rely on the machine learning library `Scikit-learn`. The first is ADI-PCA through randomized SVD (Halko et al., 2011), which approximates the SVD of M by using random projections to obtain k linearly independent vectors from the range of M, then uses these vectors to find an orthonormal basis for it and computes the SVD of M projected to this basis. The gain resides in computing the deterministic SVD on a matrix smaller than M but with strong bounds on the quality of the approximation (Halko et al., 2011).

The second variation of the ADI-PCA uses the incremental PCA algorithm proposed by Ross et al. (2008), as an extension of the Sequential Karhunen-Loeve Transform (Levy and Lindenbaum, 2000), which operates in on-line fashion instead of processing the whole data at once. For the ADI-PCA algorithm through incremental PCA, the FITS file is opened in *memmaping* mode, which allows accessing

small segments without reading the entire file into memory, thus reducing the memory consumption of the procedure. Incremental PCA works by loading the frames in batches of size b and initializes the SVD internal representation of the required lower dimensional subspace by computing the SVD of the first batch. Then it sequentially updates n/b times the PCs with each new batch until all the data is processed. Once the final PCs are obtained, the same n/b batches are loaded from disk once again and the reconstruction of each batch of frames is obtained and subtracted for obtaining the residuals, which are then de-rotated and median collapsed. A final frame is obtained as the median of the n/b median collapsed frames. A similar approach to incremental PCA, focusing on covariance update, has been proposed by Savransky (2015). In Fig. 2.7 are compared the memory consumption and total CPU time for all the variants of full-frame ADI-PCA previously discussed. These tests were performed using a 10GB (on disk) sequence and on a workstation with 28 cores and 128 GB of RAM. The results show how incremental PCA is the lightest on memory usage while randomized PCA is the fastest method. With incremental PCA, an appropriate batch size can be used for fitting in memory datacubes that otherwise would not, without sacrificing the accuracy of the results.

2.5.3 Non-negative matrix factorization for ADI

As previously discussed, the PCA-based low-rank approximation of an ADI datacube can be used to model the reference PSF for each one of its frames. In the fields of machine learning and computer vision, several approaches other than PCA have been proposed to model the low-rank approximation of a matrix (Kumar and Shneider, 2016; Udell et al., 2016). In particular, non-negative matrix factorization (NMF) aims to find a k-dimensional approximation in terms of non-negative factors W and H (Lee and Seung, 1999). NMF is distinguished from the other methods by its use of non-negativity constraints on the input matrix and on the factors obtained. For astronomical images, the positivity is guaranteed since the detector pixels store the electronic charges produced by the arriving photons. Nevertheless, the sky subtraction operation can lead to negative pixels in the background and in this case a solution is to shift all the values on the image by the absolute value of the minimum pixel, turning negative values into zeros.

NMF finds a decomposition of M into two factors of non-negative values, by optimizing for the Frobenius norm:

$$\underset{W,H}{\operatorname{argmin}} \frac{1}{2}\|M - WH\|_F^2 = \frac{1}{2}\sum_{i,j}(M_{ij} - WH_{ij})^2, \tag{2.3}$$

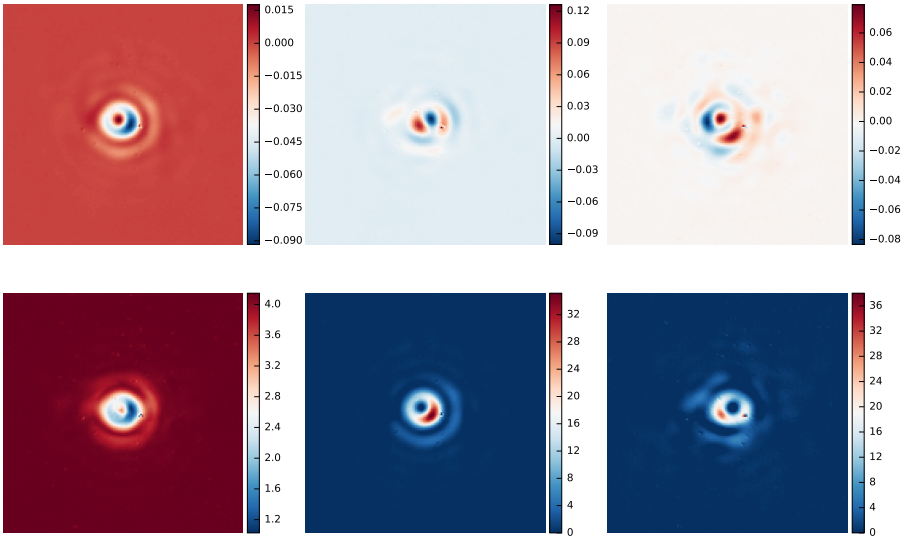


Figure 2.8: First three principal (top row) and NMF components (bottom row) for an LBT/LMIRCam (on its AGPM coronagraphic mode) dataset. The components of NMF are strictly positive.

where $W \in \mathbb{R}^{n \times k}$, $H \in \mathbb{R}^{k \times p}$, and WH is a k-rank approximation of M. Such a matrix factorization can be used to model a low-rank matrix based on the fact that $\text{rank}(WH) \leqslant \min(\text{rank}(W), \text{rank}(H))$, where $\text{rank}(X)$ denotes the rank of a matrix X. Therefore if k is small, WH is low-rank. NMF is a non-convex problem, and has no unique minimum. Therefore it is a harder computational task than PCA, and no practical algorithm comes with a guarantee of optimality (Vavasis, 2009). It is worth noting that this Frobenius-norm formulation of NMF, as implemented in `scikit-learn`, provides final images very similar to the ones from full-frame ADI-PCA. The first NMF components along with the first PCs for a same dataset are shown in Fig. 2.8. The NMF components are strictly positive and the NMF-based low-rank approximation that models the reference PSF for each frame is computed as a linear combination of these components. This NMF-based algorithm makes a useful complement to PCA-based algorithms for testing the robustness of a detection.

2.5.4 LLSG for ADI

The Local Low-rank plus Sparse plus Gaussian-noise decomposition (LLSG, see Chapter 5) has been proposed as an approach to enhance residual speckle noise suppression and improve the detectability of point-like sources in the final image. LLSG builds on recent subspace projection techniques and robust subspace models proposed in the computer vision literature for the task of background subtraction from natural images, such as video sequences (Bouwmans and Za-

hzah, 2014). The subpackage `llsg` contains an implementation of the LLSG algorithm for ADI datacubes. Compared to the full-frame ADI-PCA algorithm, the LLSG decomposition reaches a higher S/N on point-like sources and has overall better performance in the receiver operating characteristic (ROC) space. The boost in detectability applies mostly to the small inner working angle region, where complex speckle noise prevents full-frame ADI-PCA from discerning true companions from noise. One important advantage of LLSG is that it can process an ADI sequence without increasing too much the computational cost compared to the fast full-frame ADI-PCA algorithm.

2.6 DETECTION OF COMPANIONS ON 2D RESIDUAL FLUX IMAGES

`VIP` allows the production of a final residual image thanks to reference PSF subtraction techniques. Reference PSF subtraction enables the further reduction of the dynamic range in the images by modeling and subtracting the contribution from the high-flux pixels belonging to the leaked starlight and from the quasi-static speckle pattern. This happens at the expense of the subtraction of some part of the companion signal (self-subtraction) that it fitted in the reference PSF. `VIP` implements automatic detection of point-like sources relying on blob detection techniques developed in the field of computer vision: the Laplacian of Gaussian (LOG) and its approximation, the difference of Gaussians. These algorithms deal with the detection of blobs or patches of pixels sharing the same properties (such as bright or dark regions on the images). For instance, the LOG computes the Laplacian of Gaussian images with successively increasing standard deviation (related to the spatial scale of the blobs) and stacks them up in a cube, where blobs appear as local maxima. The adopted approach in the HCI community relies on visual inspection aided by the computation of S/N maps.

2.7 FLUX AND POSITION ESTIMATION FOR ADI

Once potential companions are identified in the final residual image, we proceed to characterize them. The `VIP` library implements the negative fake companion technique (NEGFC, Marois et al., 2010a; Lagrange et al., 2010) for the determination of the position and flux of companions. This implementation is contained in the subpackage `negfc`. The NEGFC consists in injecting in the data cube a negative PSF template with the aim of canceling out the companion as well as possible in the final post-processed image based on a well-chosen merit function. The PSF template is obtained from off-axis observa-

tions of the star. Injecting this negative PSF template directly in the images, before they are processed, allows to take into account the biases in photometry and astrometry induced by the post-processing algorithms. The best cancellation of the companion PSF is achieved by minimizing, in an iterative process, a well-chosen figure of merit. The intensities I_j of p pixels are extracted within a circular aperture of a radius equal to a few resolution elements (by default four resolution elements or $4 \times FWHM$), centered on the first guess position of the companion. Assuming that the noise affecting the j-th pixel value is given by $\sigma_j = \sqrt{I_j}$ (pure photon noise), we define the merit function as follows:

$$\chi^2 \propto \sum_{j=1}^{p} |I_j| \, . \tag{2.4}$$

This NEGFC function of merit can be tweaked, by changing the default parameters of the `VIP`'s NEGFC procedure. Optionally, one can minimize the standard deviation of I_j, instead of the sum of absolute values, which according to our test is better in cases when the companion is located in a region heavily populated by speckles (close to the star), or use the pixels inside a circular aperture from each residual frame thus avoiding collapsing the datacube in a single final frame. Using the $n \times p$ pixels from the residual cube helps getting rid of any bias that the collapsing method, by default a median combination of the frames, may introduce.

The steps of the NEGFC technique can be described as follows. For the chosen position/flux combination, a negative fake companion is injected in each frame of the data cube, and annular-wise ADI-PCA processing is performed on a single annulus passing through the considered companion. The pixel intensities I_j are then extracted within a circular region centered on a first guess position defined at the start of the iterative process (which means that the position of the circular aperture is fixed and does not change during the process). The estimation of the position and flux (three parameters: radius r, theta θ and flux F) is obtained by performing three consecutive procedures.

2.7.1 First guess estimation

A first guess of the flux of a companion is obtained by injecting a NEGFC in the calibrated frames while fixing r and θ and evaluating the function of merit for a grid of possible fluxes. This initial position (r and θ) is determined by visual inspection of final post-processed frames or based on prior knowledge. Only the flux is optimized during this stage, while the companion position is fixed to our first guess.

2.7.2 Nelder-Mead optimization

Although the first guess estimation results in a rough determination of the position/flux, and would constitute a valid initial set of parameters to start an Monte Carlo Markov Chain (MCMC) based Bayesian inference process (as presented in the next paragraph), it may turn out to be very time consuming due to the large number of merit function evaluations required to reach convergence in the MCMC and properly sample the posterior distributions. Thus, we refine the first guess of the position/flux of the companions for the purpose of initializing the MCMC sampling close to the highly probable solution. To this aim, we use the first position/flux estimation as an initial guess for a Nelder-Mead simplex-based optimization (Nelder and Mead, 1965) implemented into the `SciPy` Python library[5]. The adopted merit function is the one defined in Eq. 2.4, and the position (r, θ) of the NEGFC is now allowed to vary during the fit in addition to its flux. As expected, this leads to a significant improvement of the position/flux determination.

Although the randomized ADI-PCA approach is very efficient, the random process induces random variations in the merit function that can be significant compared to the variations of the merit function between two steps, especially when approaching the minimum. This can prevent the optimization process from reaching the true minimum of the merit function, or even from converging. Therefore, we strongly encourage to use the deterministic `LAPACK` SVD approach in the simplex optimization, as well as for the next step. This choice is all the more important when the companion is located in a region dominated by residual speckle noise.

2.7.3 MCMC and Bayesian parameter estimation

Although the simplex minimization leads to a significant improvement of the position/flux determination, it does not provide error bars on the three estimated parameters. More importantly, the function of merit of the NEGFC technique is not strictly convex and finding a global minimum (the exact position and flux of the candidate planet) is not guaranteed (Morzinski et al., 2015). Our approach in `VIP` is to turn our minimization function of merit into a likelihood and to use Monte Carlo methods for sampling the posterior probability density functions (PDF) for r, θ and F. This can be achieved via Markov Chain Monte Carlo (MCMC) which aims to sample the posterior probability density function (PDF), i.e. the probability of the position/flux parameters given the data cube and the prior knowledge (see e.g. Hogg et al., 2010).

5 http://www.scipy.org

The `VIP` module dedicated to NEGFC technique embeds the `emcee` package (Foreman-Mackey et al., 2013), which implements an affine-invariant ensemble sampler for MCMC proposed by Goodman and Weare (2010). Such an ensemble is composed of *walkers*, which can be considered as Metropolis-Hastings chains. The main difference between walkers and Metropolis-Hastings chains lies in the fact that the proposal distribution for a given walker depends, at a given step, on the position of all other walkers in the ensemble. Conversely, the proposal distributions involved in the Metropolis-Hastings algorithm are independent. Besides being more efficient in terms of the number of calls to the cost function, one major advantage of `emcee` is that it relies on only two calibration parameters, in comparison with the $\sim N^2$ parameters required for a Metropolis-Hastings algorithm in an N-dimensional parameter space to properly sample the PDF and speed up the process (for more details, see Foreman-Mackey et al., 2013; Goodman and Weare, 2010, and references therein). In practice, one can never be sure that a chain has actually converged, but there exist several tests to evaluate whether the chain appears to be close to convergence (or more precisely, far from non-convergence). In `VIP` we use the *Gelman-Rubin* $\hat{R}$ statistical test (Gelman and Rubin, 1992; Ford, 2006; Gelman et al., 2014) which compares, for each parameter, the variance estimated from non-overlapping parts of the chain to the variance of their estimates of the mean. A large $\hat{R}$ value may arise from slow chain mixing or multimodality (Cowles and Carlin, 1996). Conversely, a $\hat{R}$ value close to 1 indicates that the Markov chain is close to convergence. Alternatively, `VIP` offers functionality for nested sampling (Feroz and Hobson, 2008) of the NEGFC parameters, through the `nestle` `Python` library [6]. See Allison and Dunkley (2014) for a comparison of MCMC and nested sampling for Bayesian parameter estimation.

From the sampled parameters PDFs, we can infer error bars and uncertainties in our estimations, at the cost of longer computation time. This NEGFC implementation in `VIP` currently works with ADI-PCA-based post-processing algorithms. We also include a procedure for estimating the influence of speckles in the astrometric and photometric measurements, based on the injection of fake companions at various positions in the field of view. More details on the NEGFC technique, the definition of the confidence intervals, and details about the speckle noise estimation can be found in Chapter 3, where this procedure is used to derive robust astrometry for the HR8799 planets based on VLT/SPHERE data (Wertz et al., 2016).

6 https://github.com/kbarbary/nestle

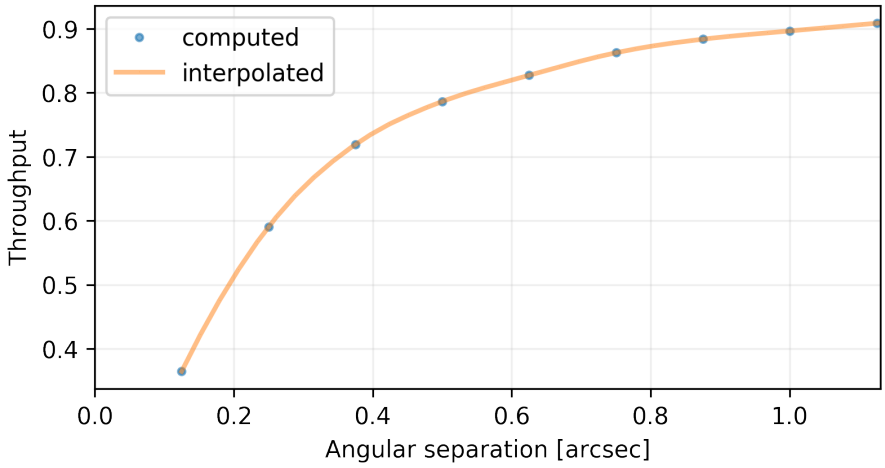
2.8 SENSITIVITY LIMITS

Sensitivity limits (in terms of planet/star contrast), often referred to as contrast curves, are commonly used in the literature for estimating the performance of high-contrast direct imaging instruments. They show the detectable contrast for point-like sources as a function of the separation from the star. `VIP` follows the current practice in building sensitivity curves from Mawet et al. (2014) and thereby applies a student-t correction to account for the effect of the small sample statistics. This correction imposes a penalty at small separations and therefore the direct comparison with contrast curves from previous works may seem more pessimistic close to the parent star. The function, contained in the subpackage `phot`, requires an ADI dataset, a corresponding instrumental PSF, and the stellar aperture photometry F_*. We suggest removing any real, high-significance companions from the data cube (for example using the NEGFC approach) before computing a contrast curve. The first step is to measure the noise as a function of the angular separation σ_r, on a final post-processed frame from the ADI datacube, by computing the standard deviation of the fluxes integrated in FWHM apertures. Then we inject fake companions to estimate empirically the throughput T_r of the chosen algorithm (i.e. the signal attenuation) at each angular separation as $T_r = F_r/F_{in}$, where F_r is the recovered flux of a fake companion after the post-processing and F_{in} is the initially injected flux of the fake companion. We define the contrast C_r as:

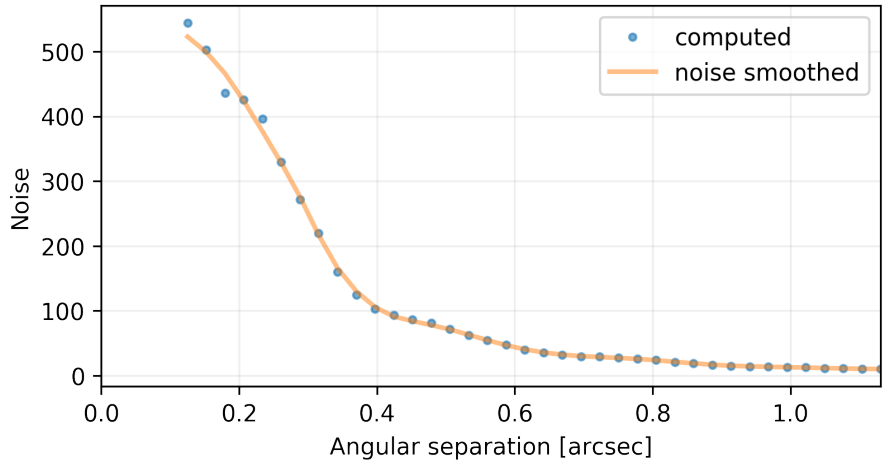
$$C_r = \frac{k \cdot \sigma_r}{T_r \cdot F_*}, \tag{2.5}$$

where k is a factor, five in case we want the five sigma contrast curve, corrected for the small sample statistics effect. An example of the output of the contrast curve calculation produce is shown in Fig. 2.9. The transmission of the instrument, if known, can be optionally included in the contrast calculation. Ruane et al. (2017) improve this methodology for computing detection limits by defining an acceptable number of false positives per annulus and computing the 95% completeness or true positive rate (instead of the typically reported 50% completeness level). This is a planned addition to `VIP`.

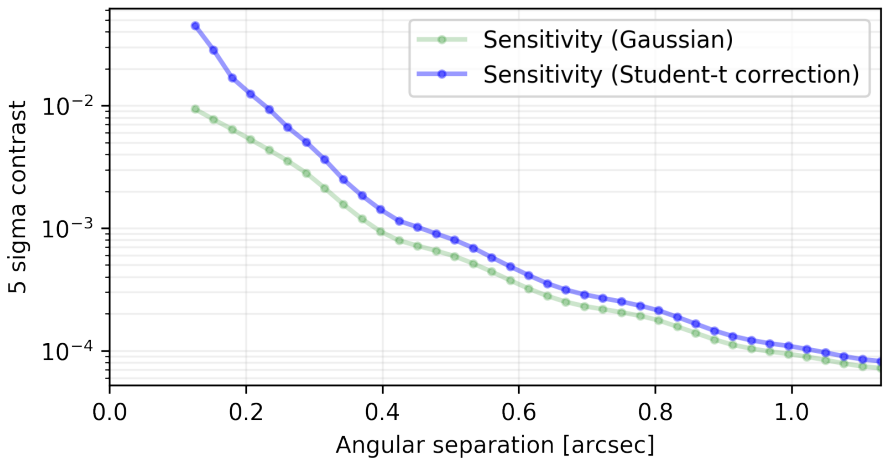
We note that contrast curves depend on the post-processing algorithm used and its tuning, as it will be shown in the panel (a) of Fig. 3.4. In signal detection theory, the performance of a detection algorithm is quantified using ROC analysis, and several meaningful figures of merit can be derived from it (Barrett et al., 2006; Lawson et al., 2012; Gomez Gonzalez et al., 2016a). These figures of merit would be better suited than sensitivity curves for post-processing algorithms performance comparison.



(a)



(b)



(c)

Figure 2.9: Full output of the contrast curve procedure computed on a A VLT/NACO dataset. Panel (a) shows the algorithm throughput (in this case ADI-PCA). Panel (b) shows the noise (computed in an annulus-wise fashion) as a function of the separation. Panel (c) shows a typical 5-sigma contrast curve with and without the small sample correction.

2.9 CONCLUSIONS

In this chapter we presented the VIP package for data processing of astronomical high-contrast imaging data. It has been successfully tested on data coming from a variety of instruments, such as Keck/NIRC2, VLT/NACO, VLT/VISIR, VLT/SPHERE and LBT/LMIRCam, thanks to our effort of developing VIP as an instrument agnostic-library. VIP implements functionalities for processing high-contrast imaging data at every stage, from pre-processing procedures to contrast curve calculations. Concerning the post-processing capabilities of VIP for the case of ADI data, it includes several types of low-rank matrix approximations for reference PSF subtraction, such as the LLSG decomposition, and PCA and NMF-based algorithms. We present, as one of several PCA enhancements, an incremental ADI-PCA algorithm capable of processing big, larger-than-memory ADI datasets. VIP includes PCA-based post-processing algorithms for observing techniques, such as RDI, multiple-channel SDI, and IFS data with wavelength and rotational diversities (mSDI+ADI). This is considered on-going work. Further development of VIP is planned, in order to improve its robustness and efficiency (for supporting big datasets in every procedure and multi-processing), and add more state-of-the-art algorithms for high-contrast imaging data processing. We propose VIP not as an ultimate solution to all high-contrast image processing needs, but as an open science exercise hoping that it will attract more users and in turn be developed by the high-contrast imaging community as a whole.

3 VIP APPLIED TO ON-SKY DATA

Contents

ABSTRACT

In this chapter I describe the application of VIP to HR8799 datasets from two different high-contrast imaging instruments. In Section 3.2 I showcase the ADI capabilities of the VIP library using a deep sequence on HR8799 taken with the LBTI/LMIRCam and its recently commissioned L-band vortex coronagraph. Using VIP I investigated the presence of additional companions around HR8799 and did not find any significant additional point source beyond the four known planets. I describe each one of the image processing steps up to the generation of sensitivity curves. In Section 3.3 I present work carried out with Olivier Wertz and Olivier Absil to characterize the HR8799 companions, focusing in the estimation of their robust astrometry. *This chapter is based on previous work published in Gomez Gonzalez et al. (2017) and Wertz et al. (2016).*

3.1 THE HR8799 SYSTEM

HR8799 is a young A5V star, located at 39 pc, hosting a multiple-planet system and a debris disk (Marois et al., 2008b). Since its discovery by Marois et al. (2008c), the HR8799 planetary system has

been and still remains one of the most intriguing among the thousands of known planetary systems. Composed of at least four giant planets in a range of angular separations of about 0″.4 to 1″.7 (Marois et al., 2010c), and of two dusty debris belts (Su et al., 2009; Hughes et al., 2011; Matthews et al., 2014; Booth et al., 2016), it has been the focus of many different studies, including spectroscopic studies to gain insights about the physics and chemistry of the four planets (Oppenheimer et al., 2013), and dynamical stability analyses to constrain their global orbital motion and estimate their masses (see e.g. Goździewski and Migaszewski, 2009; Reidemeister et al., 2009; Fabrycky and Murray-Clay, 2010; Soummer et al., 2011; Currie et al., 2012b, 2014; Goździewski and Migaszewski, 2014a; Maire et al., 2015a). Oppenheimer et al. (2013); Currie et al. (2014) presented a study of the spectra of the four companions, which concluded that they are substantially different from each other and present much redder colors than objects with similar spectral features (such as methane) currently known. These spectra do not correspond to those of any known objects, although similarities with L and T-dwarfs are present, as well as some characteristics similar to planets such as Saturn.

This dynamical approach allows the orbits of the four planet to be simultaneously constrained, but requires strong assumptions, such as coplanar (but eccentric) or circular (but not necessarily coplanar) orbits. The individual analysis of each planet offers an alternative method to constrain the orbital architecture. To this aim, nonlinear least-squares fits of Keplerian elements (semi-major axis a, eccentricity e, inclination i, longitude of ascending node Ω, argument of the periastron ω, and time of periastron passage t_p) have been performed (see e.g. Lafrenière et al., 2009; Bergfors et al., 2011; Esposito et al., 2013; Zurlo et al., 2016). Recently, Pueyo et al. (2015) proposed an in-depth analysis of the HR8799bcde orbital motion. The authors carried out a Bayesian analysis based on MCMC techniques adopting both a Metropolis Hastings algorithm (MacKay, 2003; Ford, 2005, 2006) and an affine-invariant ensemble sampler (Foreman-Mackey et al., 2013). This approach echoes the works published in Chauvin et al. (2012) for β Pictoris b, in Kalas et al. (2013) for Fomalhault b and more recently in Beust et al. (2016) for Fomalhault b and PZ Telescopii B. Among other things, Pueyo et al. (2015) discussed the coplanarity of the system, the orbital eccentricities of the planets, the possibility for mean motion resonances, and the role of HR8799d in possible dynamical interactions in the youth of this system. They also estimated the dynamical masses of HR8799bcde by computing the fraction of allowable orbits that pass the so-called close-encounter test. As pointed out in Pueyo et al. (2015), unaccounted biases and/or systematically underestimated error bars on the planets astrometry affect the MCMC results (see e.g. Givens and Hoeting, 2012) and may lead to a biased estimation of the confidence intervals for the orbital pa-

rameters. Studying the astrometric history of HR8799 reveals indeed that the errors affecting some positions are most probably underestimated, as one can readily identify pairs or sets of positions that are not consistent with each other within their error bars, or cannot be modeled with a unique orbit. This was one of the incentives of the study presented by Konopacky et al. (2016), who very recently re-reduced all the Keck/NIRC2 observations of HR8799 to come up with a self-consistent data set free from variable instrument-related biases. This consistent data set was then used to derive updated probability distributions for the elements of the planetary orbits based on Monte Carlo simulations.

An intriguing open question is the existence of an additional companion in this system (Currie et al., 2014; Maire et al., 2015b). Dynamical simulations by Goździewski and Migaszewski (2014b) suggested a fifth companion could be located at smaller separations. The warm belt truncation at ~6-10 AU is also consistent with the scenario of a planet-mass body interior to the four known planets. Currie et al. (2014) ruled out a more massive (12-13 M_J) companion at the inner edge of the warm dust belt (~6 AU), failing to find the companion predicted by Goździewski and Migaszewski (2014b). More recently, Maire et al. (2015b) obtain detection limits allowing to exclude a fifth planet slightly brighter/more massive than HR8799 b at the location of the 2:1 resonance with HR8799 e (~9.5 AU) and about twice as bright as HR8799 cde at the location of the 3:1 resonance with HR8799 e (~7.5 AU).

3.2 POST-PROCESSING OF HR8799 LMIRCAM DATA

3.2.1 Observations and image calibration

The HR8799 dataset used in this section was obtained at the Large Binocular Telescope (LBT, Hill et al., 2014) on 2013 October 17, during the first commissioning run of the L'-band annular groove phase mask (AGPM) coronagraph on the LBT Interferometer (LBTI, Hinz et al., 2014; Defrère et al., 2014). A deep ADI sequence on HR8799 was obtained in pupil-stabilized mode on the LBTI's L and M Infrared Camera (LMIRCam), equipped with its AGPM coronagraph, using only the left-side aperture. The observing sequence lasted for approximately 3.5 hours, providing a field rotation of 120° and resulting in ~17000 frames on target. The seeing was fair during the first 30 minutes (1″.2-1″.4) and good for the remaining of the observations (0″.9-1″.0). The adaptive optics loop was locked with 200 modes first and with 400 modes after 30 minutes. The off-axis PSF was regularly calibrated during the observations by placing the star away from the AGPM center (Defrère et al., 2014). The raw sequence of frames

was flat fielded and background subtracted with custom routines. The sky background subtraction was performed using the median-combination of close in time sky frames. Using a master bad pixel mask generated with sky frames taken at the end of the night, the bad pixels were subsequently fixed in each frame using the median of adjacent pixels (Defrère et al., 2014).

3.2.2 Data processing with VIP

The calibrated datacube was loaded in memory with VIP and re-centered using as a point of reference a ghost PSF present in each frame, product of a secondary reflection due to the LBTI trichroic (Skemer et al., 2014; Defrère et al., 2014). The offset between the secondary reflection and the central source was measured on the non-saturated PSF observations via a 2d Gaussian fit. VIP includes a function for aligning frames by fitting a frame region with a 2d Gaussian profile (using Astropy functionality). We used this function for fitting the secondary reflection on each frame and placing the star at the center of odd-sized square images taking into account the previously calculated offset between the reflection and the main beam.

For the bad frames rejection step, we used the VIP's algorithm based on the linear correlation of each frame with respect to a reference from the same sequence (30×30 centered sub-frames are used). The reference frame was chosen by visual inspection and in agreement with the observing log of the adaptive optics system. Ten percent of the frames were finally discarded resulting in a datacube with size 15254×391×391 (after cropping the frames to the region of interest), occupying 9.7 GB of disk space (in single float precision).

The workflow for loading data in memory and pre-processing it with VIP is as follows:

Listing 3.1: VIP: Importing the package and opening FITS files. Pre-processing steps for an ADI sequence.

```
import vip

# loading the calibrated datacube and parallactic angles
cube = vip.fits.open_fits('path_cube')
pa = vip.fits.open_fits('path_pa')

# aligning the frames
from vip.calib import cube_recenter_gauss2d_fit as recenter
cube_rec = recenter(cube, xy=cent_subim_fit, fwhm=fwhm_lbt,
                    subi_size=4, offset=offset_tuple)

# identifying bad frames
from vip.calib import cube_detect_badfr_correlation as badfrcorr
gind, bind = badfrcorr(cube_rec, frame_ref=9628, dist='pearson',
                       percentile=10, plot=False)
```

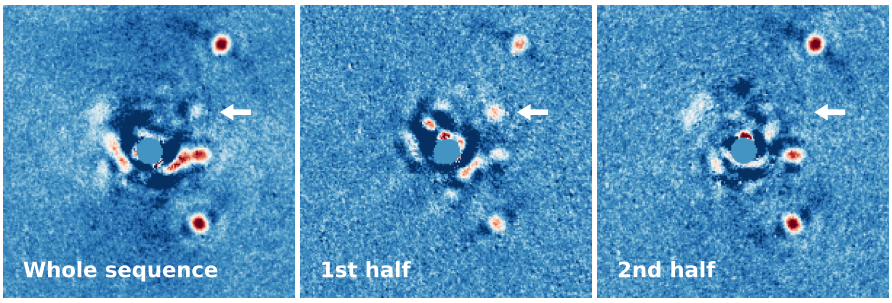


Figure 3.1: Ghost planet (shown with an arrow) due to the secondary reflection of LBTI. The left panel shows the full-frame ADI-PCA result using six PCs on the whole ADI sequence. The middle and right panels show the same processing but using only the first and second halves of the sequence.

```
# discarding bad frames
pa_gf = pa[gind]; cube_gf = cube_rec[gind]

# cropping the re-centered frames
from vip.calib import cube_crop_frames
cube_gf_cr = cube_crop_frames(cube_gf, size=391)
```

A first exploration of the full-resolution datacube with full-frame ADI-PCA showed a feature that resembled an instrumental PSF near the location of planet HR8799e, when only a few principal components were used (see left panel in Fig. 3.1). We concluded, after processing the data in two halves, that this blob was a residual artifact of the secondary reflection of LBTI, which left a PSF-like footprint due to the slow rotation in the first third of the sequence. The ghost companion appeared very bright when using the first half of the sequence and was totally absent using the second one, as shown in Fig. 3.1. Moreover, it was located at the same separation as the secondary reflection, whose offset was previously measured. For this reason, and because the adaptive optics system was locked on 200 modes during the first 5000 frames, while it locked on 400 modes for the rest of the sequence, we discarded this first batch of frames from the sequence and kept the frames with the highest quality. The rotation range of the final sequence is 100°, and the total on-source time amounts to 2.8 hours.

We processed this datacube with several ADI algorithms, and tuned their parameters for obtaining final frames of high quality, where we investigated the presence of a potential fifth companion. Other than the four known planets around HR8799, we did not find any significant detection, worth of further investigation. With the sole purpose of saving time while showcasing the `VIP` functionalities, we

then decided to sub-sample temporally our ADI sequence by mean combining each 20 frames, and thereby obtained a datacube of 499 frames. We refrained from binning the pixels and worked with the over-sampled LMIRCam images featuring a FWHM of nine pixels. The code below illustrates how these steps where done with `VIP`.

Listing 3.2: VIP: Temporal subsampling, ADI median subtraction and full-frame PCA.

```
# temporal sub-sampling of frames, average of every 20 frames
# cube_ss is a 3d numpy array with shape [499, 391, 391] and
# pa_ss a vector [499] with the corresponding paralactic angles
from vip.calib import cube_subsample
cube_ss,pa_ss = cube_subsample(cube_gf[5000:], n=20, mode='mean',
                               parallactic=pa_gf[5000:])

# ADI median subtraction using 2XFWHM annuli, 4 closest frames
# and PA threshold of 1 FWHM
fr_adi = vip.madi.adi(cube_ss, pa_ss, fwhm=fwhm_lbt,
                      mode='annular', asize=2, delta_rot=1,
                      nframes=4)

# post-processing using full-frame ADI-PCA
fr_pca = vip.pca.pca(cube_ss, pa_ss, ncomp=10)
# fr_adi and fr_pca are 2d numpy arrays with shape [391, 391]
```

Figures 3.2 and 3.3 show a non-exhaustive compilation of the ADI post-processing options with varying parameters. All the algorithms were set to mask the innermost $2\lambda/D$ region.

We observe how more complex PSF subtraction techniques ou perform the classic median subtraction approach for cleaning the innermost part of the image ($\sim 2\lambda/D$). We refrain from deriving additional conclusions about the comparison of different post-processing techniques as this is beyond the scope of the present work. Furthermore this is an exercise to be carried out using a diverse collection of datasets (from different instruments) and with appropriate metrics (defined by the whole community), such as the area under the ROC curve, in order to provide general and robust conclusions. We envision `VIP` as a library that could eventually implement all the main high-contrast imaging algorithms and become a tool suitable for benchmarking different data processing approaches under a unified open framework.

3.2.3 Sensitivity limits and discussion

The code below shows how to compute the S/N for a given resolution element, obtain an S/N map, call the NEGFC MCMC function and compute a contrast curve.

Listing 3.3: VIP: S/N map, sensitivity curves, NEGFC-MCMC.

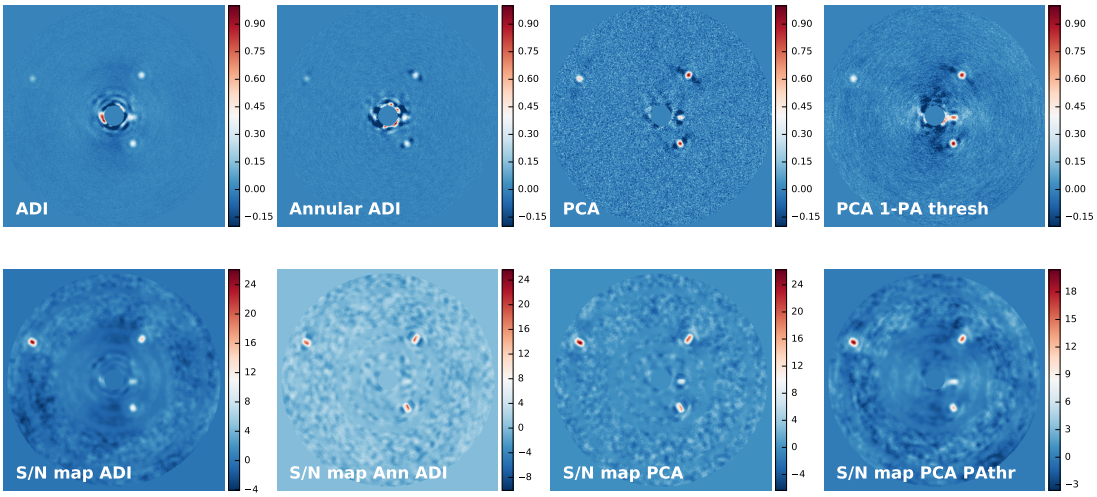


Figure 3.2: Post-processing final frames (top row) and their corresponding S/N maps (bottom row) for classical ADI, annular ADI, full-frame ADI-PCA, and full-frame ADI-PCA with a parallactic angle threshold. The final frames have been normalized to their own maximum value. No normalization or scaling was applied to the S/N maps, which feature their full range of values.

