Activity-Free Radial Velocities

Miguel Braçais
Mestrado em Astronomia e Astrofísica
Departamento de Física e Astronomia
2023

Orientador
Dr. João Faria, Investigador, Instituto de Astrofísica e Ciências do Espaço

MSc

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

“The Universe is a pretty big place. If it’s just us, seems like an awful waste of space.”

Carl Sagan

Acknowledgements

First off, I would like to thank my supervisor, Dr. João Faria, whose help and guidance were crucial in the makings of this work.

I also want to thank other investigators at IAstro: Dr. Jorge H. C. Martins, a former supervisor for one of my Physics Degree's courses, and a good friend; and Ana Rita Castro, a PhD student who was a source of inspiration for me to keep on going, and also a friend who helped me with some of my questions regarding my future after this Master's Degree.

Furthermore, I want to thank all my closest friends from our times in the Degree in Physics and in the Master's Degree in Astronomy and Astrophysics, as well as my friends from the Tuna de Ciências do Porto, Javardémica, whose environment and good disposition were essential in helping me maintain emotional balance while carrying out this work. Einstein said that, if he wasn't a physicist, he'd be a musician, and on the same note, I thank them for helping me keep the joys of music in my life.

I would also like to thank the members of the organization of volunteering, VO.U; in particular those from the project I was part of, VO.U Crescer, and the children we accompanied during these two years, as not only they also helped me keep my emotional balance, but the children, with their curiosity regarding the planets, space, and many other things of various topics, kept reminding me of why I chose this path and why I want to keep on going.

Finally, I want to thank my mom, my dad and my sister, who not only helped me keep focused on my projects and work, but also supported me in every step of the way.

UNIVERSIDADE DO PORTO

Abstract

Faculdade de Ciências da Universidade do Porto
Departamento de Física e Astronomia

MSc. Astronomy and Astrophysics

Activity-Free Radial Velocities

by Miguel Braçais

The magnetic activity of many stars can prove to be an obstacle in exoplanet detection. The detected signals of solar spots and solar plages, caused by the star's magnetic activity, can be similar to those emitted by exoplanets, and it's hard to distinguish the signals. There is one notable difference between these signals though: while stellar activity deforms the spectral lines of the stellar spectrum, the presence of planets shifts them in wavelength, excluding the transiting exoplanet observations, which also deform the spectral lines' profiles. That way, by observing the spectrum of a star, using the radial velocity technique, we can detect if the spectrum is being shifted by the action of an orbiting exoplanet. A method to study these spectra and determine their radial velocities has already been implemented before, using Hermite-Gaussian functions of first degree. What we're aiming to do with this project is to further study the method, obtain the star's radial velocity and compare our results with those obtained with other methods, and even try to further improve on the existing method. While we weren't able to obtain results with greater accuracy when compared to those obtained with the Cross-Correlation method, we did obtain results with greater precision. Even though we originally intended to also obtain the exoplanets' properties through the radial velocity, the time taken to implement the method didn't leave us with much time left for such an analysis.

Keywords: Astrophysics, Radial velocity, Stellar spectrum, Hermite-Gaussian, Absorption features

UNIVERSIDADE DO PORTO

Resumo

Faculdade de Ciências da Universidade do Porto
Departamento de Física e Astronomia

Mestrado Integrado em Engenharia Física

Titulo da Tese em Portugês

por Miguel Braçais

A atividade magnética de várias estrelas pode tornar-se um obstáculo na deteção de exoplanetas. Os sinais detetados de manchas e praias solares, causadas pela atividade magnética da estrela, podem ser semelhantes aos sinais emitidos por exoplanetas, e não é fácil distingui-los. Há no entanto uma notável diferença entre esses sinais: enquanto que a atividade solar deforma as linhas espetrais do espetro estelar, a presença de planetas transita-as em comprimento de onda, excluindo observações de trânsitos dos exoplanetas, que também deformem os perfis das linhas espetrais. Assim, através da observação do espetro de uma estrela, utilizando a técnica da velocidade radial, podemos detetar se o espetro está a ser ”shiftado”pela ação de um exoplaneta orbitante. Um método para estudar esses espetros e determinar as suas velocidades radiais já foi implementado no passado, usando funções Hermite-Gaussianas de primeiro grau. O nosso objetivo com este projeto é estudar ainda mais este método, obter a velocidade radial da estrela e comparar os nossos resultados com os obtidos através de outros métodos, e até mesmo tentar melhorar o método existente. Apesar de não termos conseguido obter resultados com maior exatidão quando comparados com os obtidos pelo método Cross-Correlation, obtivemos ainda assim resultados com maior precisão. Ainda que nós originalmente quiséssemos também obter as propriedades dos exoplanetas através da velocidade radial, o tempo que levou a implementar o método não nos deixou com tempo suficiente para essa análise.

Palavras-chave: Astrophysics, Radial velocity, Stellar spectrum, Hermite-Gaussian, Absorption features

Contents

Acknowledgements iii

Abstract v

Resumo vii

Contents ix

List of Figures xi

List of Tables xiii

1 Introduction 1

2 Method 7
  2.1 Theory 7
  2.2 Algorithm to detect and adjust spectral lines 18
  2.3 Adjustment of the H-Gs and calculation of the RVs 23

3 Applying the Method to Real Data 27
  3.1 Further changes to the algorithm 27
  3.2 Calculation of RVs order by order 31
  3.3 Combined results for all orders 35

4 Comparison of obtained RVs with other methods 39

5 Conclusion 43

Bibliography 47

List of Figures

2.1 The Hermite-Gaussian functions of the first five degrees, simulated using Gaussian curves of $\mu = 5005$ and $\sigma = 0.1$, given by Equation (2.8) 12
2.2 Overview of the evolution of the H-G coefficients of different degrees in function of the radial velocity. For smaller RVs, all coefficients but the one of the first degree become negligible, while c_1 changes in a linear fashion. This simulation of the H-G coefficients was computed for a simulated Gaussian curve with $\mu = 5005$ and $\sigma = 0.1$, to represent a typical feature found in a stellar spectrum. 14
2.3 An example of a Gaussian distribution, a Lorentzian distribution, and the Voigt profile, for comparison ($\sigma = 0.4$). 16

3.1 Section of the normalized template spectrum, showcasing the Gaussian curve fits done individually to each absorption feature. Many of the features detected have very minimal line depth. 29
3.2 Section of the normalized template spectrum, showcasing the cases where the Gaussian curve fits done to blends result in one of the features of the blend being either badly detected, or outright undetected. 30
3.3 A section of the normalized difference flux array, with the H-G curve fit in orange, and the residual difference between the difference values and the fit. This fit is taken from the first observation, order 100, and for the most part, its residual difference does not go beyond the order of magnitude of $0.1\,\mathrm{m\,s^{-1}}$. However, we can see a deviation between the difference flux values and the curve fit. While the cause of the deviation is not clear, we believe it is the reason why the RV values are not as accurate as predicted. 33
3.4 A section of the normalized difference flux array, with the H-G curve fit in orange, and the residual difference between the difference values and the fit. This fit is taken from the first observation, order 120, showcasing similar observations to those from the previous figure. This image is also an example of some of the greater deviations between the values in the difference flux and the fitted curve. 34
3.5 Examples of three distinct orders from the first observation 37

4.1 Comparison of the values obtained with the HGRV method (top) to those from the CCF method (bottom), with their respective uncertainties 40
4.2 Comparison of the values obtained with the HGRV method (top) to those from the CCF method (bottom), with the Radial Velocity axis scaled in a way to allow for a comparison of the values' variation. 40

4.3 Histogram of the RV values obtained through the means of all values from all valid orders, per observation (left); and histogram of the CCF values (right). We can see how the CCF values formed two distinct groups of values, as opposed to those from the HGRV. 41

List of Tables

1.1 Properties of Tau Ceti and its candidate orbiting planets. With the habitable zone of the system spanning an area of radius within 0.55 AU and 1.16 AU, we can place Tau Ceti f and, arguably, Tau Ceti e within the habitable zone. 5

2.1 Values for the initialization of a feature i's attributes before the Gaussian curve fit, and the bounds for the fit's results, where λ_i is the wavelength window for feature i, f_i is the corresponding set of values for the flux, y_j is the flux value located at index j such that $y_j = min(f_i)$, and x_j is the wavelength value for which we attained the minimum flux 22

Chapter 1

Introduction

The detection and study of exoplanets is one of the most prominent fields of research in astrophysics, in part thanks to its overlapping relevance with other topics, such as biology and geology, and it has become a popular subject over the last few years. The first result in the field was the discovery of two planets orbiting the pulsar PSR 1257+12, by Aleksander Wolszczan and Dale Frail in 1992 (Wolszczan and Frail [1992]), and in 1995, Michel Mayor and Didier Queloz presented the first confirmed detection of an exoplanet, orbiting a main-sequence star, 51 Pegasi (Mayor and Queroz [1995]). This discovery was part of the start of a booming interest in exoplanets, and was later recognized in 2019 for a Nobel Prize in Physics..

One of the main methods of detection of exoplanets is the Transit Photometry method (Deeg and Alonso [2018]), which focuses on the observation of stars, and the analysis of the variations in their brightness. If a planet is orbiting a star, whenever it passes in front of it in relation to the observer, when it transits the star, part of the stellar light will be blocked, which results in a temporary reduction of the observed brightness. Given the nature of planetary orbits, this means that the presence of an exoplanet will reveal itself in the periodic variation of the observed brightness of a star.

The other main method of exoplanet detection, and the main focus of this dissertation, is the Radial Velocity method, also known as Doppler spectroscopy, which requires observation and analysis of the star's spectrum. Alongside the absorption features of a star's spectrum, there may be some features which belong to an orbiting exoplanet. These features are there due to the stellar light that's reflected by the exoplanet, which contains the absorption features of said exoplanet. Furthermore, while we can detect changes in the stellar features as we observe the spectrum over an extended period of time, some of them may be caused by the star's own magnetic activity. However, while a star's magnetic phenomena cause a distortion of the spectrum's features, the presence of an exoplanet merely shifts them in relation to their wavelength, since the gravitational pull caused on the star by the planet causes a very small but periodic distortion on the stellar flux, causing it to redshift and blueshift at very small scales. The Radial Velocity method uses these shifts in the stellar spectra to calculate the star's variation in radial velocity and verify if they were caused by orbiting exoplanets. If the stellar radial velocity varies periodically, we can infer the presence of a planetary companion and estimate its orbital properties. While one cannot observe great amounts of stars simultaneously with the same instrument of observation unlike with Transit Photometry, as we can only execute the method with the flux spectrum of a single star at a time, this method has many advantages. It's fit for the detection of exoplanets around low-mass stars, which rotate more slowly, allowing for clearer spectral lines, and are more affected by the gravitational pull of their orbiting exoplanets, causing greater variations in their radial velocity. This makes this method ideal for discovering new candidates of potential Earth-like planets, which tend to be low-mass rocky planets orbiting M-type stars, although they can also be found in stars of greater mass, similar to the Sun, a G-type star,. However, Hot Jupiters, which tend to form around F- and G-type stars, cause even greater shifts with their greater mass, and are therefore easier to detect than rocky planets. For reference, Jupiter causes a shift of the Sun's velocity of $12.4\,\mathrm{ms}^{-1}$, while the Earth only causes a change of $0.1\,\mathrm{m\,s}^{-1}$.