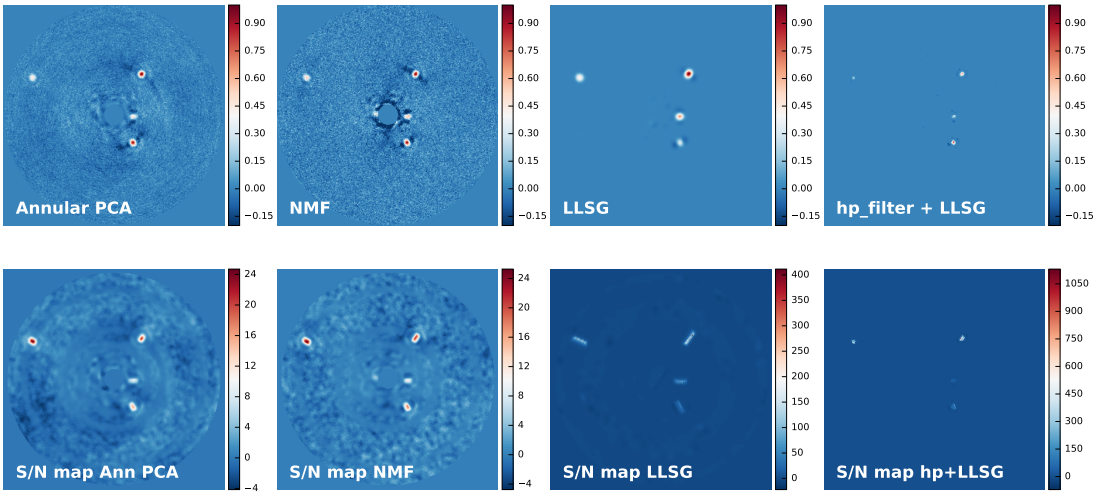
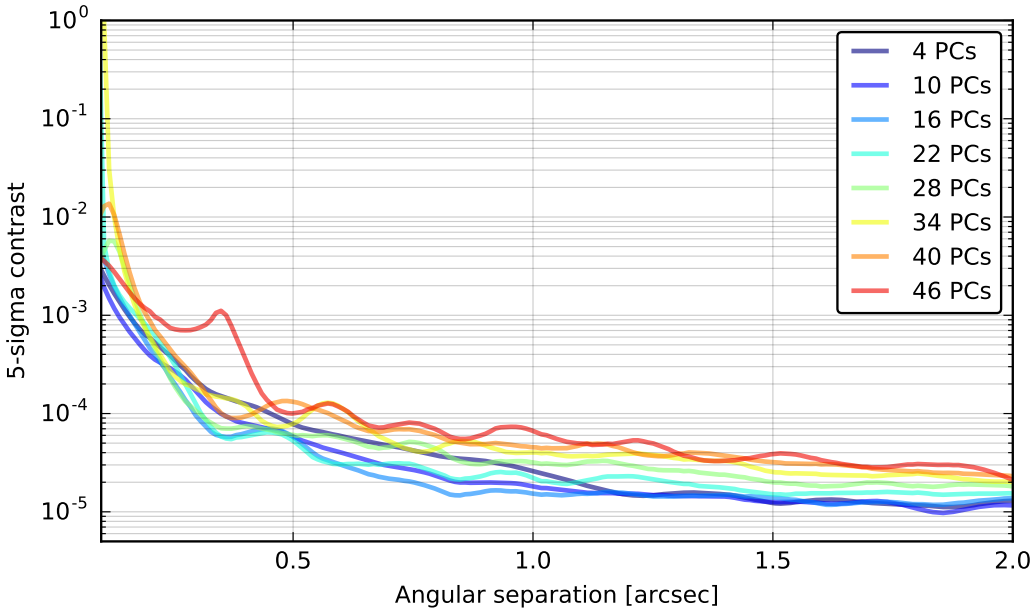
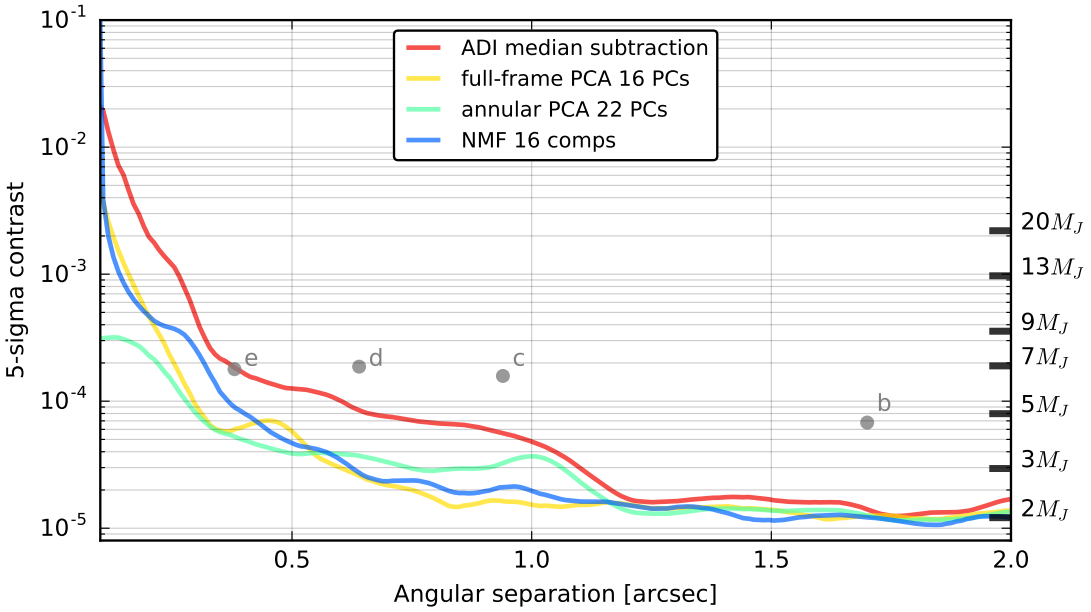


Figure 3.3: Same as Fig.3.2 for annular ADI-PCA, full-frame ADI-NMF, LLSG and high-pass filtering coupled with LLSG. We note that a high S/N does not translate in increased sensitivity to fainter companions.



(a)



(b)

Figure 3.4: (a) 5-sigma sensitivity (with the small sample statistics correction) for full-frame ADI-PCA with different numbers of PCs. (b) 5-sigma sensitivities for some of the ADI algorithms in `VIP`. The four known companions were removed before computing these contrast curves. The small sample statistics correction has been applied to these sensitivities.

```
from vip.phot import snr_ss, snrmap, contrast_curve
snr_value = snr_ss(fr_pca, source_xy=(54,266), fwhm=fwhm_lbt)

# S/N map generation
snr_map = snrmap(fr_pca, fwhm=fwhm_lbt, plot=True)

# NEGFC mcmc sampling
from vip.negfc import mcmc_negfc_sampling, confidence,
    cube_planet_free
ini_rad_theta_flux = np.array([r_0, th_0, f_0])
chain = mcmc_negfc_sampling(cube_ss, pa_ss, psfn=psf, ncomp=8,
                            plsc=pxscale_lbt, fwhm=fwhm_lbt,
                            initialState=ini_rad_theta_flux,
                            nwalkers=100, niteration_min=100,
                            niteration_limit=400, nproc=10)

# 1-sigma confidence interval calculation from the mcmc chain
val_max, conf = confidence(chain, cfd=68, gaussianFit=True,
                           verbose=True)
final_rad_theta_flux = [(r, theta, f)]
cube_emp = cube_planet_free(final_rad_theta_flux, cube_ss, pa_ss,
                            psfn=psf, plsc=pxscale_lbt)

# res_cc is a (pandas) table containing the constrast, the radii
# where it was evaluated, the algorithmic throughput and other
    values
res_cc = contrast_curve(cube_emp, pa_ss, psf_template=psf,
                        fwhm=fwhm_lbt, pxscale=pxscale_lbt,
                        starphot=starphot, sigma=5, nbranch=1,
                        algo=vip.pca.pca, ncomp=8)
```

Using the NEGFC technique, we subtracted the four known companions in our datacube and computed the sensitivity curves on this empty datacube. Panel (a) of Fig. 3.4 shows the 5-sigma sensitivity for full-frame ADI-PCA with varying principal components to exemplify the dependence on the algorithm parameters. By using `VIP`'s ADI-PCA algorithm in its annular mode and setting a different number of PCs for each separation, we could obtain the optimal contrast curve, as already shown by Meshkat et al. (2014a). Panel (b) of Fig. 3.4 shows the 5-sigma sensitivities for the available ADI algorithms in `VIP`. These sensitivity limits should be representative of the expected performance of the algorithms when applied to different data, but the result may vary, therefore preventing us from presenting more general conclusions. As expected, in panel (b) of Fig. 3.4 we observe how the median reference PSF subtraction achieves worse contrast than the rest of the algorithms. Also, we see that with annular ADI-PCA, impressive contrast is achieved at small separations (below 0.″5) and similar contrast at larger separations if it is compared to full-frame ADI-PCA. Annular ADI-PCA presents a smaller dependence on the number of principal components (the variance of the contrast curves,

when varying k, is smaller compared to full-frame ADI-PCA). For the full-frame ADI-NMF sensitivity, we used 16 components as in the case of full-frame ADI-PCA and obtained a fairly similar performance at all separations.

The contrast metric as defined in `VIP` is not well adapted to all algorithms and/or datasets, therefore we refrain from presenting such sensitivity curves for LLSG. We remind that these contrast curves were obtained on a temporally sub-sampled datacube. However, because we do not include time-smearing when injecting fake companions, we expect the results to be representative of the ultimate sensitivity based on the full (non sub-sampled) ADI sequence. The contrast-to-mass conversion for the HR8799 planets was obtained assuming an age of 40 Myr (Bowler, 2016) and using the 2014 version of the PHOENIX BTSETTL models for substellar atmospheric models described in Baraffe et al. (2015). Based on this, we can discard the presence of a fifth planet as bright as HR8799e down to an angular separation of 0.″2. Our detection limits remain in the planetary-mass regime down to our inner working angle of 0.″1, and reach a background-limited sensitivity of $2M_J$ beyond about 1.″5.

Finally, it is worth mentioning that the full-frame ADI-PCA sensitivity curve presented in the panel (b) of Fig. 3.4 (yellow curve) is slightly worse than the one shown in Maire et al. (2015b) and obtained four days later with the same instrument but without the AGPM coronagraph. In order to make a fair comparison, we re-processed this dataset with `VIP` and obtained the same results as Maire et al. (2015b) at large angular separations but more pessimistic results closer in (within 0.″5). This can be explained by the student-t correction that we apply. If we compare the contrast curves produced by `VIP` for both datasets, we observe that at small angular separations, within 1″, the AGPM coronagraph provides an improvement in contrast up to 1 magnitude.

3.3 ROBUST ASTROMETRY OF HR8799 SPHERE DATA

With the advent of second-generation high-contrast planet imagers like VLT/SPHERE, obtaining astrometric measurements of directly imaged planets is now becoming routine. It is therefore more important than ever that the methods used to derive such astrometric measurements include a careful estimation of all error sources, including systematic biases that are expected to affect even the most advanced planet imaging instruments. Here, we propose to derive the astrometry of the four HR8799 planets based on a data set obtained with SPHERE during its science verification phase in December 2014.

The astrometric position of the HR8799bcde planets based on the December 2014 SPHERE/IRDIS data set has already been determined by Zurlo et al. (2016) and Apai et al. (2016). The Zurlo et al. (2016) final astrometry was obtained from the combination of four independent image-processing pipelines, by the quadratic sum of the error bar from each data reduction pipeline plus the standard deviation associated to the individual positions. Apai et al. (2016) used their implementation of the KLIP algorithm to derive companion positions by injecting artificial planets with negative count rates, and used a manual inspection of the image quality and of the subtraction residuals to estimate the error bars. Here, we propose to go beyond these approaches and to study in details the various contributions to the astrometric error budget, in an attempt to derive more reliable error bars. Our study is also meant to explore the ultimate astrometric accuracy of a state-of-the-art instrument such as SPHERE, and to identify possible ways to improve the astrometric accuracy in future studies.

What we call *robust astrometry* consists in performing a proper evaluation of the statistical errors and systematic biases affecting the final astrometric estimation. The whole procedure consists of four steps: (i) the description and estimation of the instrumental calibration errors, (ii) the determination of the planet position with respect to the star and the related statistical error through Bayesian inference with MCMC sampling, (iii) the determination of the systematic error due to residual speckles, and (iv) the calculation of the error on the star position.

3.3.1 Observations and data reduction

3.3.1.1 *Observations*

SPHERE performs high-contrast imaging by combining an extreme adaptive optics system (Fusco et al., 2006), several coronagraphic masks, and three science sub-systems including the Infra-Red Dual-band Imager and Spectrograph (IRDIS, Dohlen et al., 2008). The observations of HR8799 were performed during five consecutive nights from 4 to 8 December 2014, using IRDIS in the broadband H filter ($1.48 - 1.77$ μm) with an apodized Lyot mask (Soummer, 2005; Carbillet et al., 2011; Guerri et al., 2011) of diameter 185 mas together with an undersized Lyot stop. A beam splitter located downstream of the coronagraphic mask produces two identical parallel beams (Beuzit et al., 2008), which results in two well separated images per acquisition, hereafter referred to as the *left* and *right* images. Each of the five observing sequences lasted for about half an hour, and consisted of 218 frames with a detector integration time (DIT) of 8 sec per frame. All observing sequences were obtained under fair seeing conditions (between $0''.8$ and $1''.5$), except on 7 December where the seeing was above $1''.5$. The sequences were acquired in pupil-stabilized mode to

take advantage of the ADI observing technique. Due to the low elevation of HR8799 as seen from Cerro Paranal (maximum altitude of 44°), the amount of parallactic angle rotation was however quite small, amounting to $8.^{\circ}7$, $8.^{\circ}5$, $8.^{\circ}3$, $8.^{\circ}1$ and $7.^{\circ}8$ for the five nights, respectively. Four elongated diffraction spots, the so-called satellite spots, were created during the whole observing sequences by injecting a waffle pattern on the deformable mirror (Langlois et al., 2012) to help with the star centering procedure, as explained further.

3.3.1.2 *Image calibration*

The IRDIS raw frames were preprocessed using the SPHERE `EsoRex` pipeline. As a first step, master dark and flat frames were created from calibration data obtained for each night of observations. Then `EsoRex` identified the outlying pixels in the master dark frame by using a sigma clipping procedure and built a static bad-pixel map. Each frame was reduced by subtracting the corresponding master dark, dividing by the master flat and interpolating the pixels flagged in the bad-pixel map. At this stage we obtained two calibrated data cubes per night, one for each side of the IRDIS detector, resulting in ten data cubes. From each data cube we discarded bad frames by measuring the correlation of each frame with a reference frame that was tagged as good by visual inspection. Only the 85% to 95% most correlated frames were kept for post-processing, depending on the night. The night of 7 December was discarded due to its poor data quality, as already proposed by Apai et al. (2016).

We deliberately chose to skip the centering of the individual frames proposed by `EsoRex`. Instead, we used `VIP` to precisely measure the position of the star and the related uncertainty for each individual frame of all data cubes by exploiting the four satellite spots. Indeed, since the satellite spots have a high S/N and are designed to be symmetric with respect to the star, one can use them to infer the position of the star. In practice, due to their wavelength-dependent elongation and to residual atmospheric dispersion, the satellite spots are not perfectly symmetric with respect to the star (Pathak et al., 2016). However, the symmetry is preserved at any given wavelength, and the spectrum-weighted astrometric position of the four satellite spots remains symmetric with respect to spectrum-weighted astrometric position of the star. To make sure to avoid the astrometric bias on the determination of the star position described by Pathak et al. (2016), the following strategy was adopted. For a given frame, we carefully fitted an asymmetric 2d Gaussian to each of the satellite spots to determine their respective centroid. Then, opposite centroids were connected by lines and the resulting intersection determined the estimated position (x, y) of the star in detector coordinates. This was done for each frame to get the offset of the star from the center of the frame. For each data cube, a histogram of these offsets was built, and global offsets were

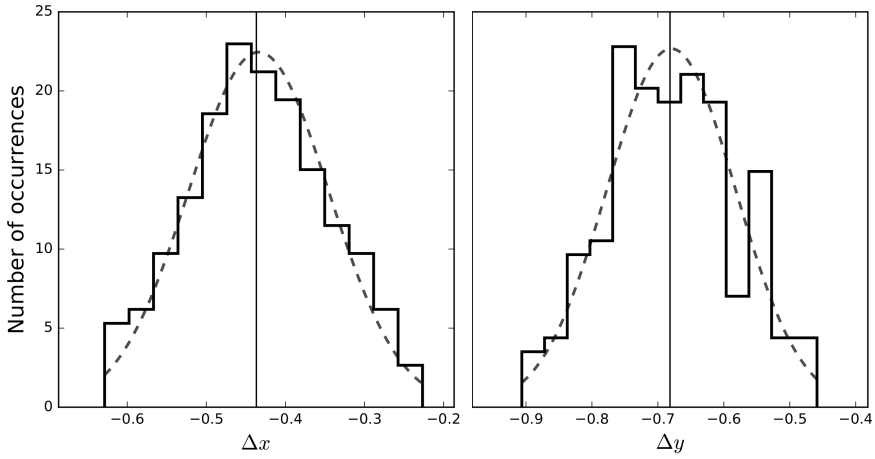


Figure 3.5: Histogram of the horizontal and vertical offsets of the star with respect to the center of the frame in the 5 December (right) data cube. The vertical line represents the median of the histogram, and was used to globally re-center the data cube. The horizontal axis is in pixels, one pixel corresponding to 12.25 mas on sky.

obtained as the median of the vertical and horizontal offsets (see Fig. 3.5). All the frames were then shifted by the same amount for each cube to cancel the global offset, and cropped to a useful field-of-view of 511×511 pixels to reduce computation time in the post-processing. Our analysis suggests that a frame-by-frame recentering of the cube would not improve the final results, because the accuracy with which the stellar position can be determined in an individual frame is generally not smaller than the width of the histogram shown in Fig. 3.5. More details about the uncertainty on the position of the star are given in Section 3.3.5.

The parallactic angles corresponding to the individual frames of each data cube were independently calculated frame by frame. The MJD time at the middle of each frame was derived from the information given by the `MJD-OBS` and `HIERARCH ESO DET FRAM UTC` header cards, which respectively give the time at the start and the end of the observing sequence, by dividing the total integration time equally into 218 parts. The parallactic angles were calculated using the spherical trigonometry formula given in Meeus (1998) based on the equatorial coordinates precessed to the epoch of the observations and corrected for nutation, aberrations, and refraction.

3.3.1.3 *Image post-processing*

We carried out the data post-processing with VIP, using its ADI capabilities to obtain final images where the planets can be identified. To optimize the determination of the astrometry, the S/N for each planet needs to be maximized. The S/N of the planets in the final, post-processed image depends mainly on the number of PCs used

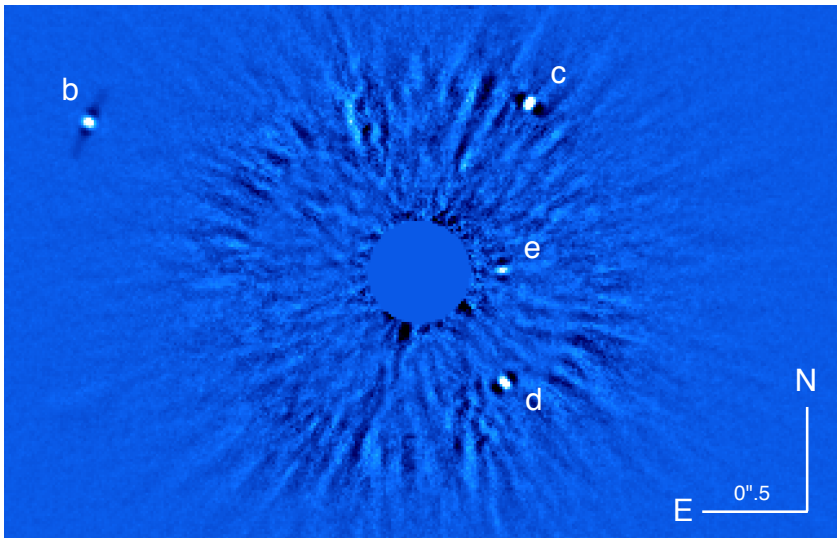


Figure 3.6: Illustration of a full-frame ADI-PCA post-processed SPHERE/IRDIS image of HR8799 acquired with broad-band H filter, left part, during the night of 4 December 2014. The central part was masked with a disk of radius of 20 pixels.

when building the reconstructed cube. A small number of PCs leads to an incomplete representation of the speckle noise, while a large number of PCs tends to capture the signal of the planet in the reconstructed cube, which leads to a lower algorithmic throughput for the planetary signal after subtraction. An optimum number of PCs can generally be found to maximize the planet S/N (Meshkat et al., 2014b). For each data cube, we thus performed a grid search on the number of PCs to maximize the mean S/N in a region of one resolution element in diameter around each companion. The optimal n_{PC} for each data cube is reported in Table 3.1.

Figure 3.6 illustrates a `VIP` post-processed image using full-frame PCA, where all the pixels of each frame are used at once to construct the reference images through SVD. The close region surrounding the host star is most affected by residual speckle noise and was masked with a disk of radius of 20 pixels to better reveal the planets in Fig. 3.6. Throughout the present analysis, we also performed annulus-wise PCA, which consists in performing PCA only for a thin annulus passing through a companion, with a typical width of a few resolution elements. Although full-frame PCA and annulus-wise PCA may lead to slightly different results, this choice does not significantly affect the final astrometry, which is dominated by other sources of error. Furthermore, annulus-wise PCA is significantly faster when performed on a single annulus, which is useful when dealing with large data cubes and/or when PCA is performed a large number of times (see Sect. 3.3.3).

3.3.2 Instrumental calibration and related errors

To derive accurate astrometric measurements from IRDIS images, various astrometric calibrations need to be performed, namely the determination of the plate scale, the orientation of the north and the optical distortion. Firstly, the plate scale, given in arcsec per pixel, depends on the characteristics of all the optical elements composing the instrument. It allows to convert positions given in pixel into arcsec. Secondly, when observing in pupil-stabilized mode, the vertical axis of the detector does not necessarily point towards north. Two contributions need to be taken into account: (i) the pupil offset, which accounts for the zero point position of the derotator and is assumed to be constant between runs, and (ii) the so-called true north, which accounts for a variation in the detector orientation with respect to the sky due to thermal or mechanical stresses, and which needs to be estimated during each observing run. Thirdly, the distortion in SPHERE/IRDIS is mainly dominated by an anamorphic magnification between the horizontal and vertical axis of the detector. This effect is due to the presence of toric mirrors in the common path of the instrument (see e.g., Zurlo et al., 2016).

Details of the observations used to derive those astrometric calibrations for IRDIS are described in Zurlo et al. (2016). We refer to that paper for the details, but we still provide the reader with the practical information used in this study. The astrometric calibrations were obtained from IRDIS observations of the globular cluster 47 Tuc acquired on 15 December 2014 with the same instrument setup and filter, and compared to the Hubble Space Telescope data of the same field, precessed to the same epoch and corrected for the differential proper motions of the individual stars. The values derived by Zurlo et al. (2016) for the plate scale and true north based on this data set have recently been revised by the SPHERE consortium, using their improved knowledge of the instrument. This revised estimation, described in Maire et al. (2016), leads to a plate scale of 12.251 ± 0.009 mas/pixel and a true north orientation of $-1.^{\circ}709 \pm 0.^{\circ}051$. These values are valid for both the left and right parts of the IRDIS detector. The pupil offset, based on commissioning and guaranteed-time data obtained on several astrometric fields, is equal to $135.^{\circ}99 \pm 0.^{\circ}11$. Finally, the IRDIS distortion measured on sky is dominated by an anamorphism of $0.60\% \pm 0.02\%$ between the horizontal and vertical directions (Maire et al., 2016). Although the SPHERE calibration plan includes the daily measurement of distortion maps based on pinhole grids, we found that the quality of the astrometric estimations does not improve by using these maps. Prior to any post-processing, we thus simply rescaled each frame of each cube by a factor 1.006 along the y axis. To take into account the uncertainty on this correction, an additional error of 0.02% on the plate scale will be considered in the following analysis.

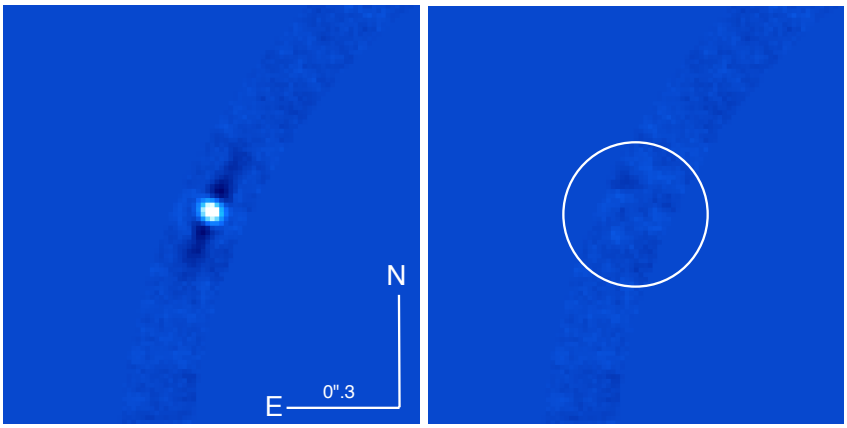


Figure 3.7: Illustration of the annulus-wise PCA post-processing and merit function evaluation used in the negative fake companion technique. *Left*. No NEGFC was injected before annulus-wise PCA processing. *Right*. A NEGFC was injected at the position and flux minimizing the merit function. The white circle illustrates the fixed circular aperture from which the pixel values I_j have been extracted to evaluate the merit function. The same color scale was used for both images.

3.3.3 Planet position and statistical error

The next step in the robust astrometry process consists in determining for each data cube the position of the planets with respect to the host star and in estimating the statistical error related purely to the photon noise of the underlying thermal background and speckles through Bayesian inference based on MCMC simulations. This step does not describe the effect of the speckles themselves on the measured planet position, which will be discussed separately in Sect. 3.3.4. Our astrometric measurements are based on the NEGFC technique described in the previous chapter of this dissertation.

The resulting post-processed images, before and after injection of a NEGFC, are represented in Fig. 3.7. Because no off-axis PSF was acquired in December 2014 with the same observing setup as for the HR8799 observations, the adopted PSF template corresponds to unsaturated off-axis images of β Pictoris obtained with SPHERE/IRDIS during science verification on 30 January 2015 (PI: A.-M. Lagrange) with the same observing mode as for HR8799 (same coronagraph, same broadband H filter, similar seeing ~ 1″). The influence of this choice will be discussed at the end of Sect. 3.3.3, together with a discussion of the effect of PSF chromatic dispersion on the measured planet position.

With the optimal number of PCs in hand (see Table 3.1), we derive a first guess on the position, which is later refined with the Nelder-Mead optimization procedure in `VIP`. The right panel of Fig. 3.7 illustrates the result of an annulus-wise ADI-PCA post-processing per-

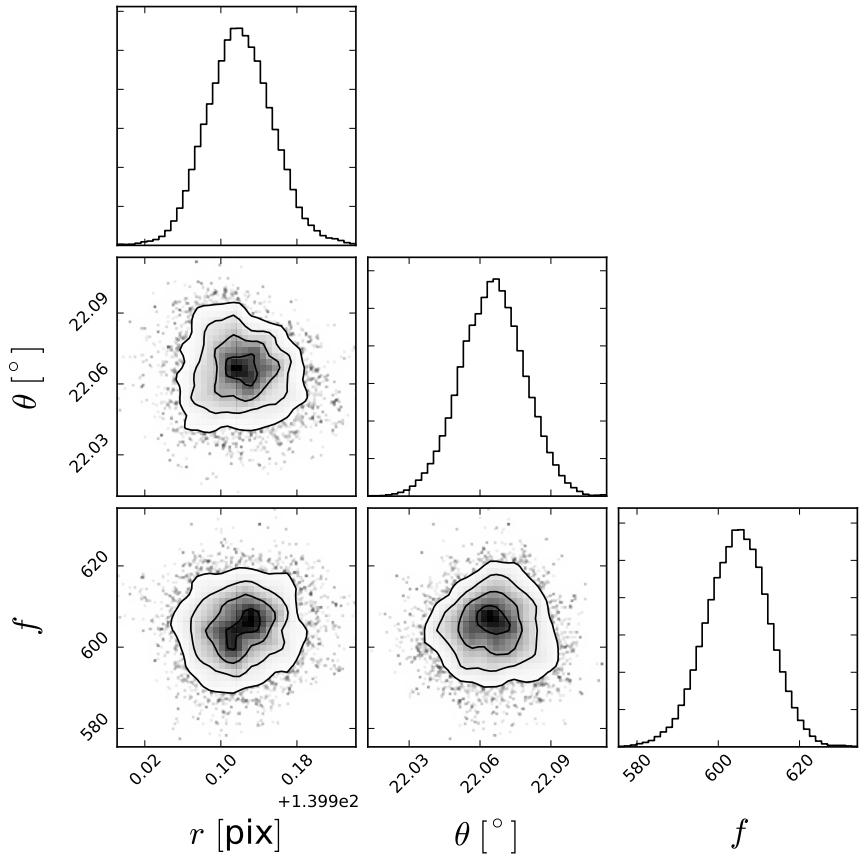


Figure 3.8: Illustration of a typical corner plot obtained from the MCMC simulations using the NEGFC technique. The target companion is HR8799b observed during the night of 6 December 2014. The radial distance r (in pixel) and azimuth θ (in degree) are detector coordinates with respect to the host star. The diagonal panels illustrate the posterior PDFs while off-axis ones illustrate the correlation between them.

formed on a single annulus passing through HR8799b, after injecting a NEGFC characterized by a position and flux minimizing the merit function. Finally, we use this estimation of the flux and position to initialize the MCMC sampling to obtain the final flux and position of the HR8799 planets.

For each data cube and each companion, we carried out MCMC simulations to sample posterior PDFs related to the planet polar coordinates (r, θ) with respect to the host star and the planet flux f. For each MCMC simulation, we used 200 walkers firstly initialized in a small ball around the solution obtained from the Nelder-Mead optimization. The chain was sufficiently close to convergence to allow Bayesian inference after typically 200 steps. In Fig. 3.8, the so-called corner plot illustrates the posterior PDFs and the correlation between the parameters (r, θ, f) for HR8799b observed during the night of 6

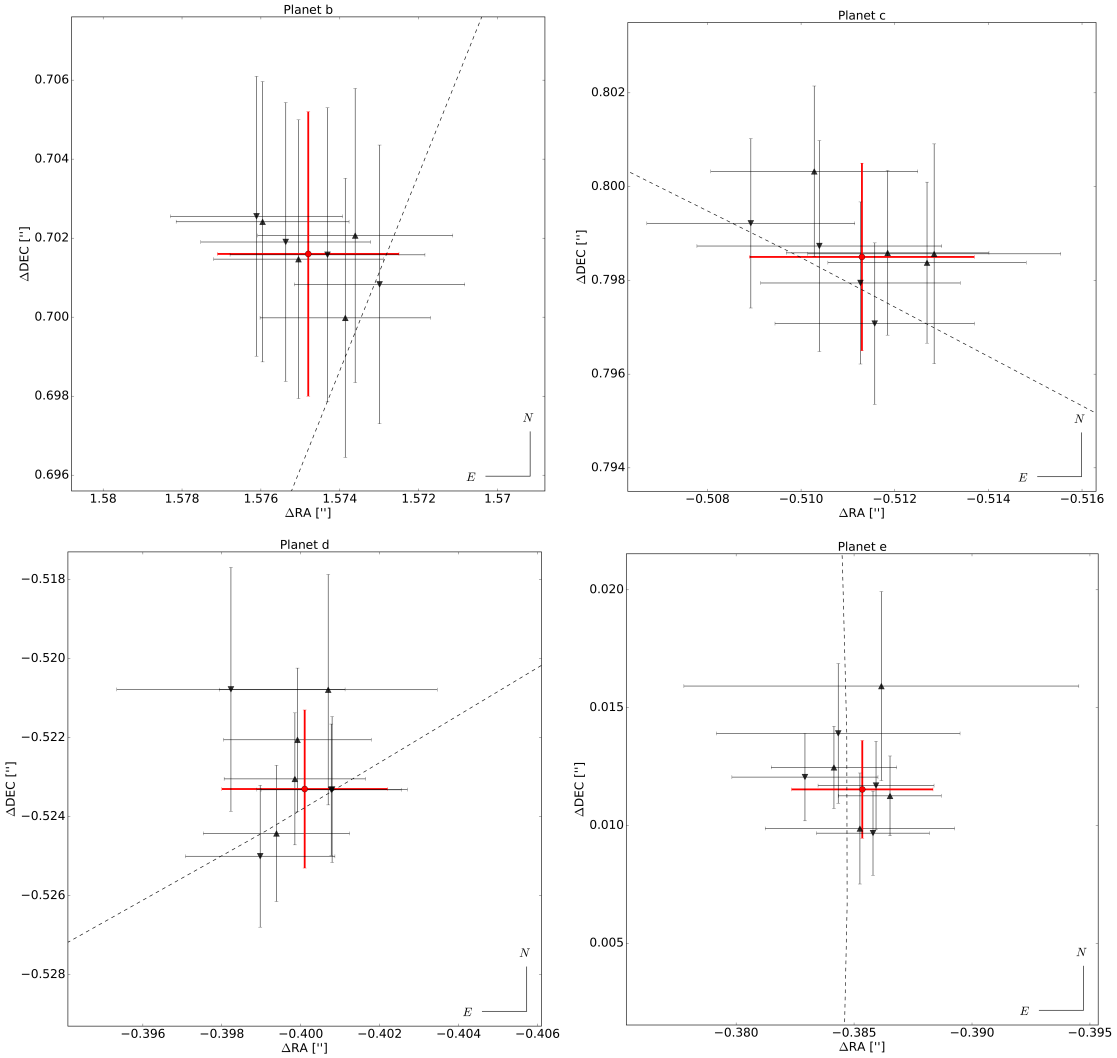


Figure 3.9: Astrometry for HR8799bcde observed during the nights of 4, 5, 6, and 8 December 2014. The positions obtained from the left (resp. right) data cubes are represented with downward (resp. upward) black triangles. The error bars on the individual data points take into account all the contributions discussed in Sects. 3.3.2, 3.3.3, 3.3.4, and 3.3.5. The red dots correspond to the final astrometric measurements for each planet, together with the final error bar discussed in Sect. 3.3.6. The dashed lines represent the best orbital solutions for each planet in terms of reduced χ^2. See Wertz et al. (2016).

December 2014. Similar results were obtained for other planets and observing nights. Although a flux estimation for each planet is obtained, we focus our analysis only on the astrometry in this section.

Taking into account the plate scale, the true north and pupil offset orientation (see Sect. 3.3.2), we have projected the HR8799bcde highly probable sets of polar coordinates onto the north and east directions. As a result, the eight HR8799bcde final positions for the four nights (left and right parts) are reported in Table 3.1 and displayed in Fig. 3.9. These positions will be used in Sect. 3.3.6 to deduce the final HR8799 astrometry for epoch 2014.93. In addition to obtaining the highly probable position/flux for a given companion, the MCMC simulations give a robust estimation of the statistical error on the astrometry (i.e., related purely to photon noise). This error, reported in columns 7 and 8 of Table 3.1, generally constitutes a minor contribution to the error budget, as discussed in the next sections.

3.3.3.1 *Influence of the template PSF*

Since a non-saturated, off-axis PSF was not obtained in the same observing mode during the nights where HR8799 was observed, we chose as a PSF template for our NEGFC analysis the closest off-axis PSF in time obtained with the same observing mode under similar weather conditions, which turned out to be an off-axis PSF of beta Pictoris obtained on 30 January 2015. The fact that both the instrument and the atmospheric conditions may have changed within the interval leads to a possible bias in our measurement of the planets position, which could vary from night to night. To evaluate this bias, we have taken a series of twelve off-axis PSFs observed in the same mode under good atmospheric conditions, obtained in 2015 in the context of the SHARDDS survey (J. Milli, personal communication). For each planet and each observing night in our HR8799 data set, we successively used the twelve off-axis PSFs as templates for the NEGFC technique, and derived the planets astrometry using the method described above. The dispersion of the astrometric measurements gives us an estimation of the bias that can be introduced by using a non-contemporaneous PSF. The observed dispersion does not depend much on the planet nor on the observing night, and has an overall standard deviation of 0.6 mas. This error bar will be added quadratically to the other error sources in Sect. 3.3.6.

3.3.3.2 *Influence of residual dispersion*

Another source of imperfection in the recovery of the planets astrometry for broadband observations is the residual atmospheric dispersion after correction by the atmospheric dispersion correctors (ADC) included in the SPHERE optical path. While the small angular separation between the star and planets ensures the residual dispersion to

Table 3.1: Final HR8799bcde astrometric measurements with respect to the host star for nights of 4, 5, 6, and 8 December 2014, derived from SPHERE/IRDIS broadband H measurements (left and right parts), in terms of RA/DEC (columns 3-4) and in polar coordinates (columns 5-6). In addition, we list the derived optimal number of principal components n_{PC} (column 2), as well as the statistical error bars (columns 7-8) and the speckle noise error bars (columns 9-10), both in polar coordinates.

Date and side	n_{PC}	ΔRA [″]	ΔDEC [″]	Δr [″]	$\Delta\theta$ [°]	$\sigma_{stat,r}$ [″]	$\sigma_{stat,\theta}$ [°]	$\sigma_{spec,r}$ [″]	$\sigma_{spec,\theta}$ [°]
HR8799b									
2014-12-04 L	3	1.5754	0.7019	1.7247	65.985	0.0003	0.007	0.0002	0.007
2014-12-04 R	2	1.5750	0.7015	1.7242	65.994	0.0003	0.008	0.0002	0.007
2014-12-05 L	7	1.5761	0.7026	1.7256	65.975	0.0003	0.009	0.0002	0.009
2014-12-05 R	6	1.5760	0.7024	1.7254	65.977	0.0003	0.008	0.0002	0.008
2014-12-06 L	6	1.5730	0.7008	1.7221	65.985	0.0004	0.013	0.0002	0.009
2014-12-06 R	6	1.5739	0.7000	1.7225	66.023	0.0004	0.013	0.0002	0.009
2014-12-08 L	4	1.5743	0.7016	1.7236	65.980	0.0003	0.013	0.0003	0.011
2014-12-08 R	4	1.5736	0.7021	1.7231	65.956	0.0003	0.008	0.0003	0.010
HR8799c									
2014-12-04 L	5	−0.5116	0.7971	0.9471	327.307	0.0002	0.014	0.0006	0.048
2014-12-04 R	6	−0.5127	0.7984	0.9488	327.293	0.0002	0.010	0.0006	0.044
2014-12-05 L	13	−0.5089	0.7992	0.9475	327.512	0.0003	0.012	0.0008	0.059
2014-12-05 R	14	−0.5103	0.8003	0.9492	327.479	0.0004	0.015	0.0008	0.053
2014-12-06 L	15	−0.5113	0.7979	0.9477	327.351	0.0005	0.020	0.0006	0.047
2014-12-06 R	18	−0.5118	0.7986	0.9485	327.342	0.0005	0.013	0.0007	0.052
2014-12-08 L	20	−0.5104	0.7987	0.9479	327.421	0.0004	0.026	0.0010	0.077
2014-12-08 R	7	−0.5128	0.7986	0.9491	327.291	0.0003	0.016	0.0012	0.088
HR8799d									
2014-12-04 L	5	−0.3990	−0.5250	0.6594	217.233	0.0012	0.024	0.0012	0.093
2014-12-04 R	5	−0.3994	−0.5244	0.6592	217.292	0.0004	0.027	0.0011	0.092
2014-12-05 L	21	−0.4008	−0.5233	0.6592	217.448	0.0006	0.035	0.0013	0.085
2014-12-05 R	21	−0.3999	−0.5221	0.6576	217.454	0.0005	0.039	0.0013	0.075
2014-12-06 L	21	−0.4008	−0.5233	0.6592	217.446	0.0005	0.022	0.0010	0.077
2014-12-06 R	20	−0.3999	−0.5230	0.6584	217.397	0.0005	0.017	0.0010	0.080
2014-12-08 L	18	−0.3982	−0.5208	0.6556	217.405	0.0004	0.030	0.0029	0.136
2014-12-08 R	46	−0.4007	−0.5208	0.6571	217.575	0.0005	0.029	0.0027	0.123
HR8799e									
2014-12-04 L	9	−0.3859	0.0117	0.3861	271.735	0.0010	0.103	0.0022	0.202
2014-12-04 R	16	−0.3852	0.0099	0.3854	271.468	0.0013	0.077	0.0039	0.292
2014-12-05 L	10	−0.3829	0.0121	0.3831	271.803	0.0006	0.044	0.0029	0.196
2014-12-05 R	12	−0.3841	0.0125	0.3843	271.859	0.0005	0.055	0.0024	0.167
2014-12-06 L	8	−0.3858	0.0097	0.3859	271.436	0.0006	0.034	0.0022	0.182
2014-12-06 R	23	−0.3865	0.0113	0.3867	271.668	0.0006	0.048	0.0019	0.159
2014-12-08 L	15	−0.3843	0.0139	0.3846	271.072	0.0016	0.186	0.0049	0.357
2014-12-08 R	11	−0.3862	0.0159	0.3865	272.360	0.0008	0.145	0.0082	0.534

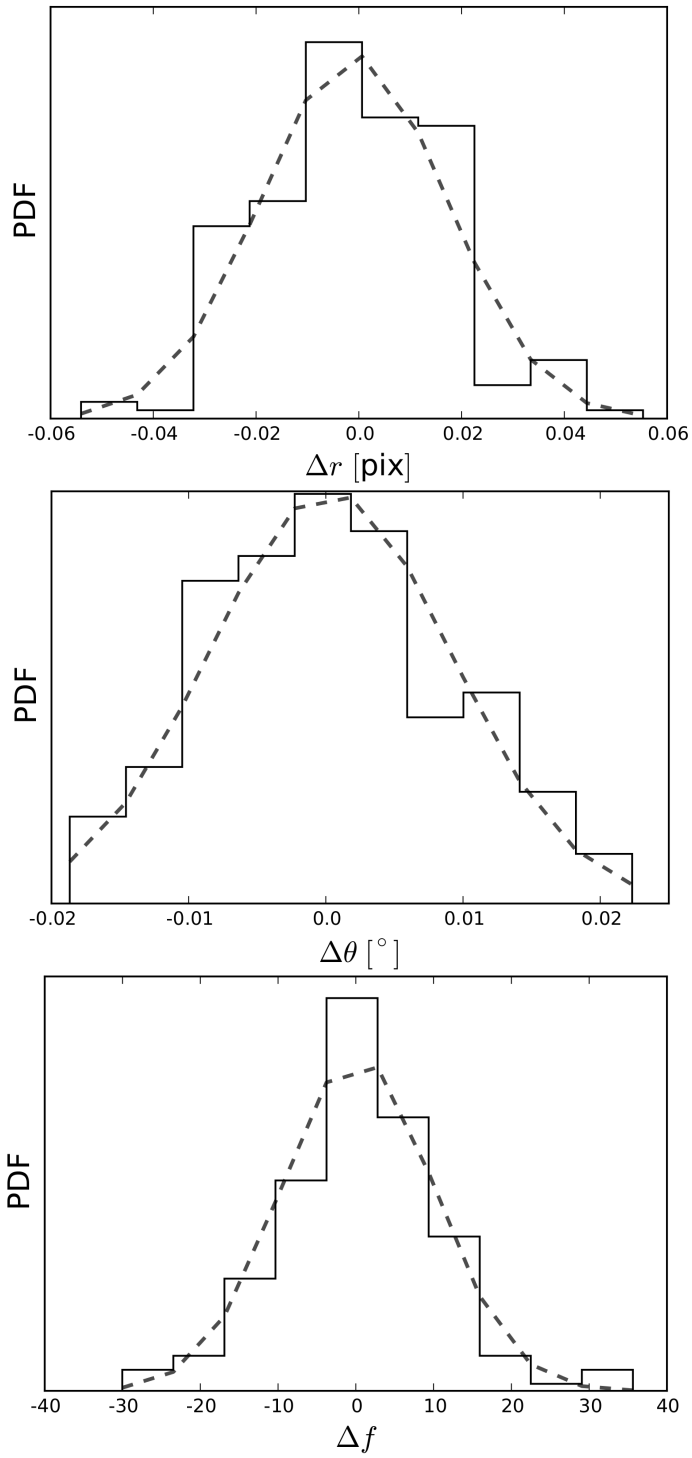


Figure 3.10: Speckle noise estimation for HR8799b observed on 6 December 2014. The histograms illustrate the offsets between the true position/flux of a fake companion and its position/flux obtained from the NEGFC technique. The dashed lines correspond to the 1D gaussian fit from which we determine the speckle noise.

be almost perfectly equal for all of them, their different spectra can result in a chromatic offset between their measured positions. The residual dispersion after correction by the SPHERE ADC has been shown to be smaller than 1.2 mas rms for zenith angles as large as the maximum of 54° encountered in the present data set (Hibon et al., 2016). Taking into account the H-band spectrum of the star and of the four planets (Bonnefoy et al., 2016), we estimate that the maximum astrometric offset between the star and planets due to residual dispersion cannot be larger than 0.25 mas in the worst case where residual dispersion shows a linear trend across the H band. This contribution is negligible in our final astrometric error budget.

3.3.4 Systematic error due to residual speckles

Performing ADI-PCA removes a large fraction of the quasi-static speckle noise and significantly improves the S/N of the companions. Although highly effective, this process is not perfect and some level of residual speckle noise remains in the post-processed images. Such noise has a major impact on photometric and astrometric measurements (Guyon et al., 2012), and needs to be taken into account in the error budget. Since speckle noise is known to have a radial dependence, we estimate its impact by injecting fake companions in the data cube at the radial distance of the real planets but for a wide range of angular positions, and by testing the ability of the Nelder-Mead optimization to find their position and flux through the NEGFC technique. The first step in this process is to create an "empty" data cube by injecting four NEGFCs characterized by the highly probable positions/fluxes derived from the previous MCMC simulations. In the empty cube, we inject a fake companion characterized by a flux f_{true} and a radial distance r_{true}, both corresponding to the highly probable solution, but at an arbitrarily chosen angular coordinate $\theta_{true,i}$. Using the NEGFC technique coupled with the Nelder-Mead optimization, we determine the position/flux (r_i, θ_i, f_i) of the fake companion. We then compute the offsets $\Delta r_i = r_{true} - r_i$, $\Delta\theta_i = \theta_{true,i} - \theta_i$, $\Delta f_i = f_{true} - f_i$ between the known position/flux characterizing the fake companion and the solution obtained from the optimization process. The same process is repeated for a series of 360 azimuths equally spaced between 0° and 360°. These 360 realizations are used to build three normalized histograms, respectively for Δr, $\Delta\theta$ and Δf. The histograms for Δr and $\Delta\theta$ are then fitted with a Gaussian function, and the standard deviations $\sigma_{spec,r}$ and $\sigma_{spec,\theta}$ of the Gaussian functions are used as an estimation of the speckle noise affecting the radial and azimuthal coordinates. A similar approach was already used by e.g. Maire et al. (2015a). We illustrate in Fig. 3.10 the three histograms for HR8799b observed on 6 December 2014. The results obtained for all the planets and data cubes are reported in column 9 and 10 in Table

3.1. It appears clear that the error induced by speckle noise increases for decreasing angular separations of the companion with respect to the host star. Indeed, the brightness of the residual speckles increases closer to the star. We also note that speckle noise is always larger than statistical noise, except for HR8799b.

Another possible way to evaluate speckle noise is to measure the influence of the number of PCs used in the PCA post-processing on the position/flux determination, as proposed e.g. by Pueyo et al. (2015). Indeed, the residual speckle pattern changes as a function of the number of PCs. To check the consistency of this method with the one proposed above, we determined the position/flux of each companion in each data cube using the NEGFC technique with the Nelder-Mead optimization using a number of PCs ranging from 5 to 90 (for a number of PCs > 90, the companion self-subtraction becomes too important to get a high S/N). We then constructed three normalized histograms, respectively for r, θ and f. As expected, the standard deviations of these histograms are similar to those deduced above.

Finally, we note that the residual speckle noise estimated here is in good agreement with the semi-empirical estimation of the astrometric accuracy based on the planet S/N proposed in the case of pure photon noise by Guyon et al. (2012, Eq. A1), if we extrapolate this relation to the speckle-dominated regime in the following way, as already proposed by Mawet et al. (2015): $\sigma_{1D}[\lambda/D] = 1/(\pi S/N)$. Using such a semi-empirical formula therefore looks like a possible way to get a quick estimation of the astrometric error bar related to speckle noise, although we recommend to go through the analysis presented in this section to obtain a robust estimation.

3.3.5 Error on the star position

Inside SPHERE, a dedicated differential tip-tilt sensor is used to obtain an image of the PSF just upstream of the coronagraph, and is used as an input for closed-loop control of the star position with respect to the coronagraph, thereby ensuring a stable star centering (Fusco et al., 2006; Baudoz et al., 2010). Based on laboratory measurements, the expected accuracy of the star centering is supposed to be around 0.5 mas on sky (Baudoz et al., 2010). As mentioned in Sect. 3.3.1.2, no individual frame centering was applied to the data cubes in our analysis, but rather a global centering of all frames in each individual cube using the same x, y offsets.

Here, we independently estimate the uncertainty on the mean star position for each data cube. The evaluation of this uncertainty is based on the histogram of the x and y offsets measured for all individual frames by the centroid plus intersection method described in Sect. 3.3.1.2. The mean position of the star in a given data cube can be obtained by a Gaussian fit of the two histograms, as illustrated

Table 3.2: Estimation of the stellar jitter in the eight data cubes.

Date and side	$\sigma_{\star,\mathrm{RA}}$ [″]	$\sigma_{\star,\mathrm{DEC}}$ [″]
2014-12-04 L	0.00076	0.00078
2014-12-04 R	0.00076	0.00078
2014-12-05 L	0.00084	0.00081
2014-12-05 R	0.00087	0.00085
2014-12-06 L	0.00077	0.00079
2014-12-06 R	0.00078	0.00080
2014-12-08 L	0.00143	0.00139
2014-12-08 R	0.00144	0.00139

in Fig. 3.5. Based on this figure, we will assume in the following discussion that the histograms follow a Gaussian distribution, so that the accuracy on the determination of the mean stellar position in a given cube is given by the standard deviation of the best-fit Gaussian divided by the square root of the number of realizations. The standard deviations of the best-fit Gaussians are given in Table 3.2 in terms of RA and DEC, by projecting the $(\sigma_{\star,x}, \sigma_{\star,y})$-error ellipses expressed in detector coordinates onto the north and east directions. We note that the derived stellar jitter estimation is slightly larger than predicted in Baudoz et al. (2010), with values varying from 0.76 mas to 1.44 mas depending on the night (i.e., around 0.1 pixel in detector coordinates). Based on these values, and taking into account the $\sim$ 200 frames present in each data cube, the error bar on the mean stellar position in any given cube amounts to less than 0.1 mas, and is therefore completely negligible in our final noise budget.

However, this contribution represents only the purely statistical error on the determination of the star position. We also need to take into account possible systematic biases on the determination of the star position based on the satellite spots. To this end, we have obtained a data set on a relatively bright star, using the waffle mode of the DM but without coronagraph. The star was mildly saturated at its center to increase the S/N on the satellite spots. We determined the center of the star based on a truncated Moffat profile, to reject the saturated part of the PSF, and compared this estimation with the prediction based on the satellite spots. We checked that the two estimations match with an accuracy better than 0.1 pixel, which represents our best estimation of an upper limit on a possible bias. This also confirms that the method proposed in Section 3.3.1.2 to determine the stellar position from the satellite spots does not lead to a major astrometric bias, even in the presence of residual atmospheric dispersion. Here, we will conservatively assume that a bias of 0.1 pixel (1.2 mas) affects our determination of the mean star position in all cubes.

3.3.6 Final astrometry

Particular care must be taken when combining the results and error bars of several astrometric measurements, especially in the presence of correlated errors. How the various error bars add up needs a specific discussion. Firstly, we note that our experimental determination of the error bar related to residual speckles inherently takes into account the contribution of photon noise. Indeed, the empirical intensity of the speckles includes the contribution of the photon noise associated to all sources of signal at any given location (stellar residuals, planet, sky emission and thermal background). This is backed up by the fact that the error bar associated to speckle noise generally dominates the error bar associated to photon noise. Only in the case of planet b are they of the same order of magnitude, which reflects the fact that residual speckles are very faint compared to residual background noise at that angular distance from the star.

Secondly, we make the conservative assumption that the errors related to speckle noise are fully correlated, not only between the left and right data cubes obtained on a same night, but also between all nights. The assumption of full correlation between the left and right data cubes is justified by the fact that the signals recorded by the two parts of the detector are almost identical (to within photon noise and some minor differential aberrations that amount to a few nm rms at most), and is backed up by the fact that the estimated error bars are almost identical for the left and right sides for most of the nights and planets (see Table 3.1). The assumption that speckle noise is fully correlated from night to night is more debatable. It is indeed expected that speckle noise will be partly correlated between successive nights, because residual speckles are often associated to non-common path aberrations in the instrument that can vary on very long timescales. To be on the conservative side, we will assume a full correlation of speckle noise in all data sets. The error bar on the final astrometry regarding speckle noise should then be computed as the median of all speckle noise-related error bars. We note however that the estimations of the speckle noise-related error bars significantly vary from one night to the other (see Table 3.1), which suggests that this noise is at least partly uncorrelated, and that our final error bars will be pessimistic.

Thirdly, we proposed in the previous section that the final error bar related to the determination of the star position is dominated by a systematic bias that can amount up to 1.2 mas, and that the variability of the PSF shape can induce a bias of up to 0.6 mas. These biases will be added quadratically to our final astrometric error bar for all planets. The same applies to instrumental calibration errors, which are supposed to affect all data cubes in the exact same way. Indeed, appropriate observations of astrometric fields were not performed on each of the five HR8799 observing nights. We therefore had to rely

Table 3.3: The final HR8799bcde astrometric measurements with respect to the star for epoch 2014.93.

Planet	Δr [″]	$\Delta\theta$ [°]	ΔRA [″]	ΔDEC [″]
HR8799b	1.7241 ± 0.0019	65.99 ± 0.13	1.5748 ± 0.0023	0.7016 ± 0.0036
HR8799c	0.9481 ± 0.0017	327.37 ± 0.16	-0.5113 ± 0.0024	0.7985 ± 0.0020
HR8799d	0.6587 ± 0.0019	217.40 ± 0.19	-0.4001 ± 0.0021	-0.5233 ± 0.0020
HR8799e	0.3855 ± 0.0030	271.71 ± 0.31	-0.3853 ± 0.0030	0.0115 ± 0.0021

on an astrometric calibration carried out by the SPHERE consortium one week later (see Section 3.3.2), which was used as a reference for all five nights. Although we could not check the stability of the calibration over a few nights, we note that the latest IRDIS astrometric calibrations by the SPHERE consortium show that the time variations of plate scale and true north are mostly within their estimated error bars, based on two years of astrometric fields observations, while the pupil offset and anamorphic factor are mostly constant (Maire et al., 2016). This suggests that our final estimation of the astrometric error bar should not include any unaccounted bias related to the variability of the IRDIS astrometric calibration. That being said, we still recommend that, in future observing programs dedicated to precise astrometric measurements, observations of standard astrometric fields be obtained during each individual night to ensure a high astrometric robustness.

Based on these assumptions, the computation of the final astrometry and related error bars proceeds as follows for each planet:

- define the final astrometry of the four planets as the weighted mean of the eight individual positions (left and right parts of the detector for the four nights), using as a weight the inverse of the variance of speckle noise;
- estimate the final error bar related to speckle noise as the median of the individual error bars on the eight astrometric measurements;
- add quadratically the contribution of speckle noise, the upper limit on the stellar centering bias, and the contribution of instrumental calibration errors to obtain the final astrometric error bars.

All these calculations are performed in polar coordinates, reflecting the fact that error bars generally have different behaviors along the radial and azimuthal directions. The last step is based on the following formulae:

$$\sigma^2_{\mathrm{tot},r} = \mathrm{PLSC}^2(\sigma^2_{r,\mathrm{spec}} + \sigma^2_{r,\star} + \sigma^2_{r,\mathrm{PSF}} + \sigma^2_{r,\mathrm{AF}} r^2) + \sigma^2_{\mathrm{PLSC}} r^2 \ , \quad (3.1)$$

$$\sigma^2_{\mathrm{tot},\theta} = \sigma^2_{\theta,\mathrm{spec}} + \sigma^2_{\theta,\star} + \sigma^2_{\theta,\mathrm{PSF}} + \sigma^2_{\theta,\mathrm{AF}} + \sigma^2_{\mathrm{PO}} + \sigma^2_{\mathrm{TN}} \ , \quad (3.2)$$

Table 3.4: Comparison between the final error bars (σ_{tot}) listed in Table 3.3 and the standard deviation of the eight positions per planet displayed in Fig. 3.9 (see also Table 3.1).

Planet	$\sigma_{tot,\Delta RA}$ [mas]	$\sigma(\Delta RA)$ [mas]	$\sigma_{tot,\Delta DEC}$ [mas]	$\sigma(\Delta DEC)$ [mas]
HR8799b	2.3	1.1	3.6	0.8
HR8799c	2.4	1.2	2.0	0.9
HR8799d	2.1	0.9	2.0	1.4
HR8799e	3.0	1.2	2.1	1.9

where r is the radial distance in pixels, $\sigma_{r,spec}$ and $\sigma_{\theta,spec}$ the final radial (pixels) and azimuthal (degrees) error bars related to speckle noise, $\sigma_{r,\star}$ and $\sigma_{\theta,\star}$ the radial (pixels) and azimuthal (degrees) stellar centering biases, $\sigma_{r,PSF}$ and $\sigma_{\theta,PSF}$ the radial (pixels) and azimuthal (degrees) error bars related to the imperfection of the PSF template in the NEGFC analysis, $\sigma_{r,AF}$ and $\sigma_{\theta,AF}$ the radial and azimuthal errors on the anamorphic factor expressed in percent, and where PLSC refers to the plate scale in ″/pixel, PO to the pupil offset and TN to the true north, both in degrees. The final astrometric positions and related error bars are given for the four planets in Table 3.3 and are illustrated in Fig. 3.9. Table 3.3 includes a projection of the error bars onto the RA and DEC directions, to comply with the usage. However, we suggest that expressing the error bars in polar coordinates is more appropriate, because polar coordinates usually correspond to the major and minor axes of the error ellipse. Another, even more appropriate way to proceed would be to specify the error ellipse by its three parameters (two axes and position angle). In the present case, the error bars are sufficiently symmetric to proceed with RA/DEC error bars, even though we note that the HR8799b error bars are significantly asymmetric, the angular error bar being twice as large as the radial one. This is mostly due to the large uncertainty on the pupil offset (0°.11, see Section 3.3.2), which severely affects planets located far away from the star.

To check the consistency of our error bars, we compared the statistical distribution of the eight individual data points obtained for each planet to the individual error bars on the eight data points. Table 3.4 shows that the final error bars are generally about twice larger than the dispersion of the individual data points. This is related to the fact that the major error sources (speckle noise, stellar position bias, instrumental calibration) are supposed to be fully correlated between individual measurements, so that the final error bar has a similar size as the individual ones. This suggests that an improvement by up to a factor two in astrometric accuracy could be achieved by improving the astrometric calibration. That said, the individual error

bars are in relatively good adequacy with the dispersion of the data points (see Fig. 3.9), although we note a significant asymmetry in the distribution of the data points towards the NE-SW direction. This asymmetry looks quite consistent between the four planets, and we therefore suggest that it comes from a time variability in the bias on the stellar position measurement (the only error source that is naturally expressed in RA/DEC), which could be related to variations in the PSF shape and/or in the diffraction pattern created by the DM on a night-to-night timescale. This variation remains within the expected amplitude of about 0.1 pixel for the star position bias.

For planet b, the main contribution to the error budget comes from the imperfect astrometric calibration and from the uncertainty on the star position, while speckle noise is negligible. This is consistent with the fact that HR8799b lies in a region that is not significantly affected by residual speckles (see Fig. 3.6). For planet c, although speckle noise significantly increases, the noise budget remains dominated by the astrometric calibration and stellar position uncertainties. The dominance of stellar centering noise in the astrometric error budget of these two planets is backed up by the fact that the dispersion in the individual astrometric measurements for planets b and c has a similar amplitude and shape (see Fig. 3.9), as expected for a global centering error. For the two inner most planets (d and e), speckle noise progressively becomes the dominant contributor to the error budget, and once again this is consistent with Fig. 3.9, where the dispersion of the astrometric data points increases significantly, especially for planet e. We finally note that our astrometric measurements are in general agreement with the astrometric measurements derived in Zurlo et al. (2016) and Apai et al. (2016) to within error bars, but that our error bars are two to three times smaller, thanks to a careful evaluation of all systematic error sources.

3.4 CONCLUSIONS

In this chapter we showcased `VIP`'s capabilities for processing ADI data, using long sequences on HR8799 taken with LBTI/LMIRCam (in its AGPM coronagraphic mode) and four shorter sequences with VLT/SPHERE/IRDIS. In the first case, we used all of `VIP`'s capabilities to investigate the presence of a potential fifth companion around HR8799 but we did not find any significant additional point-like sources. We also compared the sensitivities of coronagraphic and non-coronagraphic modes of LBTI/LMIRCam. For the VLT/SPHERE data, we obtained robust astrometry of the four known companions with a detailed analysis of the various contributions to the astrometric error budget. The resulting astrometric positions agree within 1σ with previous estimations based on the same data set (Zurlo et al.,

2016; Apai et al., 2016), with error bars two to three times smaller thanks to a careful estimation of systematic errors. The main contribution to the astrometric error depends on the angular distance from the star: the error budget is dominated by the uncertainty on the stellar position (~ 1 mas) and instrumental calibration errors for planet b, while residual speckle noise increases for smaller angular separations and becomes dominant for planet e. We note that these revised error bars match the early expectations of SPHERE in terms of astrometric accuracy (~ 2 mas), and suggest that the astrometric accuracy could even be further improved (especially for planets located outside the speckle-noise dominated regime) by a more careful IRDIS astrometric calibration and by improving upon our estimation of the bias on the star center determination using dedicated calibration programs. In practice, nothing seems to prevent SPHERE/IRDIS from reaching a 1 mas astrometric accuracy in the future based on a careful calibration plan.

The studies described in this chapter show how powerful `VIP` can be. It is also an illustration of the benefits of open-source development in science. Nowadays, `VIP` is being actively used by researchers from numerous universities and HCI research teams around the globe, which is accompanied by a growth in the developers contributing to the project (sending modification requests and proposals).

Part II

MACHINE LEARNING APPLIED TO EXOPLANET DETECTION

4 MACHINE LEARNING PERSPECTIVE

Contents

Abstract

In this Chapter, I introduce some machine learning fundamental concepts that may not be intuitive to astronomers. The motivation of this Chapter is to put high-contrast imaging data processing in a machine learning context and to connect previous Chapters, describing HCI state-of-the-art algorithms, with Chapters 5 and 6. The concepts reviewed in this Chapter will be indispensable for explaining my novel approaches to high-contrast imaging, introduced in the next two Chapters.

4.1 INTRODUCTION

Although machine learning only started to flourish in the 1990s, it has quickly become the most popular and most successful subfield of artificial intelligence, driven by the availability of faster hardware and larger datasets (benchmarks). Machine learning explores the study

and construction of algorithms that can learn from and make predictions on data. It aims to build a model from sample inputs to generate data-driven predictions. This radically differs from the classical programming approach of static program instructions and rules. Machine learning is tightly related to mathematical statistics, but unlike it, machine learning deals with larger and more complex datasets (e.g. a dataset of millions of images, each consisting of tens of thousands of pixels) for which classical statistical analysis would simply be too impractical to be possible.

Depending on the nature of the learning "signal" available to a learning system, we can classify machine learning tasks into three broad categories: unsupervised, supervised, and reinforcement learning. In the coming sections, we will focus on the first two categories of machine learning and will connect them to the tasks faced in the context of HCI data processing.

4.2 UNSUPERVISED LEARNING

Unsupervised learning techniques deal with unlabeled data or data of unknown structure. They are used to explore the structure of data to extract meaningful information without the guidance of a known outcome variable (see next Section on supervised learning) or reward function (reinforcement learning).

4.2.1 Clustering

Clustering is a technique that allows to organize a set of objects into meaningful subgroups (clusters) based on some similarity measure. It is considered an exploratory data analysis technique and it does not require having any prior knowledge of group memberships. Different clustering models differ in the particular definition of a cluster. For instance, hierarchical clustering uses a measure of dissimilarity between sets of observations, achieved by the use of a distance metric between pairs of observations and a linkage criterion specifying the dissimilarity of sets of data points. It can be agglomerative or divisive. In the agglomerative case, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. The results are usually presented using dendrograms. Another common clustering approach is the k-means algorithm, which partitions the set of observations into k pre-defined clusters, each one represented by a prototype: the centroid (average) of similar points with continuous features. The k-means algorithms works in an iterative manner with the goal of minimizing the within-cluster sum of squared errors or cluster inertia. Density-based spatial clustering of applications with noise (DBSCAN, Ester et al., 1996) is a density-based clustering

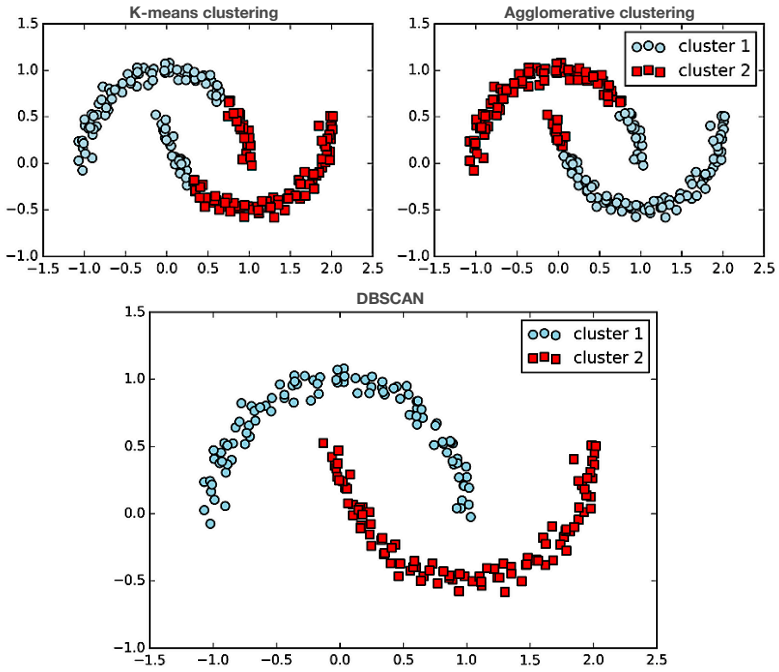


Figure 4.1: Comparison of k-means clustering, hierarchical clustering, and DBSCAN on the same dataset. Taken from Raschka (2015).

model which defines clusters as connected dense regions in the data space. DBSCAN does not require to a priori specify the number of clusters in the data. Also, it is very effective at finding non-linearly separable clusters, as illustrated in Fig. 4.1.

4.2.2 Dimensionality reduction

Unsupervised dimensionality reduction seeks to compress the input data onto a smaller dimensional subspace while retaining most of the relevant information. This is a commonly used approach for feature pre-processing as it reduces computation time and storage requirements, and enables better visualization. The most common technique in the family of dimensionality reduction algorithms is PCA. It was invented in 1901 by Karl Pearson and later developed (re-discovered) by Harold Hotelling in the 1930s. It is used to decompose a multivariate dataset in a set of successive orthogonal components that explain a maximum amount of the variance. PCA is closely related to the SVD of the input matrix X and the ED of the covariance $X^T X$, as described in Section 1.3.3.5.

PCA is also closely related to the matrix decomposition (factorization) and low-rank approximation problems. Several approaches, other than PCA, have been proposed to model the low-rank approximation of a matrix. A unified view of matrix factorization was pro-

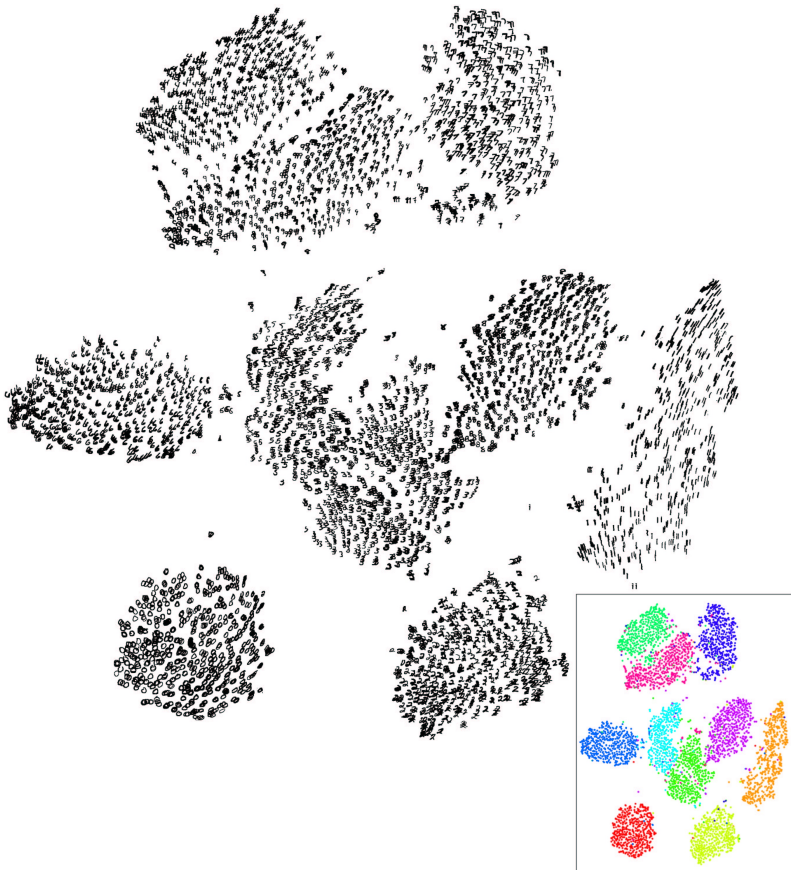


Figure 4.2: Visualization of 6,000 digits from the MNIST dataset produced with t-SNE. The subplot on the bottom-right shows the same plot, where colors represent the labels of the digits. Taken from van der Maaten and Hinton (2008).

posed by Udell et al. (2016), with their generalized low rank models (referring to the problem of approximating a data set as a product of two low dimensional factors by minimizing an objective function). This approach provides a common optimization formulation, not only for plain PCA, but also for other techniques, such as non-negative matrix factorization, matrix completion, sparse and robust PCA, k-means, k-SVD, and maximum margin matrix factorization.

Linear dimensionality reduction methods, e.g. PCA and independent component analysis, are powerful but fail to exploit non-linear important structure in the data. Non-linear dimensionality reduction or manifold learning can be thought of as an attempt to generalize linear frameworks, like PCA, to be sensitive to non-linear structure in the data. Among these methods, we count the t-distributed stochastic neighbor embedding (t-SNE, van der Maaten and Hinton, 2008). This technique is particularly well suited for embedding high-dimensional data into a space of two or three dimensions, enabling effective visu-

alizations. Fig. 4.2 shows a visualization of a subset of examples from the MNIST dataset[1] using the t-SNE.

4.2.3 Unsupervised learning in HCI: low-rank approximations and dictionary learning

As we saw in Section 1.3.3.5, PCA finds a low rank matrix that minimizes the approximation error, in the least-squares sense, to the original data set. It turns out that the approach of modeling a reference PSF (for differential imaging) using low-rank approximations can be considered as unsupervised learning. Over the last few years, the PCA-based approach has become the standard way to process HCI datasets. NMF (2.5.3) is another example of such low-rank approximation methods, which I proposed as an alternative to PCA-based reference PSF subtraction. In Chapter 5, I will explore other low-rank modeling concepts, such as robust PCA and low-rank plus sparse approximations, and describe a new post-processing method for ADI sequence based on low-rank plus sparse decompositions.

Another interesting approach, that I applied to HCI reference PSF modeling (for the first time), is dictionary learning (Olshausen and Field, 1996). This technique consists in learning to express (code) a set of input vectors, such as image patches, as linear combinations of a small number of vectors chosen from a large dictionary. Dictionary learning is a matrix factorization problem that consists in finding a (usually overcomplete) dictionary that will perform good at sparsely encoding the fitted data (Mensch et al., 2016). It can be interpreted as a generalization of the task of image approximation and reference PSF modeling in terms of a "basis", sharing common ideas with PCA (notion of basis) and LOCI (reconstruction on patches) types of HCI algorithms.

Dictionary learning is an optimization problem solved by alternatively updating the sparse code considering the dictionary fixed, and then updating the dictionary to best fit the sparse code:

$$\begin{aligned} (U^*, V^*) &= \underset{U,V}{\operatorname{argmin}} \frac{1}{2} \|X - UV\|_2^2 + \alpha \|U\|_1 \\ \text{s.t.} \, \|V_k\|_2 &= 1, \quad \text{for all} \quad 0 \leqslant k < n_{comps}, \end{aligned} \tag{4.1}$$

where X is a matrix whose rows are vectorized overlapping patches (from the input images), V is the dictionary and U the sparse code.

1 The MNIST dataset (modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for benchmarking machine learning algorithms. It consists of 60,000 training images and 10,000 testing images, each one of 20x20 pixels (grayscale) with a corresponding label.

Sparse coding aims to find a representation of the data as a linear combination of as few dictionary atoms as possible. The reconstruction of the target images can be formulated as:

$$\begin{aligned} &\min \|T - UV\|_2^2 \\ &\text{s.t. } \|U\|_0 \leqslant k, \end{aligned} \tag{4.2}$$

where T is a matrix whose rows are vectorized overlapping patches from the target images (it can be different from X or be the same) and V is the dictionary learned with Equation 4.1 from X. The formulation of Equation 4.2 corresponds to solving the sparse coding U with orthogonal matching pursuit. k is the number of atoms used to reconstruct our images. In a similar way, the sparse coding can be formulated with an L_1 penalty on the code.

4.2.3.1 *Application of dictionary learning to HCI data*

For the task of RDI sequence processing we are interested in producing a reference PSF model for a target sequence using a separate set of reference images (observations of a different star). Here I show the application of dictionary learning to the task of RDI post-processing using several VLT/SPHERE datasets[2]. One of the datasets (HD 206893) was chosen as the target, since it contains a newly discovered companion, which appears with a high S/N after a simple ADI-PCA processing. We concatenated four different data cubes and obtained a reference cube of 2541 frames. We learned an overcomplete dictionary (500 atoms) from 330,000 random 15x15 patches grabbed from the reference cube. A subset of the atoms of this dictionary is shown in panel (a) of Fig. 4.3.

It is interesting to notice how the dictionary captures the nature of the signal, which is composed of blob-like structures, Gabor-like filters and all sorts of oriented gradients. Panel (b) illustrates how sparse coding reconstructs a patch as a sparse linear combination of k atoms (20 in this case) from the dictionary. With more atoms used, we obtain smaller residuals. Each frame from the target sequence is decomposed in overlapping patches, for which we create approximations in terms of the atoms of the reference dictionary (Eq. 4.2). The number of non zero coefficients (or the number of atoms used) is tuned as a function of the distance from the center of the image (more aggressive towards the center). Panel (c) shows the final ADI image after reconstructing all the patches, from each one of the images of the sequence, and subtracting their reconstructions.

This approach results in an aggressive least squares approximation of the images and large self-subtraction (most of the companion signal is fitted in the reference PSF). It is also very computationally intensive (~ 8 hours on a 56-cores server for obtaining the dictionary

2 Provided by Julien Milli from his SHARDDS survey.

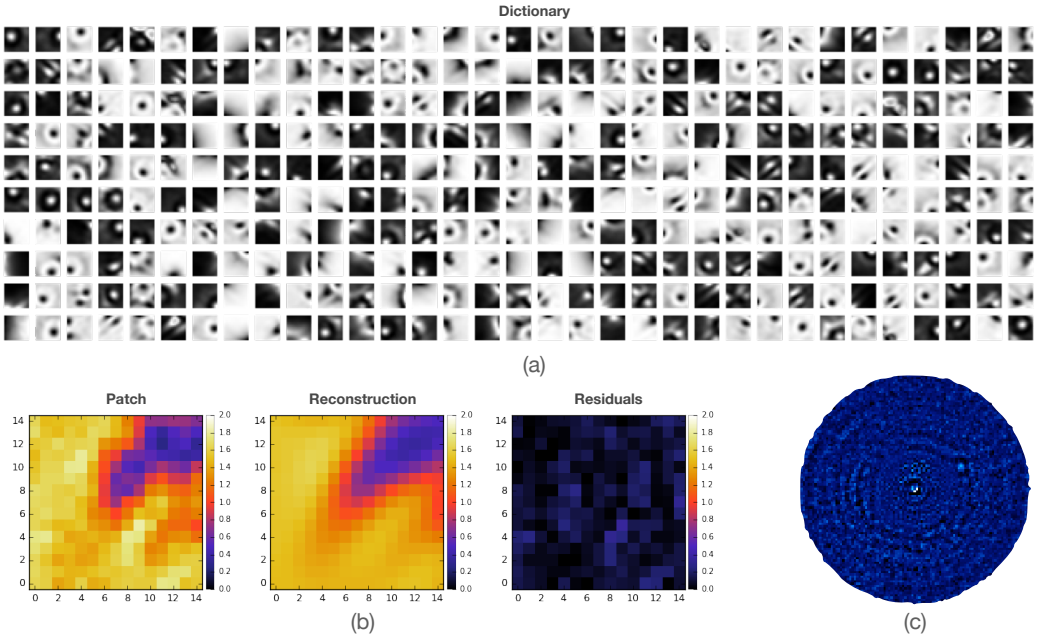


Figure 4.3: Dictionary learning and sparse coding for the task of HCI reference PSF modeling. (a) Subset of atoms from an overcomplete dictionary. (b) Reconstruction of a patch, from one image of the ADI sequence, using 100 atoms from the dictionary. (c) Final ADI sequence image, after the reconstruction (using dictionary learning and sparse coding) has been subtracted.

described earlier) when compared to a simple full-frame ADI-PCA post-processing (a few seconds). The reconstructive approach of dictionary learning and sparse coding could be complemented with a discriminative model to classify the patches in terms of the atoms used or the error of the reconstructions. This line of research was abandoned to pursue the promising field of deep learning (see Chapter 6).

4.3 SUPERVISED LEARNING

Supervised learning aims to learn a model from labeled training data in order to make predictions about unseen or future data. The term supervised refers to the need of a ground truth or set of examples (from now on *samples*) where the desired output signals (targets or *labels*) are already known. This mapping from input data *features* (input variables) to targets is done by observing many examples of inputs and targets.

Supervised learning problems can be further grouped into regression and classification problems, depending on the nature of the labels. For a classification problem, the ground truth is a category (e.g. the corresponding digit for the MNIST samples) while for a regression problem the ground truth is a real value. An intermediate kind of problem is the ordinal classification, which aims to predict an ordinal variable (a ranking). Among the most important families of supervised machine learning algorithms, we count the linear methods, support vector machines (SVM), nearest neighbor methods, and tree-based methods.

Linear models were largely developed in the pre-computer age of statistics. The linear methods model or approximate the relationship between the input features and the labels as linear. Different algorithms exist for estimating the coefficients involved in the linear combination of the input variables (e.g. least squares). The simplicity of these methods allows an interpretable description of how the inputs affect the output. See Hastie et al. (2009) for a detailed review of linear methods for regression and classification.

SVM methods (Boser et al., 1992) aim to construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or outliers detection. This hyperplane optimally separates the classes (in the classification case) when it has the largest distance to the nearest data point of any classes. This extends to non-linear cases by using the so-called kernel trick, which consists in projecting the original input space into a higher-dimensional space where the separation is presumably easier.

Nearest neighbor methods are non-parametric, meaning that there is no training stage. K-nearest neighbors implements a simple func-

tion of the training data. At test time it finds a predefined number (k) of training samples closest in distance to the new point, and predicts the label from these. K-nearest neighbors can construct highly irregular and complex decision boundaries and achieve high accuracy at the expense of a high computational cost.

Decision trees (Breiman et al., 1984) and tree-based methods are also non-parametric, having the ability of modeling arbitrarily complex relationships between inputs and outputs, without any prior information or assumptions about the underlying functions. Tree-based methods are also suited for classification and regression tasks. Decision trees can be combined with variance reducing techniques (see next Section) such as bagging (bootstrap aggregation) and boosting, giving rise to state-of-the-art algorithms such as random forests (Breiman, 2001) and gradient boosted trees (Friedman, 2001).

4.3.1 Bias-variance trade-off and generalization error

The generalization performance of a supervised learning algorithm relates to its prediction capability on independent (unseen) test data. This performance is extremely important as it guides the choice of a learning model and the choice of its hyperparameters[3] (Hastie et al., 2009). It is very important to avoid training and evaluating a learning model on the same input data X since it would produce a very optimistic error value due the model simply memorizing X. This would not inform on how well it generalizes to new, unseen data.

Cross-validation is a well-known model validation technique used in statistics. It allows to assess how the results of a model generalize to an independent dataset. The most simple case is to randomly divide the input data X into two subsets, one for training the supervised learning model (*training set*) and the other for validating the predictions of the trained model (*test set*).