The visible variations in radial velocity for planets in inclined orbits in relation to the observer's line of sight are much smaller and harder to detect. Moreover, this method is limited by the observations' signal-to-noise ratio (SNR), requiring a high value of that ratio for more precise results. Even still, for planets orbiting very bright stars or reflecting a lot of light, we can measure their mass and obtain information about their composition. More so, this method allows us to directly measure the exoplanet's eccentricity, and although the radial velocity of a star only allows for an estimation of the orbiting planet's minimum mass, due to the fact that we only measure the star's movement along our line-of-sight, with enough brightness either coming from a star or being reflected by a planet, we can obtain the planet's own radial velocity by distinguishing the spectral lines of each and separately analyzing their shifts, allowing us to infer the inclination of the planet's orbit. This is because, through the binary mass function, we know the relation between the masses of two orbiting bodies, their distances to the center of mass, and their orbiting velocities, and therefore, with the mass of the star and the radial velocities of both the star and the exoplanet, we can obtain the exoplanet's mass as well. Even if, overall, the Transit method has proven to be the most effective of the two, the usage of both methods not only yields more precise measurements but also allows for a clearer picture of the detected exoplanet's properties, such as its mass and radius. Rodler et al. [2012]

When it comes to analyzing the observed spectra in order to obtain information on the RV values and their shifts, there have been multiple methods over the years to perform that analysis. One of the most popular is the Cross-Correlation technique, in which the observed spectrum is shifted along the template spectrum, and the degree of similarity between them is determined by making use of the Cross-Correlation Function (CCF), where the highest similarity value reveals the number of pixels by which the observed spectrum is shifted in relation to the template spectrum, allowing us to calculate the difference of RV between them. This method is intuitive and conceptually easy, and several authors have suggested improvements to further increase its precision (Zucker [2003]) (Prieto [2007]). Another popular technique is the absorption cell calibration technique, which employs the use of a glass cell filled with gas, placed in the path of the incoming starlight and imprinting the observed spectra with its own spectral absorption features, which serve as a grid against which the stellar line shifts can be detected, and therefore are used to track these changes. While hydrogen fluoride gas (HF) had been used before (Campbell and Walker [1979]), it was later discovered that Iodine gas (I_2) yielded results with greater precision, around the order of $\mathrm{m\,s^{-1}}$ (Marcy and Butler [1992]) (Butler et al. [1996]), and it's currently regarded as the gas of choice for this method, the Iodine Cell method. The Simultaneous Reference technique is another method that yields high-precision RV measurements, which consists of the insertion of a template spectrum simultaneously with the observed spectrum through the use of a Thorium-Argon lamp (ThAr). Two channels of the spectrograph are used, one for the scientific data, and another, the simultaneous-reference channel, is used to monitor instrumental drifts and do wavelength calibrations to the observed data. This is done because the observed data is subjected to certain effects, them being the varying index of air refraction and mechanical flexures, and with this method, since the two channels will follow nearly similar paths in the spectrograph, the effects will disturb them in approximately the same way, allowing us to measure the instrumental drift. We do this by comparing, at the end of the observation session, the initial ThAr spectrum to the ThAr spectrum obtained simultaneously with the observed stellar flux, and we then subtract said drift from the scientific data channel to obtain the stellar variations without the drift effects (Lovis and Fischer [2010]) (Wright [2018])

Regardless of the technique used, many discoveries were made with this method in those past few years, starting with 51 Pegasi, but also including the discovery of the first Multi-Planet System of the binary star Upsilon Andromedae by the Lick Observatory (Butler et al. [1999]), whose observations spanned for a total of 11 years to confirm the presence of three exoplanets; the prediction of the time of transit of multiple exoplanets; the discovery of planets around stars with masses greater and smaller than solar-type stars, and revealing to us the types of planets that orbit around those other types of stars, including the reveal of brand new planet types (Lovis and Fischer [2010]).

Up until very recently, the RV precision achieved by state-of-the-art spectrographs was limited to about $1\,\mathrm{m\,s^{-1}}$. High precision had been reached with the aid of the High Accuracy Radial Velocity Planet Searcher (HARPS) spectrograph in 2003 (Mayor et al. [2003]), and it set a new standard in RV precision, but newer spectrographs would be required in order to achieve even more precise measurements. At that time, various techniques had already been developed for exoplanet discovery, including Doppler measurements, transit photometry, direct imaging, and microlensing. Ground-based transit surveys demonstrated the existence of transiting extra-solar planets and ultimately managed to achieve 1 mmag of photometric precision, but the Kepler space mission was a dramatic improvement over these ground-based photometry methods, confirming that most of the stars in our galaxy have planetary systems, with a high frequency in small planets.

In 2016, a paper was published describing the state of the art regarding the radial velocity method (Fischer et al. [2016]). It described the current advancements that had already been achieved, as well as the current obstacles. The RV measurements used to be carried out by either the iodine cell method or the cross-correlation technique. Estimates of errors include the estimated single measurement precision (SMP), which helps indicate the ”on-sky” precision; and the long-term velocity root-mean-square (rms), which exposes the systematic instrumental errors and limitations in the analysis techniques for the treatment of stellar jitter. In the year of 2015, some of the new space missions included the Transiting Exoplanet Survey Satellite (TESS), the CHaracterizing ExOPlanet Satellite (CHEOPS) and the PLAnetary Transits and Oscillation (PLATO) of stars. In July of the same year, the second workshop for Extreme Precision Radial Velocities (EPRV) was held at Yale University, where they discussed and examined the current state of the Doppler precision and the recent advances of the time, and work to achieve the 10 cm/s level of precision.

However, while this document provides a great level of insight into the state of the radial velocity method field, much more work was done afterwards. In particular, one method has been developed, one that aims to be much more mathematically precise in the calculation of stellar RVs - the Hermite-Gaussian Radial Velocity (HGRV) method, which utilizes first-degree Hermite-Gaussian curves. First proposed by Parker Holzer, a PhD student at Yale University in the Department of Statistics & Data Science (Holzer et al. [2020], this article is the basis and focus of our work, with which we aim to not only replicate this method of stellar data treatment and exoplanet detection, but also a look into the obstacles that presented to ourselves along the way, how we managed to mitigate them, the influence on the results of some of our decisions and choices regarding the treatment of some of the data, and hopefully to further push interest in developments and advancements of this method.

In this work, we will be analyzing data obtained by the ESPRESSO spectrograph installed at the ESO's VLT (Pepe et al. [2021]). We will focus on observations of the star HD10700, also known as Tau Ceti. This is a G-class star with a spectrum similar to our Sun's, although with less than 80% of its mass, and at a distance of 3.7 parsecs. Its observed radial velocity is of about $-17\,\mathrm{km\,s^{-1}}$. The star has at least four candidate orbiting planets, two of which are located on the star's habitable zone, Tau Ceti e and f (Tuomi, M. et al. [2013], Feng et al. [2017]). This star is commonly used by several RV surveys as a standard, in order to track low-amplitude instrumental variations over time. As such, there is a large collection of public observations.

Star	Mass ($M_\odot$)	Radius ($R_\odot$)	RV ($km\ s^{-1}$)
Tau Ceti	0.783 ± 0.012	0.793 ± 0.004	-16.68 ± 0.05

Exoplanet	Mass ($M_\oplus$)	Semimajor Axis (AU)	Eccentricity
Tau Ceti g	$\geq 1.75\ +0.25/-0.40$	$0.133+0.001/-0.002$	0.06 ± 0.13
Tau Ceti h	$\geq 1.83\ +0.68/-0.26$	0.243 ± 0.003	$0.23+0.16/-0.15$
Tau Ceti e	$\geq 3.93\ +0.83/-0.64$	0.538 ± 0.006	$0.18+0.18/-0.14$
Tau Ceti f	$\geq 3.93\ +1.05/-1.37$	$1.334+0.017/-0.044$	$0.16+0.07/-0.16$

Table 1.1: Properties of Tau Ceti and its candidate orbiting planets. With the habitable zone of the system spanning an area of radius within 0.55 AU and 1.16 AU, we can place Tau Ceti f and, arguably, Tau Ceti e within the habitable zone.

In Chapter 2, we will discuss the theory behind the HGRV method, as well as explain in detail the process of the functionality and development of all the algorithm's components. In Chapter 3, we will apply the method to the data obtained from Tau Ceti and in Chapter

4 we will compare our results with those obtained using other methods. Finally, we will discuss our conclusions and closing thoughts in Chapter 5.

Chapter 2

Method

2.1 Theory

Perhaps it is important, before delving further, to clarify what radial velocity is. Radial velocity, by general definition, is the velocity of an object relative to a defined point of reference, or observer, taken as the center with radial coordinates, measuring the rate of the object's change in distance from the observer. In the context of Astrophysics and Astronomy, Earth tends to be the observer when we want to measure a star's radial velocity, so these values measure how fast a star is moving away from our planet or, in the case of negative radial velocities, how fast they are approaching.

Much of what we currently know about the motion of exoplanets around stars can be inferred from what we already know about the inner workings of binary star systems - after all, in the simplest form of both cases, we have two bodies orbiting their common center of mass (CM). We can understand their orbits through the variation of their radial velocity, and from there, calculate their relative masses. In the cases when only one of the stars is bright enough to properly measure its spectrum, we can define the mass function f that relates the semi-amplitude K of the binary system's orbit with its properties, such as the orbital period P, eccentricity e and inclination i:

$$f \equiv \frac{PK^3(1-e^2)^{\frac{3}{2}}}{2\pi G} = \frac{M_{unseen}^3 \sin^3 i}{(M_{unseen} + M_{seen})^2} \tag{2.1}$$

with G being Newton's constant, and M_{seen} and M_{unseen} being the masses of the seen and unseen stars, respectively. Furthermore, we can define the orbit's phase and orientation with respect to the observer's line of sight with the additional variables φ and ω. The radial velocity of the system's CM in relation to the observer is denoted as γ. With these parameters, we can define the orbit of the observed star around the CM of the system, caused by the movement of the unseen star, for which the orbital parameters are inferred from the system's motion. However, most of the unseen star's orbital elements are identical to the seen star's, except for the semi-amplitude, in which K is scaled by a factor of the masses' ratio, and its orientation, where ω differs by π since the stars are opposite to each other in their orbits:

$$K_{seen} = \frac{M_{unseen}}{M_{seen}} K_{unseen} \tag{2.2}$$

$$\omega_{seen} = \omega_{unseen} - \pi \tag{2.3}$$

These relations between two orbiting stellar objects can be applied to many other cases, including those where a star is being orbited by an exoplanet. In this scenario, since the mass of an exoplanet is much smaller than that of a star, we can approximate the mass ratio as $M_{unseen}/M_{seen} = M_{planet}/M_{*} \ll 1$, allowing us an approximation of equation 2.1 into a function that relates a Doppler shift's amplitude to the planet's orbital properties and mass:

$$K \approx \left(\frac{2\pi G}{PM_{*}^{2}}\right)^{1/3} \frac{M_{planet} \sin i}{\sqrt{1-e^{2}}} \tag{2.4}$$

With a weak dependence on the orbit's period and eccentricity (for low values), it's clear that the strongest dependence comes from the product of the planet's mass and the sine of the orbit's inclination, referred to as the planet's minimum mass, as the planet's true mass is actually larger than the aforementioned value by a geometric factor of $1/sini$. There is also a dependence on stellar mass, which shows that the radial velocity method should be more sensitive for low-mass stars.

From this, we simply need to look into the Doppler-shifted spectra of stars to deduce the parameters of their orbiting planet. A stellar spectrum is a representation of the intensity of a star's radiation flux, or brightness, in function to the values of wavelength it emits at. The narrow dips and valleys we can see in the spectra correspond to the absorption features that are characteristic to every chemical element existent within the star's atmosphere, from atoms and ions to molecules, and thanks to atomic physics, we know the values of the wavelengths at which each element absorbs radiation, their rest wavelengths, λ_{rest}. The star's motion causes the lines in observed spectra to be Doppler shifted to their observed wavelength values, λ_{obs}. We can use the Doppler formula to calculate the relative radial speed v_r between the star and the observer that measured the emitted light, through the redshift value z:

$$z \equiv \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}} = \frac{\lambda_{obs}}{\lambda_{rest}} - 1 = \frac{1}{\gamma(1 + v_r/c)} - 1 \tag{2.5}$$

with γ here being the relativistic factor $1/(1-(v/c)^2)$, c being the speed of light, and v being the relative scalar speed between the star's frame and the observer, though not necessarily in the radial direction. However, the stellar spectra itself pose some obstacles to this simple method.

The blending and overlap of features are very frequent in stellar spectra, and it's caused by the high frequency of such features. Meanwhile, the remainder of the spectrum, the line of base flux from which the features dip from, is referred to as the continuum. The continuum of stellar spectra is not flat, which stems from two main reasons, besides background noise and stellar jitter: one being the black-body effect, and the other being an error of instrumental origin, known as the blaze function. The formerly mentioned effect relates to the way a star's spectrum is shaped. A black body is an object that, ideally, is non-reflective and fully opaque, absorbing all radiation that hits it, with the radiation it emits in thermal equilibrium with its surroundings. A star can be considered a black body in this regard, as we can assume its temperature remains constant. The radiation emitted by a black body has a spectrum that only depends on the body's temperature, in accordance with Planck's Law, and it has a maximum for a specific wavelength, after which the wavelength spectrum becomes inversely related to its intensity, which explains its continuum's characteristic shape. The latter, the blaze function, relates to how the diffraction grating of a spectroscopic tool must be shifted relatively to the center-edge interference of each of the grating's facets. We can, however, divide these effects out, normalize the continuum and diminish the effect of those effects Planck [1901].