Let us suppose a target variable Y (labels of X), a relationship relating Y to X such as $Y = f(X) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$, and a prediction model $\hat{f}(X)$. The expected squared prediction error at a point x_0 is (Hastie et al., 2009):

$$\mathrm{Err}(x_0) = \mathrm{E}[(Y - \hat{f}(x_0))^2], \tag{4.3}$$

which can be decomposed into bias and variance components:

$$\begin{aligned} \mathrm{Err}(x_0) &= \sigma_\epsilon^2 + \left[\mathrm{E}\hat{f}(x_0) - f(x_0)\right]^2 + \mathrm{E}\left[\hat{f}(x_0) - \mathrm{E}\hat{f}(x_0)\right]^2 \\ &= \text{Irreducible Error} + \text{Bias}^2 + \text{Variance}. \end{aligned} \tag{4.4}$$

The first term, irreducible error, is the variance of the target around its true mean $f(x_0)$ (noise in the true relationship that cannot be avoided). The second term, squared bias, is the amount by which the average

3 Hyperparameters are the tuning parameters of a machine learning algorithm.

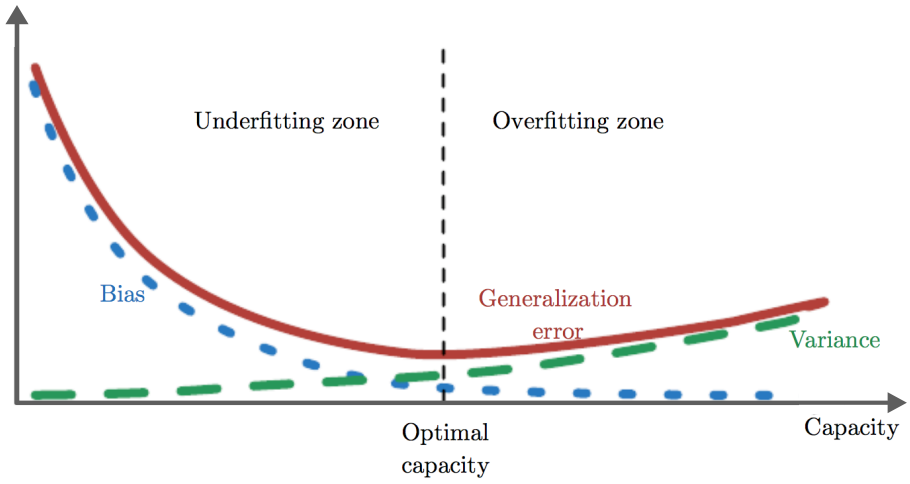


Figure 4.4: Shape of the generalization error curve and the relationship with the bias-variance trade-off. Taken from Goodfellow et al. (2016).

of our estimate differs from the true mean (error from erroneous assumptions in the learning algorithm). The third term, variance, is the expected squared deviation of $\hat{f}(x_0)$ around its mean (error from sensitivity to small fluctuations in the training set). The bias-variance trade-off (or dilemma) is the problem of simultaneously minimizing the bias and variance that prevent supervised learning algorithms from generalizing beyond their training set (see Fig. 4.4 and 4.5).

From this trade-off, we obtain the concepts of *underfitting* and *overfitting*. Underfitting originates when the supervised learning algorithm misses the relevant relationships between features and target outputs. In this case, the model has a high bias. Model overfitting is caused by a high variance, which is caused when the algorithm models random noise in the training data, limiting the prediction power on unseen new data (lack of generalization). We can control whether a model is more likely to overfit or underfit by altering its complexity or *capacity* (ability to fit a wide variety of functions). While simpler functions are more likely to generalize, or to achieve a small gap between the training and test error, we still need to choose a sufficiently complex hypothesis to achieve low training error.

Fig. 4.4 shows the typical U-shaped behavior of the generalization error as a function of the model capacity. Fig. 4.5 shows the approximation of a non-linear function with three models. The left most model is a linear one and is not sufficient to fit the training data (blue dots), which causes underfitting. The model in the middle is a linear regression with polynomial features of degree four, which fits the data very well. The model on the right uses a polynomial of degree 15 and clearly overfits the data (it learns the noise on the training dataset).

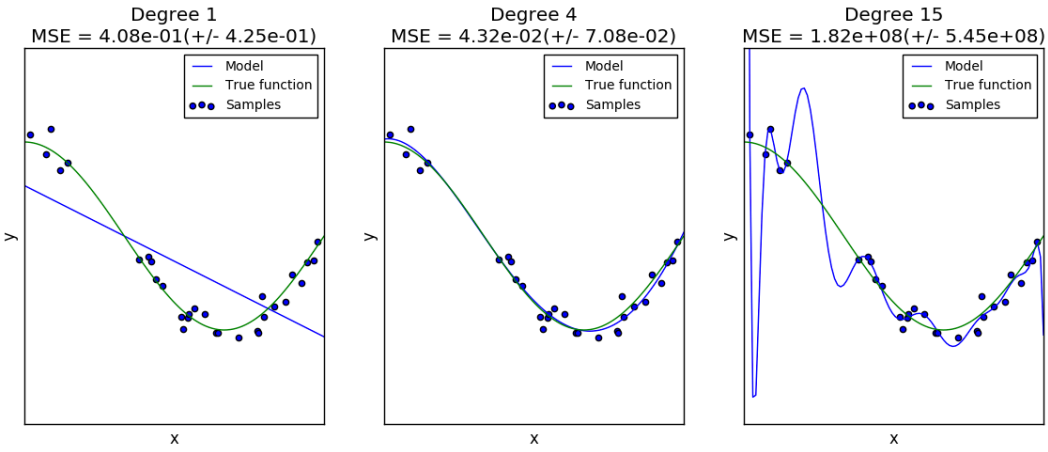


Figure 4.5: The approximation of a non-linear function with models of increasing complexity. The left model is underfitting while the right model is overfitting the data points. Taken from the `scikit-learn` documentation.

When evaluating the optimal hyperparameters of a learning model, a better solution is to split X into three parts: a training set (used to fit the models), a validation set (used for model selection based on their prediction error), and a test set (for assessment of the generalization error of the final model). The strategy of keeping a held-out subset of data is sensitive to the random partition of X. This can be addressed with k-fold cross-validation, where a test set is held out for final evaluation and the training set is split into k smaller sets. For each one of the fold, a model is trained on $k-1$ of the folds (as the training set) and validated by computing a performance metric on the remaining data. Fig. 4.5 shows an example of a regression model, evaluated using a mean squared error metric using a ten-fold cross-validation procedure. Variations or the cross-validation procedure are possible, such as Leave-p-out or stratified k-fold cross-validation.

The higher the validation error, the less likely the model generalizes well beyond the training data. Adjusting model representational capacity is not the only way to decrease the generalization error. We can, for instance, give the learning algorithm a preference for one function over another. A modification made to a learning algorithm to reduce its generalization error, but not its training error, is called *regularization*.

4.3.2 Supervised learning in HCI

It is worth noting that in the context of HCI, only one discriminative model has been proposed for the task of exoplanet detection (on multiple-channel SDI data): the DS4 Detect algorithm, an extension of the S4 algorithm (Fergus et al., 2014). Indeed, the task of exoplanet

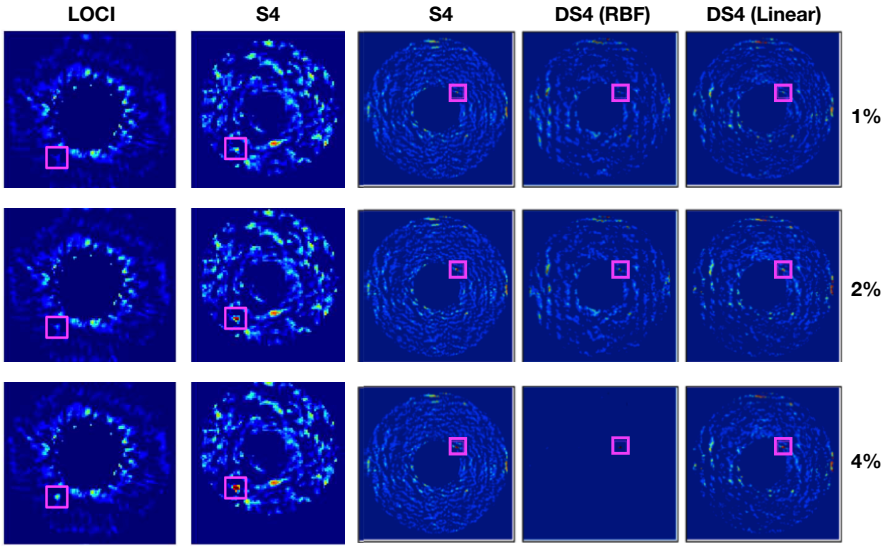


Figure 4.6: S4 and DS4 algorithms results. The first two columns (from the left side) correspond to one location and compare the S4 algorithm with damped LOCI. The remaining columns show a second injection location and compare S4 with DS4 using a linear SVM and a radial basis function (RBF or Gaussian) kernel SVM. Adapted from Fergus et al. (2014) and from slides presented by the same author at the Keck Institute for Space Studies Workshop on exoplanet imaging (2016).

signal detection can be formulated as a supervised learning problem. DS4 Detect adopts a discriminative approach based on SVMs, trained on a labeled dataset composed of negative samples taken directly from the input data and positive samples generated by creating synthetic companions. Unfortunately, there is no publication describing the details of this algorithm or robustly assessing its performance. It only appears in the discussion section of Fergus et al. (2014). Fig. 4.6 shows a comparison of S4, DS4 and LOCI on a few companion injections. The companions are injected at 1%, 2% and 4% of relative brightness with respect to the speckles. Judging by the final detection maps provided by the authors, the detection capacity is similar to the original S4 algorithm and to damped LOCI, which might explain why it has not been extensively used by the GPI campaign (Ruffio et al., 2017).

4.4 DEEP LEARNING

Deep learning is a specific subfield of machine learning. The "deep" in the term refers to the use of successive layers of representations. These layered representations are learned via models called "neural networks", which consist in stacked layers (see Fig. 4.7) one after the

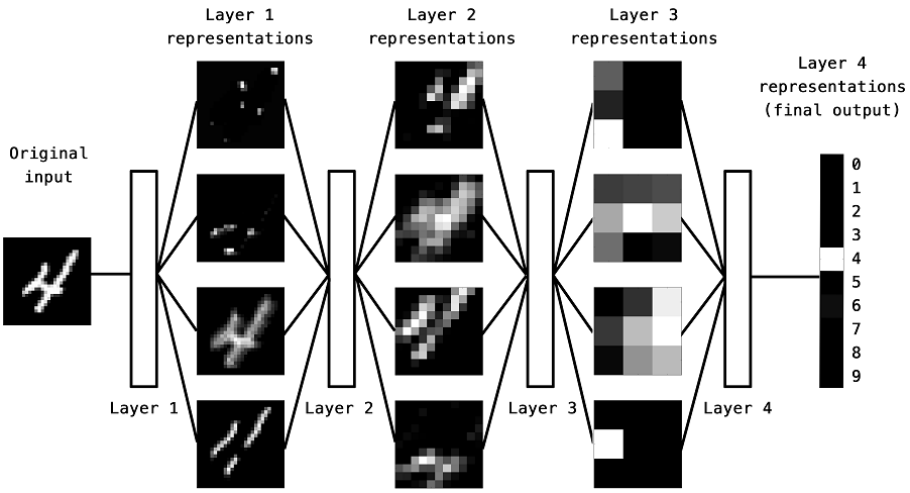


Figure 4.7: Diagram of a three-layer deep network for digit recognition (classification) showing its deep representations. Taken from Chollet (2017).

other. These neural networks are loosely based on the brain's use of layers of neurons working together, although there is no evidence that the brain implements anything like the learning mechanisms used in modern deep learning models. Deep learning is merely a mathematical framework for learning representations from data. Deep neural networks can be used for supervised or unsupervised learning tasks. In a supervised learning context, a deep neural network creates a mapping from inputs X to targets Y by applying a sequence of simple data transformations.

4.4.1 Neural networks

For understanding neural networks, it is important to review the way an artificial neuron or perceptron (Rosenblatt, 1958) works. This computational model receives multiple inputs, which are multiplied by a continuous valued weight. The perceptron then returns a single output, 1 if the weighted sum of the inputs is above some threshold, or 0 if otherwise. The function that determines the output of a neuron is known as the *activation function*. In the case of Rosenblatt's perceptron (see Fig. 4.8), it is the thresholding operation or step function (as shown in the top-left panel of Fig. 4.9). The perceptron has a single layer of weights and can only express linear decision surfaces. Also, with a single output it cannot handle the task of classification into several categories. A function with multiple outputs can be learned by forming a layer with multiple perceptrons. Artificial neural networks can be thought as composition of such layers of perceptrons (also called units).

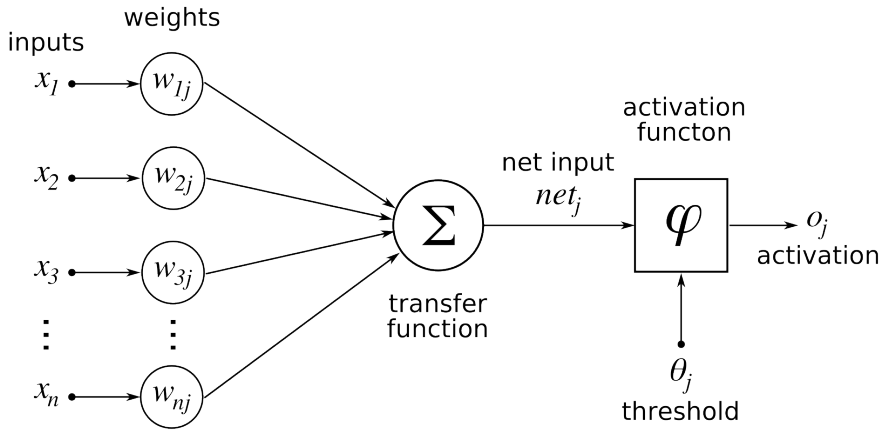


Figure 4.8: Artificial neural network diagram. Credits: Wikimedia commons.

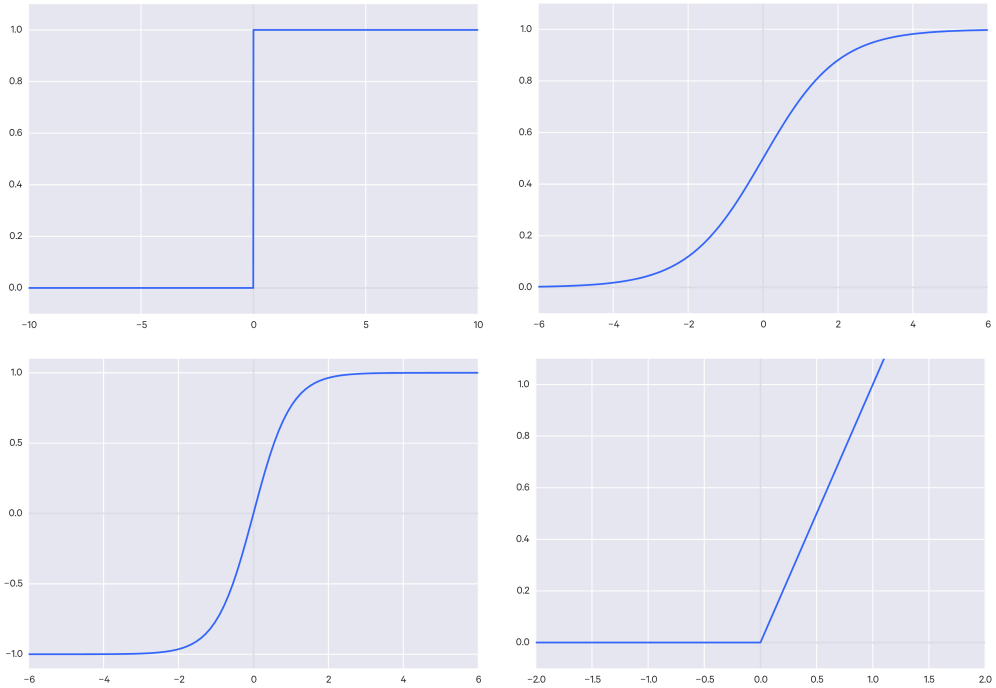


Figure 4.9: Common activation functions. Top-left: Step function. Top-right: Sigmoid function. Bottom-left: Tanh function. Bottom-right: ReLU activation function.

Other activation functions can be used, such as the sigmoid (see top-right panel of Fig. 4.9), hyperbolic tangent (see bottom-left panel of Fig. 4.9) and ReLU (see bottom-right panel of Fig. 4.9) functions. The sigmoid is a smoothed, differentiable threshold function and its output is a continuous function of its input. ReLU is the simplest non-linear activation function (computationally cheap), which works by setting any negative input value to zero, and is the preferred choice for training deep neural networks. The softmax function is commonly used as the output's layer activation function in the case of multiclass classification. The output of this function is positive and sums to one, being equivalent to a categorical probability distribution.

Deep neural networks can be seen as a composition of simple linear data transformations (matrix multiplications and summations), specified by the layer's weights, and non-linearities introduced by the activation functions. Introducing non-linearity extends the kinds of functions that neural networks can represent. The weights or parameters of the network are initialized with random values and the process of adjusting these parameters is the whole goal of training the deep neural network. Learning the weights (typically millions of them) requires to define a loss function, which measures the distance between the network output Y' and the true labels Y. A gradient-based optimization procedure is then used in order to minimize this loss.

4.4.2 Optimization of neural networks

Optimization is the task of minimizing some function $f(x)$ by altering x. Let $f(x)$ be the loss function (also called objective, cost or error function). Consider a function $y = f(x)$ and its derivative $f'(x)$. This derivative gives the slope of $f(x)$ at the point x. Gradient descent optimization consists in reducing $f(x)$ by moving x in small steps with opposite sign of the derivative (see Fig. 4.10 for an example). Critical points are those for which the derivative is zero. A local minimum is a point where $f(x)$ is lower than at all neighboring points, while a global minimum is a point where $f(x)$ obtains its absolute lowest value. In deep learning, optimization of functions with many local minima (not optimal) and saddle points is the rule. Usually, finding a value of f that is very low (not necessarily minimal) is enough. For multidimensional functions, we must use the concept of partial derivatives. The partial derivative $\frac{\partial}{\partial x_i} f(x)$ measures how f changes as only the variable x_i increases at point x. The vector containing all of the partial derivatives of f is called the gradient (derivative with respect to a vector), and is denoted $\nabla_x f(x)$. In this multidimensional context, a critical point is reached when every element of the gradient is zero (Goodfellow et al., 2016).

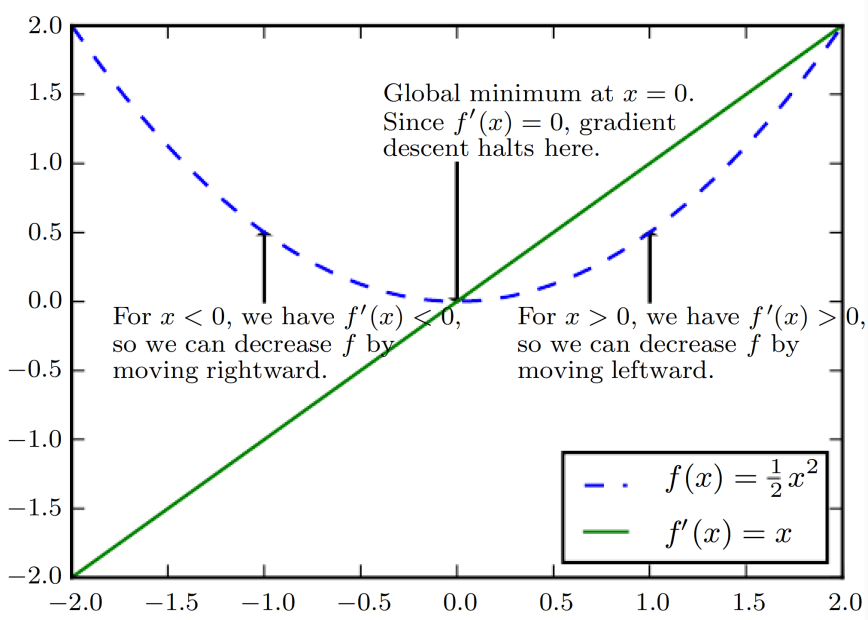


Figure 4.10: Gradient descent technique. Taken from Goodfellow et al. (2016).

The optimization of deep networks, with a large number of parameters, is accomplished with *mini-batch stochastic gradient descent*. It works by drawing a random batch from X and Y, performing a *forward pass* (running it through the network) to obtain Y′, computing the loss score on this batch and the gradient of the loss with regard to the parameters of the network (*backward pass*), and changing the parameters in the direction opposite to the gradient (Chollet, 2017). The aim of this process is to lower the loss on the batch by a small step, also called *learning rate*. The whole process of learning the weights (that minimize the loss) is made possible by the fact that neural networks are chains of differentiable tensor operations. Therefore it is possible to use the *backpropagation* method, by applying the chain rule of derivation to find the gradient function mapping the current parameters and current batch of data to a gradient value. Fig. 4.11 shows a schematic view of a deep neural network with the backward and forward passes.

The the learning rate is important, because if it is too small, the process would take many iterations (and it could get stuck in a local minimum), whereas, if the step is too large, it would never converge. A way to avoid getting trapped in local minima is to use *momentum*, which draws inspiration from physics, and updates the weights based not only on the current gradient value but also based on the previous weight updates. Other methods have been proposed to automatically

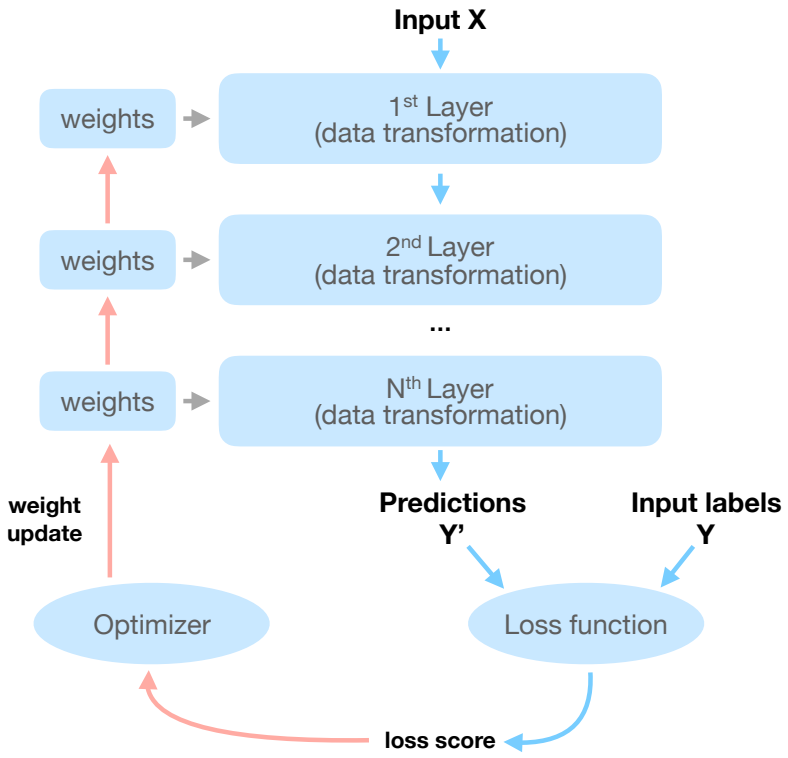


Figure 4.11: Schematic view of the deep neural network training process. The blue and red arrows show the directions of the forward and backward passes correspondingly. Adapted from Chollet (2017).

tune the learning rates, such as Adagrad (Duchi et al., 2011), RMSprop (Dauphin et al., 2015) and Adam (Kingma and Ba, 2014). For an overview of gradient descent optimization algorithms, see Ruder (2016). Training a deep network usually requires thousands of training samples and tens of epochs (passes over the training dataset). The network is trained once the loss score is minimal and the predictions Y' are optimal.

Although mini-batch stochastic gradient descent works very well in practice, it is important to keep in mind that we have little theoretical understanding of the very general non-convex optimization problems involved in deep learning[4] (Goodfellow et al., 2016).

As described in the previous Section, overfitting is a challenge for the generalization capacity of unsupervised learning algorithms. Effective ways of regularizing deep neural networks are dropout (Srivastava et al., 2014), data augmentation techniques, and early stopping. Dropout consists in randomly setting half of the activations (outputs) of a hidden layer to zeros. It acts like a form of model aver-

4 This is one of the topics that according to Yann LeCun, one of the founding fathers of deep learning, are yet to be solved. See his answer on Quora.

aging over all possible instantiations of the model prototype, and it is shown to deliver significant gains in performance in a number of applications. Data augmentation is the process of creating synthetic training data to make a machine learning model generalize better. This is well suited for classification problems, and has been very effective for object recognition tasks. Early stopping is a simple strategy for training large models with optimal representational capacity, without overfitting the task (exploiting the U-shape of the validation loss curve). Other regularization techniques applied to deep learning are weight decay, addition of noise to the weights, parameter sharing and ensemble methods (see Goodfellow et al. (2016) for a complete review of these techniques).

Deep learning has led to breakthroughs in the most challenging areas of machine learning, specially those dealing with perceptual data: near-human image classification and speech recognition, improved machine translation and text-to-speech conversion, autonomous driving, processing natural language queries, amongst other tasks (Chollet, 2017).

4.4.3 Convolutional neural networks

Convolutional neural networks (CNNs, LeCun et al., 1989; Krizhevsky et al., 2012) are a type of deep learning model for processing data that having a grid-like topology (e.g. images), and are almost universally used in computer vision. CNNs can also work with time series and 3D input data. They are called convolutional because of the use of convolutions, linear operations which do not correspond precisely to the definition of convolution as used in other fields such as engineering or pure mathematics (cross-correlation from a signal processing perspective). CNNs use these convolutions instead of general matrix multiplication in at least one of their layers (Goodfellow et al., 2016). See Dumoulin and Visin (2016) for an overview of convolution arithmetic in deep learning. Fig. 4.12 shows an illustration of a 2D convolution. These convolutional layers compose the first stage of a CNN. In a second stage, each linear activation (output of the convolutional layer) is run through a non-linear activation function, such as the rectified linear activation function. In a final stage, a pooling function is used to modify the output of the layer further. The pooling layers provide invariance to small translation of the input signal. *Max pooling* is the most common pooling used and consists in splitting the input in non-overlapping patches and outputting the maximum value of each patch.

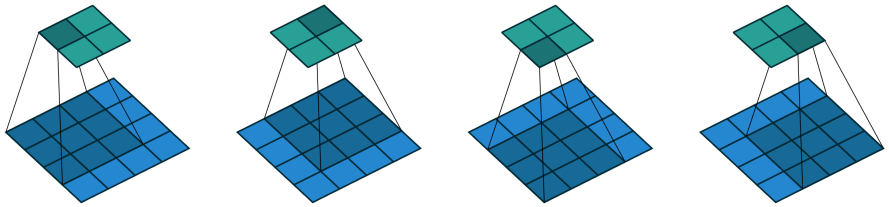


Figure 4.12: Illustration of the convolution of a 3×3 kernel over a 4×4 input using unit strides. Taken from Dumoulin and Visin (2016).

4.4.4 Recurrent neural networks

A powerful specialization of the neural networks framework, recurrent neural networks (RNN, Rumelhart et al., 1986), are designed to process one-dimensional, sequential data. These networks can handle long sequences, which would be impractical for networks without sequence-based specialization. RNNs build on the idea of sharing parameters across different parts of a model. These networks use previous outputs as inputs to form a recurrence, where a time state contains a summary of the past states. The most effective sequence models used in practical applications are called gated RNNs. Long short-term memory (LSTM, Hochreiter and Schmidhuber, 1997) is a type of such recurrent network, that has proven stable and powerful for modeling long-range dependencies. The major innovation of LSTM is its memory cell, which essentially acts as an accumulator of the state information. The cell is accessed, written and cleared by several controlling gates. LSTM networks have broken records for tasks such as improved machine translation, large-vocabulary speech recognition and text-to-speech synthesis.

4.4.5 Deep learning in HCI

Although neural networks have been used in astronomy since the early nineties (Odewahn et al., 1992; Bertin and Arnouts, 1996; Tagliaferri et al., 2003), the use of deep learning has started to spread only in the last couple of years. CNNs are becoming more and more common for image-related tasks, such as galaxy morphology prediction (Dieleman et al., 2015), astronomical image reconstruction (Flamary, 2016), photometric redshift prediction (Hoyle, 2016), and star-galaxy classification (Kim and Brunner, 2017). Other deep neural network architectures such as autoencoders (Frontera-Pons et al., 2017) have been proposed for feature-learning in spectral energy distributions of galaxies. Very recently, generative adversarial networks (Schawinski et al., 2017) were used for image reconstruction and compared to conventional deconvolution techniques. Fig. 4.13 shows an overview of

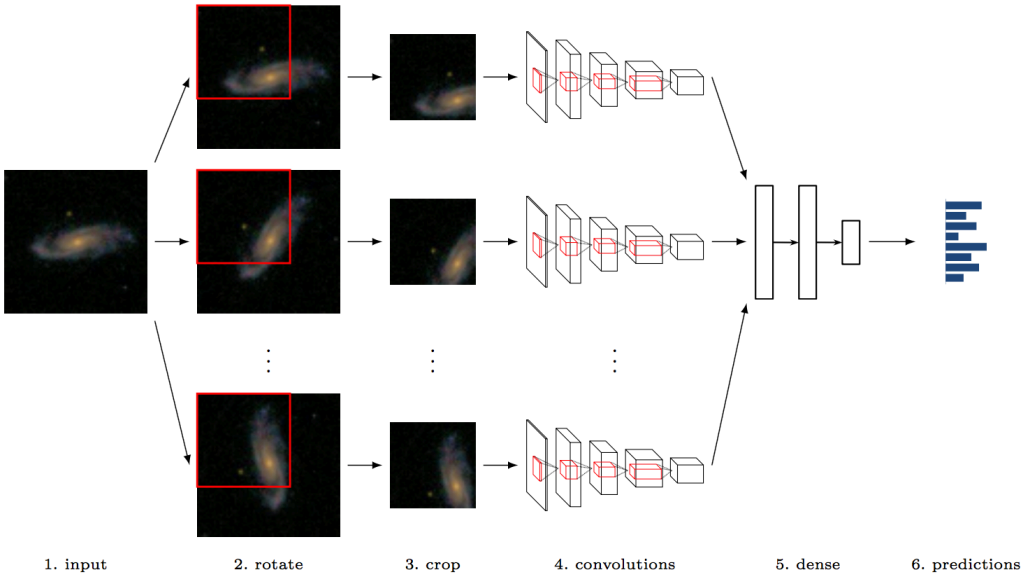


Figure 4.13: Schematic overview of a neural network architecture for galaxy morphology classification, exploiting galaxies rotational symmetry. Taken from Dieleman et al. (2015).

the CNN used by Dieleman et al. (2015) for winning the Galaxy Zoo[5] competition of galaxy morphology classification.

In the context of HCI, no deep learning solutions have been proposed to date. In Chapter 6, we describe a novel supervised learning framework for exoplanet detection in ADI sequences, which applies deep neural networks to HCI for the first time. We also evaluate its performance and compare it to state-of-the-art differential imaging approaches such as ADI-PCA and LLSG.

5 https://www.galaxyzoo.org/

5 LOW-RANK PLUS SPARSE DECOMPOSITIONS APPLIED TO HCI

Contents

ABSTRACT

In this chapter I present my first proposed algorithm for post-processing ADI sequences which builds on unsupervised learning techniques describes previously, such as PCA. Inspired by recent advances in machine learning algorithms such as robust PCA, we propose a localized subspace projection technique that surpasses current PCA-based post-processing algorithms in terms of the detectability of companions at near real-time speed, a quality that will be useful for future direct imaging surveys. The main idea of this novel approach is to use randomized low-rank approximation methods recently proposed in the machine learning literature, coupled with entry-wise thresholding to decompose an ADI image sequence locally into low-rank, sparse, and Gaussian noise components (LLSG). This local three-term decomposition separates the starlight and the associated speckle noise from the planetary signal, which mostly remains in the sparse term. The performance of LLSG was tested on a long ADI sequence obtained on β Pictoris with VLT/NACO. Compared to a standard AID-PCA approach, LLSG decomposition reaches a higher signal-to-noise ratio and has an overall better performance in the receiver operating characteristic space. This three-term decomposition

brings a detectability boost compared to the full-frame standard ADI-PCA approach, especially in the small inner working angle region where complex speckle noise prevents ADI-PCA from discerning true companions from noise. *This chapter is based on previous work published in Gomez Gonzalez et al. (2016a).*

5.1 INTRODUCTION

Plain PCA extracts a lower-dimensional basis that is optimal in the least-squares sense. More recently, alternatives to the least-squares criterion have been proposed in the field of computer vision to consider other objectives, such as sparsity of the noise or robustness to outliers. In this chapter, we propose one implementation of these algorithms applied to ADI image sequences. In our approach we decompose the images locally into low-rank, sparse, and Gaussian noise components to enhance residual speckle noise suppression and improve the detectability of point-like sources in the final combined image.

Throughout this chapter we use $\mathrm{rank}(X)$ to denote the rank of a matrix X, and $\mathrm{card}(X)$ to denote the cardinality (l_0-pseudo norm or number of non-zero elements) of X.

5.2 SUBSPACE PROJECTION AND LOW-RANK PLUS SPARSE DECOMPOSITIONS

The problem of matrix low-rank approximation has been studied extensively in recent years in many different fields, such as natural language processing, bioinformatics, and computer vision. In particular for image analysis, there are multiple tasks that can be achieved using low-rank modeling, such as image compression, denoising, restoration, alignment, face recognition, and background subtraction (or foreground detection in video sequences) (Zhou et al., 2014). The applicability of low-rank approximations is guided by the fact that the latent structure of high-dimensional data usually lies in a low-dimensional subspace. If we consider a sequence of n images and a matrix $M \in \mathbb{R}^{n \times p}$ whose columns are vectorized versions of those images, the above statement can be expressed as $M = L + E$, where L has low rank and E is a small perturbation matrix. An estimate of L is given by a best low-rank approximation of M in the least-squares sense:

$$\min_{L} \|M - L\|_F^2, \text{ subject to } \mathrm{rank}(L) \leqslant k, \tag{5.1}$$

where $\|X\|_F^2 = \sqrt{\sum_{ij} X_{ij}^2}$ denotes the Frobenius norm of a matrix X, and k is the rank of the low-rank approximation L. This can be

solved analytically through SVD (Eckart and Young, 1936; Candès et al., 2009):

$$M = U\Sigma V^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T, \tag{5.2}$$

where the vectors u_i and v_i are the left and right singular vectors, and σ_i the singular values of M. Choosing the first k left singular vectors forms an orthonormal basis for the low-dimensional subspace that captures most of the variance of M. This procedure corresponds to PCA (Hotelling, 1933), as it is usually called in statistics.

In computer vision, for the task of segmentation of an image sequence into background and foreground pixels, PCA was proposed by Oliver et al. (2000), who modeled the background pixels using an eigenspace model. Each image is approximated by its projection onto the first k principal components. They noted that, because they do not appear at the same location in the n sample images and are typically small, moving objects do not make a significant contribution to the PCA model. The foreground pixels are found by subtracting from each image its low-rank PCA approximation and thresholding the pixel values in the residual images.

In astronomy PCA has proven to be effective for modeling time- and position-dependent PSF variations of the Sloan Digital Sky Survey and later for the Advanced Camera for Surveys on the Hubble Space Telescope (see Jee et al., 2007). In the context of reference PSF subtraction for high-contrast imaging, a PCA-based approach has been proposed independently by Soummer et al. (2012) and Amara and Quanz (2012). The problem of modeling and subtracting a reference PSF with the purpose of detecting a moving planet in an ADI image sequence has a lot in common with the segmentation of video sequences into background and foreground pixels (e.g., for video surveillance and detection of moving objects), since the reference PSF and quasi-static speckles can be modeled using a low-rank PCA approximation. The orthogonal basis formed by the first principal components (PCs) is learnt from the ADI sequence itself, which adds complications to the low-rank approximation task because some part of the foreground signal is absorbed in the background model. This relates to the fact that PCA gives a suitable low-rank approximation only when the term E (foreground signal) is small and independent and identically distributed Gaussian (see section 1.1 in Candès et al., 2009). This is unfortunately not the case for moving planets in ADI images.

5.2.1 Robust PCA

In recent computer vision literature, several subspace projection algorithms exploiting the low-rank structure of video sequences have

been proposed to solve the weaknesses of the basic PCA model and provide more versatile and robust background models (Bouwmans and Zahzah, 2014). The most notable is the family of robust PCA (RPCA) algorithms, which model the data as the superposition of low-rank and sparse components, containing the background and the foreground pixels, respectively. One of the first approaches for solving this decomposition was proposed by Candès et al. (2009), with an algorithm called principal component pursuit (PCP). PCP aims to decompose M into low-rank plus sparse ($L + S$) matrices by solving the following problem:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \text{ , subject to } L + S = M, \tag{5.3}$$

where L is low-rank, S contains sparse signal of arbitrarily large magnitude, $\|S\|_1 = \sum_{ij} |S_{ij}|$ is the l_1-norm of S, and $\|L\|_*$ denotes the nuclear norm of L or sum of its singular values. The nuclear norm and the l_1-norm are the convex relaxations of rank(L) and card(S) and provide the best computationally tractable approximation to this problem. Under rather weak assumptions, this convex optimization recovers the low-rank and sparse components that separate the varying background and the foreground outliers. Important limitations of this algorithm are its high computational cost and the assumption that the low-rank component is exactly low-rank, and the sparse component is exactly sparse, contrary to what we find in real data, which is often corrupted by noise affecting a large part of the entries of M (Bouwmans and Zahzah, 2014). In the ideal case, when applying such decomposition to an ADI image sequence, the reference PSF would be captured by the low-rank component and the small moving planets (realizations of the instrumental PSF) by the sparse component. In real ADI coronagraphic images, the reference PSF, composed of the stellar PSF and speckles, is never exactly low-rank owing to the quasi-static component of the speckle noise. Therefore the exact decomposition into low-rank and sparse components does not exist, and the S component recovered by PCP becomes polluted by residual noise from the quasi-static speckles that will produce a final image resembling the results of standard PCA.

5.2.2 GoDec

Several modifications of PCP have been proposed to address its limitations with real data for the problem of foreground detection (see Bouwmans and Zahzah, 2014, for a complete review). Beyond PCP, there are different approaches to RPCA via low-rank plus sparse matrix decomposition.

Among them, GoDec (Go Decomposition, Zhou and Tao, 2011b) is a convenient approach, in terms of computational cost, to the decom-

position of M. It proposes a three-term decomposition (instead of the typical low-rank plus sparse one):

$$M = L + S + G, \text{rank}(L) \leqslant k, \text{card}(S) \leqslant c, \tag{5.4}$$

where G is a dense noise component, and k and c the constraints on the rank of L and the cardinality of S. GoDec produces an approximated decomposition of M, whose exact low-rank plus sparse decomposition does not exist because of additive noise G, restricting rank(L) and card(S) in order to control model complexity. This three-term decomposition can be expressed as the minimization of the decomposition error:

$$\min_{L,S} \|M - L - S\|_F^2, \text{ subject to } \text{rank}(L) \leqslant k, \text{ card}(S) \leqslant c. \tag{5.5}$$

The optimization problem of Eq. (5.5) is tackled by alternatively solving the following two subproblems until convergence, when the decomposition error reaches a small error bound ($=10^{-3}$):

$$\begin{cases} L_t = \underset{\text{rank}(L)\leqslant k}{\text{argmin}} \ \|M - L - S_{t-1}\|_F^2; \\ S_t = \underset{\text{card}(S)\leqslant c}{\text{argmin}} \ \|M - L_t - S\|_F^2. \end{cases} \tag{5.6}$$

In Eq. (5.6), L_t can be updated via singular value hard thresholding of $M - S_{t-1}$ (via SVD in each iteration), and S_t via entry-wise hard thresholding of $M - L_t$. It must be noted that singular value hard thresholding is equivalent to the truncation of the number of PCs in the PCA low-rank approximation.