Another source of noise that contributes to a decrease in precision is the presence of telluric lines in the spectrum. Since the instruments used to observe the stars are situated on Earth, our atmosphere influences the observed data. Besides the flux spectra we obtain from observed stars with their absorption features, our atmosphere's own absorption lines, which are what we refer to as telluric lines, also contaminate the observed spectra. This is what's referred to as telluric contamination, and it varies with many factors, such as altitude and weather conditions, with their variability being what makes it so difficult to remove. Since these lines are more common at higher wavelength values, in the red and near Infra-Red regions, the contamination will be more noticeable at the later orders of wavelength in observed spectra than the earlier ones. They're also narrower than the other absorption features in most cases, and many advancements have been made over the years to further mitigate their effect on observed spectra. However, the only way to guarantee their complete removal is by performing the observations with an instrument located in space, outside our atmosphere.

Taking a template spectrum of a star, a noiseless and normalized spectrum with no effects applied to it, we can take another spectrum from the same star, a Doppler-shifted one, and calculate the difference between both to obtain the difference flux. This is useful because one of the main factors of the HGRV method is that it can estimate such Doppler shifts using the first-degree Hermite-Gaussian function, fitted to the difference flux spectrum. Since a Doppler shift only occurs horizontally, in the wavelength axis, very little information will be obtained from the normalized continuum, as opposed to sloped lines, such as those from absorption features. In order to work with such data, an algorithm must be computed in order to correctly locate and detect stellar features, both standalone features and blends of multiple of them, and obtain their properties. The process of computation of this algorithm will be described further below.

Given how spectral features will be our main sources of data for this method, we shall focus on them in particular. Let's first assume that the shape of the inverted absorption features in the spectrum can be approximated to the shape of a Gaussian curve (Gray [2005]). With x as the wavelength of the observed light, and f(x) as the corresponding normalized light flux value, then, according to special relativity, f(ξ x) will be the value of the normalized Doppler-shifted light flux, with the inverse of ξ being referred to as the Doppler factor, and defined as

$$\xi = \frac{1 + v_r/c}{\sqrt{1 - (v/c)}} \tag{2.6}$$

with c defined as the speed of light, v as the absolute speed of the source, and v_r as the star's radial velocity. While the Earth's motion around the Solar System's barycenter can introduce relativistic effects to the observations, they can be corrected (Wright and Eastman [2014], Blackman et al. [2017], Blackman et al. [2020]), and since the values of RVs are well under the speed of light's, we can simplify the formula into

$$\xi = 1 + \frac{v_r}{c} \tag{2.7}$$

Since we're assuming that the features are shaped like a Gaussian, let us take the Hermite-Gaussian functions, which are defined as

$$\psi_n(x) = \frac{1}{\sqrt{2^n n! \sqrt{\pi}}} H_n(x) e^{-x^2/2} \tag{2.8}$$

with $H_n(x)$ as the physicist's Hermite polynomial of n'th degree

$$H_k(s) = k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m}{m!(k-2m)!} (2s)^{k-2m} \tag{2.9}$$

where $\lfloor k/2 \rfloor$ is the floor function that finds the closest integer number equal to or less than $k/2$.

The Hermite-Gaussian curves have other properties of their own, due to the unique attributes of the physicist's Hermite polynomials as opposed to the probabilist's polynomials, defined respectively as $H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}$ and $He_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$. For instance, as stated in Johnston [2014], we can show that

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2} dx = \sqrt{\pi} 2^n n! \mathbb{1}\{m = n\} \tag{2.10}$$

where $\mathbb{1}\{A\}$ is an indicator, similar to the Kronecker delta function, that equals 1 when the event A within the brackets is true, and zero when otherwise. Therefore, we can show, using the functions 2.8 and 2.10, that

$$\int_{-\infty}^{\infty} \psi_n(x) \psi_m(x) dx = \mathbb{1}\{m = n\} \tag{2.11}$$

From the same work of Johnston [2014] on weighted Hermite polynomials, we can also find that the set of Hermite-Gaussian functions forms a full orthonormal basis of the set of every square-integrable, real-valued functions, $L^2(\mathbb{R})$, and the same applies to sets of generalized Hermite-Gaussian functions with variables to specify their general location in the x axis and their scale, through the variables μ and σ, for any $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. Thus, we can redefine the H-G as

$$\psi_n(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2^n n! \sqrt{\pi}}} H_n\left(\frac{x - \mu}{\sigma}\right) e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.12}$$

Therefore, for an $L^2(\mathbb{R})$ function $g(x)$, we can decompose it as

$$g(x) = \sum_{n=0}^{\infty} c_n \psi_n(x; \mu, \sigma) \tag{2.13}$$

with c_n being the coefficient associated with the generalized H-G function of the corresponding degree.

With that, taking $f(x)$ as a Gaussian curve with a defined center μ and width σ, and $g(x;\xi) = f(x) - f(\xi x)$ the difference between the Gaussian and its Doppler-shifted version. If we decompose this difference using equation 2.13, and only using $n = 1$, i.e. the first degree of Hermite equations, then we obtain Holzer's first theorem:

Theorem 2.1. For any $\sigma \in \mathbb{R}^+$ and any $\mu, \xi \in \mathbb{R}$,and for $g(x;\xi) = e^{-\frac{-(x-\mu)^2}{2\sigma^2}} - e^{-\frac{-(\xi x-\mu)^2}{2\sigma^2}}$, we have that

$$\lim_{\xi \to 1} \frac{\int_{-\infty}^{\infty} (g(x;\xi) - c_1(\xi)\psi_1(x;\mu,\sigma))^2 dx}{\int_{-\infty}^{\infty} (g(x;\xi))^2 dx} = \left(1 + \frac{2\mu^2}{3\sigma^2}\right)^{-1} \tag{2.14}$$

This theorem was proven in Holzer et al. [2020] and it states that, as ξ approaches 1, at small RV values, the proportion of the difference between $g(x;\xi)$ and the ψ_1 curve that was

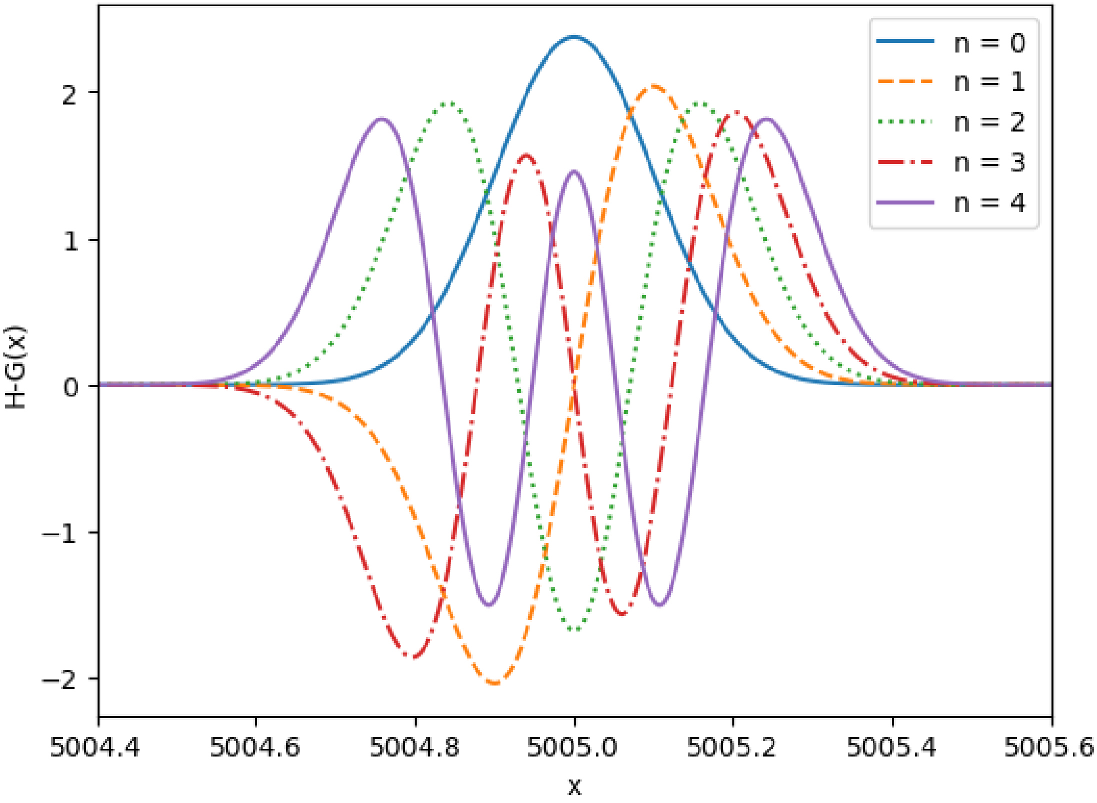


Figure 2.1: The Hermite-Gaussian functions of the first five degrees, simulated using Gaussian curves of $\mu = 5005$ and $\sigma = 0.1$, given by Equation (2.8)

modelled with the same σ and μ as the Gaussian curve, approaches $\left(1+\frac{2\mu^2}{3\sigma^2}\right)^{-1}$, and given that, for real data, $\mu \gg \sigma$, this means that this value is close to zero. What this means is that Doppler shifting a Gaussian absorption feature at a small RV can be considered to be equivalent to adding a constant multiple of ψ_1 to the feature, meaning that the subtraction between both Gaussian curves results in a constant multiple of a Hermite-Gaussian. This is because, despite the multiplicative nature of the Doppler shift, this approximation for small RV values allows for the assumption that the shift is additive instead.

To find the RV limit for which this theorem remains valid, Holzer et al. [2020] stated some Lemmas to solve for the coefficients for Equation 2.13 to solve.

Lemma 2.2. For $I_k(a,b,c) := \int_{-\infty}^{\infty} u^k e^{-\left(au^2+bu+c\right)} du$ where $a > 0$, we have that

$$I_0(a,b,c) = \sqrt{\frac{\pi}{a}} e^{\left(\frac{b^2}{4a} - c\right)} \tag{2.15}$$

$$I_1(a,b,c) = -\frac{\sqrt{\pi} b}{2a^{3/2}} e^{\left(\frac{b^2}{4a} - c\right)} \tag{2.16}$$

and for $k \geq 2$,

$$I_k(a,b,c) = -\frac{b}{2a} I_{k-1}(a,b,c) + \frac{k-1}{2a} I_{k+2}(a,b,c) \tag{2.17}$$

With this purely mathematical proof, Holzer then created the second lemma to find the solutions of Lemma 2.2.

Lemma 2.3. For $g(x;\xi) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} - e^{-\frac{(\xi x-\mu)^2}{2\sigma^2}}$ decomposed as $g(x) = \sum\limits_{n=0}^{\infty} c_n \psi_n(x;\mu,\sigma)$, and for $I_k(a,b,c)$ as defined in Lemma 2.2, for $\epsilon = \xi - 1$, we have that

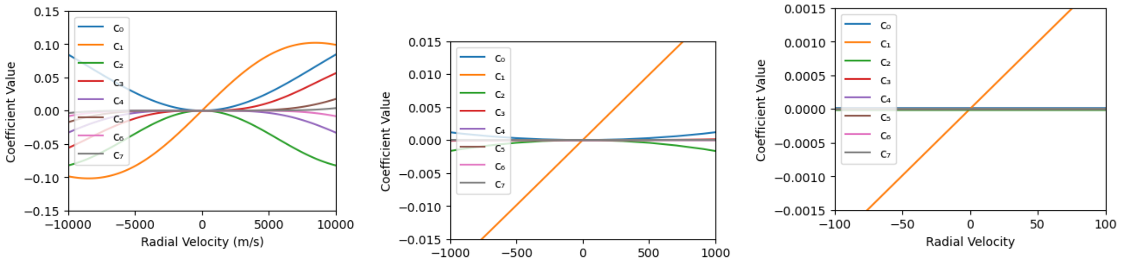
$$c_0(\epsilon) = \sqrt{\sigma\sqrt{\pi}} - \frac{1}{\sqrt{\sigma\sqrt{\pi}}} I_0\left(\frac{1+\epsilon+\frac{\epsilon^2}{2}}{\sigma^2}, -\frac{2\mu+\epsilon\mu}{\sigma^2}, \left(\frac{\mu}{\sigma}\right)^2\right) \tag{2.18}$$

and, for all $k \geq 1$,

$$c_k(\epsilon) = -\sqrt{\frac{\sigma k! 2^k}{\sqrt{\pi}}} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m}{4^m m!(k-2m)!} I_{k-2m}\left(1+\epsilon+\frac{\epsilon^2}{2}, \frac{\epsilon\mu}{\sigma}(1+\epsilon), \frac{1}{2}\left(\frac{\epsilon\mu}{\sigma}\right)^2\right) \tag{2.19}$$

With these lemmas, we can calculate the coefficients of the n'th H-Gs as functions of RV, and while all of them equal to zero with a null RV, since the $g(x;\xi)$ function in Theorem 2.1 becomes a zero-function at that point, when RV approaches zero, at the window of magnitude below 100 ms^{-1}all coefficients besides c_1 become negligible, while c_1

evolves almost linearly with v_r within the values of magnitude 500 m s^{-1}. This means that, for stars whose orbiting exoplanets cause RV variations with a semi-amplitude lesser than 100 m s^{-1}, we can only take the Hermite-Gaussian coefficient of first degree into consideration when modelling Gaussian stellar feature that suffered a Doppler shift due to an orbiting exoplanet, as we can observe in Figure 2.2



(a) RV = [-10; 10] km s^{-1}

(b) RV = [-1; 1] km s^{-1}

(c) RV = [-100; 100] m s^{-1}

Figure 2.2: Overview of the evolution of the H-G coefficients of different degrees in function of the radial velocity. For smaller RVs, all coefficients but the one of the first degree become negligible, while c_1 changes in a linear fashion. This simulation of the H-G coefficients was computed for a simulated Gaussian curve with $\mu = 5005$ and $\sigma = 0.1$, to represent a typical feature found in a stellar spectrum.

With the knowledge that the c_1 coefficient is the most dominant one within the above-defined RV value limits, we should now define the approximation error made by ignoring the other coefficients, by introducing the standardized approximation error, $D(\phi||\varphi)$, which gives the proportion of the squared function φ that remains to be modelled after we approximate it with ϕ:

$$D(\phi||\varphi) = \frac{\int_{-\infty}^{\infty} (\varphi(x) - \phi(x))^2 dx}{\int_{-\infty}^{\infty} \varphi(x)^2 dx} \tag{2.20}$$

For our theorem, we'll take the case where $\phi = g(x;\xi)$ and $\varphi = c_1(\xi)\psi_1(x;\mu,\sigma)$, as Holzer wrote the following Lemmas:

Lemma 2.4. For $g(x;\xi) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} - e^{-\frac{(\xi x-\mu)^2}{2\sigma^2}}$ decomposed as $g(x;\xi) = \sum_{n=0}^{\infty} c_n(\xi)\psi_n(x;\mu,\sigma)$, we have

$$D(g(x;\xi)||c_1(\xi)\psi_1(x;\mu,\sigma)) = 1 - \frac{c_1^2(\xi)}{\int_{-\infty}^{\infty}(g(x;\xi))^2 dx} \tag{2.21}$$

Lemma 2.5. $\lim_{\xi \to 1} \frac{c_1^2(\xi)}{\int_{-\infty}^{\infty}(g(x;\xi))^2 dx} = \frac{1}{1+\frac{3\sigma^2}{2\mu^2}}$

These Lemmas are, once again, proven in Holzer et al. [2020].

With a small enough v_r, and a big enough μ/σ ratio, as well as an absorption feature with a shape close enough to a Gaussian, we can do a least-squares fitting of the first-degree Hermite-Gaussian curve to the difference between a template spectrum and a Doppler-shifted spectrum, and then link the fitted coefficient to an RV value. According to Lemma 2.3:

$$c_1(\epsilon) = \frac{\sqrt{\sqrt{\pi}}}{\sqrt{2\sigma}} \epsilon \mu (1+\epsilon) \tilde{h}(\epsilon) \tag{2.22}$$

$$\lim_{\epsilon \to 0} \frac{\delta}{\delta \epsilon} c_1(\epsilon) = \frac{\mu \sqrt{\sqrt{\pi}}}{\sqrt{2\sigma}} \tag{2.23}$$

If we use Equation 2.7 with $\epsilon = \xi - 1$, we find that $v_r(\epsilon) = c\epsilon$ and $lim_{\epsilon \to 0} \frac{\delta}{\delta \epsilon} v_r(\epsilon) = c$ Therefore, going back to 2.23, we can define the proportionality constant, which is

$$\lim_{\epsilon \to 0} \frac{\delta}{\delta v_r} c_1(v_1(\epsilon)) = \frac{\mu \sqrt{\sqrt{\pi}}}{c\sqrt{2\sigma}} \tag{2.24}$$

As such, the proportionality at low RV values is defined as

$$c_1 = \frac{\mu \sqrt{\sqrt{\pi}}}{c\sqrt{2\sigma}} v_r \tag{2.25}$$

However, we cannot get precise enough RVs from a single absorption feature. So, we need to be able to use as many features from the stellar spectrum as possible. Rather than fitting a single Hermite-Gaussian around a single feature, we can fit a sum of these features to the difference spectrum. It has to be taken into account that every feature has its own center, width and depth, a new variable that has to be taken into consideration with the involvement of multiple features. Since Doppler-shifting any Gaussian, regardless of amplitude, only multiplies the resulting coefficients by the same amplitude, which in this case is the line depth, then we can use Equation 2.25 to compute the model of the difference spectrum at pixel i as a function of wavelength:

$$y_i = v_r \sum_{j=1}^{n} \frac{\sqrt{\sqrt{\pi}} d_j \mu_j}{c\sqrt{2\sigma_j}} \psi_1(x_i; \mu_j, \sigma_j) + \epsilon_i \tag{2.26}$$

where the sum goes over all n features, x_i is the wavelength at pixel i, d_j is the line depth of the j'th feature, and ϵ_i is an independent uncertainty with expectation 0.

Finally, it is important to delve into this model's misspecifications. Many absorption features in a stellar spectrum are described by the Voigt profile, itself described by the convolution of a Cauchy-Lorentz distribution and a Gaussian distribution:

$$V(x;\sigma,\gamma) \equiv \int_{-\infty}^{\infty} G(x';\sigma)L(x-x';\gamma)dx' \tag{2.27}$$

where $G(x';\sigma)$ is the centered Gaussian distribution, and $L(x-x';\gamma)$ is the centered Lorentzian distribution:

$$G(x;\sigma) = \frac{e^{-x^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \tag{2.28}$$

$$L(x;\gamma) = \frac{\gamma}{\pi(x^2+\gamma^2)} \tag{2.29}$$

where γ is a scale parameter that's defined by the Full Width Half Measure (FWHM) as $FWHM = 2\gamma = 2\sqrt{2\ln 2}\sigma$.

Since, for the values of σ that the absorption features usually have, the Voigt profile has a similar shape to that of the Gaussian distribution, that is what allows the assumption that the shapes of the absorption features are Gaussian in nature, the basis for the HGRV method. However, that does not necessarily hold true at all times, as some features may have a non-negligible Lorentzian component, and some may even have a line depth too

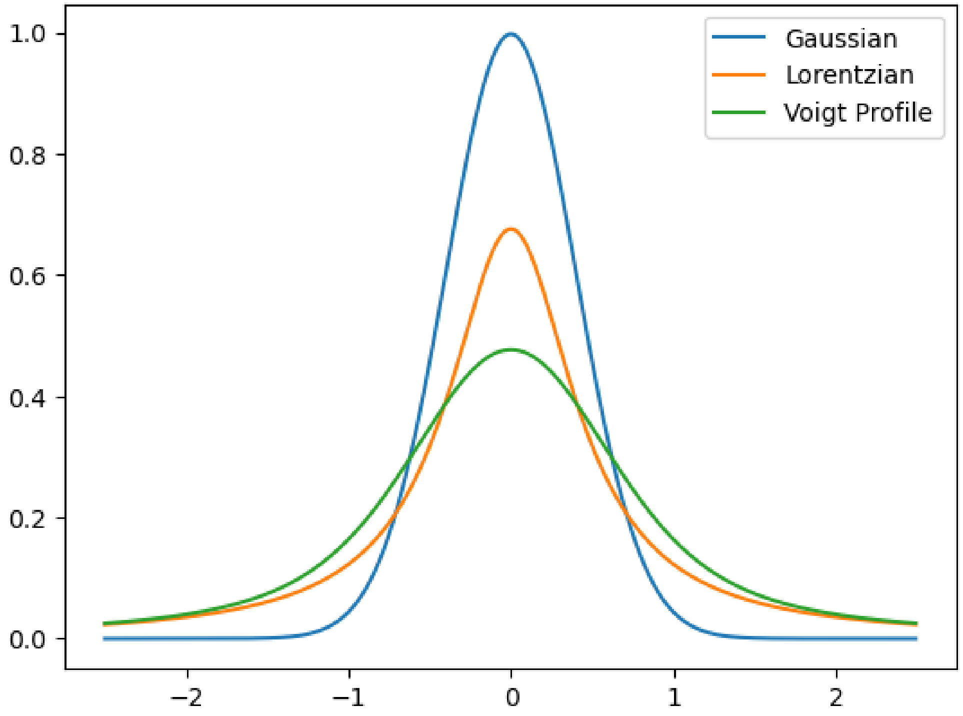


Figure 2.3: An example of a Gaussian distribution, a Lorentzian distribution, and the Voigt profile, for comparison ($\sigma = 0.4$).

great to even follow the Voigt profile. Holzer et al. [2020] showed that the obtained RV values would consistently have a bias that would be three orders of magnitude lower.

2.2 Algorithm to detect and adjust spectral lines

Our algorithm was written using the Python language. The first step in the creation of the algorithm that would be used for this project consisted in the creation of a function that would allow me to generate a Hermite-Gaussian curve, using equation (2.12).

```
def HG_single(x, mu, sigma):
    return -hermite(1)((x-mu)/sigma) *
        e**(-(x-mu)**2/(2*sig**2)) /
        np.sqrt(sigma*2**n*factorial(n)*np.sqrt(np.pi))
```

where $hermite(1)(x)$ is the Hermite function of the 1^{st} degree in x, mu stands as the variable for μ, and $sigma$ as that for σ.

Then, we wrote a function that fits a Hermite-Gaussian curve of the first degree to the difference between the template spectrum and a shifted spectrum, returning the deduced value for the radial velocity. For that, we wrote, inside said function, a ”helper” function of wavelength, with an RV value as the only variable. In this function, we would at first, for one feature, obtain the value of its center and width, μ and σ, to calculate the Hermite-Gaussian curve with the associated coefficient in accordance to Equation (2.25):

```
def HG_list(x, y_dif, mu, sigma, n):
    def HG_helper(x, vr):
        y = HG_single(x, mu, sigma, n) * np.sqrt(np.sqrt(np.pi)) *
            1/(sig*np.sqrt(2*np.pi)) * mu / (c*np.sqrt(2*sigma))
        return y*vr

    popt, pcov = curve_fit(HG_helper, x, y_dif, v_i)
    vr = popt[0]
    y = HG_helper(x, vr)

    return y, vr
```

with c as the speed of light in SI units, and where we ended up using a fixed value for the initial guess of the radial velocity used for the Doppler shifted curve. While we were testing these functions, we merely used a single Gaussian curve, generated by us using values for center and width that are typically found in absorption features, by using a function found in the *scipy.special* Python package, $norm.pdf(x)$, upon which we'll touch in a bit. With this first generated feature correctly fitted, we moved on to the computation of the feature detection algorithm, to first try and replicate the described algorithm in Holzer et al. [2020].