A randomized and improved version of GoDec was proposed by the same authors with SSGoDec. In this approximated RPCA algorithm, the cardinality constraint is modified by introducing an l_1 regularization, which induces soft-thresholding when updating S (Zhou and Tao, 2013). The soft-thresholding operator $\mathcal{S}_\gamma$ with threshold γ applied to the elements of a matrix X can be expressed as

$$\mathcal{S}_\gamma X = \text{sgn}(X_{ij}) \max\left(|X_{ij}| - \gamma, 0\right). \tag{5.7}$$

The reduced computational cost of SSGoDec mostly comes from using, on each iteration, the bilateral random projections (BRP, Zhou and Tao, 2011a) of M instead of singular value thresholding for its low-rank approximation. BRP is a fast randomized low-rank approximation technique making use of M's left and right random projections, $Y_1 = MA_1$ and $Y_2 = M^T A_2$, where $A_1 \in \mathbb{R}^{n\times k}$ and $A_2 \in \mathbb{R}^{m\times k}$ are random matrices. The rank-k approximation of M is computed as

$$L = Y_1 (A_2^T Y_1)^{-1} Y_2^T. \tag{5.8}$$

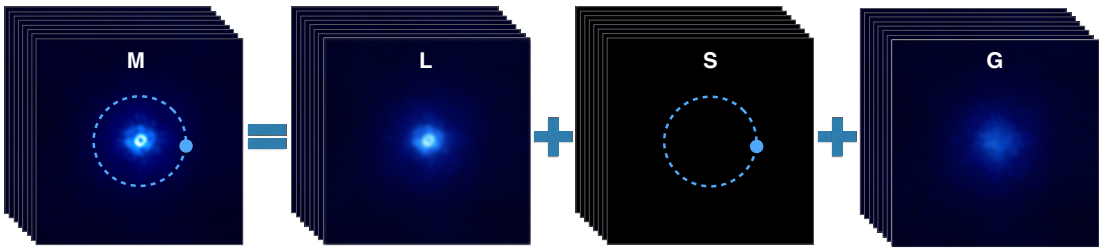


Figure 5.1: LLSG decomposition of ADI data (left) into low-rank (middle left) plus sparse (middle right) plus Gaussian noise (right) terms. In the ideal case, this decomposition separates the reference PSF and quasi-static speckle field from the signal of the moving planets, which stays in the sparse component.

The computation of this approximated L requires less floating-point operations than the SVD-based approximation. The bounds of the approximation error in BRP are close to the error of the SVD approximation under mild conditions (Zhou and Tao, 2011a).

5.3 LOCAL LOW-RANK PLUS SPARSE DECOMPOSITION OF ADI DATASETS

Restricting the cardinality of $M - L_t$ while operating on whole images is problematic in the presence of multiple companions, as the dimmest one could get severely subtracted from the data (especially for close-in companions), or bright speckles could turn into false positives. We find that applying a local three-term decomposition, which exploits the geometrical structure of ADI image sequences, can alleviate the problem of a global thresholding and in addition provides a better low-rank approximation for the given patch.

These ideas were put together to build an ADI post-processing algorithm for boosting point-like source detection, the Local Low-rank plus Sparse plus Gaussian-noise decomposition (LLSG). A schematic illustration of this decomposition is shown in Fig. 5.1. The algorithm follows four main steps:

1. the images of the cube are broken into patches, specifically in quadrants of annuli of width $2\lambda/D$ (see Fig. 5.2);
2. each of these quadrants is decomposed separately as in Eq. (5.6), alternatively updating its L and S components for a fixed number of iterations;
3. for each patch, the S component of the decomposition is kept;

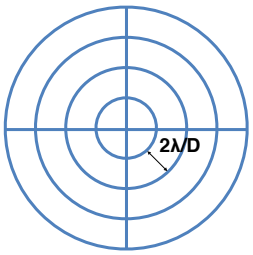


Figure 5.2: Quadrants of annuli used for partitioning the images for LLSG.

4. all frames are rotated to a common north and are median combined in a final image.

The soft-thresholding will enforce the sparsity of the S component throughout the iterations. Instead of using a common threshold parameter γ, we use a different one for each patch. Values of γ that are too high will remove the signal of companions too much along with the residual speckle noise, while values that are too low will not lead to much improvement over ADI-PCA processing, therefore hindering the detection of potential very faint companions. Instead of leaving this as a free parameter, our algorithm defines γ for each patch as the square root of the median absolute deviation of the entries in S_t. This thresholding can be scaled up or down safely by the user with a tunable parameter in case it is needed. Partitioning the frames using quadrants of annuli does not increase the computational cost and alleviates the problem of applying entry-wise soft thresholding globally (on the whole frames), thereby giving better results in the case of several companions with different brightnesses or when very bright speckles are present.

Among the free parameters of LLSG, the rank (low integer value) is certainly the most important one. This parameter is equivalent to the number of PCs in the ADI-PCA algorithm and defines the size of the low-rank approximation of our dataset (L term). Values of the rank that are too high cause too much planetary signal to be absorbed by the low-rank term, whereas a low value produces a noisier sparse term. The sweet spot depends on data. The sparsity level (for scaling the soft-thresholding, by default is equal to one) is the second parameter of LLSG, which controls how sparse the S term is and how much noise goes into the G term. It usually does not require user intervention since it is internally defined for each image patch. The third parameter of LLSG is the number of iterations. A small number of iterations is enough (the default is ten) to achieve good decomposition according to our tests on several datasets, but it can be fine-tuned by the user. The number of iterations affects the running time of LLSG

and generally by doubling the number of iterations we double the computation time.

As explained before, the sparse component is the main product of this algorithm, where potential companions will have a higher signal-to-noise ratio (S/N). The two other components in this decomposition, the low-rank and the noise, can serve as estimates of the total noise of our data. The L component will contain the starlight and most of the static and quasi-static structures, while G will capture the small and dense residual noise that was not captured by the low-rank approximation. An implementation of LLSG for ADI data is provided in the `VIP` library (see Chapter 2).

This idea of a three-term decomposition with some modifications, e.g. a different partitioning of the images, could be used as well for spectrally dispersed data obtained with an integral field spectrograph (IFS). After rescaling IFS data, the companions will appear to move radially through the speckle noise field. Therefore, LLSG can be a good choice for decomposing the image sequence and capturing potential planets in the sparse term.

5.4 APPLICATION TO REAL DATA

5.4.1 Data used

The application of the LLSG decomposition to real data gives a first taste of its capabilities. In this Chapter, we use the data set of β Pic and its planetary companion β Pic b (Lagrange et al., 2010) obtained on January 2013 with VLT/NACO in its AGPM coronagraphic mode (Absil et al., 2013). The observations made at L′ band were performed under poor weather conditions, nevertheless β Pic b could be seen on the real-time display thanks to the excellent peak starlight extinction provided by the AGPM. The total on-source integration time was 114 min with a parallactic angle ranging from $-15°$ to $68°$. A clean cube was obtained after basic preprocessing steps, such as flat fielding, bad pixel removal, bad frames removal, recentering of frames, and sky subtraction. After temporal subsampling, by averaging 40 successive frames, a new cube of 612 individual frames with 8 sec of effective integration time was created (for details see Absil et al., 2013). As a final step the central 161x161 pixels were cropped on each frame.

5.4.2 Results

Figure 5.3 shows the final post-processed frames using full-frame ADI-PCA and the three terms of the LLSG decomposition. We clearly see how LLSG can separate the starlight and quasi-static speckles from the planetary signal. The sparse term is where most of the sig-

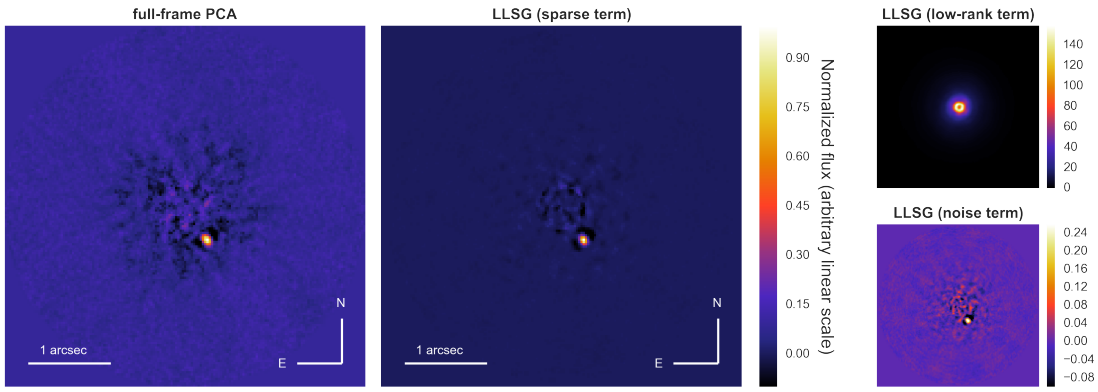


Figure 5.3: Final result of post-processing with ADI-PCA (left) and the three terms of the LLSG decomposition (middle and right) for β Pic NACO data. 38 PCs were used to maximize the S/N for β Pic b in the full-frame ADI-PCA approach. For the LLSG decomposition a rank of 10 gave a significant improvement on b's S/N (see Fig. 5.4). All the frames were normalized to the maximum value of the LLSG sparse frame (middle panel).

nal of β Pic b is present. In the following discussion, we use the S/N between the planet signal and the background pixels to compare the performance of the two algorithms for the task of detection of point-like sources.

For details on the S/N definition, see Section 2.3 of Chapter 2. The use of the S/N as a metric for comparing algorithms can become problematic when, in some cases, the noise can be almost totally suppressed, making the S/N infinite. In this scenario, if a companion is present, a clear detection through visual vetting can be claimed. We have encountered this situation when processing other datasets of better quality (conditions of observation and/or better wavefront sensing). We also note that, the S/N of a point-like source depends on the choice of the aperture sizes and on the position of the apertures themselves, especially at small angular separations where the small sample statistics effect becomes dominant (Quanz et al., 2015). Throughout this Chapter we use an aperture size of 4.6 pixels, which is the Gaussian FWHM measured on the off-axis PSF of β Pic.

The positioning of the apertures is done in an automatic way and is the same for each realization, when measuring S/N on the final frames of the compared algorithms. As an example, we only have 24 background apertures (n_2) for the case of β Pic b (using the FWHM as an approximation for the value of the λ/D parameter). In spite of these limitations, we stick to the use of the S/N for its practicality for the task of detecting point-like sources.

Figure 5.4 shows the S/N maps corresponding to full-frame ADI-PCA and the sparse term of the LLSG algorithm. With LLSG the S/N

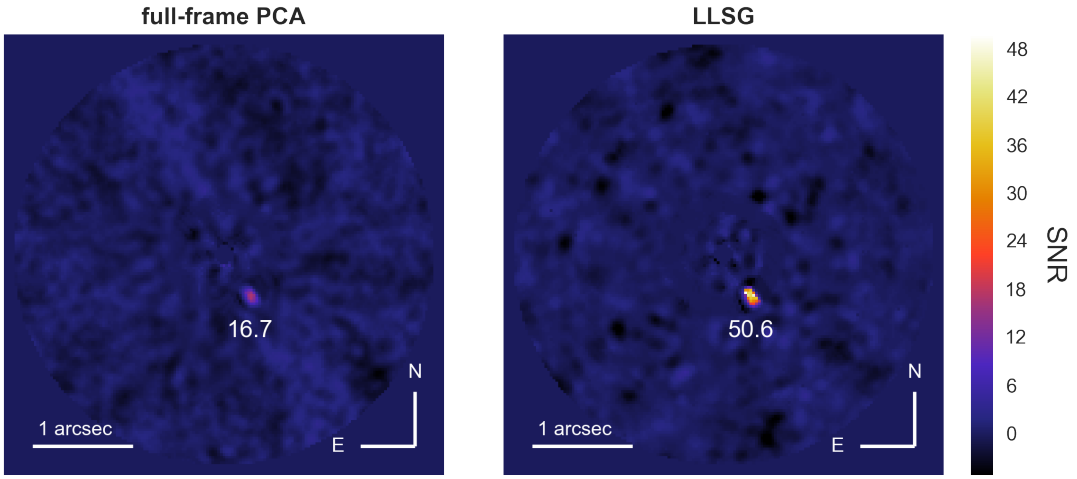


Figure 5.4: S/N maps for ADI-PCA (left) and LLSG (right) computed from the final frames shown in Fig. 5.3 (left and middle panels). Planet β Pic b has roughly three times higher S/N when processed with our LLSG decomposition.

of β Pic b is roughly three times higher than with full-frame ADI-PCA thanks to the small amount of residual noise in S. To maximize the S/N with full-frame ADI-PCA, we varied the number of PCs in the interval from 1 to 100, measuring at every step the S/N at the location of β Pic b. The highest S/N (=16.7) was achieved with 38 PCs. In the case of LLSG, the best compromise between residual noise subtraction and companion signal recovery was obtained with a rank equal to 10. The default number of iterations worked well for this ADI sequence. To validate if the increased S/N translates in the ability to detect fainter planets, we perform a ROC analysis comparing LLSG to full-frame ADI-PCA (see Section 5.5.2 and Section 6.7).

As we can see in Fig. 5.3, roughly 25% of the planetary signal leaks into the LLSG noise term. However, this is less than the amount of companion signal absorbed in the PCA low-rank approximation, when using the 38 PCs that maximized the S/N of β Pic b. In this case, the leaking into the G term does not hinder the goal of LLSG for improving the detectability of a point-like source. In the following section, we test whether this holds true for more complicated scenarios with fainter companions. Nevertheless, the ultimate goal is to avoid any signal loss. This will be the subject of future work.

5.5 SIMULATIONS WITH SYNTHETIC COMPANIONS

5.5.1 Single test case

The use of on-sky data with simulated companions allows us to probe the performance of the detection algorithms with planets at different locations on the image plane and with varying brightness. This enables us to test how LLSG deals with a fainter and closer-in companion than β Pic b, which presented a rather easy scenario. To obtain a data cube without any companion, β Pic b was subtracted using the negative fake companion technique (Lagrange et al., 2010; Absil et al., 2013), which uses the off-axis PSF as a template to remove the planet from each frame by optimizing the position and flux of the injected negative candidate (see Section 2.7). This optimization is performed by minimizing the sum of the absolute values of the pixels in a 4×FWHM aperture on the ADI-PCA processed final frame. We used the downhill simplex minimization algorithm (as implemented in VIP) for this purpose, which is enough to obtain a planet-free cube.

In the empty cube, we injected the normalized off-axis PSF to create four synthetic companions and compared the results of full-frame ADI-PCA and our approach (see Figs. 5.5 and 5.6). The companions a(285°), b(185°), c(5°), and d(85°) were injected at 2, 5, 8, and 13 λ/D from the center, respectively. The brightness of the fake companions was scaled as a function of the local noise before injection. The noise was measured as the standard deviation of the fluxes of the resolution elements inside the corresponding annulus in the classical ADI processed frame (which means it has been median-subtracted, derotated, and median-combined). The injected PSFs were scaled at 0.5, 5, 5, and 7 times the noise of the respective annulus.

In this particular example of processing a cube with several injected synthetic companions, we encounter a first problem with full-frame ADI-PCA: it is not possible to optimize the S/N for each individual companion at the same time by adjusting the number of PCs used. For the innermost injected planet, we need to use 13 PCs for full-frame ADI-PCA in order to reduce the residual speckle noise and achieve the best possible S/N. As done previously, the number of PCs was varied from 1 to 40, each time measuring the S/N at the location where the innermost planet was injected. This number of PCs may not be optimal for farther companions, which could achieve higher S/Ns with a smaller number of PCs. The optimal number of PCs in general decreases when the planet is farther away from the star in the photon-noise limited regime, since the planets have more rotation and the speckle noise is not dominant.

A better strategy in this case is to use the ADI-PCA low-rank approximation annulus-wise (see middle panel in Fig. 5.5). In this case, it is even possible to apply a frame-rejection criterion based on a par-

allactic angle threshold (Absil et al., 2013). The motivation behind this is that an annular ADI-PCA low-rank approximation will capture the background and speckle noise in a better way for a given patch. Furthermore, keeping only the frames where the planet has rotated by at least 1λ/D in our ADI-PCA reference library, we prevent the planetary signal from being captured by the low-rank approximation and subsequently subtracted from the science images, thereby increasing the S/N in the final frame. We provide a parallelized implementation of this algorithm in the VIP repository. For the innermost planet, located at 2λ/D, we can obtain a maximum S/N of 3.2 after optimizing the number of PCs (by testing from 15 to 35 PCs) and using 2λ/D wide annuli. For LLSG we kept the same rank as we used before and slightly reduced the sparsity level to achieve the highest S/N for the innermost fake companion.

As seen in the S/N maps shown in Fig. 5.6, our LLSG algorithm provides a gain of a factor three in S/N at 2λ/D with respect to full-frame ADI-PCA, resulting in a clear detection instead of the false negative in the case of full-frame ADI-PCA (even after careful optimization of the ADI-PCA truncation and knowledge of the coordinates of the planet). For the three other synthetic companions, located farther from the star, the S/N becomes between two and three times higher compared to full-frame ADI-PCA. The annular version of ADI-PCA does not provide much improvement over full-frame ADI-PCA in the small inner working angle region, even for an ADI sequence that has a large parallactic angle rotation range (~80°) and after adjusting the number of PCs. In this single simulation, we show some practical disadvantages of a full-frame ADI-PCA and the gain in S/N we obtain by using local versions of ADI-PCA and the proposed three-term decomposition.

5.5.2 Performance

Of course, based on a single realization, we cannot characterize the detection performance of the algorithms. More exhaustive approaches are needed, such as the use of receiver operating characteristic (ROC) curves. The performance of a detection algorithm is quantified using ROC analysis, and several meaningful figures of merit can be derived from it (Barrett et al., 2006; Lawson et al., 2012). ROC and localization ROC curves are widely used tools in statistics and machine learning for visualizing the performance of a binary classifier system in a true positive rate (TPR) - false positive rate (FPR) plot as a decision threshold τ varies. The ultimate goal of high-contrast imaging, as for any signal detection application, is to maximize the TPR while minimizing the FPR, which can be achieved by maximizing the area under the curve (AUC) in the ROC space (Mawet et al., 2014). In general the goal of a classifier in the ROC space is to be as close as possible

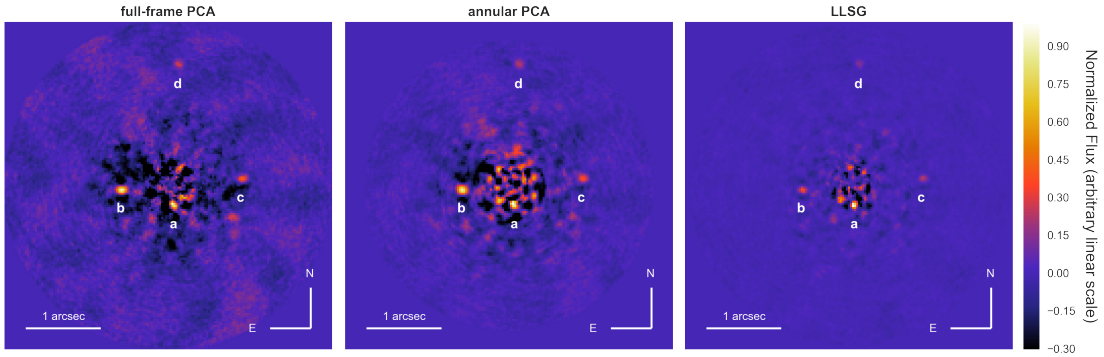


Figure 5.5: Final result of post-processing with ADI-PCA (left), annular ADI-PCA (middle), and our LLSG decomposition (right). The images were normalized to their own maximum value. Four synthetic companions a(285°), b(185°), c(5°), and d(85°) were injected at 2, 5, 8, and 13 λ/D from the center, respectively. The injected PSFs were scaled at 0.5, 5, 5, and 7 times the noise of the respective annulus. We used 13 PCs when applying full-frame ADI-PCA and 25 PCs for the annular ADI-PCA (applying the same number of PCs in every annulus) in order to maximize the S/N of the innermost fake companion in each case.

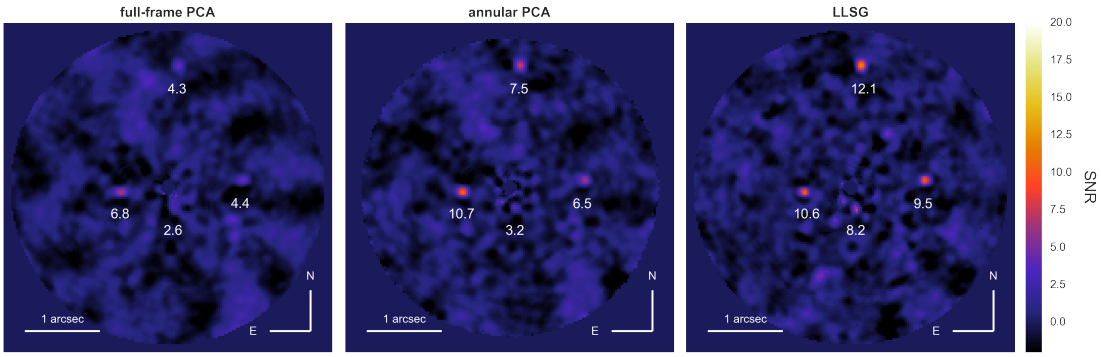


Figure 5.6: S/N maps for full-frame ADI-PCA (left), annular ADI-PCA (middle), and our LLSG decomposition (right) showing the values of each fake companion S/N. With our algorithm the four injected companions are clearly revealed. The S/N of the fake companion at 2λ/D is clearly at the level of the noise (false negative) in the case of full-frame ADI-PCA and its annular version. With LLSG we reach a peak S/N that is three times higher. For the rest of synthetic companions, the S/N obtained with LLSG is up to three times higher than the one obtained with full-frame ADI-PCA.

to the upper lefthand corner (perfect classifier, referred to as "oracle" with FPR=0 and TPR=1) and away from the random classifier line (TPR=FPR).

Building this plot for any exoplanet detection algorithm requires a large number of fake companion simulations, especially if a single planet is injected for each realization as in our case. One hundred realizations were made per annulus, centered at 2, 4, 6, 8, 10, and 14 λ/D, in a random way, meaning that there is a 50% chance that the datacube contains a synthetic planet. In the case of an injection, the fake companion has a random azimuthal position in the given annulus and a random brightness, scaling the PSF (ranging from 0.5 times to 5 times the noise in the considered annulus) as described previously. We consider localization ROC curves for which the decision of whether there is a planet or not is tied to a given position in the image plane.

The detection decision is based on comparing the value of the peak S/N of a given resolution element with a threshold τ. We call the peak S/N here the maximum S/N value obtained from shifting the center of the test resolution element inside a λ/D circular aperture centered on the considered coordinates. This is equivalent to taking the maximum S/N value in a λ/D circular aperture, centered on the considered coordinates, from an S/N map. We find this is in practice better than using the S/N of a resolution element centered on some given injection coordinates, because the maximum S/N for a point-like source (blob) will usually be shifted by a small amount because of post-processing. A true positive (TP) means that, in the case of an injection, the tested resolution element has a peak S/N $\geqslant \tau$. A false positive (FP) arises in case of a non-injection, when a random resolution element inside the considered annulus has a peak S/N $\geqslant \tau$. It is important to notice that we inspect only one resolution element each time instead of the total number of resolution elements in the image (even for the FP counts) to preserve the 50-50 prior we described previously. We vary τ from 0 to 8 in steps of 0.1 in order to have enough points in our empirical ROC curve.

The TPR and FPR for these ROC curves are the averaged TPR and FPR over all brightnesses and the tested annuli. The ROC curves are shown in Fig. 5.7. It is important to emphasize that every point, for every τ, of the LLSG decomposition ROC curve is higher than the one for ADI-PCA, which means that the LLSG detection algorithm is closer to the perfect classifier. The full range of values of FPR (up to one) in our ROC curves is not fully covered even when testing unrealistically low values of τ. In this case calculating the AUC becomes problematic, and using other metrics derived from a ROC curve becomes more suitable for comparing algorithms (classifiers). An example of such a metric is the Euclidean distance to the upper lefthand corner or "oracle" (Braham et al., 2014). In the case of ADI-PCA, the

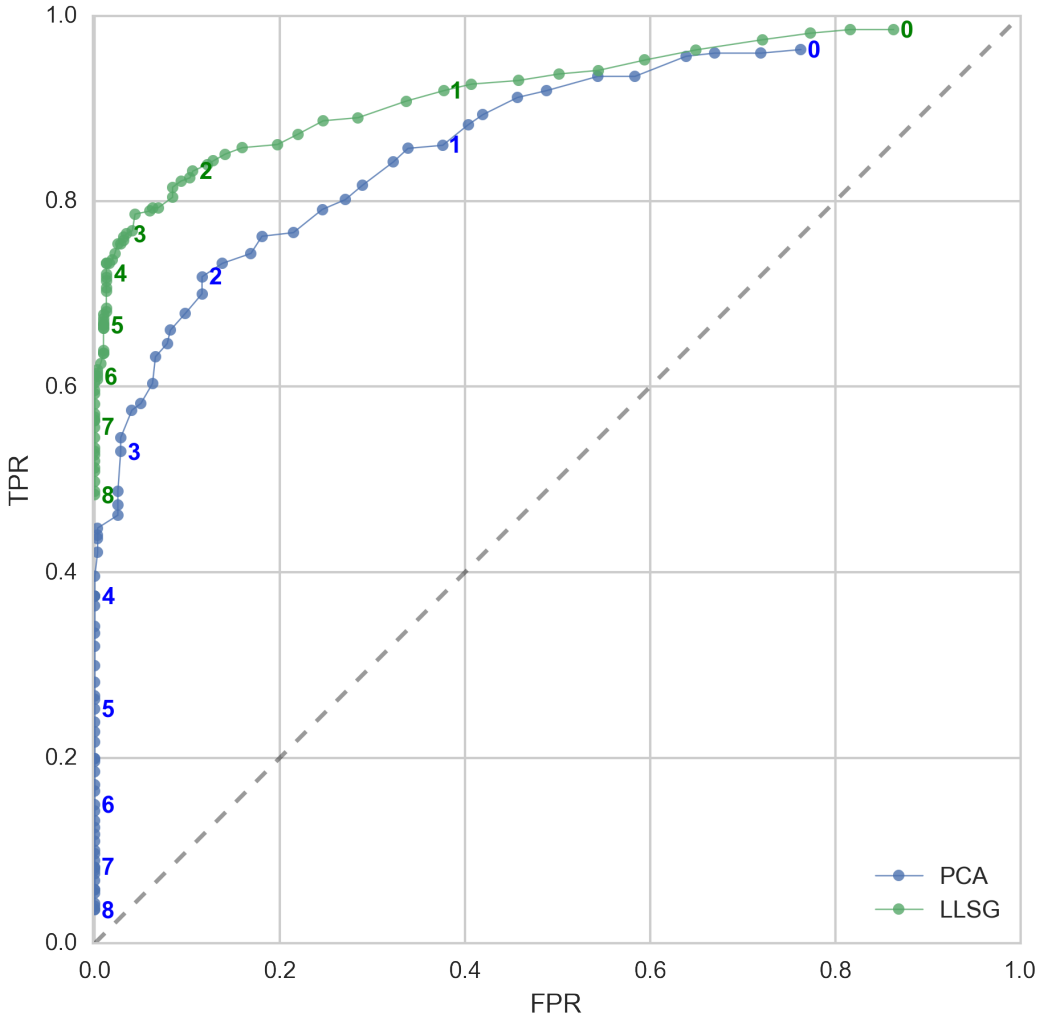


Figure 5.7: ROC curves for our LLSG decomposition and full-frame ADI-PCA. The S/N thresholds τ are shown for integer values. Our algorithm ROC curve is close to the oracle (perfect classifier) in the upper left corner.

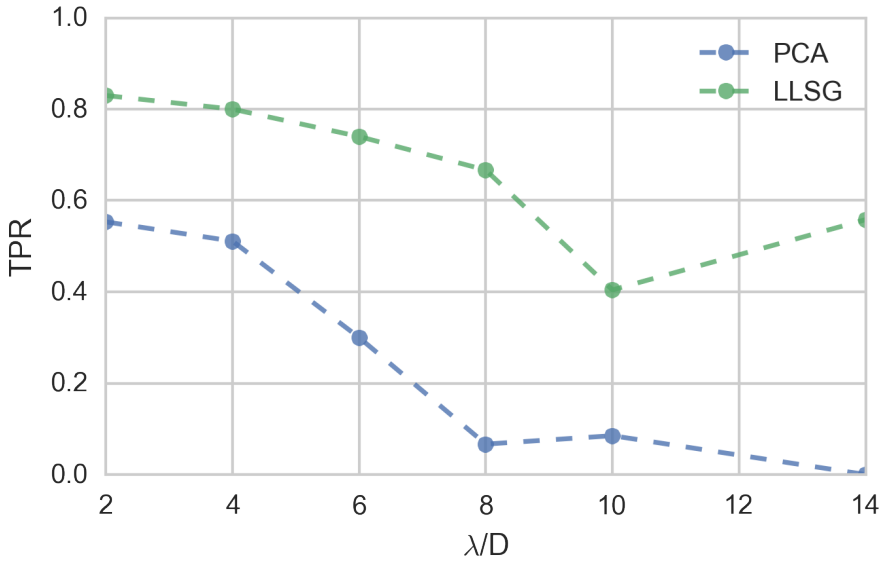


Figure 5.8: TPR as a function of the distance from the star for an S/N threshold $\tau = 5$.

minimum distance to the upper lefthand corner is 0.3, while for our algorithm it is 0.2, which again confirms its superiority.

For generating these ROC curves, we used fifteen PCs, which corresponds to 90% of the explained variance of M, a common approach for choosing the number of PCs for ADI-PCA in machine learning and statistics. The rank of the three-term decomposition was set to fifteen, and the number of iterations was set to ten. Tuning parameters instead of having them fixed for all the realizations could lead to minor improvements in the ROC curves. Tuning the parameters would also increase the complexity in the procedure of generating the ROC curves and would in general be a less fair approach.

The TPR or sensitivity is generally a more relevant measure than the FPR (which is generally fixed for a given S/N threshold), especially for surveys and for obtaining planet population constraints (Mawet et al., 2014). Therefore it is important to evaluate the TPR as a function of the distance from the star for an S/N threshold of 5, which is equivalent to 5-σ under the assumption of nearly Gaussian residuals in the final images. The TPR for both algorithms for the tested annuli and $\tau = 5$ are shown in Fig. 5.8. The TPR for the LLSG decomposition is higher for each one of the tested annuli compared to ADI-PCA. It is especially interesting how at $2\lambda/D$, where the speckle noise is dominant, the TPR for our algorithm reaches 83% instead of the 55% achieved by ADI-PCA.

Another great advantage of the LLSG decomposition over more expensive algorithms is that its computational cost is comparable to that of full-frame ADI-PCA. For instance, it can process the $612 \times 161 \times$

161 ($\sim 15.8 \times 10^6$) pixel datacube used in our simulations in about ten seconds (when using only one process), whereas full-frame ADI-PCA (equivalent to KLIP or pynpoint implementations) using the LAPACK optimized multithreading library can do it in four seconds. This timing depends, as explained before, on the number of iterations for the three-term decomposition.

It is important to clarify that LLSG is an algorithm for improving detection of faint exoplanets, which decomposes the images, separating the static and quasi-static structures from the moving planets. This process penalizes the signal of the potential companions, and in consequence the final LLSG frames cannot be used for estimating in a robust way the position or flux of those companions. We still need to rely on the injection of negative companion candidates, as we described in a previous section, to calibrate the photometry and astrometry of potential detections, as well as their uncertainties.

In the case of ADI data, the range of rotation (parallactic angles) affects the efficiency of post-processing algorithms when searching for potential companions. With small rotation, the signal of a planet remains more static through the sequence of frames (this effect gets worse in the innermost part of the frames), and a low-rank approximation based algorithm will fail to retrieve it. This effect combines with other factors, such as the number of frames and the PSF decorrelation rate during the sequence, and will limit different post-processing algorithms in different ways. Better understanding of the correlation between these various factors will be useful for choosing algorithms and for designing optimal observing runs.

5.6 CONCLUSIONS

In this chapter we have shown, for the first time, how recent subspace projection techniques and robust subspace models proposed in the computer vision literature can be applied to ADI high-contrast image sequences. In particular our implementation of a randomized low rank-approximation recently proposed in the machine learning literature coupled with entry-wise thresholding allowed us to decompose an ADI image sequence locally into low-rank, sparse, and noise components. LLSG brings a detectability boost compared to full-frame ADI-PCA approach at all positions of the field of view as can be seen in the ROC curves with averaged TPR and FPR, and in the plot of the TPR as a function of distance. This is especially important because it allows us to access the small inner working angle region ($\sim 2\lambda/D$ for this dataset), where complex speckle noise prevents ADI-PCA from finding faint companions.

One important advantage of this algorithm is that it can process a typical $612 \times 161 \times 161$ pixel cube without sacrificing too much of

the computational cost compared to the fast full-frame ADI-PCA approach. That the patches can be processed separately leads to real-time processing if coupled with parallelism to exploit modern multi-core architectures, making this algorithm suitable for coming survey pipelines.

We have shown the potential of low-rank plus sparse decompositions and, in particular, the LLSG decomposition for high-contrast imaging. More expensive formulations of these decompositions coupled with a fine-tuned model of the noise could lead to even better reference PSF subtraction for exoplanet detection than the one we proposed in the present Chapter and will be the focus of future work.

6 SUPERVISED DETECTION OF EXOPLANETS

Contents

Abstract

Post-processing algorithms play a key role in pushing the detection limits of high-contrast imaging (HCI) instruments. Differential imaging post-processing approaches enable the production of science-ready images relying on unsupervised learning techniques, such as low-rank approximations. In this chapter we present SODIRF and SODINN, two new approaches for detecting exoplanets in angular differential imaging sequences. Both algorithms share a novel paradigm that enables the reformulation of HCI post-processing as a supervised learning problem building on well-established machine learning techniques. We show, through receiver operating characteristic curves analysis, the improved performance of both SODIRF and SODINN compared to state-of-the-art ADI post-processing algorithms. In particular, from one to two λ/D, SODINN improves the TPR by a factor of $\sim$3.5 with respect to ADI-PCA when working at the same false positive level. SODINN brings the possibility of re-processing existing HCI databases to maximize their scientific return and improve the demographics of directly imaged exoplanets. *This Chapter is largely based on a paper submitted for publication in A&A.*

6.1 INTRODUCTION

In the last decade, direct imaging of exoplanets has become a reality thanks to advances in optimized wavefront control, specialized coronagraphs, innovative observing techniques and dedicated post-processing algorithms (Bowler, 2016; Milli et al., 2016). The amount of available archival high-contrast imaging (HCI) data has increased rapidly with the advent of second generation instruments (e.g., VLT/-SPHERE, GPI). However, the HCI community has slowly adopted the latest developments in data management and machine learning, compared to fields such as computer vision, biology, and medical sciences.

The computational power and data storage increase in the last decade has enabled the emergence of data-driven discovery methods in sciences (Ball and Brunner, 2010), in parallel to the popularization of machine learning and data science fields of study. Data-driven models are especially important in HCI, if we consider the sheer amounts of data that modern high-contrast imaging instruments are producing. Learning artificial neural networks is an algorithmic approach proposed a few decades ago in the machine learning community, inspired by our understanding of the biology and structure of the brain. Only recently, with graphics processing unit (GPU) computing going mainstream, larger amounts of data, and the use of deep architectures (with increased number of layers and neurons), deep learning has produced impressive results across a wide range of applications in computer vision and language understanding (Goodfellow et al., 2016). Deep learning has allowed to re-examine many computer vision tasks by removing the necessity of hand-crafted features, beating every single performance metric (Xie et al., 2017). Deep learning techniques are also gaining popularity for tackling a variety of astronomical problems (Dieleman et al., 2015; Schawinski et al., 2017).

6.2 STATE-OF-THE-ART IMAGE PROCESSING TECHNIQUES FOR HCI

The HCI planet hunter pipeline includes the generation of a science-ready final image, where potential exoplanets are flagged by visual inspection aided with by the computation of signal-to-noise (S/N) maps[1]. In the case of angular differential imaging (ADI, Marois et al., 2006) data, the generation of a final image relies on differential imaging techniques. The purpose of these reference PSF subtraction techniques is to reduce the image dynamic range, by modeling and subtracting the contribution from the high-flux pixels belonging to the

1 We use the definition of S/N by Mawet et al. (2014), as explained in Section 2.3.

leaked starlight and from the quasi-static speckle noise. This results in residual final images where, unfortunately, part of the companion signal is fitted in the reference PSF and subtracted (this is called companion self-subtraction in HCI). Among the reference PSF subtraction techniques we count LOCI (Lafrenière et al., 2007), principal component analysis (PCA) based algorithms (Soummer et al., 2012; Amara and Quanz, 2012), and LLSG (Gomez Gonzalez et al., 2016a). All these approaches use different types of low-rank approximation to model the reference PSF. A slightly different approach is taken by ANDROMEDA (Mugnier et al., 2009; Cantalloube et al., 2015), which uses maximum likelihood estimation on residual images obtained by pairwise subtraction within the ADI sequence.

The exoplanet detection problem is critical, as it triggers the subsequent steps such as the characterization (determination of position and flux) of detected companions. As we pointed out earlier, the detection of potential companions lacks automation. The detectability of companions by visual inspection is limited by human perception biases and the procedure of computing S/N maps is upper bounded by the performance of the reference PSF subtraction techniques. Moreover, the S/N metric does not deal with the truthfulness of potential blobs. Other procedures for detection of companions such as the Laplacian of Gaussian (see Section 2.6) or matched filtering (Ruffio et al., 2017) suffer from the same problem. For a review of general purpose source detection techniques on astronomical images, see Masias et al. (2012). More robust approaches with higher sensitivities to dim companions are needed. The ultimate goal is to produce a per-pixel likelihood or probability of companion presence for each ADI sequence. A step in this direction is ANDROMEDA but its performance has not been thoroughly measured (contrast curves show its performance to be at the same level as full-frame ADI-PCA). A promising approach is the use of discriminative models, as it has been proposed for the case of multiple-channel SDI data by Fergus et al. (2014): the DS4 Detect algorithm, an extension of the S4 algorithm (see Section 4.3.2). Unfortunately, there is no study describing the details of this algorithm or assessing its performance.

In the following sections we describe a machine learning method for detecting exoplanets in ADI datasets using a supervised learning framework and deep neural networks. Our approach brings a great improvement in terms of sensitivity and of false positive rate.

6.3 FROM UNSUPERVISED TO SUPERVISED LEARNING

Differential imaging post-processing approaches rely on unsupervised learning techniques, such as low-rank approximations, to enable the

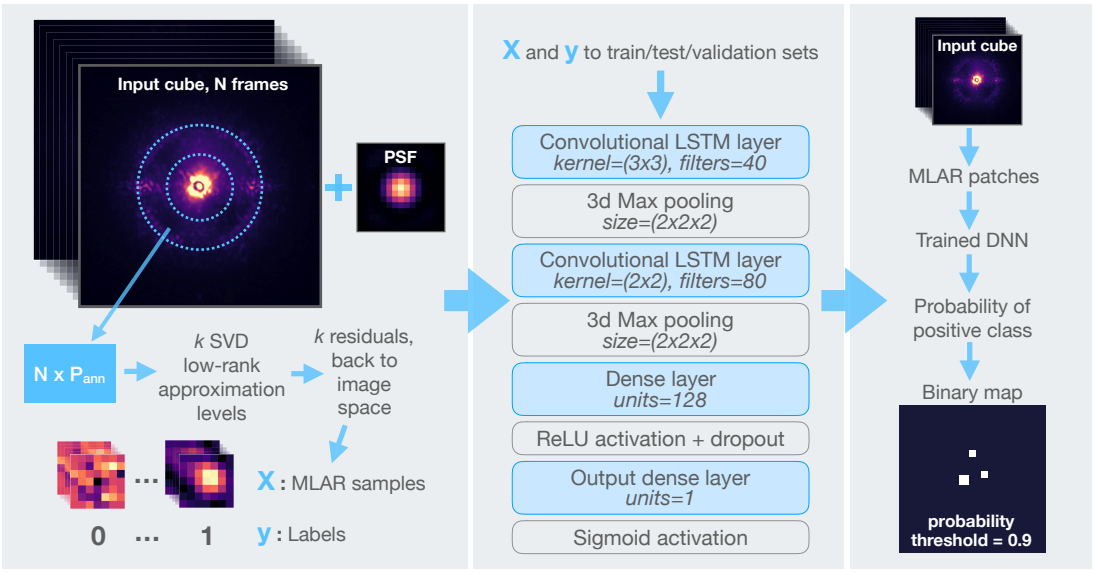


Figure 6.1: The three stages of SODINN's framework. Panel (a) illustrates the labeled data generation step. The ADI sequence and off-axis PSF template are examples of VLT/SPHERE data. Panel (b) illustrates the model training step. Panel (c) concerns the evaluation of the trained model on the original cube and shows the schematic representation of SODINN's output detection map.

production of final residual images. The detection ability of these techniques, which is directly connected to the true positive rate for a given ADI dataset, depends on a variety of factors: the number of frames in the sequence, the total range of field rotation, the distance of a companion to its parent star, and the aggressiveness of the differential imaging subtraction approach.