In a list of ordered wavelength values $\Gamma = (x_0, x_1, ..., x_n)$ and their corresponding flux values $\tau = (\tau_0, \tau_1, ..., \tau_n)$ from a stellar spectrum, and for a certain pixel index i, we have a value of wavelength x_i and one of flux τ_i. For a wavelength region of size m, we observe the wavelength points of the spectrum in a window from x_{i-m} to x_{i+m}, and the same for the flux points in a window within τ_{i-m} and τ_{i+m}. We will separate these points into four separate arrays of size m:

$$\Gamma_{l,i} = (x_{i-m+1}, x_{i-m+2}, ..., x_i)^T$$

$$\Gamma_{r,i} = (x_i, x_{i+1}, ..., x_{i+m-1})^T$$

$$\tau_{l,i} = (\tau_{i-m+1}, \tau_{i-m+2}, ..., \tau_i)^T$$

$$\tau_{r,i} = (\tau_i, \tau_{i+1}, ..., \tau_{i+m-1})^T$$

Afterwards, take the arrays of corresponding regions and do two linear, least-squares regressions, with their curves defined as:

$$\tau_{l,i} = \beta_{0,l}\mathbb{1}_m + \beta_{1,l}\Gamma_{l,i} + \epsilon$$

$$\tau_{r,i} = \beta_{0,r}\mathbb{1}_m + \beta_{1,r}\Gamma_{r,i} + \epsilon'$$

with ϵ, ϵ' being the regression's uncertainty errors, the coefficients of degree 0 β_0 being the $y-$intercepts, and those of degree 1 β_1 being the regressions' slopes. We do this in order to estimate the coefficients $\beta_{0,l}$ and $\beta_{1,l}$ for the region left of pixel i, and $\beta_{0,r}$ and $\beta_{1,r}$ for the region to its right.

When we find the coefficient $\beta_{1,l}$ to be negative while $\beta_{1,r}$ is positive with enough statistical significance, this means that x_i is found to be a statistically significant minimum - the minimum of an absorption feature. This process is done automatically in our algorithm by the use of the function $linregress(x, y, alternative)$ from the $scipy.stats$ Python package, with the $alternative$ variant being set as $"less"$ for the regression of the left region and $"greater"$ for that of the right region, since that makes the function only take in the values where the slope of the regression is lower than zero and greater than zero, respectively.

From the results of these functions, we can obtain the regressions' p-values, p_l, i and p_r, i, which we'll need for the next part of the feature-finding process. P-values are statistical numbers obtained from testing the first-degree coefficients when equal to zero against themselves when in their proper conditions, i.e. testing $\beta_{1,l} = 0$ against $\beta_{1,l} < 0$ and $\beta_{1,r} = 0$ against $\beta_{1,r} > 0$, and they serve to conclude if there is a statistically strong relationship between the two variables in the linear regression, or if the slope coefficients are actually null. When the p-values are below a given small probability, then the coefficients are valid. In our algorithm, we'll start testing the validity of the features we found by initializing a new index $j = m$, with an upper bound $u = 0$. While j is located within the entirety of Γ with a margin of value m at the beginning and end, i.e. while $j \leq length(\Gamma) - m + 1$, which will allow us to have a look at all features and their slopes from both sides, then, for a coefficient of significance α, chosen by us, we'll test if the two p-values at index j, $p_{l,j}$ and $p_{r,j}$, are lower than half of the value of α, the coefficients are very likely to not be zero. As such, we can start computing the wavelength margins of the feature we're indexing over with j, by creating two lists for the left and right, and taking the maximum value from the former and the minimum value from the latter where their respective p-values are equal or greater to another significance coefficient, η, whose value is also chosen by us:

$$k_{max} = max\{k \in \{u, u+1, ..., j\} : p_{l,k} \geq \eta\}$$

$$k_{min} = min\{k \in \{j, j+1, ..., length(\Gamma)\} : p_{r,k} \geq \eta\}$$

With k_{max} and k_{min}, we can calculate the wavelength bounds of the feature we're analyzing, and progress with the search on the other features, associating new values to j and u:

$$\left(\frac{x_{k_{max}} + x_{k_{max}-m}}{2}, \frac{x_{k_{min}} + x_{k_{min}+m}}{2} \right)$$

$$j \leftarrow \left\lfloor (k_{min+m/2}) \right\rfloor, u \leftarrow j$$

Other additions were made to the algorithm to make sure we would ignore cases when the p-values were greater than η, so that we would not obtain null values for k_{max} and k_{min}. However, since the wavelength values we calculated for our bounds were in no way guaranteed to fit in the values of our wavelength list, we needed to find a way to find the indexes of the values in the list closest to the ones we calculated. We did so by subtracting our calculated bounds from the wavelength list, turning all values within the result into their absolute values, and then finding the index of the minimum value - the index of the value closest to zero would be the index of the correct bound value. After that, we can isolate the feature within these bounds, and fit a Gaussian curve to it in order to obtain its central wavelength, standard deviation (width), and later on, also its line depth and offset.

With this algorithm done, we were ready to try it on some simulated spectra. While we at first tried to randomly create a small spectrum segment with features using a list of our own values for μ and σ, we decided to test the algorithm by use of the functions from the Python package *Spectrum*. With it, we were able to randomly generate spectra with an amount of features similar to those of the real spectra we would later use. There were other complications that we had to fix, such as an improved version of the Gaussian curve and fit of the absorption features that would give us their line depth, essential for working with multiple features, and an issue where the recovered radial velocities were over 5 to 6 times the expected, due to an earlier error in how we defined the Gaussian function. That issue was due to a misinterpreted definition of the function $norm.pdf(x)$, which computes a normalized Gaussian curve over the list of values x, it being

$$norm.pdf(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

However, since we wanted to define the central wavelength and standard deviation of each curve, we used an alternate form of the function, $norm.pdf(x, \mu, \sigma)$, which instead returned

$$norm.pdf(x; \mu, \sigma) = \frac{e^{-y^2/2}}{\sigma\sqrt{2\pi}}$$

$$y = \frac{x - \mu}{\sigma}$$

While the changes regarding y do not affect the function overall, the appearance of σ in the denominator causes a noticeable change. Therefore, when using that form of the $norm.pdf$ function, we also multiplied it by σ to nullify those changes:

```
def HG_single(x, m, sig):
    fac = (x-m)/sig
    return hermite(1)(fac) * norm.pdf(x, m, sig) * sig /
        np.sqrt(sig*2*factorial(1)*np.sqrt(np.pi))
```

With these mistakes out of the way, we were presented with a new challenge, the overlapping features. The curves that were fitted around those features, at the time, had a high error factor, so we followed the suggestions of the paper, and we executed comparisons between the features' Root-Mean Square, and the fitted curve:

$$RMS(\hat{\tau}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tau_i - \hat{\tau}_i)^2} \tag{2.30}$$

If the difference was too high, then that meant what we initially assumed as a single Gaussian feature was actually a double Gaussian feature, so we simply had to fit two Gaussian curves at once, which would result in two sets of attributes for each of the two features:

```
def gauss_func(x, m, sig, c, off):
    return  off - c*norm.pdf(x, m, sig) * sig

def double_gauss_func(x, m1, sig1, c1, m2, sig2, c2, off):
    return off - (c1*norm.pdf(x, m1, sig1) * sig1 +
        c2*norm.pdf(x, m2, sig2) * sig2)
```

However, the criteria for what's considered a high difference between values of RMS consists of whether the difference is higher than the quadruple of the median of all obtained RMS', which meant we first had to fit a single Gaussian curve over all the features, obtain their RMS, and then realize these comparisons afterwards, but the results were very positive. While cases of three overlapping features were possible, they were infrequent enough for us to dismiss them. By this point, we also added another variable to the Gaussian curves to be detected and returned, the continuum line's offset, to further reduce the possibility of errors. We also added, as per the article's recommendations, recommended initial values and bounds for the attributes of the features we're fitting, to avoid obtaining any poorly fitted curves. The originally intended initialization for these values, as well as their bounds, can be easily summarized in Table 2.1. However, as we'll later see, some changes had to be made in response to the data we would work with.

Attribute	Initial Value	Bounds
d_i	1 - y_j	[0, 1]
μ_i	x_j	[min(λ_i), max(λ_i)]
σ_i	width(λ_i)/5	[0, width(λ_i)]

Table 2.1: Values for the initialization of a feature i's attributes before the Gaussian curve fit, and the bounds for the fit's results, where λ_i is the wavelength window for feature i, f_i is the corresponding set of values for the flux, y_j is the flux value located at index j such that $y_j = min(f_i)$, and x_j is the wavelength value for which we attained the minimum flux

With this, we tested the algorithm for multiple radial velocities which led to computing multiple shifted spectra. We realized that the RVs that were obtained from the algorithm were consistently slightly smaller in absolute value compared to the generated RVs. We deduced it was a systematic error caused by the intrinsic imperfections of our approximations, like using a Gaussian curve model to fit the spectral lines. Another issue that had been noticeable up to this point was how our algorithm ignored some of the more prominent lines in the curves. Our assumption was that the m window for the feature detection was too big for the number of points in those features, making it so that the algorithm would ignore them entirely. With real data, however, as we ended up seeing further ahead, this issue was not as noticeable, and the detection of the absorption features and their attributes was deemed to be satisfactory.

We attempted another technique to further improve our values, this time to the spectra we were using as the template, in the case we would not have any given template spectra to use. In that case, we would need to compute our own from the shifted spectra we would have. Therefore, depending on how shifted the spectra were in relation to each other, we would do one of the following: we either simply calculate a mean of all shifted spectra, and use the result as the final template spectrum, or we use that resulting spectrum as a first approximation, and use it to obtain approximations of the radial velocities. With those, we then ”re-shift” the shifted spectra with the corresponding values of obtained RVs towards their approximate unshifted position and calculate a mean of those resulting spectra to obtain a better template spectrum, with which we calculate the final difference spectra to then calculate the final RV values. Both of these methods, as they were initially planned, used simple unweighted means of the spectra to create a template. While we were going to further perfect them however, given how the data we had available included template spectra, this was an approach that ended up being unnecessary.

2.3 Adjustment of the H-Gs and calculation of the RVs

With the template and observed spectra obtained, as well as their features’ attributes, we can then create the difference spectrum, and use the values obtained to fit Hermite-Gaussian curves over every feature in the difference spectrum. However, the functions we first wrote to define the Hermite-Gaussian functions and fit the curves on them required a few changes first. One of the main changes done was fixating the degree of the Hermite function as 1 to avoid using too many unnecessary variables since we would not be using curves of any other degree. And, as it was mentioned prior to this section, we multiplied the *norm.pdf* function by sigma to cut that variable from the formula, since it does not appear in the Gaussian component of Equation (2.8). However, that leaves out a constant included with the *norm.pdf* function, $\frac{1}{\sqrt{2\pi}}$. The reason why I did not remove it comes from the fact that it would be introduced in the next function, where we would sum all of the features, now multiplied by their respective line depth, and fit a curve to the resulting list of values:

```
def HG_list(x, y_subtraction, list_m, list_sig, list_const,
    list_off, v_guess):
    def HG_helper(x, vr):
        y = np.zeros_like(x)
        for m, sig, con, off in zip(list_m, list_sig, list_const, list_off):
            c1 = con * np.sqrt(np.sqrt(np.pi)) * m / (c*np.sqrt(2*sig))
            y -= (HG_single(x, m, sig) * c1)
        return y*vr

    popt, pcov = curve_fit(HG_helper, x, y_subtraction,
        p0 = v_guess)
    vr = popt[0]
    y = HG_helper(x, vr)

    return y, vr
```

with said value lists being for μ, σ, the line depth (represented as the variable con) and the offset as the function's input, the latter of which we could not find a way to incorporate into the formula, and including an initial guess of the RV for the curve fit as well. The inclusion of the constant carried over from the previous formula makes it so that our current definition of the difference flux is the following:

$$y = v_r \sum_{j=1}^{n} \frac{\sqrt{\sqrt{\pi}}\mu}{c\sqrt{2\sigma}} d \tag{2.31}$$

with $d = \frac{con}{\sqrt{2\pi}}$ defined as the line depth, making con an unnormalized line depth.