Our approach consists in a reformulation of the exoplanet detection task as a supervised learning problem. Supervised learning requires labeled data (or ground truth) in order to train a discriminative model that can make predictions. Depending on the model used, two algorithms are proposed: SODIRF – Supervised exOplanet detection via Direct Imaging with Random Forests – and SODINN – Supervised exOplanet detection via Direct Imaging with deep Neural Networks.

The first stage, and challenge, is the generation of a labeled dataset from a single ADI image sequence. As we show in Chapter 6.4 this procedure relies on a technique called data augmentation, which is widely used for training deep neural networks. Once the model is trained on this labeled dataset, it can be applied to the same ADI sequence for evaluation without risk of overfitting. Fig. 6.1 shows a diagram of our novel framework for SODINN's variant.

The fact that SODIRF and SODINN can be trained on a labeled dataset created from a given ADI sequence means that these models are fine-tuned to each ADI sequence (Braham and Van Droogenbroeck, 2016). We have tested SODINN on ADI sequences from different instruments, using two very different datasets in terms of their characteristics. The first dataset, a VLT/NACO sequence on β Pic (Absil et al., 2013), consists of 612 frames with 8 sec of effective integration time, and has a total field rotation of 83 degrees. The second dataset, a VLT/SPHERE sequence on V471 Tau (Hardy et al., 2015), consists of 50 frames with 64 sec of integration time, and has a total field rotation of 30 degrees. Throughout this Chapter and for simplicity, we assume that $1 \times \text{FWHM} = 1\lambda/D = 4\,\text{pxs}$.

6.4 GENERATION OF A LABELED DATASET

In our proposed framework, the generation of a labeled dataset requires a transformation of the ADI image sequence that suits better a supervised learning problem and enables us to create two distinguishable classes. In order to achieve a different view of the data, we work on patches (instead of full frames) which is motivated by the fact that the exoplanet signal spatial scale is small compared to the frame size, and facilitates the creation of a large labeled dataset even from a single ADI sequence, as explained hereafter.

Our labeled dataset is generated in the form of 3D residual patches at several Singular Value Decomposition (SVD) approximation levels, which we will call from now on, Multi-level Low-rank Approximation Residual (MLAR) patches. These MLAR samples are built in the following way. Consider a matrix $M \in \mathbb{R}^{n \times p}$ whose rows contain the pixels inside a centered annulus of a given width. n is the number of frames in the ADI sequence. Recall that SVD is a matrix factorization such that:

$$M = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T, \tag{6.1}$$

where the vectors u_i and v_i are the left and right singular vectors, and σ_i the singular values of M. SVD is involved in several least-squares problems, such as finding the best low-rank approximation of M in the least-squares sense, i.e.,

$$\underset{X}{\text{argmin}} \, \|M - X\|_F^2, \tag{6.2}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. As it was mentioned in Subsection 1.3.3.5, we form a low-dimensional subspace B, capturing most of the variance of M, by keeping k right singular vectors. The residuals are given by:

$$res = M - MB^T B. \tag{6.3}$$

These residuals are later reshaped to image space, de-rotated and median combined as the usual ADI workflow dictates. In general, the larger the value of k the better the reconstruction, which produces smaller residuals (with less energy or standard deviation).

Instead of choosing one k value for estimating the low-rank approximation of M and obtaining a residual flux image (which is the goal of PCA-based approaches), we choose multiple k values sampling different levels of reconstruction. The MLAR patches are obtained by cropping square patches, of odd size and about twice the size of the FWHM, from the combined final residual frames obtained for each k.

These 3D samples, or MLAR patches, can be understood as computing annulus-wise PCA residual patches at different numbers of principal components. Working with these MLAR patches, we replace the ADI temporal information with the patch evolution as a function of the approximation level. As mentioned earlier, the SVD processing is always computed in an annulus-wise fashion. Defining the values of k, relies on the cumulative explained variance ratio (CEVR). Let $\hat{M}$ be the matrix M, from which its temporal mean has been subtracted, and $\hat{\sigma}_i$ the singular values of $\hat{M}$. The explained variance ratio for the j^{th} singular vector is defined as:

$$\frac{\hat{\sigma}_j{}^2}{\sum_i \hat{\sigma}_i{}^2}, \tag{6.4}$$

where i goes from one to $\min(n, p)$. It measures the variance explained by each singular vector and the CEVR measures the cumulative explained variance up to the k^{th} singular vector. Sensible values for k lie within the interval from 0.6 to 0.95 CEVR (for one example, see Fig. 6.2), based on our experience with PCA-based methods applied to HCI. The number of steps in this interval can be tuned, although the general rule is that more steps in the MLAR patches lead to more expressive samples that generally lead to higher classification power of a discriminative model. In our tests, with 8 to 20 steps, we could train accurate models. It is worth noting that our initial tests, working with 2D raw patches, were not successful and this motivated the constructions of the MLAR samples.

Using this data transformation, we generate labeled MLAR samples from our two classes, one containing the signature of a companion (*positive* class c^+) and the other representing the background and speckle diversity (*negative* class c^-). Therefore, we generate labels $y \in \{c^-, c^+\}$ associated to the MLAR samples. We emphasize that generating a large labeled dataset is possible, in the task of ADI processing, thanks to the possibility of generating synthetic companions and the exploitation of the rotation diversity of an ADI sequence.

The creation of the *positive* class relies on the injection of off-axis PSF templates, a procedure accepted within the HCI community for assessing the sensitivity limits of image processing techniques. The

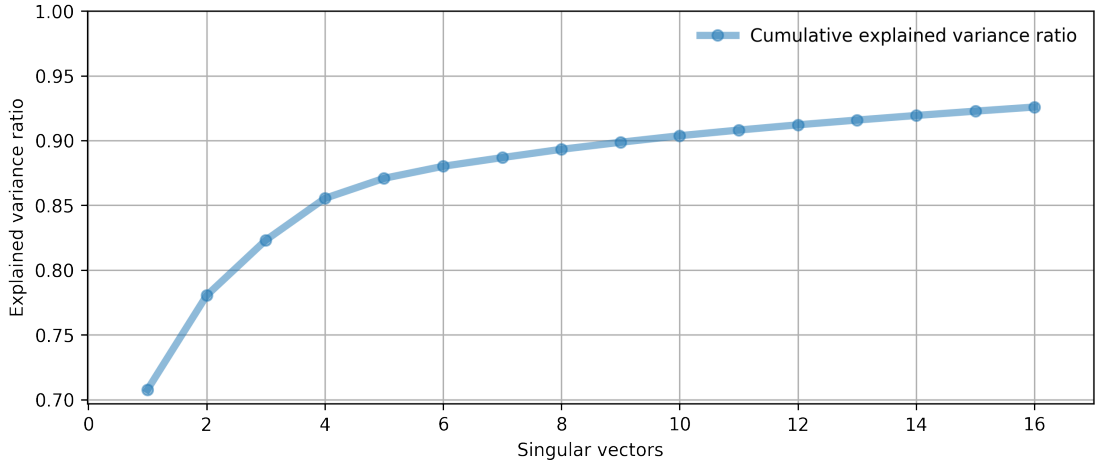


Figure 6.2: Generation of a labeled dataset: determination of approximations levels. We show the cumulative explained variance ratio (as defined from Eq. 6.4) for the first 16 singular vectors.

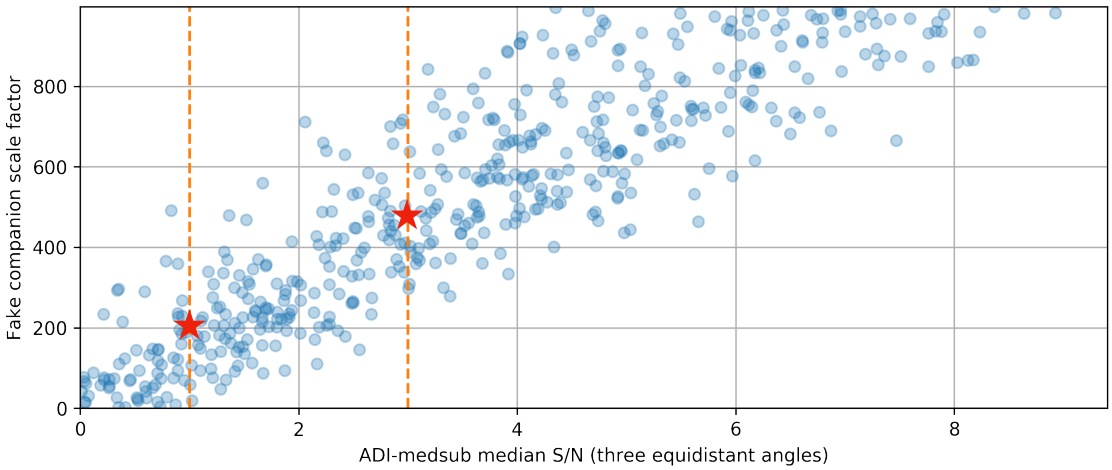


Figure 6.3: Generation of a labeled dataset: determination of flux intervals. We show the median S/N of injected companions, in a ADI-median subtracted residual frame, as a function of the scaling factor (from a large flux interval [5,1000]). The red star signs denote the lower and upper bounds of the companion injections for generating MLAR samples of the *positive* class.

injection consists in the addition of the PSF template at a variable frame location with a random brightness from a predefined interval. This interval is carefully chosen to avoid class overlap, which occurs when a same MLAR sample (or very similar) appears as a valid example of both classes. This can happen when the lower bound of our brightness interval is too low, in which case the signature of the companion signal is almost indistinguishable from the one of the background polluted with quasi-static speckles. Sensible lower and upper bounds are set with respect to the achieved S/N, usually in the interval [1,3], on a final residual frame obtained through classical ADI median subtraction (see Fig. 6.3). These flux intervals are defined in an annulus-wise fashion and are therefore related to the radial flux profile of the images.

The generation of the samples from *negative* class, representing the background and speckles, relies on the exploitation of the rotation associated to an ADI sequence and common machine learning data augmentation techniques[2]. The generation of a large number of *negative* samples faces two main difficulties. First, the fact that with a single ADI image sequence we obtain a single realization of the residual noise (in a PCA-based differential imaging context). Second, the number of patches we can grab from a given annulus is limited, and orders of magnitude smaller than the number of samples that are needed in the labeled dataset. If we feed these samples to the network, it would quickly memorize them, and that would produce strong overfitting. Our dedicated data augmentation process addresses these issues and can be summarized in the following steps:

1. We randomly grab up to ten percent of the MLAR patches in the given annulus.

2. We use negative parallactic angles when derotating the residual images (after reshaping to image space the residuals obtained in expression 6.3) to obtain final median combined images that preserve the noise correlation and keep the same statistical properties, while blurring any astrophysical signal. We grab all the available MLAR patches from the given annulus.

3. We randomly pick samples from the two previous subsets and average them to produce new samples.

4. Finally, we perform random rotations and small shifts of the MLAR samples obtained in the previous three steps to create even more diversity.

Optionally, a chosen area (circular aperture) of the ADI frame sequence can be masked to conceal a known, true companion. The cor-

2 This refers to the process of creating synthetic data and adding these to the training set in order to make a machine learning model generalize better (see section 7.4 of Goodfellow et al. (2016)).

responding patches are then ignored by the generator of labeled data. In the end, the patches of the *positive* class contain the injected companions and the patches of the *negative* class contain augmented samples. Thanks to this strategy, we avoid showing the samples (MLAR patches) from the original ADI sequence to our classifiers, thus reducing model overfitting. Note that the pixels in each slice of the MLAR sample are normalized in the interval [0,1], bringing them to the same value range. As a sanity test we inject fake planets in the ADI sequence we use to generate the labeled dataset, without masking the injected companions. Afterwards, we check whether the trained models are able to detect these pre-existing companions or not. In our tests, the injected companions (a proxy for the situation when we have a new dataset with a pre-existing unknown exoplanet) can be recovered every single time, which demonstrates that our approach prevents overfitting at the labeled dataset generation stage. Therefore, we conclude that our framework can be applied to new ADI datasets and the performance assessment shown in Chapter 6.7 is fair. Note that having access to multiple datasets taken with the same instrument (survey data), would allow us to train a more general model and depend less strongly on our data augmentation procedure.

In panels (a) and (b) of Fig. 6.4, we show a few examples of the resulting MLAR samples composing our labeled data set. The patch size was set to seven pixels. The MLAR *positive* samples shown in panel (b) clearly illustrate the exoplanet PSF morphological distortion introduced by differential imaging post-processing, as a function of the aggressiveness (analogous to the number of principal components used in a PCA-based post-processing approach). This is related to the well-known problem in HCI of companion self-subtraction. The PSFs of the companions clearly degrade as the CEVR increases (they eventually vanish), which affects the positions of the PSF centroids.

We use the `VIP Python` library (Gomez Gonzalez et al., 2017) for low-level image operations and the generation of labeled datasets. The generation of the MLAR samples is done on CPU in a parallelized way and the SVD computations use the randomized SVD algorithm proposed by Halko et al. (2011) to speed up the computation time. We use the above described procedure to generate a balanced labeled dataset of up to half a million MLAR samples. Here again the general rule is that more samples are better for the discriminative power of our models. A thorough analysis of the influence of the labeled dataset size on the performance of our discriminative models has yet to be performed.

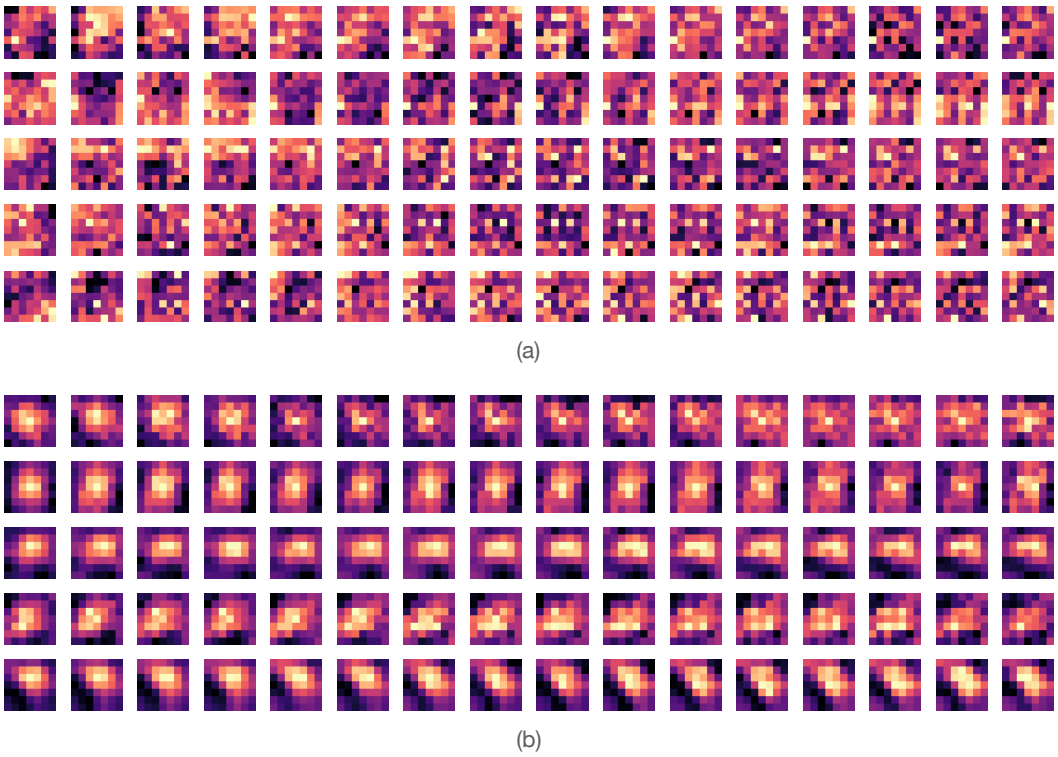


Figure 6.4: (a) Each row corresponds to a random MLAR sample from the *negative* class (background and speckles). (b) Each row corresponds to a MLAR sample from the *positive* class (exoplanet signal). The *positive* samples are shown, from top to bottom, with increasing flux. Every slice of the MLAR sample is normalized in the interval [0,1]. The CEVR for these MLAR samples are shown in Fig. 6.2.

6.5 DISCRIMINATIVE MODEL

The role of the discriminative model is to disentangle two kinds of signal signatures. The *positive* class c^+ corresponds to the signal from the exoplanets and the *negative* class c^- to the speckles and background. The goal of the classifier is to learn a mapping from the input MLAR samples to their corresponding labels and to generate predictions on new samples $\hat{y} \in \{c^-, c^+\}$. This is possible thanks to the fact that the footprint of a true companion in the MLAR patches is different from the one of a speckle or a background area. In the following Sections, we propose two ways of exploiting the structure of our MLAR samples, one using random forests (SODIRF) and a more sophisticated one using deep neural networks (SODINN).

6.5.1 Random forest based approach

In the case of SODIRF, we need a 2D matrix of samples versus features suitable for training the random forest classifier. The feature matrix is constructed by vectorizing the MLAR samples and stacking them in a matrix. A random forest (Breiman, 2001) is an ensemble learning method that fits a multitude of decision tree classifiers on various sub-samples of the dataset (with bootstrapping), and uses averaging to improve the predictive accuracy. A random forest also controls over-fitting to the training dataset, reducing the generalization error, if compared to single decision trees (with higher variance). We implemented SODIRF with the machine leaning library `scikit-learn`. This implementation of a random forest combines the decision tree classifiers by averaging their probabilistic prediction. We trained the random forest by using 100 fully developed trees to form the ensemble model. The model was trained using a simple train-test splitting procedure and reached over 99.5% test accuracy.

Random forests can be efficiently trained on CPUs, in just a few minutes, exploiting modern multi-processor architectures. This is different from deep neural networks which require last generation GPUs and more computing time to be trained. They differ not only in terms of the computational cost but also in terms of performance, as we show in Section 6.7.

6.5.2 Deep neural network based approach

As we discussed in Chapter 4, a recurrent neural network (Rumelhart et al., 1986) is a class of neural network suited for sequence modeling. Long-short term memory (Hochreiter and Schmidhuber, 1997) networks are a special kind of RNN, capable of learning long term dependencies. They are widely used in machine translation, large-vocabulary speech recognition and text-to-speech synthesis. On

the other hand, convolutional neural networks (LeCun et al., 1989; Krizhevsky et al., 2012) are the preferred solution for processing data that has a grid-like topology, e.g. images, and are almost universally used in computer vision.

Convolutional LSTM networks (Shi et al., 2015) combine convolutional and LSTM architectures, which makes them very efficient for modeling spatio-temporal dependencies and correlations. A convolutional LSTM layer is similar to an LSTM layer, with the difference that the input and recurrent transformations are both convolutional. As we mentioned before, in our framework we replace the time axis with the MLAR dimension. SODINN makes use of a deep neural network model that exploits these 3D samples using convolutional LSTM cells. The MLAR samples are directly fed to the neural network, without sacrificing their spatio-temporal structure as in the case of SODIRF.

We have implemented SODINN's neural network classifier with the highly modular and minimalist `Keras` library (Chollet et al., 2015) using its `Tensorflow` (Abadi et al., 2015) backend. The networks were trained on a dedicated NVIDIA DGX-1 deep learning system with eight P100 GPUs. As shown in Fig. 6.1, the architecture consists of two convolutional LSTM layers, each one followed by a 3D max pooling layer with sizes $2\times2\times2$. The first convolutional LSTM layer uses a 3×3 kernel with 40 filters, while the second features a 2×2 kernel with 80 filters. These are followed by two fully connected dense layers, the first with 128 hidden units (on which we apply dropout (Srivastava et al., 2014)) and the last consisting of a sigmoid unit. The network weights are initialized randomly using a Xavier uniform initializer and are learned by back-propagation with a binary cross-entropy loss function:

$$\mathcal{L} = -\sum_{n} (y_n \ln(\hat{y}_n) + (1 - y_n) \ln(1 - \hat{y}_n)), \tag{6.5}$$

where y_n is the true label of the n^{th} MLAR sample and $\hat{y}_n = p(c^+ \mid \text{MLAR sample})$ is the probability that the n^{th} MLAR sample belongs to the *positive* class.

The network is trained after splitting the labeled data in train, test (ten percent of the initial samples), and validation sets. An Adam optimization strategy (Kingma and Ba, 2014) is used with a learning rate of 0.003 and mini-batches of 64 training samples. We include an early stopping condition monitoring the validation loss. Usually, our model is trained with 15 epochs (passes of the stochastic gradient descent optimizer through the whole train set) reaching 99.9% validation accuracy.

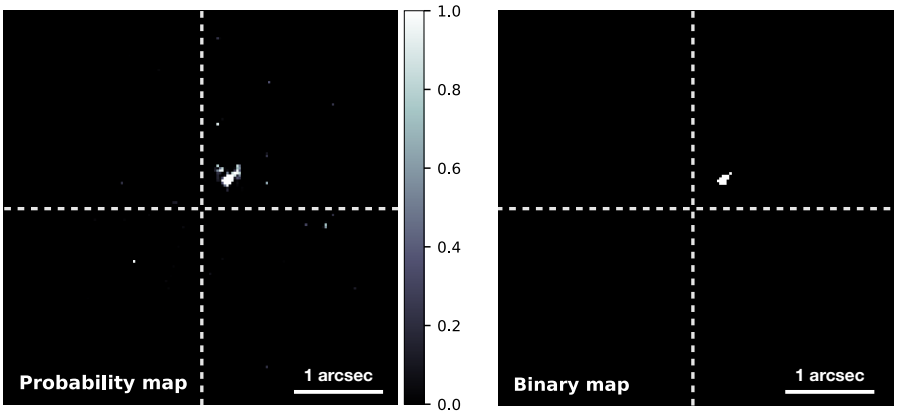


Figure 6.5: SODINN's probability map (left) and binary detection map (right), thresholded at 99%, for the VLT/NACO β Pic dataset.

6.6 PREDICTION STAGE

Once the models are trained, they are applied to the input data cube. First, we perform the same transformations to the input ADI sequence to obtain MLAR patches centered on each one of the pixels of the frame. The discriminative model then classifies these MLAR patches, assigning a probability of membership to the *positive* class, $p(\hat{y} = c^+ \mid \text{MLAR sample})$. The outputs of both SODIRF and SODDIN are a probability map and a binary map, obtained after thresholding the probabilities, as exemplified in Fig. 6.5 for the VLT/NACO dataset. We can see how SODINN's binary map clearly reveals the presence of β Pic b, without false positives, for this probability threshold.

For comparison, in a differential imaging PCA-based approach, one would tune the number of principal components that works best for a companion at a given radial distance, resulting in a residual flux image (a single realization with a single value of k) where one could visually identify companions or use a S/N metric (taking into account the noise in a given annulus). In our framework, grabbing MLAR patches centered on each pixel of the frame, enables the estimation of a class probability in a detection map. We then threshold this map at a desired level of *positive* class probability. This detection criterion is evaluated independently for each pixel on the frame, contrary to the S/N estimation approach.

6.7 PERFORMANCE ASSESSMENT

After the test on a known companion, we can proceed with testing the performance (detection capacity) of the trained models by injecting fake companions. In this section we will focus on SODINN. Using the V471 Tau VLT/SPHERE dataset, we inject four companions (using

Table 6.1: Parameters for the fake companions of Figs. 6.6 and 6.7.

FC	Radius	Angle	Flux (ADUs)	Contrast	ADI-PCA S/N
1	1.5 λ/D	170°	9000	2.5×10^{-4}	3.2
2	1.75 λ/D	230°	7000	1.9×10^{-4}	5.9
3	2.5 λ/D	0°	1500	4.2×10^{-5}	1.3
4	5 λ/D	90°	400	1.1×10^{-5}	2.7

the input off-axis PSF), from one to five λ/D. The parameters of the synthetic companions are shown in Table 6.1. The quoted S/N is the best one obtained, at the location of each companion, after trying all possible numbers of principal components. The first, third and fourth companions are pretty much at the level of the speckle noise at these separations (see Fig. 6.6). The shapes of their PSFs are hard to distinguish from surrounding noise and the S/N values (obtained in a 1×FWHM aperture centered at the injection positions) are small. Only the second companion has a S/N over five, which is due to the fact that it was injected on top of a bright speckle (purposely). The visual inspection would not be definitive for such a companion. As shown in Fig. 6.7, SODINN outperforms the full-frame ADI-PCA approach by recovering the four companions at a high (99%) probability without any false positive. Tests with known and injected companions are a first attempt to measure performance. Unfortunately, it is not possible to judge the performance of a detection algorithm based on a few realizations of such tests (this is discussed in detail in Appendix A). We can adopt a more robust approach using a signal detection theory.

A receiver operating characteristic (ROC) curve is a useful tool for assessing the performance of binary classifiers (see Section 5.5.2). In general, ROC curves allow us to study the performance of a binary classifier system in a true positive rate (TPR $= p(\hat{y} = c^+ \mid y = c^+)$ - false positive rate (FPR $= 1 - p(\hat{y} = c^- \mid y = c^-)$ space, as a decision threshold τ varies. In other words, they can assess the TPR (also called sensitivity) and the FPR at the same time. This, by principle, differs from the sensitivity or contrast curves widely used in HCI. In Fig. 6.8, we illustrate the task of a binary classifier in a signal detection context and the effect of choosing a detection threshold. By varying this threshold, we can adjust the FPR that we are willing to accept for a specific sensitivity. The ability of separating the class is inherent to the model; that is what a ROC curve shows. HCI as a signal detection problem seeks to simultaneously maximize the sensitivity to companions and minimize the number of false detections (FPR).

In this study, we choose to build our ROC curves in a TPR (percentage of detected fake companions) vs mean per-frame false positives, instead of a TPR vs FPR space. The total number of false positives is counted on the whole detection map, and is averaged for each τ.

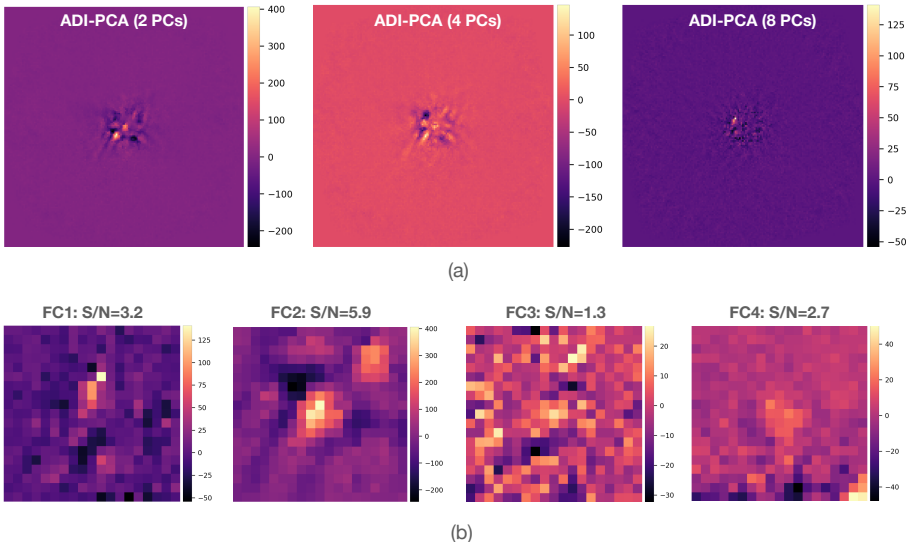


Figure 6.6: Injection of four synthetic companions (parameters detailed in Table 6.1) in the V471 Tau VLT/SPHERE ADI sequence. (a) Three final residual frames with ADI-PCA with 2, 4 and 8 PCs. (b) Cropped frames centered on the injected companions (after optimizing the number of PCs to maximize the S/N of each companion).

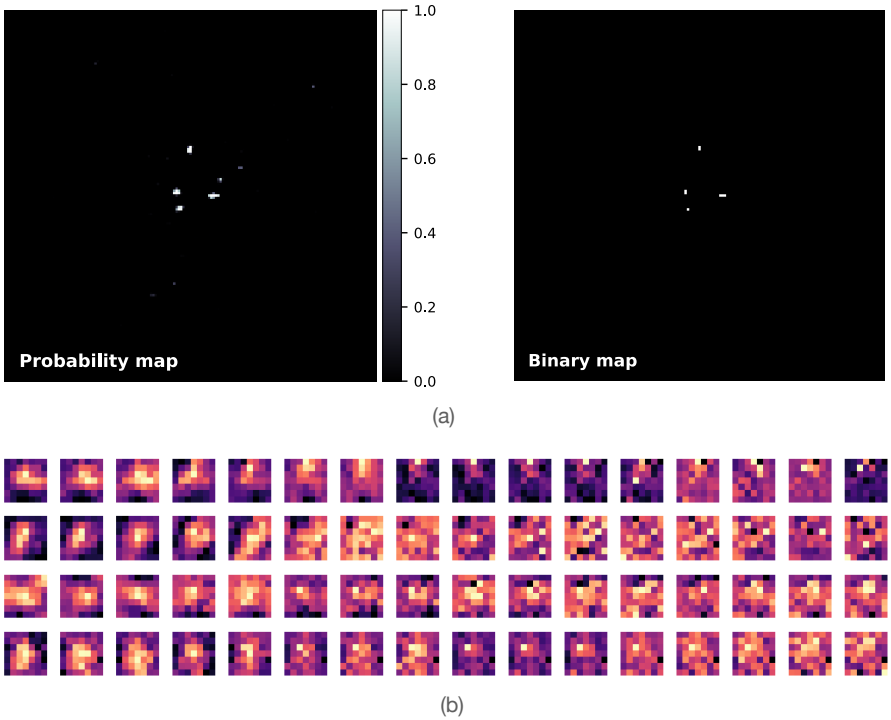


Figure 6.7: Injection of four synthetic companions (parameters detailed in Table 6.1) in the V471 Tau VLT/SPHERE ADI sequence. (a) SODINN's probability and binary maps where the four planets are clearly detected (without false positives). (b) The MLAR patches centered on each one of the fake companions.

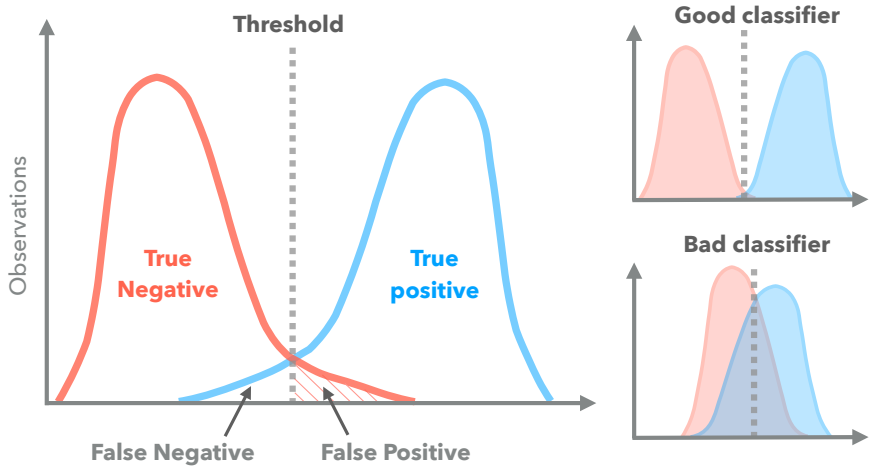


Figure 6.8: Behavior of a binary classifier in a signal detection theory context.

This reflects better the goal of a planet hunter and facilitates interpretation of the performance simulations. This strategy differs from the one used in Chapter 5. The ROC curves are built separately for different annuli with a tuned uniform flux distribution for the injection of fake companions. Having ROC curves for different separations from the star better illustrates the algorithm performance at different noise regimes. When interpreting the results, it is important to compare the ROC curves for different algorithms to each other, for a given annulus, considering that the TPR depends on the brightness of the injected companions, while the mean per-frame false positives do not. It is also important to examine the shape of the curves. For instance, it is preferable to have a steeper curve, which means that such classifier does better in minimizing the number of FP while it increases its sensitivity.

We compare SODINN and SODIRF to classical ADI median subtraction, full-frame ADI-PCA and LLSG on the VLT/SPHERE V471 Tau dataset, a challenging ADI sequence with few frames and mild rotation. As mentioned earlier, differential imaging approaches (unsupervised learning), i.e. ADI median subtraction, ADI-PCA and LLSG, do not generate a prediction (probability) but rather a residual image to look at. We obtain detection maps for these approaches by building S/N maps and thresholding them at several values of τ. For each injection of a fake companion, a new data cube is built and processed with each of the five algorithms. In the case of SODINN and SODIRF, the discriminative models are not retrained for each injection of a fake companion. Two principal components, or 0.7 CEVR of the full-frame data cube, are used for ADI-PCA, and the same value is used as the rank parameter of LLSG (no other hyperparameters were tuned). S/N maps were built for the resulting residual frames and thresholded at different values of τ: 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5,

Table 6.2: Parameters used for the ROC curves of Fig. 6.10.

Panel	Radius	Flux (ADUs)	Contrast
(a)	1-2 λ/D	U(3000,7000)	8.5×10^{-5} to 1.9×10^{-4}
(b)	2-3 λ/D	U(1000,5000)	2.9×10^{-5} to 1.4×10^{-4}
(c)	4-5 λ/D	U(50,350)	1.4×10^{-6} to 1.0×10^{-5}
(d)	4-5 λ/D	U(150,450)	4.3×10^{-6} to 1.3×10^{-5}

5. For SODINN and SODIRF, we thresholded the probability map at several levels: 0.1, 0.2, 0.3, 0.4, 0.5, 0.59, 0.69, 0.79, 0.89, 0.99. Fig. 6.9 illustrates one single realization of a companion injection and a single threshold value. The ROC curves, built for three different annuli, are shown in Fig. 6.10. Brightnesses, contrasts and distances, for all the injected companions, are shown in Table 6.2. MLAR samples of 16 steps, in the interval 0.5-0.95 CEVR, are used for training SODINN and SODIRF. A total of 100 injections are performed for building each one of the ROC curves.

Reading the ROC curves in Fig. 6.10 is straightforward: panel (a), for an annulus from one to two λ/D, shows that a blob (at least one active pixel inside a 3×3 pixels box) sticks out above the detection threshold at the position of the fake companion injection in 12%, 16%, 18%, 5% and 44% of the cases for ADI median subtraction, ADI-PCA, LLSG, SODIRF and SODINN respectively, when setting the detection thresholds at their highest values (S/N of five for ADI median subtraction, ADI-PCA and LLSG, 99% probability for SODINN and SODIRF). At the same time, an average of ~0.8 false positives is present in the ADI median subtraction and ADI-PCA S/N maps, while LLSG, SODIRF and SODINN show no spurious blobs. The ROC curves for larger separations consistently show SODINN's improved performance with respect to other approaches (see panels (b), (c) and (d) of Fig. 6.10). The performance of LLSG seems to degrade at larger separations, most probably due to the lack of hyperparameter tuning. SODIRF's sensitivity improves with the separation and starts to match the performance of SODINN. In Appendix A.1 we provide more details about the construction of the ROC curves and the assessment of post-processing algorithms for exoplanet detection.

6.8 CONCLUSIONS

This study illustrates the potential of machine learning in HCI for the task of exoplanet detection. We propose a novel paradigm for detecting point-like companions by reformulating HCI post-processing as a supervised learning problem, building on well-established machine learning techniques. Instead of relying on unsupervised learn-

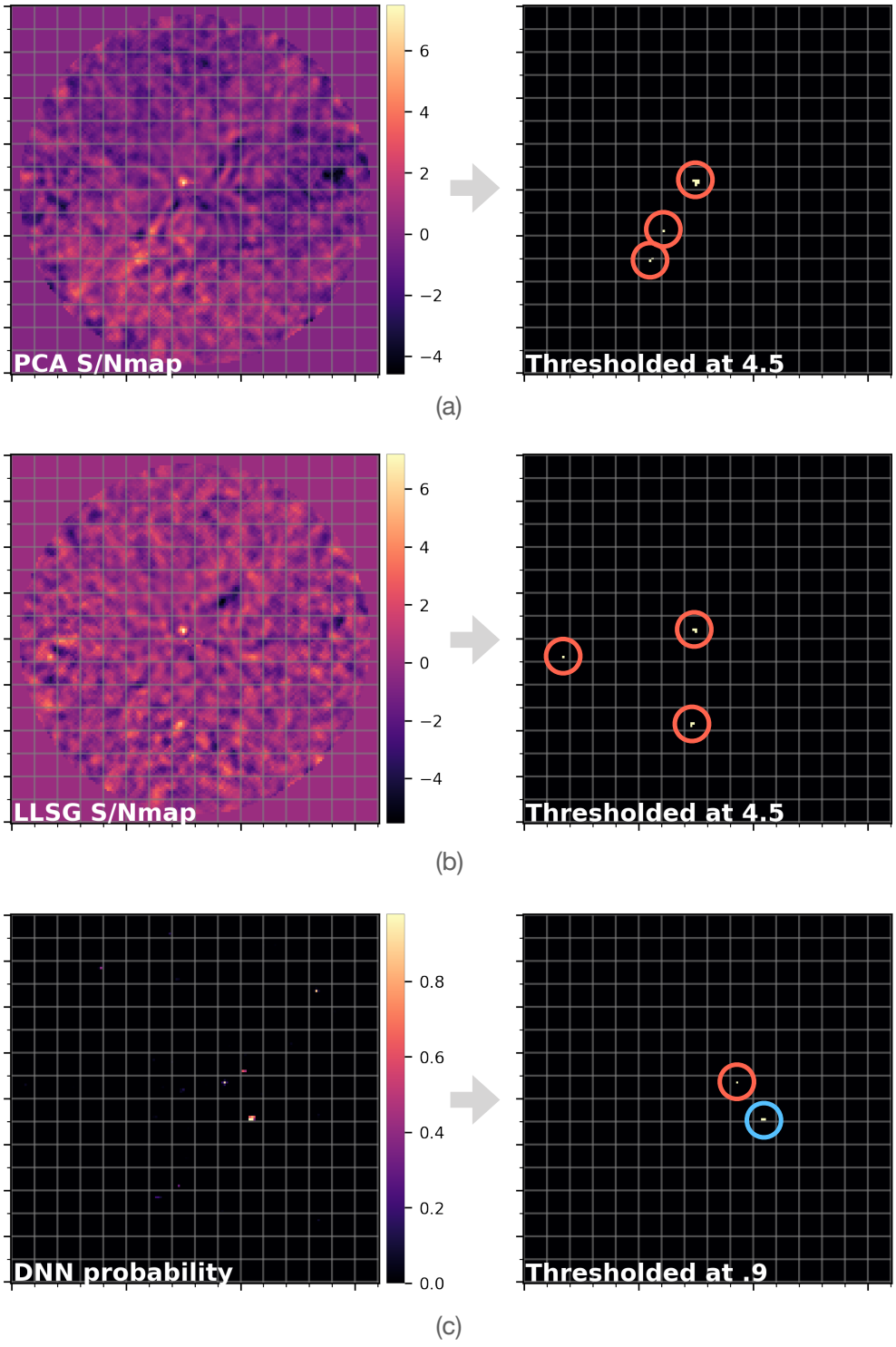


Figure 6.9: Example of one single injection, showing the detection maps of ADI-PCA, LLSG and SODINN. False positives are shown with red circles while true positives are shown with blue ones, for a given threshold value. These are the counts done per injection and later used for building the ROC curves. Fig. A.3 is an extended version of this figure.

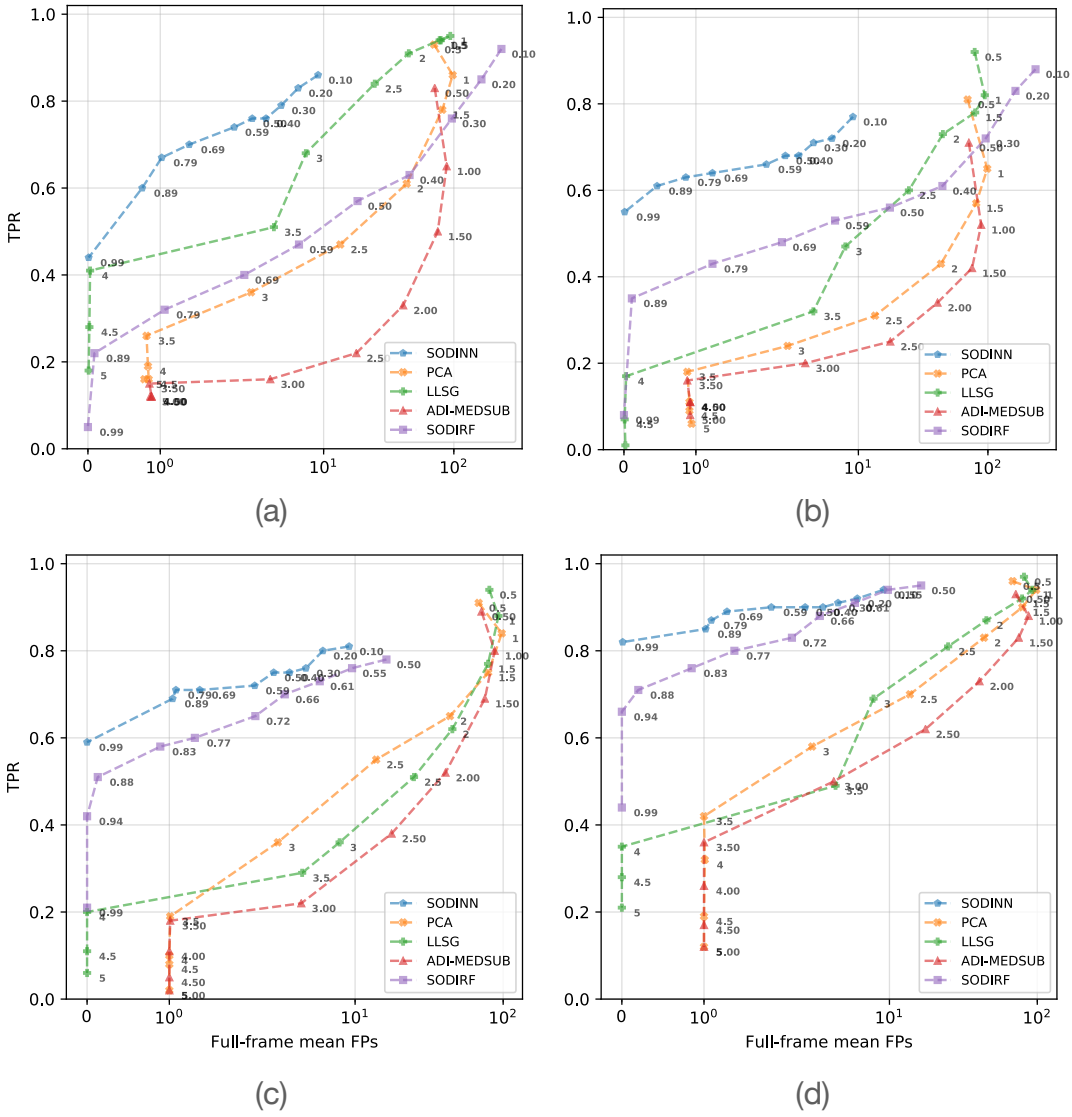


Figure 6.10: ROC curves for the VLT/SPHERE V471 Tau dataset, comparing ADI median subtraction, ADI-PCA, LLSG, SODIRF and SODINN (a) for the 1-2 λ/D annulus, (b) for the 2-3 λ/D annulus, (c) and (d) for the 4-5 λ/D annulus with two different flux distributions, as shown in Table 6.2.

ing techniques, as most of the state-of-the-art ADI post-processing algorithms, we generate labeled datasets (using data augmentation techniques) and train discriminative models that classify each pixel of the image, assigning a probability of containing planetary signal. We present two approaches that differ in the type of discriminative model used: SODIRF and SODINN. The former employs a random forest classifier while the latter features a more advanced deep neural network model, which exploits better the structure of the labeled MLAR samples.

In order to assess the detection capabilities of our approaches, we perform a ROC analysis comparing both SODINN and SODIRF to techniques such as ADI-PCA and LLSG. The performances of both algorithms are beyond what ADI-PCA and ADI median subtraction can offer. In particular, the performance of SODINN lies in a separate zone of the ROC space: from one to two λ/D, it improves the TPR by a factor of $\sim$3.5 with respect to ADI-PCA when working at the same false positive level (one false positive per-image). Moreover, this is achieved in the case of a challenging ADI sequence, with mild rotation and few frames, from a last-generation HCI instrument – VLT/SPHERE (see Appendix A.1 for a deeper discussion of the ROC curves performance assessment). The fact that these models are versatile and can be fine-tuned to each specific ADI sequence opens great possibilities of re-processing existing databases, from first- and second-generation HCI instruments, to maximize their scientific return.

Although in this Chapter we only addressed single ADI datasets, our framework's true potential is in the context of surveys, where the data from different observations could be used to generate a larger and more diverse labeled dataset. This would allow us to train more general, and perhaps more accurate, neural network models. The exploitation of SODINN for surveys will be the focus of a future study. Other interesting venues of future research are the inclusion of the companion brightness into the model, the extension to other HCI observing techniques (beyond ADI), and the use of generative neural networks for complementing the data augmentation process.

The simultaneous increase in sensitivity, which translates in deeper detection limits (the ability to detect companions at higher contrasts), and reduction of the per-image false positives clearly indicate that our supervised approach SODINN is a very promising HCI exoplanet detection technique. Considering that ADI remains the most common HCI observing strategy and the large reservoirs of archival data, SODINN could potentially improve the demographics of directly imaged exoplanets at all separations, including those in the inner vicinity (1-2 λ/D) of their parent stars where ADI signal self-subtraction and speckle noise are the strongest.

7 CONCLUSIONS

In this dissertation, I have presented the results of my research on the development of novel approaches to data processing in high-contrast imaging. Data processing constitutes a critical component of high-contrast imaging, as advanced image processing algorithms are bound to push the detection limits of high-contrast imaging instruments. The consensus, after ten years of high-contrast imaging, is that massive planets, such as those of HR8799, are rare at wide separations. From indirect methods, we know that super-Earths and rocky planets are much more common than giant planets. For this reason, the development of new image processing techniques is of key importance for maximizing the scientific return, specially at short separations, of existing first and second generation high-contrast instruments. The ultimate goal of post-processing procedures is to improve the planet to star contrast (increasing the ability to detect faint companions), reduce the image dynamic range and combat speckle noise (increasing the ability to detect close-in companions). Most of these differential imaging algorithms model a reference PSF, by exploiting the diversity introduced with a given observing technique, to produce a residual final image. Unfortunately, part of the companion signal is fitted in this reference PSF and subtracted together with the speckle pattern. New detection methods must address these issues and detection biases of differential imaging. On the other hand, during the last five years, we have witnessed the emergence of data-driven discovery methods in sciences in parallel to the popularization of machine learning and data science fields of study. The high-contrast imaging community has slowly adopted the latest developments in data management and machine learning for analyzing the increasing amount of available data. This dissertation attempted to fill in this gap and developed at the interface of various fields such as computer science, machine learning, statistics and astrophysics.

In the first part of this dissertation, I have presented my contribution to the field of high-contrast imaging in terms of scientific software development. Concretely, I have developed the Vortex Image Processing (`VIP`) library, an open-source `Python` library, which contains a large number of state-of-the-art and novel pre- and post-processing algorithms for the analysis of high-contrast imaging data. `VIP` was inspired, in the scientific context, by existing high-contrast imaging pipelines written in commercial languages. In the software developing context, it was inspired by the open source `Python` community, its ethos and practices. `VIP` is being actively used by several

high-contrast research teams and has been cited more than a dozen times to date. In the first part of this manuscript, I also showcased the capabilities of `VIP`, presenting examples of its application to two real on-sky HR8799 datasets from two high-contrast ground instruments. In these two published studies, `VIP` was used for investigating the presence of a fifth companion in this system, producing sensitivity limits, and obtaining robust astrometry of the four known companions with high astrometric accuracy.

In the second part of this dissertation, I have presented my developments in terms of ADI image processing algorithms, with a special emphasis on the task of exoplanet detection. These original approaches were developed drawing ideas from computer vision and machine learning fields, and their performances were evaluated in a robust signal detection theory framework. The first approach, the LLSG decomposition, is a method that improves over PCA-based low-rank approximation methods for reference PSF subtraction. It builds on recent algorithmic developments in computer vision and decomposes the image sequence into three components: a low-rank, a sparse and a Gaussian noise term. The motivation behind this is to isolate the signal of moving companions in the sparse term, facilitating the detection task and boosting the S/N of potential companions. A comparison of LLSG and ADI-PCA performances in a ROC space shows that LLSG improves in TPR while preserving the same FPR.

The second approach takes high-contrast imaging post-processing from an unsupervised to a supervised learning framework. This novel data-driven supervised detection approach exploits unique characteristics of ADI datasets, which enables the generation of large labeled datasets to train discriminative models. Two algorithms are proposed in this framework: SODIRF and SODINN. The former uses a well-known random forest classifier while the latter implements cutting-edge developments in deep learning. In particular, SODINN's performance lies in a separate zone of the ROC space when compared to state-of-the-art techniques such as ADI-PCA. Using a challenging VLT/SPHERE dataset, the performances of SODIRF and SODINN were compared to ADI median subtraction, ADI-PCA and LLSG algorithms. From one to two λ/D, SODINN improved the TPR by a factor of $\sim$3.5 when working at the same false positive level. This supervised detection framework offers the possibility of re-processing existing archival databases to maximize their scientific return and potentially improve the demographics of exoplanets detected through high-contrast imaging.

There are several clear and exciting directions forward following the results of this dissertation. Regarding `VIP`, it is rewarding that more users are getting on board and that some are turning into co-developers, implementing enhancements and fixing issues. `VIP` started

as an experimental playground and therefore its API[1] was not designed or planned in advance. This is due to the fact that scientific programming is highly nonlinear and exploratory. Besides the consolidation of other observing techniques and the addition of new algorithms, the implementation of a better API would be the best way to benefit and enlarge the user community. As I discussed in the Appendix section of this dissertation, the high-contrast community needs to create a database of standardized datasets. I plan to contribute to the consolidation of such a database and to carry out thorough performance algorithm comparisons and data challenges. `VIP` will be a great asset for this project.

Another interesting research direction would be the extension of my supervised detection framework, by embedding the flux and subpixel position of exoplanets into the model (in order to characterize detected candidates), and its adaptation to other observing techniques such as multiple-channel SDI. I plan to contribute to the integration of machine learning methods to high-contrast imaging at different stages of the observing pipeline (beyond post-processing). Finally, I will focus on the application of the results presented in this dissertation to large ground-based surveys, in the near future. The harvest of large archival databases with new powerful tools, such as SODINN, will certainly be a thrilling enterprise.

1 Application Programming Interface.

Part III

APPENDIX

A ASSESSING ALGORITHM PERFORMANCE FOR HCI EXOPLANET DETECTION

In this Appendix I present some guidelines that, from my experience developing HCI algorithms for exoplanet detection, should be used for performance assessment. The steps mentioned below also build on the practices adopted by the HCI community, but are by no means exhaustive. This discussion focuses on the detection task, where we are interested in evaluating the sensitivity or ability to detect planets of varying contrast (brightness with respect to the star). In this context, the ultimate goal is to maximize the TPR while minimizing the FPR (Mawet et al., 2014; Gomez Gonzalez et al., 2016a). The proposed approaches, ordered by increasing level of robustness and computational complexity, are the following:

1. Tests on fully synthetic datasets employing simple metrics, e.g. S/N.
2. Tests using on-sky data with known companions (with simple metrics, e.g. S/N).
3. Tests using on-sky data with injected fake companions.
4. Tests on datasets from different instruments and with different parameters, e.g. rotation/wavelength range, integration time, number of frames, weather condition, wavefront control system performance, coronagraphic solution.
5. Comparative studies with respect to the state-of-the-art algorithms using sensitivity metrics, e.g. contrast curves.
6. Comparative studies using signal detection approaches, e.g. ROC curves.
7. Comparative studies with ROC curves, using several datasets (as in point 4).

Point 1 refers to using modeled image sequences with synthetic companions, which is an approach widely used for studying the performance of future instruments. This should best be avoided for comparing algorithms for which on-sky real data exist. Synthetic images will be a naive approximation of real data, no matter how complex the models of the instrument and aberrations introduced by the atmosphere are.

Point 2 is an important sanity check used when building post-processing algorithms. Regarding the S/N metric, I used the definition proposed by Mawet et al. (2014) throughout this dissertation. This definition addresses the small sample statistic issue and could be considered the standard S/N definition in HCI (although the recent work of Ruffio et al. (2017) neglects its use). As it was discussed in Chapter 5, the S/N has notable disadvantages. The S/N value for a point-like source depends on the choice of the aperture sizes and on the position of the apertures themselves, especially at small angular separations where the small sample statistics effect becomes dominant (Quanz et al., 2015). Also, the S/N by definition requires noise in the images, as is the case of residual flux images. When applied to frames where the noise has been severely suppressed (as it can happen with LLSG) or to sparse detection maps (including binary maps) the S/N metric produces extremely high values or even becomes infinite. An interesting solution, proposed by Pairet et al. (in prep.), uses simple statistics (such as the mean and standard deviation) in the temporal direction to produce a S/N map. This S/N map does not suffer from the small sample statistic effect and is suited for sparse image sequences (e.g. LLSG).

Point 3 is the most common test used in papers describing HCI algorithms. The injection of fake companions at different locations and brightnesses enables the generation of diverse planet-star-contrast configurations. Testing with a limited number of fake injections is a good way to showcase a new algorithmic approach. It is critical to note that this is far from a robust approach and should be followed by other procedures. For instance, the fact that an algorithm works well in a single dataset (from a given instrument and with certain characteristics) does not guarantee that it will have a similar performance on data taken with another instrument or under different conditions. This is proposed in Point 4. It helps to assess whether an algorithm generalizes to other dataset instances.

Point 5 refers to the use of contrast curves for the task of algorithm performance assessment. Contrast curves are a widely used tool in HCI instrumentation design that has been also adopted to measure performance of algorithms, detection limits for single observations and whole planet searching campaigns. A contrast curve is not the optimal choice, in the case of algorithm assessment, because the contrast definition does not address the TPR and FPR trade-off (see Section 2.8). The main problem is that the contrast curves only evaluate the sensitivity or TPR while making strong (most of the time unverifiable) assumptions about the nature of the noise, which is closely related to the FPR. Nevertheless, when several algorithms are compared, using the same contrast evaluation procedure on a single dataset, a first approximation to HCI algorithmic performance can be obtained.

In Point 6, I propose the use of ROC curves as the best tool for evaluating the performance, on a TPR vs FPR space, of HCI detection algorithms. Unfortunately, in this case of using a single dataset, there are no guarantees that the performance of an algorithm will behave in a similar way for different data cubes. Details about the construction of a single ROC curve will be discussed in the following Section. Point 7 is the final step for ensuring a robust assessment of the general performance of a detection algorithm for HCI, a ROC analysis on several datasets (see Section A.2 of this Appendix). See Piérard et al. (submitted) for a discussion on averaging the performance of binary classifiers in a ROC space, from a signal detection theory point of view.

A.1 BUILDING ROC CURVES FOR HCI EXOPLANET DETECTION ALGORITHMS

ROC curves are commonly used tools for assessing the performance of binary classifiers. The planet detection problem can be seen as a binary classification and therefore ROC curves can be used for our purposes. They have been proposed in the context of HCI for comparing the detection performance of post-processing algorithms (Barrett et al., 2006; Lawson et al., 2012; Gomez Gonzalez et al., 2016a). This is the approach I used to compare LLSG to the state-of-the-art, which was the first time a ROC curve was built for comparing algorithm performance in the HCI literature.

As I mentioned before, a ROC curve shows the TPR-FPR trade-off as a function of a detection threshold. It is important to understand that the relative ROC performance of two different algorithms changes due to several factors: the dataset used (which has a set of characteristics as mentioned in the previous Section of this Appendix), hyper-parameter tuning of each algorithm (see Fig. A.1 for an example), noise regime or separation from the star, and contrast of the injected companions (as shown in Fig. A.2). There is no shortcut to avoiding the dependence on these factors, unless the metric makes strong assumptions about the data and noise distributions (which is undesirable). I argue that a data-driven approach, using standardized databases (see Section A.2) and high performance computing simulations, is the most fair, robust, and reliable performance assessment approach.

In this dissertation, I presented ROC curves in Chapters 5 and 6. The approaches used in these two Chapters are different. For instance, the ROC curve from Chapter 5 was built by averaging the TPR and FPR over all brightnesses and the tested annuli or distances. The resulting ROC curves give a global sense of the performance but it is difficult to pinpoint, for instance, the sensitivity as a function of the

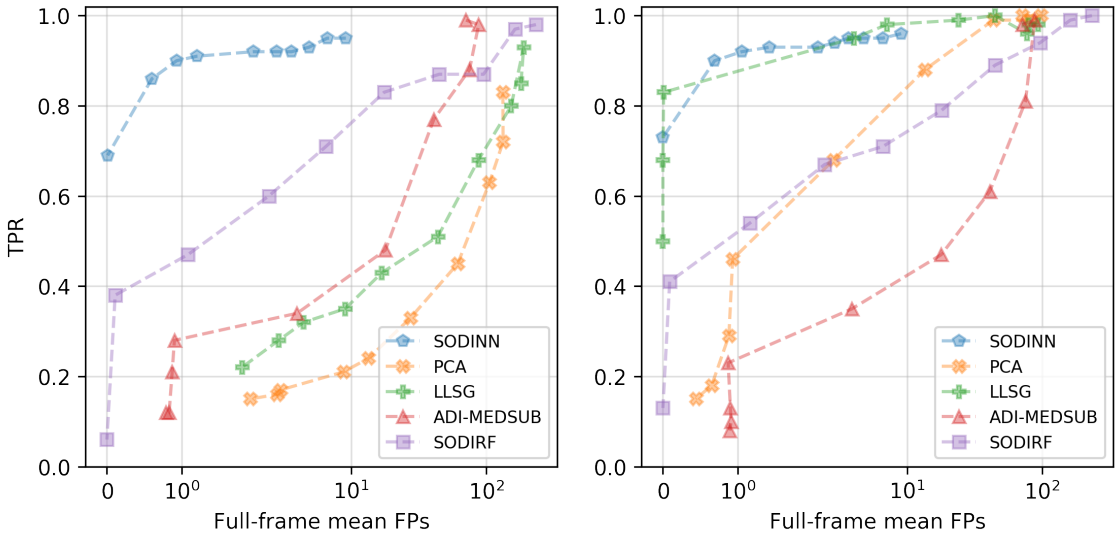


Figure A.1: Example of ROC curves when changing the algorithms hyper-parameters. These were computed for the same dataset of Fig. 6.10. For the left panel, the number of PCs for ADI-PCA and the rank for LLSG were set to 7. For the right panel, a value of 2 was used. Notice that the locations of the curves for SODINN, SODIRF and ADI median subtraction remain the same with respect to each other in both panels. On the other hand, the performance of ADI-PCA and LLSG is worst when too aggressive hyper-parameters are used (see how they move upward in the right panel). This exemplifies the pitfalls of comparative studies using ROC curves, and how easy it is to obtain wrong performances and present unfair conclusions.

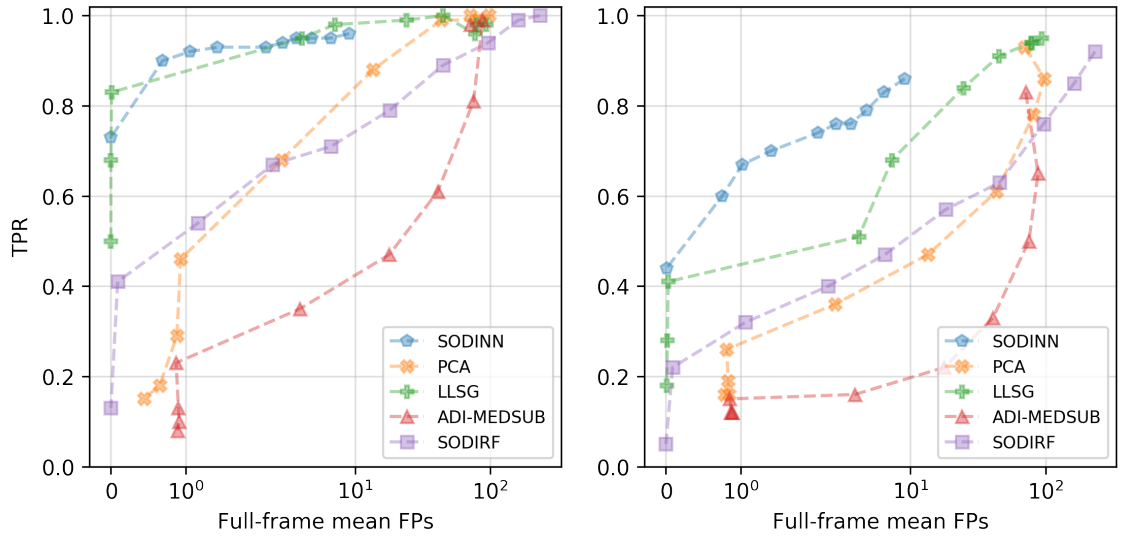


Figure A.2: Example of ROC curves with two different contrast intervals. These were computed for the same dataset of Fig. 6.10. The right panel shows ROC curves built with a higher planet to star contrast (fainter injected companions) with respect to the left panel. This explains the degraded performance. Notice that the ROC curve of SODINN is on top in every case.

separation from the star. Also, the ROC curves in Chapter 5 evaluate the presence of a false positive at a single random location for each companion injection, whereas for the ROC curves of Chapter 6 the whole frame is considered.

This has been partially addressed by Jensen-Clem et al. (submitted) who propose a performance map that plots the TPR as a function of separation and astrophysical flux ratio (planet to star contrast). The authors discuss the issues related to contrast curves, and aim to present a metric tool for representing the performance of a high dynamic range exoplanet imaging system without any assumptions about the distribution of the noise.

I argue that a data-driven approach to the calculation of ROC curves is the best method for assessing the performance of different HCI algorithms. The ROC curves shown in In Chapter 6 (the case of a single dataset) are generated in the following way:

1. An on-sky dataset is chosen. Any high S/N or known companion is removed using the NEGFC technique.

2. A separation from the star ($1\times$FWHM annulus) and a planet to star contrast interval (the brightness of the injected companions) are selected. Also a list of detection thresholds (τ) is defined.

3. A large enough number of data cubes are built with a single injected companion at the selected separation and within the chosen contrast interval.

4. The data cubes are processed with each algorithm involved in the performance assessment/comparison. Panel (a) of Fig. A.3 shows the resulting residual flux frames for the reference PSF subtraction approaches. Panel (b) shows the resulting probability maps of SODIRF and SODINN. S/N maps are produced from the residual flux frames (see panel (b) of Fig. A.3).

5. Detection maps are obtained by thresholding the S/N and probability maps for different values of τ (see panels (c), (d) and (e) of Fig. A.3). For each detection map and for each τ, a true positive is counted if a blob is recovered at the injection location. False positives are other significant blobs at any other location in the detection map.

6. For each τ, the true positives and the number of false positives are averaged.

After the first step, we assume that the chosen dataset is empty or free of astrophysical exoplanetary signal. This is done based on visual vetting performed on a reference PSF-subtracted residual image. As shown in Chapter 6, the PSF-subtraction methods have limited sensitivity and therefore we cannot have 100% probability of having an

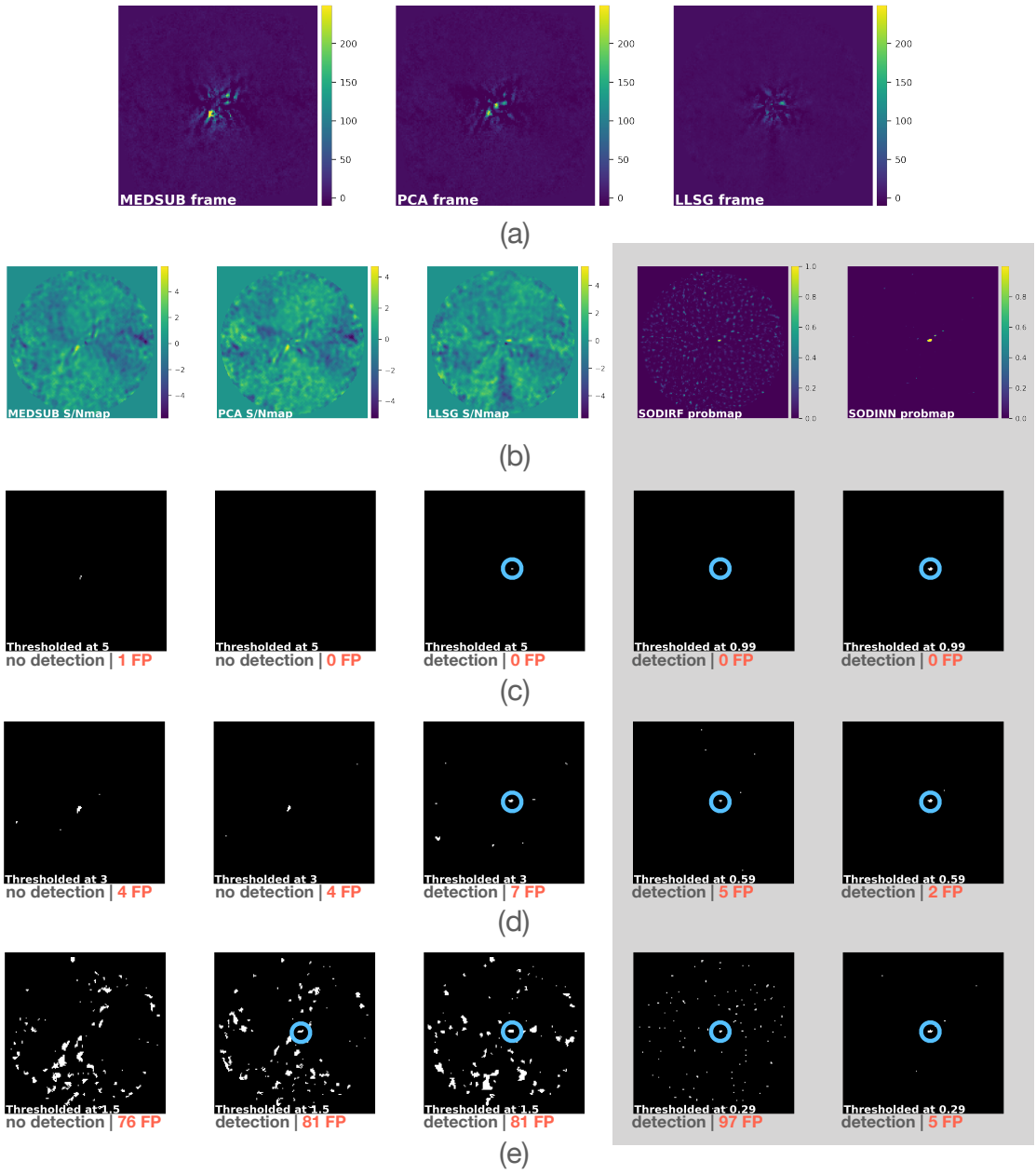


Figure A.3: Example of the final frames and S/N maps of ADI median subtraction, ADI-PCA and LLSG. Probability and detection maps of SODIRF and SODINN. Panel (a) groups the final residual frames for the reference PSF subtraction approaches. Panel (b) shows the S/N maps obtained from the previous residual flux frames and the probability maps of SODIRF and SODINN. Panels (c), (d) and (e) show the detection maps obtained from the thresholded S/N and probability maps of panel (b). The detected fake companion is shown with a blue circle on the detection maps. The detection state and the number of FPs are also shown next to each detection map. Notice that the number of FPs grows when the detection threshold is decreased (also that SODINN controls the FPs). A large number of these injections (with varying flux and position) are performed in order to build the ROC curves.

empty dataset. We have to work under this assumption and will flag any potential companion as a false positive, in the next steps. In the last step, averaging the number of false positives (instead of assuming a static noise realization per τ and counting once) addresses small fluctuations in this value, caused by the interaction of an injected companion with the false positives at the same separation (which biases the S/N).

A.2 DATA CHALLENGES AND STANDARDIZED HCI DATABASES

Citizen or crowd-sourced science initiatives are popular ways of engaging a large public in science projects (see e.g. Zooniverse project[1] and Scistarter [2]). They have also become popular in Astronomy with projects such as GalaxyZoo[3], Planet Hunters[4], and Disk Detective[5]. For instance Disk Detective has produced a published study that includes external (non-scientific) collaborators[6]. Preparing such projects is an involved process, because the problem must be well focused (narrow) and carefully defined. These projects are suitable for large amounts of data (surveys, where a repetitive task requires little brainpower).

A more relevant type of data challenge for HCI are the Kaggle[7] competitions. The Kaggle data science community focuses on data challenges of all shapes and sizes, open to skilled data scientists and machine learning experts (to anyone in fact). I proposed to create a Kaggle competition focused on HCI exoplanet detection[8]. For such a competition, the task must be defined as a supervised learning problem, and proposed along with appropriate metrics of evaluation (contrast vs separation, S/N, accuracy, sensitivity, ROC curves, AUC, others). This could be a relatively easy way to get access to brainpower from the worlds most skilled data scientists, either for free (the incentive is the status or ranking in the Kaggle community) or for an offered reward. An important issue here is that the data (sequence of calibrated images, PSF, and other related information) has to be made public. An example of the remarkable success of such competitions is the paper published by Dieleman et al. (2015) as a (winning) solution to the Galaxy Zoo Kaggle competition[9].

1 www.zooniverse.org
2 www.scistarter.com
3 www.galaxyzoo.org
4 www.planethunters.org
5 www.diskdetective.org
6 www.arxiv.org/abs/1610.05293
7 www.kaggle.com
8 At the KISS workshop on Exoplanet Imaging, Caltech, Pasadena, US (August 2016).
9 www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Another way of carrying out data challenges is by inviting the HCI community to participate and apply home-made methods to a subset of agreed datasets. This was proposed by Lawson et al. (2012). This brings the issue of creating a standardized database of HCI datasets, which I also proposed in the context of HCI[10]. The creation of such database is critical for the task of *fair* and robust comparison of algorithmic performances. Doing so, different processing algorithms could be tested under the same conditions. This has to be done as a community effort and such initiative has been started by the SAG19 study group led by Dimitri Mawet. What I propose here is to learn from the experience of the computer vision community, which has vast experience carrying out such challenges within their own field. Great examples of such databases for worldwide challenges are the Detection change[11] video database and the ImageNet[12] database.

Such a standardized database for HCI must consist of a set of datasets, agreed on by experts in the community, that are representative of the wide range of conditions and scenarios possible (e.g. rotation/wavelength range, integration time, number of frames, weather conditions, wavefront control system performance, coronagraphic solution). The datasets must be carefully labeled and the ground truth or set of labels must be provided. Also, a set of metrics should be chosen, as in the case of a Kaggle competition. With this approach, researchers would be able to use the standardized database and the metrics, accepted by the community, when assessing the performance of HCI algorithms.

10 At the KISS workshop on Exoplanet Imaging, Caltech, Pasadena, US (August 2016).
11 www.changedetection.net
12 www.image-net.org

B | REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Absil, O., Milli, J., Mawet, D., Lagrange, A.-M., Girard, J., Chauvin, G., Boccaletti, A., Delacroix, C., and Surdej, J. (2013). Searching for companions down to 2 AU from β Pictoris using the L′-band AGPM coronagraph on VLT/NACO. *A&A*, 559:L12.

Absil, P.-A., Mahony, R. E., and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds.* Princeton University Press.

Allison, R. and Dunkley, J. (2014). Comparison of sampling techniques for Bayesian parameter estimation. *MNRAS*, 437:3918–3928.

Amara, A. and Quanz, S. P. (2012). PYNPOINT: an image processing package for finding exoplanets. *MNRAS*, 427:948–955.

Apai, D., Kasper, M., Skemer, A., Hanson, J. R., Lagrange, A.-M., Biller, B. A., Bonnefoy, M., Buenzli, E., and Vigan, A. (2016). High-cadence, High-contrast Imaging for Exoplanet Mapping: Observations of the HR 8799 Planets with VLT/SPHERE Satellite-spot-corrected Relative Photometry. *ApJ*, 820:40.

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O. (2013). Astropy: A community Python package for astronomy. *A&A*, 558:A33.

Ball, N. M. and Brunner, R. J. (2010). Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics D*, 19:1049–1106.

Baraffe, I., Homeier, D., Allard, F., and Chabrier, G. (2015). New evolutionary models for pre-main sequence and main sequence low-mass stars down to the hydrogen-burning limit. *A&A*, 577:A42.

Barrett, H. H., Myers, K. J., Devaney, N., Dainty, J. C., and Caucci, L. (2006). Task performance in astronomical adaptive optics. *Proc. SPIE*, 6272:1W.

Baudoz, P., Dorn, R. J., Lizon, J.-L., Fusco, T., Dohlen, K., Charton, J., Beuzit, J.-L., Puget, P., Mouillet, D., Felt, M., Wildi, F., Barrufolo, A., Kasper, M., and Hubin, N. (2010). The differential tip-tilt sensor of SPHERE. In *Ground-based and Airborne Instrumentation for Astronomy III*, volume 7735 of *Proc. SPIE*, page 77355B.

Bergfors, C., Brandner, W., Janson, M., Köhler, R., and Henning, T. (2011). VLT/NACO astrometry of the HR 8799 planetary system. L'-band observations of the three outer planets. *A&A*, 528:A134.

Berry, R. and Burnell, J. (2000). *The handbook of astronomical image processing*.

Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *A&AS*, 117:393–404.

Beust, H., Bonnefoy, M., Maire, A.-L., Ehrenreich, D., Lagrange, A.-M., and Chauvin, G. (2016). Orbital fitting of imaged planetary companions with high eccentricities and unbound orbits. Their application to Fomalhaut b and PZ Telecopii B. *A&A*, 587:A89.

Beuzit, J.-L., Feldt, M., Dohlen, K., Mouillet, D., Puget, P., Wildi, F., Abe, L., Antichi, J., Baruffolo, A., Baudoz, P., Boccaletti, A., Carbillet, M., Charton, J., Claudi, R., Downing, M., Fabron, C., Feautrier, P., Fedrigo, E., Fusco, T., Gach, J.-L., Gratton, R., Henning, T., Hubin, N., Joos, F., Kasper, M., Langlois, M., Lenzen, R., Moutou, C., Pavlov, A., Petit, C., Pragt, J., Rabou, P., Rigal, F., Roelfsema, R., Rousset, G., Saisse, M., Schmid, H.-M., Stadler, E., Thalmann, C., Turatto, M., Udry, S., Vakili, F., and Waters, R. (2008). SPHERE: a 'Planet Finder' instrument for the VLT. In *Ground-based and Airborne Instrumentation for Astronomy II*, volume 7014 of *Proc. SPIE*, page 701418.

Bonnefoy, M., Zurlo, A., Baudino, J. L., Lucas, P., Mesa, D., Maire, A.-L., Vigan, A., Galicher, R., Homeier, D., Marocco, F., Gratton, R., Chauvin, G., Allard, F., Desidera, S., Kasper, M., Moutou, C., Lagrange, A.-M., Antichi, J., Baruffolo, A., Baudrand, J., Beuzit, J.-L., Boccaletti, A., Cantalloube, F., Carbillet, M., Charton, J., Claudi,

R. U., Costille, A., Dohlen, K., Dominik, C., Fantinel, D., Feautrier, P., Feldt, M., Fusco, T., Gigan, P., Girard, J. H., Gluck, L., Gry, C., Henning, T., Janson, M., Langlois, M., Madec, F., Magnard, Y., Maurel, D., Mawet, D., Meyer, M. R., Milli, J., Moeller-Nilsson, O., Mouillet, D., Pavlov, A., Perret, D., Pujet, P., Quanz, S. P., Rochat, S., Rousset, G., Roux, A., Salasnich, B., Salter, G., Sauvage, J.-F., Schmid, H. M., Sevin, A., Soenke, C., Stadler, E., Turatto, M., Udry, S., Vakili, F., Wahhaj, Z., and Wildi, F. (2016). First light of the VLT planet finder SPHERE. IV. Physical and chemical properties of the planets around HR8799. *A&A*, 587:A58.

Booth, M., Jordán, A., Casassus, S., Hales, A. S., Dent, W. R. F., Faramaz, V., Matrà, L., Barkats, D., Brahm, R., and Cuadra, J. (2016). Resolving the planetesimal belt of HR 8799 with ALMA. *MNRAS*, 460:L10–L14.

Boser, B. E., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Computational Learing Theory*, pages 144–152.

Bouwmans, T. and Zahzah, E.-H. (2014). Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34.

Bowler, B. P. (2016). Imaging Extrasolar Giant Planets. *PASP*, 128(10):102001.

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*.

Braham, M., Lejeune, A., and Van Droogenbroeck, M. (2014). A physically motivated pixel-based model for background subtraction in 3D images. In *International Conference on 3D Imaging (IC3D)*, pages 1–8, Liège, Belgium.

Braham, M. and Van Droogenbroeck, M. (2016). Deep background subtraction with scene-specific convolutional neural networks. In *IWSSIP*, pages 1–4, Bratislava, Slovakia.

Brandt, T. D., McElwain, M. W., Turner, E. L., Abe, L., Brandner, W., Carson, J., Egner, S., Feldt, M., Golota, T., Goto, M., Grady, C. A., Guyon, O., Hashimoto, J., Hayano, Y., Hayashi, M., Hayashi, S., Henning, T., Hodapp, K. W., Ishii, M., Iye, M., Janson, M., Kandori, R., Knapp, G. R., Kudo, T., Kusakabe, N., Kuzuhara, M., Kwon, J., Matsuo, T., Miyama, S., Morino, J.-I., Moro-Martín, A., Nishimura, T., Pyo, T.-S., Serabyn, E., Suto, H., Suzuki, R., Takami, M., Takato, N., Terada, H., Thalmann, C., Tomono, D., Watanabe, M., Wisniewski, J. P., Yamada, T., Takami, H., Usuda, T., and Tamura, M. (2013). New Techniques for High-contrast Imaging with ADI: The ACORNS-ADI SEEDS Data Reduction Pipeline. *ApJ*, 764:183.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Breiman, L., Friedman, J., Olshen, A., and C.J., S. (1984). *Classification and regression trees*. Wadsworth.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2009). Robust principal component analysis? *CoRR*, abs/0912.3599.

Cantalloube, F., Mouillet, D., Mugnier, L. M., Milli, J., Absil, O., Gomez Gonzalez, C. A., Chauvin, G., Beuzit, J.-L., and Cornia, A. (2015). Direct exoplanet detection and characterization using the ANDROMEDA method: Performance on VLT/NaCo data. *A&A*, 582:A89.