Having obtained the lists of Gaussian spreads (σ_i), centers (μ_i) and amplitudes (d_i), we defined a function within the main HG_list function, one also of flux in function of the wavelength, defined above as the HG_helper, with the Radial Velocity as its variable. We first recursively create a list of flux values to which we keep adding up the calculated Hermite-Gaussian curves, using the values of the spectral line's attribute lists. Then, we fit the helper function to this newly created list of flux values to obtain the value of its only variable, the Radial Velocity of the shifted spectrum. During our testing work with generated spectra, only one major adjustment had to be made to this part of the algorithm, to correct the way this function was defined. It was also during this time that we ran into some issues regarding the shape of the obtained difference flux. With out first test runs, the continuum was not flat and instead showed large oscillations of a periodic nature, not too dissimilar to a large Hermite-Gaussian curve of the first degree. As it turned out, the feature detection algorithm, which at the time still did not have bounds for the fits of detected fits implemented, had obtained severely deviating values for some of the features, causing the large H-G curves to deform the continuum of the difference spectrum. We fixed this by introducing proper bounds for the values of the attributes of the spectral lines, in our spectral line detection algorithm.

Finally, in order to allow the algorithm to work with multiple shifted spectra, we added a for cycle to go over the entire list of spectra. We also took this chance to compute a graph generator that would show the template spectrum and the current shifted spectrum, as well as their difference spectrum and the fitted Hermite-Gaussian curves. I also took the liberty of using the obtained RVs to calculate new H-Gs, to see how would they compare to the ones obtained from the difference spectrum. One other thing must be taken into consideration, however: the signal-to-noise ratio of the values of each observation. Holzer et al. [2020] had realized multiple simulations of spectra with a known RV value, using the HGRV method on them to calculate the error that comes from modelling alone, by calculating the RMS of the obtained RV value of all n simulations against the fixed value:

$$\widehat{RMS}(\hat{v}_r) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{v}_{r,i} - v_r)^2} \tag{2.32}$$

We can also estimate the standard deviation (SD) and bias as

$$\widehat{SD}(\hat{v}_r) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{v}_{r,i} - \overline{v}_r)^2} \tag{2.33}$$

$$\widehat{Bias}(\hat{v}_r) = \overline{v}_r - v_r \tag{2.34}$$

with $\overline{v}_r$ as the average estimated velocity. Holzer et al. [2020] explored how these values, simulated to imitate those obtained from the star 51 Pegasi, varied with the SNR and the given radial velocity. With that, he uncovered that the HGRV method, in the small RV regime, can obtain a precision of less than $0.3\,\mathrm{m\,s^{-1}}$ for an SNR of approximately 250 or higher, while the bias is only above $0.1\,\mathrm{m\,s^{-1}}$ for RV values in the order of 100. Furthermore, expanding on the bias introduced by the model misspecification, the combination of multiple non-Gaussian absorption features in the HGRV method does not cause a bias amplification. In fact, the SD contributes more to the RMS than the bias, which is also shown to be somewhat proportional to the RV value.

Chapter 3

Applying the Method to Real Data

3.1 Further changes to the algorithm

Applying our method to real ESPRESSO observations posed some additional challenges. The data was stored in $.fits$ files, which required the package astropy.io to properly unpack and explore. As we previously stated, we were working with observations of the star *Tau Ceti*, of which we have 30 observations done on the 29th of November 2022, 15 of which were done from 1:13:02 to 1:30:28, and the latter 15 from 3:15:49 to 3:33:15. There is also a template spectrum available, which was built by median-combining individual observations. The spectra are provided in so-called S2D files, separated by the 170 Echelle orders of ESPRESSO. The wavelengths in the provided files were already corrected for the relativistic effects introduced by the Earth's motion, the values of which can be checked in the headers. For each order, we would obtain the corresponding template, and compare it with all of the provided observed spectra. Since the spectra cover a wavelength range of 3700-7900Å, the first few orders will be overfilled with features and their continuum line will be very sloped, and the spectral regions at longer wavelengths will have telluric features. It is also important to remember that, while we created our algorithm to work with velocities in the SI units (meters per second), the values within the real data files are set for radial velocities in kilometers per second, so a conversion of values was also taken into consideration, which involved a simple rescaling of the obtained RV values.

Each of the spectra of every order first needed to undergo proper treatment for better handling and for obtaining more precise results. For that, we need to first normalize the spectra, in order to remove the continuum in each order. For this, we decided to calculate a polynomial fit to the continuum, and then use it to normalize the spectra itself. However, many areas of the template spectrum have blanked out sections in several orders, with values that were set to zero. This is to mask regions of very low SNR, regions contaminated by telluric lines or cosmic rays, for example. Having these values set to 0 interferes with the calculation of the polynomial fit, so we use the provided masks, available in the .fits files, to convert the masked values into *NaN* values which can then be ignored using a *Numpy* function that determines which elements of an array are finite values or otherwise. We can then calculate the polynomial fit using only the finite values:

```
temp_wave = hdr_temp[1].data[order]
temp_y = hdr_temp[2].data[order]

maskbad = hdr_temp[3].data[order] == 1
temp_y[maskbad] = np.nan

idx = np.isfinite(temp_wave) & np.isfinite(temp_y)
cont_fit = np.polyfit(temp_wave[idx], temp_y[idx], 2)

temp_y = temp_y/np.polyval(cont_fit, temp_wave)
```

in which *hdr_temp* is the *.fits* file from which we obtained the wavelength and flux arrays, as well as the mask that was used. *idx* is a list of Boolean values, which is used to select the non-*NaN* values. The same process was used for the normalization of the observed spectra although, as we will later see, that led to some oversights.

With all of this information, we could now fit the Gaussian curves to the features of the template individually, then obtain their properties, and use them to calculate the respective H-Gs after calculating the difference flux for each observation and order. As we can see in Figure (3.1), many of the features found have a very minimal line depth, barely managing to distinguish themselves from the continuum line, as a result of the significance levels we had set. Most of our attempts to change these values within a reasonable range to try and ignore these minimal features did not prove to be very fruitful. While they could simply be a result of noise, many of these oscillations remain consistent among orders, hence why we decided to keep them for the rest of our analysis. We can also see in Figure (3.2) some of the cases where the features in a blend are not correctly detected. Besides this however, much like the first part of the algorithm, many changes had to be made to not only accommodate the algorithm for the format of real data we were now using, but to also correct and mitigate any errors that appeared during that process.

First, during the time when we were still testing the algorithm with a single real spectrum, we focused on being able to fit the new format of the available data, so the code

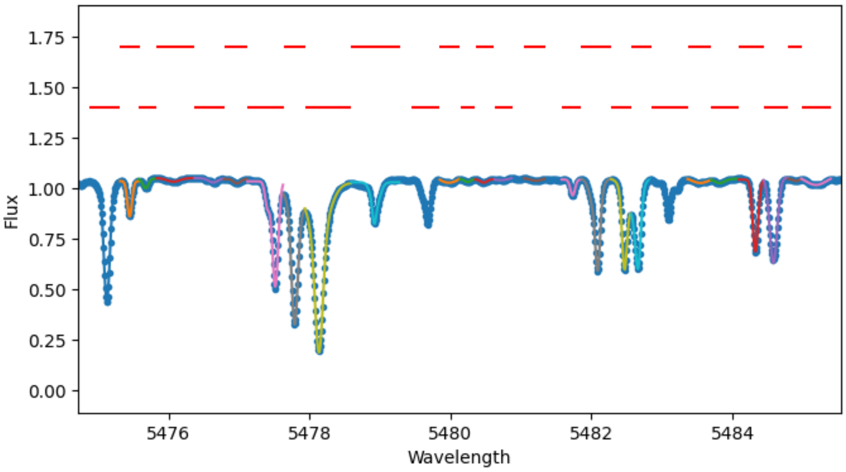


Figure 3.1: Section of the normalized template spectrum, showcasing the Gaussian curve fits done individually to each absorption feature. Many of the features detected have very minimal line depth.

was changed to take in the values from the previously created lists, so we could analyze the features of the spectrum, at each of its orders. This means we also had to remove the flux values at the wavelength ranges that were obscured by the template's mask, using the same method that was used to first treat the template spectrum. Afterwards, we have to normalize the observed spectrum's continuum and mask out the contaminated regions, using the same method we had for the template. However, the continuum varies substantially from order to order and from one spectrum to the other, so using a polynomial fit of a fixed degree proved to be unreliable, and it led to clear errors of normalization. As such, for each order, we calculate the first and second-degree fits of the continuum, and then we calculate the Root Mean Square of the residuals in order to choose the fit to normalize the spectrum. If none of the fits were suitable, we assumed the continuum was closer to a constant in shape, and simply divided the spectrum by its maximum value. Once the individual spectra are continuum normalized, we proceeded with the feature-finding algorithm.

Another detail regarding the algorithm had to do with the RV of the template spectrum. The template was calculated in Earth's rest frame and so had an RV value equal to 0. But the individual spectra have RVs around the systemic velocity of Tau Ceti which is around $-16.7\,\mathrm{km\,s^{-1}}$. Since our RV calculation algorithm is only valid for RVs smaller than $100\,\mathrm{m\,s^{-1}}$, we had to bring our template closer to the observed spectra, which meant we had to Doppler shift the template spectrum by a value close enough to the observed

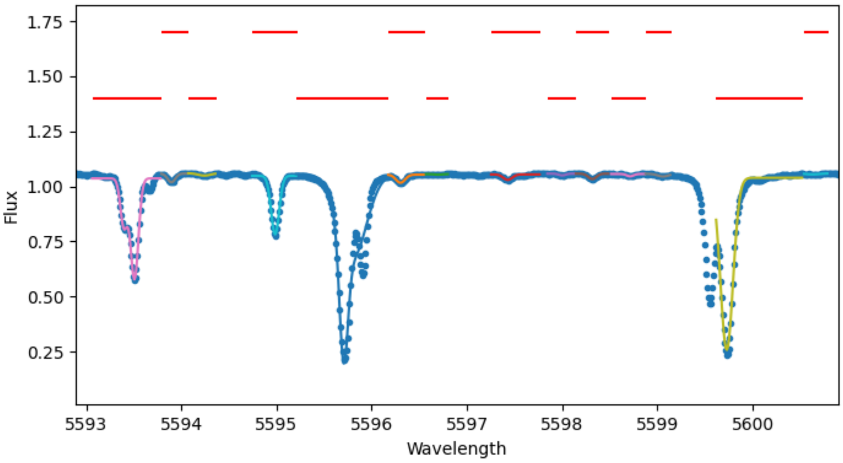


Figure 3.2: Section of the normalized template spectrum, showcasing the cases where the Gaussian curve fits done to blends result in one of the features of the blend being either badly detected, or outright undetected.

spectrum's RVs. If we shift the template to a value too close to the observed spectrum's RV, then the difference flux would be nonexistent, so we agreed on a value close enough to that of the RV that would bring the template's RV closer to those of the observation spectra. We did so with such a value that their difference would be under the limit for which our algorithm would still be valid. Then, once we obtain the features, calculate the difference flux and obtain the RV values from the curve fit to the H-Gs, we now have the values of the normalized RV shifts, to which we add the value used to Doppler shift the template spectra in order to obtain the true RV values.

We also needed to include in the algorithm the calculation of the uncertainties on the derived RV values. We started by obtaining the flux uncertainties, associated with each spectra, from their respective $.fits$ file. After treating the list of values with the flux spectrum's mask and normalizing it to the same continuum fit as the spectrum's, we also shift its values in accordance with the shift done to the template. However, it must be done in a different way, since for the uncertainties, we only want to change the index of the values, without changing the values themselves. So, for the flux error, we instead used this:

```
flux_err = doppler_shift(template_wavelength, flux_err,
    -rv*1e3, kind = 'nearest')
```

with the function $doppler\_shift(x, y, v, kind)$ being different from the one used in the first method, and instead allowing us to specify the kind of interpolation we want to do for the shift. We used the ”nearest” interpolation to keep the exact uncertainty, just with changed positioning. We then incorporate the flux uncertainties into the least squares fit of the H-G functions, which allows us to obtain a statistical estimate of the uncertainty in the derived RV. Some care had to be taken to mask out the previously set NaN values before the H-G fit.

3.2 Calculation of RVs order by order

At first, we had set the initial Doppler shift done to the template spectrum and its flux errors to an amount equal to $9/10$ of the symmetric of the CCF-obtained RV of the spectrum that was observed at 1:30:29, to see if the process worked. While the RV values obtained were still lower than those obtained through the CCF method, they were close in order of magnitude, so we were on the right track. What we needed was to bring the template spectrum closer to the observed spectra so that the RV difference value between them is below the $100\,\mathrm{m\,s^{-1}}$ limit. Therefore, we set the Doppler shift of the template to $199/200$ of the symmetric of the CCF-obtained RV of the spectrum that was observed at 1:30:29, a shift of approximately $16.58\,\mathrm{km\,s^{-1}}$, which led to a vast improvement of the values, since it equated to a difference between template and observed spectra of roughly $83.33\,\mathrm{ms^{-1}}$, which is the value we used for the initial guess of the RV for the H-Gs curve fit.

In order to locate the absorption features on the template spectrum, we defined the algorithm values by following some of the recommendations from Holzer et al. [2020], setting $m = 20$, $\alpha = 0.01$ and $\eta = 0.05$. We set a lower value for m, the size of the search window, in order to avoid skipping spectral lines. These values remained the same as we obtained the features of the template spectra over all 170 orders. However, when we set the initial values and bounds for the attributes of the Gaussian fits for the absorption features, we realized we had to increase the limit for the line depth, mostly due to the features in the earlier orders being relatively deeper and more asymmetric, with a continuum that was difficult to fully normalize. However, any features whose set initial values went over the normal limits, such as too wide or deep features, or those with negative values for the attributes, would be cut. With that set, the feature-finding process went normally for all orders of templates, getting the wavelength bounds of the features and their attributes as well, and saving them in distinct files.

From there, we ran the second part of the algorithm to go over the 30 $.fits$ files over each order, where we compared the respective template data with all observed spectra from that order, calculating the difference flux and using the Hermite-Gaussian formulas to fit the curve on it. After we obtained the values, we noticed that some of the RV values had inconsistently high errors associated with them, particularly those from orders 146, 147, 162 and 163. A look into the observed spectra revealed that, while we were normalizing them correctly, we were only masking the values that did not appear on the template spectra after their normalization. Because of that, some of the spectra, which had more irregular regions in their spectra - which corresponded to the masked sections of the templates - led to more irregular continuum lines, and therefore to curved difference flux spectra instead of flat, which increased the values of the RV errors drastically. So, we masked the observed spectra using the template masks before normalizing them, yielding values with smaller uncertainties. Some of the orders still had slightly higher error values, mostly due to frequent masked regions in their spectra, leading to fewer features to fit. In Figures (3.3) and (3.4), we can see examples of the H-G fits done to the difference flux after all the corrections.

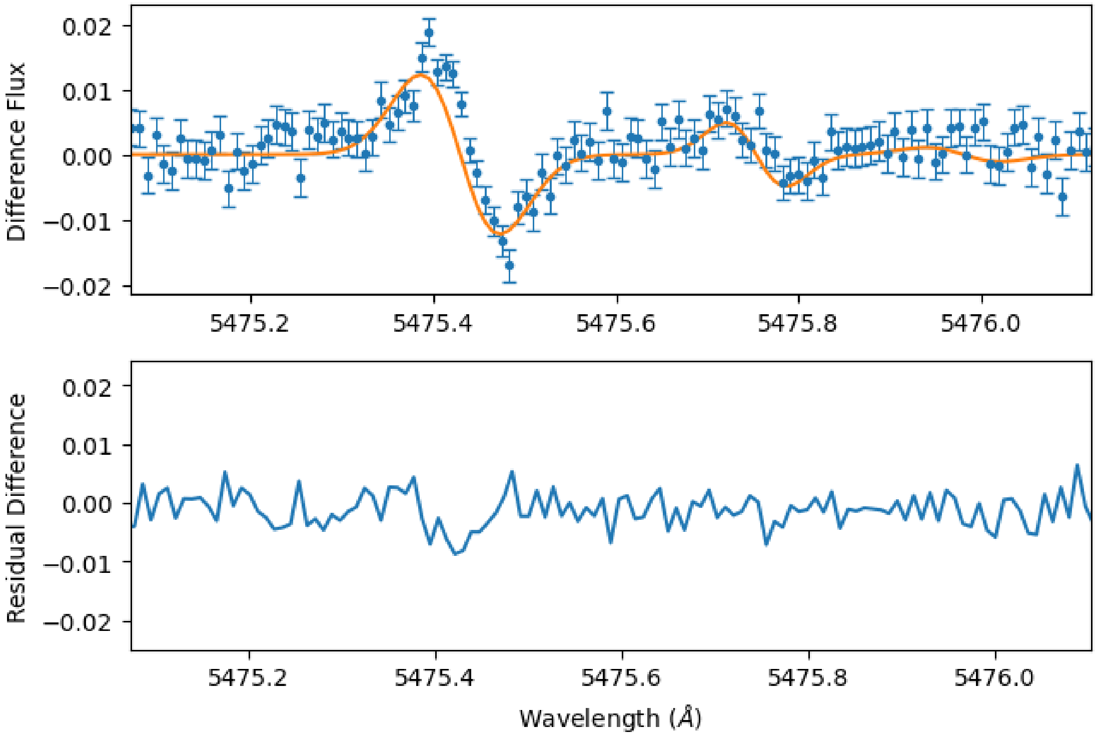


Figure 3.3: A section of the normalized difference flux array, with the H-G curve fit in orange, and the residual difference between the difference values and the fit. This fit is taken from the first observation, order 100, and for the most part, its residual difference does not go beyond the order of magnitude of $0.1\,\mathrm{m\,s^{-1}}$. However, we can see a deviation between the difference flux values and the curve fit. While the cause of the deviation is not clear, we believe it is the reason why the RV values are not as accurate as predicted.

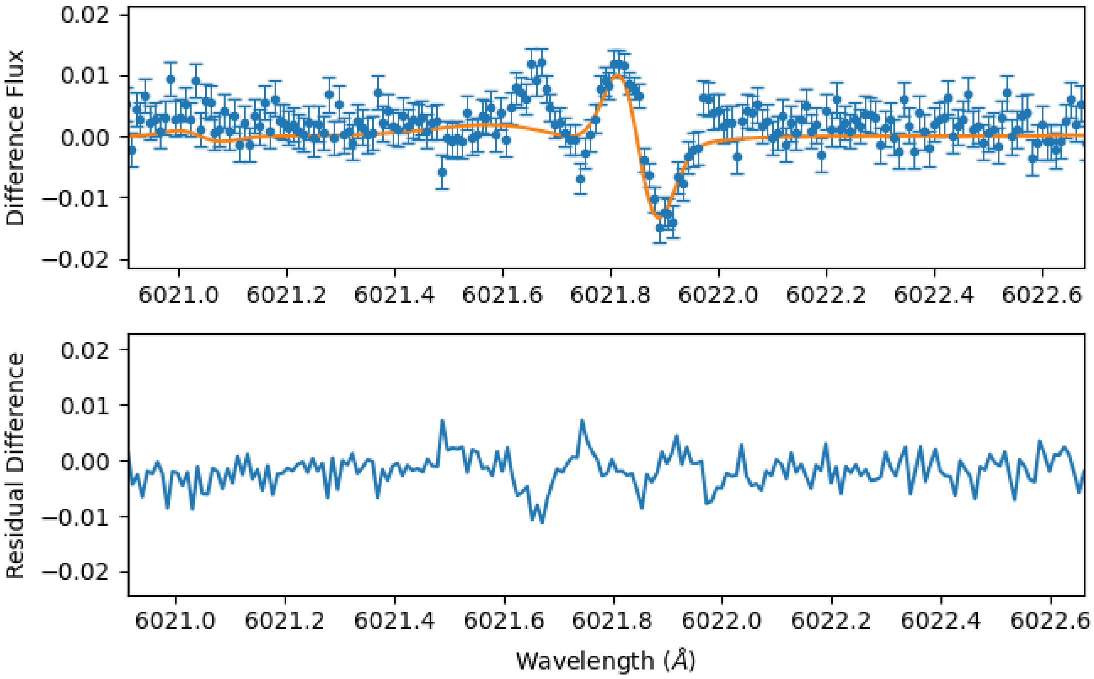


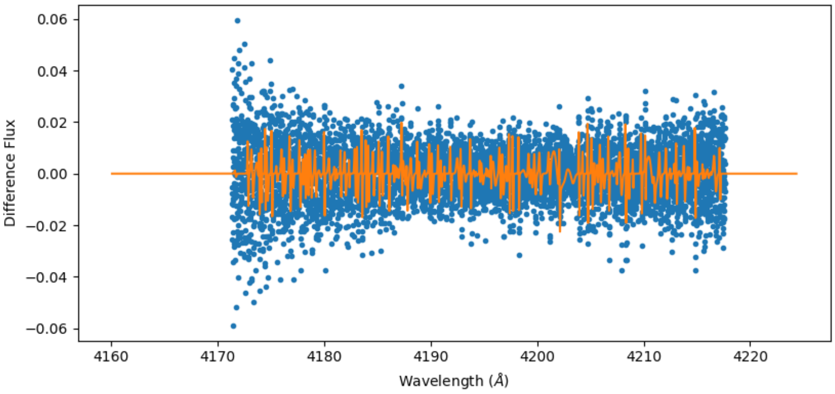
Figure 3.4: A section of the normalized difference flux array, with the H-G curve fit in orange, and the residual difference between the difference values and the fit. This fit is taken from the first observation, order 120, showcasing similar observations to those from the previous figure. This image is also an example of some of the greater deviations between the values in the difference flux and the fitted curve.

3.3 Combined results for all orders

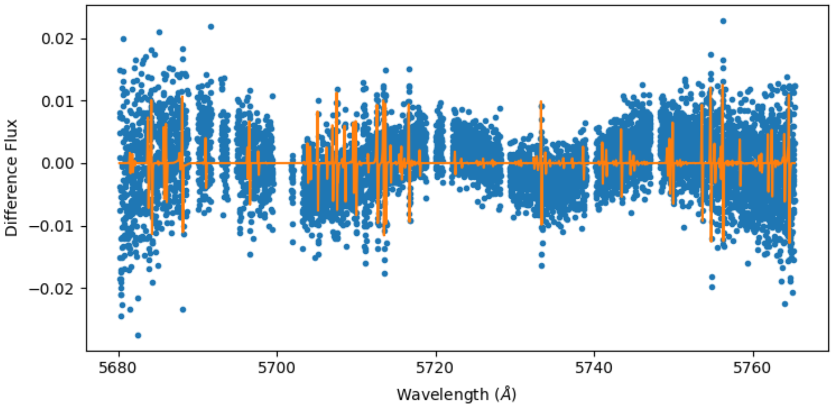
Once we obtained the Radial Velocity values, before we could combine the results, we decided to check some of the information about the observed spectra from which we obtained them. Each file has a signal-to-noise ratio (SNR) for each of the orders that it covers. As expected, the earlier orders, overfilled with features and stellar noise, and all before the stellar spectrum's flux peak, have a very low SNR as opposed to the later orders, although there's a decrease in the last values, and while it is possibly due to the presence of telluric features, the lesser amount of spectral lines could also play a part in that. Calculating the mean of the SNR of all 30 observed spectra per order, we'll take into account the RVs that came from orders with an SNR greater than 250, as recommended in Holzer et al. [2020]. With the orders of enough SNR, we then proceeded to calculate the mean of the RVs from the 30 observations per order, as well as of their errors, ignoring the orders for which their templates were fully masked out.