Carbillet, M., Bendjoya, P., Abe, L., Guerri, G., Boccaletti, A., Daban, J.-B., Dohlen, K., Ferrari, A., Robbe-Dubois, S., Douet, R., and Vakili, F. (2011). Apodized Lyot coronagraph for SPHERE/VLT. I. Detailed numerical study. *Experimental Astronomy*, 30:39–58.

Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M. (2000). Detection of Planetary Transits Across a Sun-like Star. *ApJ*, 529:L45–L48.

Chauvin, G., Lagrange, A.-M., Beust, H., Bonnefoy, M., Boccaletti, A., Apai, D., Allard, F., Ehrenreich, D., Girard, J. H. V., Mouillet, D., and Rouan, D. (2012). Orbital characterization of the β Pictoris b giant planet. *A&A*, 542:A41.

Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., and Lowrance, P. (2005). Giant planet companion to 2MASSW J1207334-393254. *A&A*, 438:L25–L28.

Chollet, F. (2017). *Deep Learning with Python*. Manning Publications Company.

Chollet, F. et al. (2015). Keras. https://github.com/fchollet/keras.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904.

Currie, T., Burrows, A., Girard, J. H., Cloutier, R., Fukagawa, M., Sorahana, S., Kuchner, M., Kenyon, S. J., Madhusudhan, N., Itoh, Y., Jayawardhana, R., Matsumura, S., and Pyo, T.-S. (2014). Deep Thermal Infrared Imaging of HR 8799 bcde: New Atmospheric Constraints and Limits on a Fifth Planet. *ApJ*, 795:133.

Currie, T., Fukagawa, M., Thalmann, C., Matsumura, S., and Plavchan, P. (2012a). Direct Detection and Orbital Analysis of the Exoplanets HR 8799 bcd from Archival 2005 Keck/NIRC2 Data. *ApJ*, 755:L34.

Currie, T., Fukagawa, M., Thalmann, C., Matsumura, S., and Plavchan, P. (2012b). Direct Detection and Orbital Analysis of the Exoplanets HR 8799 bcd from Archival 2005 Keck/NIRC2 Data. *ApJ*, 755:L34.

Dauphin, Y. N., de Vries, H., Chung, J., and Bengio, Y. (2015). Rmsprop and equilibrated adaptive learning rates for non-convex optimization. *CoRR*, abs/1502.04390.

Defrère, D., Absil, O., Hinz, P., Mawet, D., Kuhn, J., Mawet, D., Mennesson, B., Skemer, A., Wallace, K., Bailey, V., Downey, E., Delacroix, C., Durney, O., Forsberg, P., Gomez, C., Habraken, S., Karlsson, M., Kenworthy, M., Leisenring, J., Montoya, M., Pueyo, L., Skrutskie, M., and Surdej, J. (2014). L'-band AGPM vector vortex coronagraph's first light on LBTI/LMIRCAM. In Apai, D. and Gabor, P., editors, *Search for Life Beyond the Solar System. Exoplanets, Biosignatures and Instruments*, page P4.75.

Dieleman, S., Willett, K. W., and Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *MNRAS*, 450:1441–1459.

Dohlen, K., Langlois, M., Saisse, M., Hill, L., Origne, A., Jacquet, M., Fabron, C., Blanc, J.-C., Llored, M., Carle, M., Moutou, C., Vigan, A., Boccaletti, A., Carbillet, M., Mouillet, D., and Beuzit, J.-L. (2008). The infra-red dual imaging and spectrograph for SPHERE: design and performance. In *Ground-based and Airborne Instrumentation for Astronomy II*, volume 7014 of *Proc. SPIE*, page 70143L.

Duchi, J. C., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159.

Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for deep learning. *CoRR*, abs/1603.07285.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

Esposito, S., Mesa, D., Skemer, A., Arcidiacono, C., Claudi, R. U., Desidera, S., Gratton, R., Mannucci, F., Marzari, F., Masciadri, E., Close, L., Hinz, P., Kulesa, C., McCarthy, D., Males, J., Agapito, G., Argomedo, J., Boutsia, K., Briguglio, R., Brusa, G., Busoni, L., Cresci, G., Fini, L., Fontana, A., Guerra, J. C., Hill, J. M., Miller, D., Paris, D., Pinna, E., Puglisi, A., Quiros-Pacheco, F., Riccardi, A., Stefanini, P., Testa, V., Xompero, M., and Woodward, C. (2013). LBT observations of the HR 8799 planetary system. First detection of HR 8799e in H band. *A&A*, 549:A52.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press.

Fabrycky, D. C. and Murray-Clay, R. A. (2010). Stability of the Directly Imaged Multiplanet System HR 8799: Resonance and Masses. *ApJ*, 710:1408–1421.

Fergus, R., Hogg, D. W., Oppenheimer, R., Brenner, D., and Pueyo, L. (2014). S4: A Spatial-spectral model for Speckle Suppression. *ApJ*, 794:161.

Feroz, F. and Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384:449–463.

Fischer, D. A., Howard, A. W., Laughlin, G. P., Macintosh, B., Mahadevan, S., Sahlmann, J., and Yee, J. C. (2014). Exoplanet Detection Techniques. *Protostars and Planets VI*, pages 715–737.

Fitzgerald, M. P. and Graham, J. R. (2006). Speckle Statistics in Adaptively Corrected Images. *ApJ*, 637:541–547.

Flamary, R. (2016). Astronomical image reconstruction with convolutional neural networks. *CoRR*, abs/1612.04526.

Ford, E. B. (2005). Quantifying the Uncertainty in the Orbits of Extrasolar Planets. *AJ*, 129:1706–1717.

Ford, E. B. (2006). Improving the Efficiency of Markov Chain Monte Carlo for Analyzing the Orbits of Extrasolar Planets. *ApJ*, 642:505–522.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *PASP*, 125:306–312.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29.

Frontera-Pons, J., Sureau, F., Bobin, J., and Le Floc'h, E. (2017). Unsupervised feature-learning for galaxy SEDs with denoising autoencoders. *ArXiv e-prints*.

Fusco, T., Rousset, G., Sauvage, J.-F., Petit, C., Beuzit, J.-L., Dohlen, K., Mouillet, D., Charton, J., Nicolle, M., Kasper, M., Baudoz, P., and Puget, P. (2006). High-order adaptive optics requirements for direct detection of extrasolar planets: Application to the SPHERE instrument. *Optics Express*, 14:7515.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.*, 7(4):457–472.

Gillon, M., Triaud, A. H. M. J., Demory, B.-O., Jehin, E., Agol, E., Deck, K. M., Lederer, S. M., de Wit, J., Burdanov, A., Ingalls, J. G., Bolmont, E., Leconte, J., Raymond, S. N., Selsis, F., Turbet, M., Barkaoui, K., Burgasser, A., Burleigh, M. R., Carey, S. J., Chaushev, A., Copperwheat, C. M., Delrez, L., Fernandes, C. S., Holdsworth, D. L., Kotze, E. J., Van Grootel, V., Almleaky, Y., Benkhaldoun, Z., Magain, P., and Queloz, D. (2017). Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature*, 542:456–460.

Givens, G. H. and Hoeting, J. A. (2012). *Computational statistics*, volume 710. John Wiley & Sons.

Gomez Gonzalez, C. A., Absil, O., Absil, P.-A., Van Droogenbroeck, M., Mawet, D., and Surdej, J. (2016a). Low-rank plus sparse decomposition for exoplanet detection in direct-imaging ADI sequences. The LLSG algorithm. *A&A*, 589:A54.

Gomez Gonzalez, C. A., Wertz, O., Absil, O., Christiaens, V., Defrère, D., Mawet, D., Milli, J., Absil, P.-A., Van Droogenbroeck, M., Cantalloube, F., Hinz, P. M., Skemer, A. J., Karlsson, M., and Surdej, J. (2017). VIP: Vortex Image Processing Package for High-contrast Direct Imaging. *AJ*, 154:7.

Gomez Gonzalez, C. A., Wertz, O., Christiaens, V., Absil, O., and Mawet, D. (2015). VIP: Vortex Image Processing Package for High-Contrast Direct Imaging.

Gomez Gonzalez, C. A., Wertz, O., Christiaens, V., Absil, O., and Mawet, D. (2016b). VIP: Vortex Image Processing pipeline for high-contrast direct imaging of exoplanets. Astrophysics Source Code Library.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invarianve. *Commun. Appl. Math. Comput. Sci.*, 5(1):65.

Goździewski, K. and Migaszewski, C. (2009). Is the HR8799 extrasolar system destined for planetary scattering? *MNRAS*, 397:L16–L20.

Goździewski, K. and Migaszewski, C. (2014a). Multiple mean motion resonances in the HR 8799 planetary system. *MNRAS*, 440:3140–3171.

Goździewski, K. and Migaszewski, C. (2014b). Multiple mean motion resonances in the HR 8799 planetary system. *MNRAS*, 440:3140–3171.

Guerri, G., Daban, J.-B., Robbe-Dubois, S., Douet, R., Abe, L., Baudrand, J., Carbillet, M., Boccaletti, A., Bendjoya, P., Gouvret, C., and

Vakili, F. (2011). Apodized Lyot coronagraph for SPHERE/VLT: II. Laboratory tests and performance. *Experimental Astronomy*, 30:59–81.

Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R. (2008). Efficient subpixel image registration algorithms. *Opt. Lett.*, 33(2):156–158.

Guyon, O. (2005). Limits of Adaptive Optics for High-Contrast Imaging. *ApJ*, 629:592–614.

Guyon, O., Bendek, E. A., Eisner, J. A., Angel, R., Woolf, N. J., Milster, T. D., Ammons, S. M., Shao, M., Shaklan, S., Levine, M., Nemati, B., Pitman, J., Woodruff, R. A., and Belikov, R. (2012). High-precision Astrometry with a Diffractive Pupil Telescope. *ApJS*, 200:11.

Hagelberg, J., Ségransan, D., Udry, S., and Wildi, F. (2016). The Geneva Reduction and Analysis Pipeline for High-contrast Imaging of planetary Companions. *MNRAS*, 455:2178–2186.

Halko, N., Martinsson, P.-G., and Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288.

Hardy, A., Schreiber, M. R., Parsons, S. G., Caceres, C., Retamales, G., Wahhaj, Z., Mawet, D., Canovas, H., Cieza, L., Marsh, T. R., Bours, M. C. P., Dhillon, V. S., and Bayo, A. (2015). The First Science Results from Sphere: Disproving the Predicted Brown Dwarf Around V471 Tau. *ApJ*, 800:L24.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning*. Springer Verlag.

Hibon, P., Vigan, A., Dohlen, K., Milli, J., Girard, J., Beuzit, J.-L., and Mouillet, D. (2016). Characterization of the Atmospheric Dispersion Correctors from VLT/SPHERE. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Proc. SPIE*, pages 9908–115.

Hill, J. M., Ashby, D. S., Brynnel, J. G., Christou, J. C., Little, J. K., Summers, D. M., Veillet, C., and Wagner, R. M. (2014). The Large Binocular Telescope: binocular all the time. In *Ground-based and Airborne Telescopes V*, volume 9145 of *Proc. SPIE*, page 914502.

Hinz, P., Bailey, V. P., Defrere, D., Downey, E., Esposito, S., Hill, J., Hoffmann, W. F., Leisenring, J., Montoya, M., McMahon, T., Puglisi, A., Skemer, A., Skrutskie, M., Vaitheeswaran, V., and Vaz, A. (2014). Commissioning the lbti for use as a nulling interferometer and coherent imager. volume 9146, pages 91460T–91460T–10.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. 9(8):1735–1780.

Hogg, D. W., Bovy, J., and Lang, D. (2010). Data analysis recipes: Fitting a model to data. *ArXiv e-prints*.

Hotelling, H. (1933). Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology*, 24:417–441.

Hoyle, B. (2016). Measuring photometric redshifts using galaxy images and Deep Neural Networks. *Astronomy and Computing*, 16:34–40.

Hughes, A. M., Wilner, D. J., Andrews, S. M., Williams, J. P., Su, K. Y. L., Murray-Clay, R. A., and Qi, C. (2011). Resolved Submillimeter Observations of the HR 8799 and HD 107146 Debris Disks. *ApJ*, 740:38.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95.

Hunziker, S., Quanz, S. P., Amara, A., and Meyer, M. R. (2017). A PCA-based approach for subtracting thermal background emission in high-contrast imaging data. *ArXiv e-prints*.

Jee, M. J., Blakeslee, J. P., Sirianni, M., Martel, A. R., White, R. L., and Ford, H. C. (2007). Principal Component Analysis of the Time- and Position-dependent Point-Spread Function of the Advanced Camera for Surveys. *PASP*, 119:1403–1419.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python.

Kalas, P., Graham, J. R., Fitzgerald, M. P., and Clampin, M. (2013). STIS Coronagraphic Imaging of Fomalhaut: Main Belt Structure and the Orbit of Fomalhaut b. *ApJ*, 775:56.

Kenworthy, M. A., Codona, J. L., Hinz, P. M., Angel, J. R. P., Heinze, A., and Sivanandam, S. (2007). First On-Sky High-Contrast Imaging with an Apodizing Phase Plate. *ApJ*, 660:762–769.

Kim, E. J. and Brunner, R. J. (2017). Star-galaxy classification using deep convolutional neural networks. *MNRAS*, 464:4463–4475.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Konopacky, Q. M., Marois, C., Macintosh, B. A., Galicher, R., Barman, T. S., Metchev, S. A., and Zuckerman, B. (2016). Astrometric Monitoring of the HR 8799 Planets: Orbit Constraints from Self-consistent Measurements. *AJ*, 152:28.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kumar, N. K. and Shneider, J. (2016). Literature survey on low rank approximation of matrices. *CoRR*, abs/1606.06511.

Lafrenière, D., Marois, C., Doyon, R., and Barman, T. (2009). HST/NICMOS Detection of HR 8799 b in 1998. *ApJ*, 694:L148–L152.

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., and Artigau, É. (2007). A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging. *ApJ*, 660:770–780.

Lagrange, A.-M., Bonnefoy, M., Chauvin, G., Apai, D., Ehrenreich, D., Boccaletti, A., Gratadour, D., Rouan, D., Mouillet, D., Lacour, S., and Kasper, M. (2010). A Giant Planet Imaged in the Disk of the Young Star β Pictoris. *Science*, 329:57–.

Langlois, M., Vigan, A., Moutou, C., Dohlen, K., Costille, A., Le Mignant, D., Martinez, P., Mouillet, D., Boccaletti, A., Moeller-Nilsson, O., Sauvage, J.-F., Mugnier, L., Feldt, M., Gry, C., Wildi, F., and Beuzit, J.-L. (2012). Infrared differential imager and spectrograph for SPHERE: performance status with extreme adaptive optics before shipment to ESO/VLT. In *Adaptive Optics Systems III*, volume 8447 of *Proc. SPIE*, page 84473B.

Lawson, P. R., Poyneer, L., Barrett, H., Frazin, R., Caucci, L., Devaney, N., Furenlid, L., Gładysz, S., Guyon, O., Krist, J., Maire, J., Marois, C., Mawet, D., Mouillet, D., Mugnier, L., Pearson, I., Perrin, M., Pueyo, L., and Savransky, D. (2012). On advanced estimation techniques for exoplanet detection and characterization using ground-based coronagraphs. In *Adaptive Optics Systems III*, volume 8447 of *Proc. SPIE*, page 844722.

LeCun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E., and Hubbard, W. (1989). Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27(11):41–46.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative factorization. *Nature*, 401:788.

Levy, A. and Lindenbaum, M. (2000). Sequential Karhunen-Loeve basis extraction and its application to images. *IEEE Trans. Image Processing*, 9(8):1371–1374.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Maire, A.-L., Langlois, M., Dohlen, K., Lagrange, A.-M., Gratton, R., Chauvin, G., Desidera, S., Girard, J. H., Milli, J., Vigan, A., Zins, G., Delorme, P., Beuzit, J.-L., Claudi, R. U., Feldt, M., Mouillet, D., Puget, P., Turatto, M., and Wildi, F. (2016). SPHERE IRDIS and IFS astrometric strategy and calibration. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Proc. SPIE*, pages 9908–34.

Maire, A.-L., Skemer, A. J., Hinz, P. M., Desidera, S., Esposito, S., Gratton, R., Marzari, F., Skrutskie, M. F., Biller, B. A., Defrère, D., Bailey, V. P., Leisenring, J. M., Apai, D., Bonnefoy, M., Brandner, W., Buenzli, E., Claudi, R. U., Close, L. M., Crepp, J. R., De Rosa, R. J., Eisner, J. A., Fortney, J. J., Henning, T., Hofmann, K.-H., Kopytova, T. G., Males, J. R., Mesa, D., Morzinski, K. M., Oza, A., Patience, J., Pinna, E., Rajan, A., Schertl, D., Schlieder, J. E., Su, K. Y. L., Vaz, A., Ward-Duong, K., Weigelt, G., and Woodward, C. E. (2015a). The LEECH Exoplanet Imaging Survey. Further constraints on the planet architecture of the HR 8799 system. *A&A*, 576:A133.

Maire, A.-L., Skemer, A. J., Hinz, P. M., Desidera, S., Esposito, S., Gratton, R., Marzari, F., Skrutskie, M. F., Biller, B. A., Defrère, D., Bailey, V. P., Leisenring, J. M., Apai, D., Bonnefoy, M., Brandner, W., Buenzli, E., Claudi, R. U., Close, L. M., Crepp, J. R., De Rosa, R. J., Eisner, J. A., Fortney, J. J., Henning, T., Hofmann, K.-H., Kopytova, T. G., Males, J. R., Mesa, D., Morzinski, K. M., Oza, A., Patience, J., Pinna, E., Rajan, A., Schertl, D., Schlieder, J. E., Su, K. Y. L., Vaz, A., Ward-Duong, K., Weigelt, G., and Woodward, C. E. (2015b). The LEECH Exoplanet Imaging Survey. Further constraints on the planet architecture of the HR 8799 system. *A&A*, 576:A133.

Marois, C., Correia, C., Galicher, R., Ingraham, P., Macintosh, B., Currie, T., and De Rosa, R. (2014). GPI PSF subtraction with TLOCI: the next evolution in exoplanet/disk high-contrast imaging. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9148 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0.

Marois, C., Doyon, R., Nadeau, D., Racine, R., and Walker, G. A. H. (2003). Effects of Quasi-Static Aberrations in Faint Companion Searches. In Aime, C. and Soummer, R., editors, *EAS Publications Series*, volume 8 of *EAS Publications Series*, pages 233–243.

Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., and Nadeau, D. (2006). Angular Differential Imaging: A Powerful High-Contrast Imaging Technique. *ApJ*, 641:556–564.

Marois, C., Lafrenière, D., Macintosh, B., and Doyon, R. (2008a). Confidence Level and Sensitivity Limits in High-Contrast Imaging. *ApJ*, 673:647–656.

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., and Doyon, R. (2008b). Direct Imaging of Multiple Planets Orbiting the Star HR 8799. *Science*, 322:1348.

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., and Doyon, R. (2008c). Direct Imaging of Multiple Planets Orbiting the Star HR 8799. *Science*, 322:1348–.

Marois, C., Macintosh, B., and Véran, J.-P. (2010a). Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline. In *Adaptive Optics Systems II*, volume 7736 of *Proc. SPIE*, page 77361J.

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., and Barman, T. (2010b). Images of a fourth planet orbiting HR 8799. *Nature*, 468:1080–1083.

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., and Barman, T. (2010c). Images of a fourth planet orbiting HR 8799. *Nature*, 468:1080–1083.

Martinez, P., Kasper, M., Costille, A., Sauvage, J. F., Dohlen, K., Puget, P., and Beuzit, J. L. (2013). Speckle temporal stability in XAO coronagraphic images. II. Refine model for quasi-static speckle temporal evolution for VLT/SPHERE. *A&A*, 554:A41.

Masias, M., Freixenet, J., Lladó, X., and Peracaula, M. (2012). A review of source detection approaches in astronomical images. *MNRAS*, 422:1674–1689.

Matthews, B., Kennedy, G., Sibthorpe, B., Booth, M., Wyatt, M., Broekhoven-Fiene, H., Macintosh, B., and Marois, C. (2014). Resolved Imaging of the HR 8799 Debris Disk with Herschel. *ApJ*, 780:97.

Mawet, D., Absil, O., Delacroix, C., Girard, J. H., Milli, J., O'Neal, J., Baudoz, P., Boccaletti, A., Bourget, P., Christiaens, V., Forsberg, P., Gonte, F., Habraken, S., Hanot, C., Karlsson, M., Kasper, M., Lizon, J.-L., Muzic, K., Olivier, R., Peña, E., Slusarenko, N., Tacconi-Garman, L. E., and Surdej, J. (2013). L'-band AGPM vector vortex coronagraph's first light on VLT/NACO. Discovery of a late-type companion at two beamwidths from an FoV star. *A&A*, 552:L13.

Mawet, D., David, T., Bottom, M., Hinkley, S., Stapelfeldt, K., Padgett, D., Mennesson, B., Serabyn, E., Morales, F., and Kuhn, J. (2015). Discovery of a Low-mass Companion Around HR 3549. *ApJ*, 811:103.

Mawet, D., Milli, J., Wahhaj, Z., Pelat, D., Absil, O., Delacroix, C., Boccaletti, A., Kasper, M., Kenworthy, M., Marois, C., Mennesson, B., and Pueyo, L. (2014). Fundamental Limitations of High Contrast Imaging Set by Small Sample Statistics. *ApJ*, 792:97.

Mawet, D., Pueyo, L., Lawson, P., Mugnier, L., Traub, W., Boccaletti, A., Trauger, J. T., Gladysz, S., Serabyn, E., Milli, J., Belikov, R., Kasper, M., Baudoz, P., Macintosh, B., Marois, C., Oppenheimer, B., Barrett, H., Beuzit, J.-L., Devaney, N., Girard, J., Guyon, O., Krist, J., Mennesson, B., Mouillet, D., Murakami, N., Poyneer, L., Savransky, D., Vérinaud, C., and Wallace, J. K. (2012). Review of small-angle coronagraphic techniques in the wake of ground-based second-generation adaptive optics systems. In Clampin, M. C., Fazio, G. G., MacEwen, H. A., and Oschmann, J. M., editors, *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, volume 8442 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 04.

Mawet, D., Riaud, P., Absil, O., and Surdej, J. (2005). Annular Groove Phase Mask Coronagraph. *ApJ*, 633:1191–1200.

Mawet, D., Serabyn, E., Liewer, K., Burruss, R., Hickey, J., and Shemo, D. (2010). The Vector Vortex Coronagraph: Laboratory Results and First Light at Palomar Observatory. *ApJ*, 709:53–57.

Mawet, D., Serabyn, E., Stapelfeldt, K., and Crepp, J. (2009). Imaging the Debris Disk of HD 32297 with a Phase-Mask Coronagraph at High Strehl Ratio. *ApJ*, 702:L47–L50.

Mayor, M. and Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, 378:355–359.

McKinney, W. (2010). Data structures for statistical computing in python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56.

Meeus, J. (1998). *Astronomical algorithms*. Richmond, Willmann-Bell.

Mensch, A., Mairal, J., Thirion, B., and Varoquaux, G. (2016). Dictionary learning for massive matrix factorization. *CoRR*, abs/1605.00937.

Meshkat, T., Kenworthy, M. A., Quanz, S. P., and Amara, A. (2014a). Optimized Principal Component Analysis on Coronagraphic Images of the Fomalhaut System. *ApJ*, 780:17.

Meshkat, T., Kenworthy, M. A., Quanz, S. P., and Amara, A. (2014b). Optimized Principal Component Analysis on Coronagraphic Images of the Fomalhaut System. *ApJ*, 780:17.

Milli, J., Mawet, D., Mouillet, D., Kasper, M., and Girard, J. H. (2016). Adaptive Optics in High-Contrast Imaging. *Astronomy at High Angular Resolution*, 439:17.

Morzinski, K. M., Males, J. R., Skemer, A. J., Close, L. M., Hinz, P. M., Rodigas, T. J., Puglisi, A., Esposito, S., Riccardi, A., Pinna, E., Xompero, M., Briguglio, R., Bailey, V. P., Follette, K. B., Kopon, D., Weinberger, A. J., and Wu, Y.-L. (2015). Magellan Adaptive Optics First-light Observations of the Exoplanet β Pic b. II. 3-5 μm Direct Imaging with MagAO+Clio, and the Empirical Bolometric Luminosity of a Self-luminous Giant Planet. *ApJ*, 815:108.

Mugnier, L. M., Cornia, A., Sauvage, J.-F., Rousset, G., Fusco, T., and Védrenne, N. (2009). Optimal method for exoplanet detection by angular differential imaging. *Journal of the Optical Society of America A*, 26:1326.

Muterspaugh, M. W., Lane, B. F., Kulkarni, S. R., Konacki, M., Burke, B. F., Colavita, M. M., Shao, M., Hartkopf, W. I., Boss, A. P., and Williamson, M. (2010). The Phases Differential Astrometry Data Archive. V. Candidate Substellar Companions to Binary Systems. *AJ*, 140:1657–1671.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., and Zumach, W. A. (1992). Automated star/galaxy discrimination with neural networks. *AJ*, 103:318–331.

Oliver, N., Rosario, B., and Pentland, A. (2000). A Bayesian computer vision system for modeling human interactions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):831–843.

Olshausen, B. and Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Oppenheimer, B. R., Baranec, C., Beichman, C., Brenner, D., Burruss, R., Cady, E., Crepp, J. R., Dekany, R., Fergus, R., Hale, D., Hillenbrand, L., Hinkley, S., Hogg, D. W., King, D., Ligon, E. R., Lockhart, T., Nilsson, R., Parry, I. R., Pueyo, L., Rice, E., Roberts, J. E., Roberts, Jr., L. C., Shao, M., Sivaramakrishnan, A., Soummer, R., Truong, T., Vasisht, G., Veicht, A., Vescelus, F., Wallace, J. K., Zhai, C., and Zimmerman, N. (2013). Reconnaissance of the HR 8799 Exosolar System. I. Near-infrared Spectroscopy. *ApJ*, 768:24.

Pathak, P., Guyon, O., Jovanovic, N., Lozi, J., Martinache, F., Minowa, Y., Kudo, T., Takami, H., Hayano, Y., and Narita, N. (2016). A high precision technique to correct for residual atmospheric dispersion in high-contrast imaging systems. *ArXiv e-prints*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.,

Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pepe, F., Ehrenreich, D., and Meyer, M. R. (2014). Instrumentation for the detection and characterization of exoplanets. *Nature*, 513:358–366.

Pueyo, L. (2016). Detection and Characterization of Exoplanets using Projections on Karhunen Loeve Eigenimages: Forward Modeling. *ApJ*, 824:117.

Pueyo, L., Crepp, J. R., Vasisht, G., Brenner, D., Oppenheimer, B. R., Zimmerman, N., Hinkley, S., Parry, I., Beichman, C., Hillenbrand, L., Roberts, L. C., Dekany, R., Shao, M., Burruss, R., Bouchez, A., Roberts, J., and Soummer, R. (2012). Application of a Damped Locally Optimized Combination of Images Method to the Spectral Characterization of Faint Companions Using an Integral Field Spectrograph. *ApJS*, 199:6.

Pueyo, L., Soummer, R., Hoffmann, J., Oppenheimer, R., Graham, J. R., Zimmerman, N., Zhai, C., Wallace, J. K., Vescelus, F., Veicht, A., Vasisht, G., Truong, T., Sivaramakrishnan, A., Shao, M., L. C. Roberts, J., Roberts, J. E., Rice, E., Parry, I. R., Nilsson, R., Lockhart, T., Ligon, E. R., King, D., Hinkley, S., Hillenbrand, L., Hale, D., Dekany, R., Crepp, J. R., Cady, E., Burruss, R., Brenner, D., Beichman, C., and Baranec, C. (2015). Reconnaissance of the hr 8799 exosolar system. ii. astrometry and orbital motion. *The Astrophysical Journal*, 803(1):31.

Pueyo, L., Soummer, R., Hoffmann, J., Oppenheimer, R., Graham, J. R., Zimmerman, N., Zhai, C., Wallace, J. K., Vescelus, F., Veicht, A., Vasisht, G., Truong, T., Sivaramakrishnan, A., Shao, M., Roberts, Jr., L. C., Roberts, J. E., Rice, E., Parry, I. R., Nilsson, R., Lockhart, T., Ligon, E. R., King, D., Hinkley, S., Hillenbrand, L., Hale, D., Dekany, R., Crepp, J. R., Cady, E., Burruss, R., Brenner, D., Beichman, C., and Baranec, C. (2015). Reconnaissance of the HR 8799 Exosolar System. II. Astrometry and Orbital Motion. *ApJ*, 803:31.

Quanz, S. P., Amara, A., Meyer, M. R., Girard, J. H., Kenworthy, M. A., and Kasper, M. (2015). Confirmation and Characterization of the Protoplanet HD 100546 b - Direct Evidence for Gas Giant Planet Formation at 50 AU. *ApJ*, 807:64.

Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., and Marois, C. (1999). Speckle Noise and the Detection of Faint Companions. *PASP*, 111:587–594.

Rameau, J., Chauvin, G., Lagrange, A.-M., Thébault, P., Milli, J., Girard, J. H., and Bonnefoy, M. (2012). High-contrast imaging of the

close environment of HD 142527. VLT/NaCo adaptive optics thermal and angular differential imaging. *A&A*, 546:A24.

Raschka, S. (2015). *Python Machine Learning*. Packt Publishing.

Reidemeister, M., Krivov, A. V., Schmidt, T. O. B., Fiedler, S., Müller, S., Löhne, T., and Neuhäuser, R. (2009). A possible architecture of the planetary system HR 8799. *A&A*, 503:247–258.

Rodigas, T. J., Weinberger, A., Mamajek, E. E., Males, J. R., Close, L. M., Morzinski, K., Hinz, P. M., and Kaib, N. (2015). Direct Exoplanet Detection with Binary Differential Imaging. *ApJ*, 811:157.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Ross, D. A., Lim, J., Lin, R.-S., and Yang, M.-H. (2008). Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3):125–141.

Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., and Labeyrie, A. (2000). The Four-Quadrant Phase-Mask Coronagraph. I. Principle. *PASP*, 112:1479–1486.

Rousset, G., Fontanella, J. C., Kern, P., Gigan, P., and Rigaut, F. (1990). First diffraction-limited astronomical images with adaptive optics. *A&A*, 230:L29–L32.

Ruane, G., Mawet, D., Kastner, J., Meshkat, T., Bottom, M., Femenía Castellá, B., Absil, O., Gomez Gonzalez, C., Huby, E., Zhu, Z., Jenson-Clem, R., Choquet, É., and Serabyn, E. (2017). Deep Imaging Search for Planets Forming in the TW Hya Protoplanetary Disk with the Keck/NIRC2 Vortex Coronagraph. *AJ*, 154:73.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *ArXiv e-prints*.

Ruffio, J.-B., Macintosh, B., Wang, J. J., Pueyo, L., Nielsen, E. L., De Rosa, R. J., Czekala, I., Marley, M. S., Arriaga, P., Bailey, V. P., Barman, T., Bulger, J., Chilcote, J., Cotten, T., Doyon, R., Duchêne, G., Fitzgerald, M. P., Follette, K. B., Gerard, B. L., Goodsell, S. J., Graham, J. R., Greenbaum, A. Z., Hibon, P., Hung, L.-W., Ingraham, P., Kalas, P., Konopacky, Q., Larkin, J. E., Maire, J., Marchis, F., Marois, C., Metchev, S., Millar-Blanchaer, M. A., Morzinski, K. M., Oppenheimer, R., Palmer, D., Patience, J., Perrin, M., Poyneer, L., Rajan, A., Rameau, J., Rantakyrö, F. T., Savransky, D., Schneider, A. C., Sivaramakrishnan, A., Song, I., Soummer, R., Thomas, S., Wallace, J. K., Ward-Duong, K., Wiktorowicz, S., and Wolff, S. (2017). Improving and Assessing Planet Sensitivity of the GPI Exoplanet Survey with a Forward Model Matched Filter. *ApJ*, 842:14.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. and Mcclelland, J. L., editors, *Parallel Distributed Processing*, volume 1, pages 318–362. MIT Press.

Sauvage, J.-F., Fusco, T., Petit, C., Costille, A., Mouillet, D., Beuzit, J.-L., Dohlen, K., Kasper, M., Suarez, M., Soenke, C., Baruffolo, A., Salasnich, B., Rochat, S., Fedrigo, E., Baudoz, P., Hugot, E., Sevin, A., Perret, D., Wildi, F., Downing, M., Feautrier, P., Puget, P., Vigan, A., O'Neal, J., Girard, J., Mawet, D., Schmid, H. M., and Roelfsema, R. (2016). SAXO: the extreme adaptive optics system of SPHERE (I) system overview and global laboratory performance. *Journal of Astronomical Telescopes, Instruments, and Systems*, 2(2):025003.

Savransky, D. (2015). Sequential Covariance Calculation for Exoplanet Image Processing. *ApJ*, 800:100.

Schawinski, K., Zhang, C., Zhang, H., Fowler, L., and Santhanam, G. K. (2017). Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *MNRAS*, 467:L110–L114.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. NIPS, pages 802–810, Cambridge, MA, USA. MIT Press.

Skemer, A. J., Hinz, P., Esposito, S., Skrutskie, M. F., Defrère, D., Bailey, V., Leisenring, J., Apai, D., Biller, B., Bonnefoy, M., Brandner, W., Buenzli, E., Close, L., Crepp, J., De Rosa, R. J., Desidera, S., Eisner, J., Fortney, J., Henning, T., Hofmann, K.-H., Kopytova, T., Maire, A.-L., Males, J. R., Millan-Gabet, R., Morzinski, K., Oza, A., Patience, J., Rajan, A., Rieke, G., Schertl, D., Schlieder, J., Su, K., Vaz, A., Ward-Duong, K., Weigelt, G., Woodward, C. E., and Zimmerman, N. (2014). High contrast imaging at the LBT: the LEECH exoplanet imaging survey. In *Adaptive Optics Systems IV*, volume 9148 of *Proc. SPIE*, page 91480L.

Soummer, R. (2005). Apodized Pupil Lyot Coronagraphs for Arbitrary Telescope Apertures. *ApJ*, 618:L161–L164.

Soummer, R. and Aime, C. (2004). Statistics of pinned speckles in direct and coronagraphic high-contrast imaging. In Bonaccini Calia, D., Ellerbroek, B. L., and Ragazzoni, R., editors, *Advancements in Adaptive Optics*, volume 5490 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 495–503.

Soummer, R., Brendan Hagan, J., Pueyo, L., Thormann, A., Rajan, A., and Marois, C. (2011). Orbital Motion of HR 8799 b, c, d Using

Hubble Space Telescope Data from 1998: Constraints on Inclination, Eccentricity, and Stability. *ApJ*, 741:55.

Soummer, R., Pueyo, L., and Larkin, J. (2012). Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages. *ApJ*, 755:L28.

Sparks, W. B. and Ford, H. C. (2002). Imaging Spectroscopy for Extrasolar Planet Detection. *ApJ*, 578:543–564.

Spergel, D. and Kasdin, J. (2001). A Shaped Pupil Coronagraph: A Simpler Path towards TPF. In *American Astronomical Society Meeting Abstracts*, volume 33 of *Bulletin of the American Astronomical Society*, page 1431.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.

Student, W. S. G. (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

Su, K. Y. L., Rieke, G. H., Stapelfeldt, K. R., Malhotra, R., Bryden, G., Smith, P. S., Misselt, K. A., Moro-Martin, A., and Williams, J. P. (2009). The Debris Disk Around HR 8799. *ApJ*, 705:314–327.

Tagliaferri, R., Longo, G., Milano, L., Acernese, F., Barone, F., Ciaramella, A., Rosa, R. D., Donalek, C., Eleuteri, A., Raiconi, G., Sessa, S., Staiano, A., and Volpicelli, A. (2003). Neural networks in astronomy. *Neural Networks*, 16(3):297 – 319. Neural Network Analysis of Complex Scientific Data: Astronomy and Geosciences.

Udell, M., Horn, C., Zadeh, R., and Boyd, S. P. (2016). Generalized low rank models. *Foundations and Trends in Machine Learning*, 9(1):1–118.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., and the scikit-image contributors (2014). scikit-image: image processing in Python. *PeerJ*, 2:e453.

Vavasis, S. A. (2009). On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377.

Wahhaj, Z., Cieza, L. A., Mawet, D., Yang, B., Canovas, H., de Boer, J., Casassus, S., Ménard, F., Schreiber, M. R., Liu, M. C., Biller, B. A., Nielsen, E. L., and Hayward, T. L. (2015). Improving signal-to-noise in the direct imaging of exoplanets and circumstellar disks with MLOCI. *A&A*, 581:A24.

Wells, D. C., Greisen, E. W., and Harten, R. H. (1981). FITS - a Flexible Image Transport System. *A&AS*, 44:363.

Wertz, O., Absil, O., Gómez González, C. A., Milli, J., Girard, J. H., Mawet, D., and Pueyo, L. (2016). VLT/SPHERE robust astrometry of the HR8799 planets at milliarcsecond-level accuracy Orbital architecture analysis with PyAstrOFit. *ArXiv e-prints*.

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and Wilson, P. (2014). Best practices for scientific computing. *PLoS Biol*, 12(1):e1001745.

Wolszczan, A. and Frail, D. A. (1992). A planetary system around the millisecond pulsar PSR1257 + 12. *Nature*, 355:145–147.

Xie, D., Zhang, L., and Bai, L. (2017). Deep learning in visual computing and signal processing. *Applied Comp. Int. Soft Computing*, vol. 2017:ID 1320780, 13 pages.

Zhou, T. and Tao, D. (2011a). Bilateral random projections. *CoRR*, abs/1112.5215.

Zhou, T. and Tao, D. (2011b). Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In Getoor, L. and Scheffer, T., editors, *Proceedings of the 28th International Conference on Machine Learning*, ICML, pages 33–40, New York, NY, USA. ACM.

Zhou, T. and Tao, D. (2013). Unmixing incoherent structures of big data by randomized or greedy decomposition. *CoRR*, abs/1309.0302.

Zhou, X., Yang, C., Zhao, H., and Yu, W. (2014). Low-rank modeling and its applications in image analysis. *ACM Comput. Surv.*, 47(2):36:1–36:33.

Zurlo, A., Vigan, A., Galicher, R., Maire, A.-L., Mesa, D., Gratton, R., Chauvin, G., Kasper, M., Moutou, C., Bonnefoy, M., Desidera, S., Abe, L., Apai, D., Baruffolo, A., Baudoz, P., Baudrand, J., Beuzit, J.-L., Blancard, P., Boccaletti, A., Cantalloube, F., Carle, M., Cascone, E., Charton, J., Claudi, R. U., Costille, A., de Caprio, V., Dohlen, K., Dominik, C., Fantinel, D., Feautrier, P., Feldt, M., Fusco, T., Gigan, P., Girard, J. H., Gisler, D., Gluck, L., Gry, C., Henning, T., Hugot, E., Janson, M., Jaquet, M., Lagrange, A.-M., Langlois, M., Llored, M.,

Madec, F., Magnard, Y., Martinez, P., Maurel, D., Mawet, D., Meyer, M. R., Milli, J., Moeller-Nilsson, O., Mouillet, D., Origné, A., Pavlov, A., Petit, C., Puget, P., Quanz, S. P., Rabou, P., Ramos, J., Rousset, G., Roux, A., Salasnich, B., Salter, G., Sauvage, J.-F., Schmid, H. M., Soenke, C., Stadler, E., Suarez, M., Turatto, M., Udry, S., Vakili, F., Wahhaj, Z., Wildi, F., and Antichi, J. (2016). First light of the VLT planet finder SPHERE. III. New spectrophotometry and astrometry of the HR 8799 exoplanetary system. *A&A*, 587:A57.

Dissertation submitted in partial fullfilment of the requirements for the degree of Doctor of Science