Looking more in-depth at the results, two main observations regarding the structure of the observed data were made. The first 90 orders, roughly within the range of [3800 Å - 5200 Å], both the left and right sections of the data sets are made out of null values, while past that order, in [5200 Å - 7900 Å], the observed values span the entire wavelength range. The other observation made was that the results from the orders seemed to group together in consecutive pairs. This is due to the fact that, for every two orders, two observations on the same wavelength range were made, leading to pairing results of great similarity, although since the similarities were not exact, we decided to not separate them. Looking back at the spectra of lower SNR, relative to those of higher SNR, and as it can be seen in Figure (3.5), we can see a much higher frequency of features, as well as some much wider features, blending with multiple ones of regular width at once. Plus, before the corrections that were needed for the method of normalizing the observed spectra, some abnormalities were found in the spectrum of some of the orders with higher SNR, due to the peculiar bands and features in the respective observed spectra. In some orders, one or two negative absorption features could be spotted, but since these abnormal flux spikes were not present in the pairing orders of the same wavelength range, we can assume these are actually a result of either an instrumental error or an error that comes from stray flux emissions. We also noticed, upon making a direct comparison between the template spectrum's masks and the observed spectra, that many regions in many of the later orders were vastly cut down, but of greater note are the regions within the ranges [5870 Å - 6000 Å], [6460 Å -

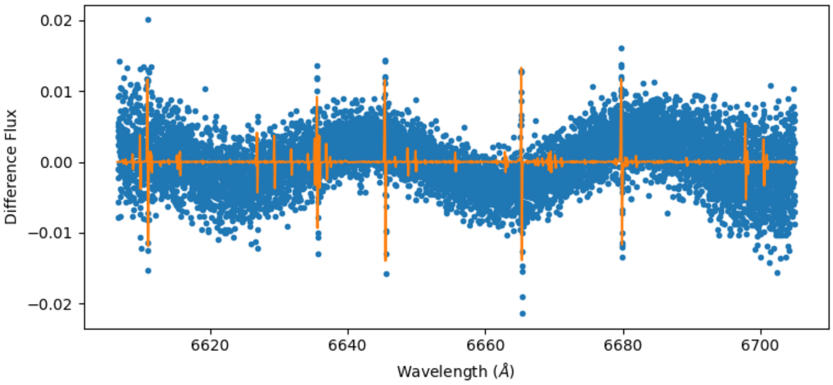
6600 Å], [6980 Å - 7430 Å], with this last region having entire orders cut out entirely, with an unusually high-frequency rate of absorption features, and [7590 Å - 7750 Å] also having many peculiarities. These last two regions also stuck out to us due to the noticeable odd change in the shape and frequency of their features, within the ranges [6870 Å - 6940 Å] and [7590 Å - 7700 Å].



(a) Order 40 - We can see how the left and right sections have NaN values, as well as the large amount of spectral lines



(b) Order 110 - We see some breaks in the middle of the difference flux due to the masks of the template. We can also see a fluctuation of the values



(c) Order 150 - At higher orders, the amount of spectral lines is very low

Figure 3.5: Examples of three distinct orders from the first observation

Chapter 4

Comparison of obtained RVs with other methods

In order to test the practicality and precision of the HGRV method, we decided to test the RV values we obtained by comparing them to the CCF-obtained values of the same observed spectra, present in the header of the $.fits$ files, as we can see in Figure (4.1). We did so by calculating the mean of all RVs and their errors across the selected orders per each of the 30 observations, to then directly compare them to the CCF-obtained RV values. We can observe that the values we obtained deviate from those obtained through the CCF, with both groups of values being outside the others' uncertainty range. However, this deviation is less than 0.5%, with all of the values still remaining in the same order of magnitude. Plus, as we can observe in Figure (4.2), the variation range of our values is also in the same order as that of the CCF values, despite a few outliers. Even still, the uncertainties of our values are roughly 20 times greater than those obtained with the CCF method.

We also wanted to take a look into the dispersion of the RV values we obtained and compare it to that of the CCF values, so we created two distinct histograms. The first, seen in Figure (4.3) was created with the 30 values we obtained from calculating the mean of all RV values obtained for all valid orders, per each of the observations, directly comparing it with the histogram created with the 30 RV values from the CCF. We can see that, while the RV values obtained from the CCF reveal a tendency towards two separate values, the HGRV values show a more cohesive dispersion, peaking at a single value, which means that for these observations, the HGRV method obtains values with a greater precision.

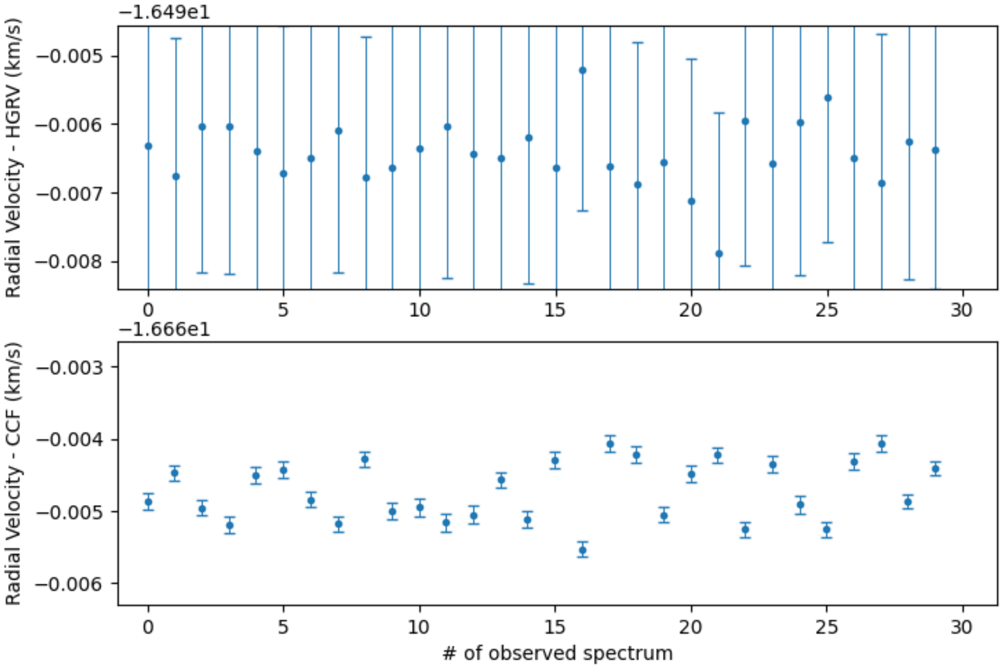


Figure 4.1: Comparison of the values obtained with the HGRV method (top) to those from the CCF method (bottom), with their respective uncertainties

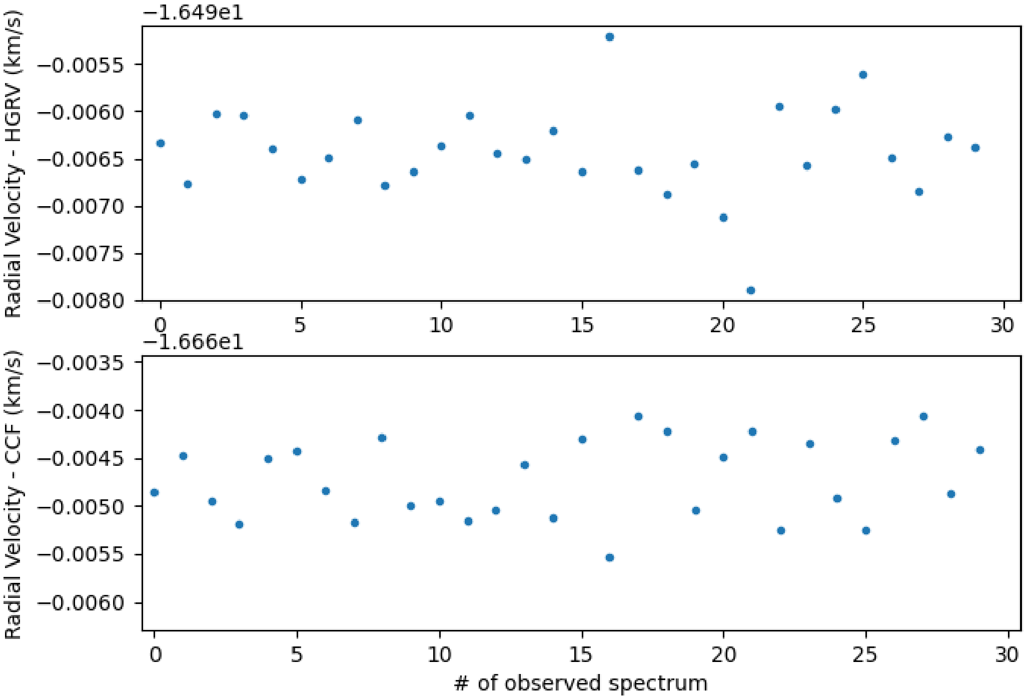


Figure 4.2: Comparison of the values obtained with the HGRV method (top) to those from the CCF method (bottom), with the Radial Velocity axis scaled in a way to allow for a comparison of the values' variation.

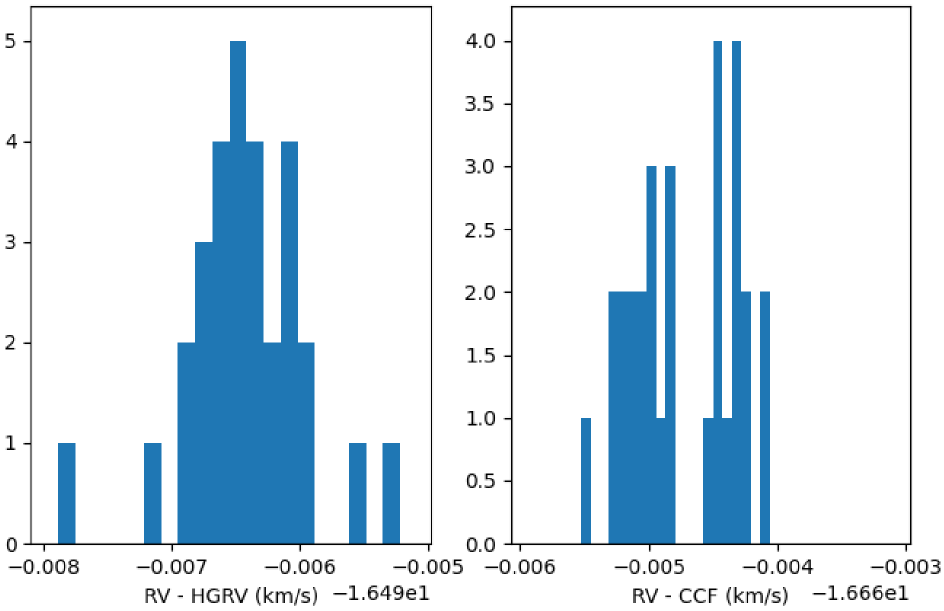


Figure 4.3: Histogram of the RV values obtained through the means of all values from all valid orders, per observation (left); and histogram of the CCF values (right). We can see how the CCF values formed two distinct groups of values, as opposed to those from the HGRV.

Chapter 5

Conclusion

In this work, we looked into the Radial Velocity method for detecting exoplanets. We examined the Hermite-Gaussian Radial Velocity method, proposed by Holzer et al. [2020], which showed promising results in terms of precision, and in doing so, we obtained a deeper understanding of the theory behind the method, and of the simulations that showcased the limits for which the method was valid. Our algorithm searches a template spectrum for spectral lines, which we assume are Gaussian in shape, to then obtain their properties. After that, we calculate the difference in flux between our template and the observation spectrum, use the properties of the features to define their respective Hermite-Gaussian curves of the first degree, and fit them in the difference flux, which will give us the observed spectrum's relative radial velocity to our template spectrum.

We then proceeded to put the method into practice, gaining more insight into the practical work needed to execute it, from developing the algorithm that finds absorption lines in a stellar spectrum and distinguishes them from each other within blends, obtaining their attributes, to the functions necessary to fit Hermite-Gaussian functions of first degree on the data that results from the difference between the template spectrum and the observed spectra. This was the most time-consuming part of our work, as we spent a long time not only testing our functions with randomly generated spectra, which had their own limitations that were not present in real spectra, as well as realizing many corrections to our formulas, caused by oversights during their initial development. Besides that, we also tried to come up with methods to tackle some of the method's obstacles, like the tackling of blended features and the construction of our own templates from the observed spectra, although some of these methods ended up at least partially unused in some areas, not deemed necessary or worthwhile, mostly due to time constraints.

When we moved to real data, we realized more work needed to be done to not only treat the data so it could be used but also change the algorithm to accommodate the format of the data. The observed spectra were all divided in the 170 echelle orders of ESPRESSO, so the algorithm had to obtain the template's features and use them to fit the H-Gs to the difference flux for each of them. Moreso, the template spectrum had many masked regions, so we needed to use the masks of each order to remove the corresponding sections of the observed spectra before normalizing their continuum. From this, we were not only able to obtain the radial velocities of each observation for each order, which all had different values of SNR, but we also implemented the template spectrum's uncertainties in the curve fits, so as to properly obtain the errors of the RV values. This part of our work, while not the most time-consuming, was the one that required the most changes and corrections, as multiple oversights of the results of individual parts of our algorithm resulted in poorly normalized spectra, bad H-G fits, or drastic increases in the RV errors.

While we were not able to achieve more accurate results than those obtained through the Cross-Correlation Function method, we did obtain a more precise array of values. The lack of accuracy could be attributed to a number of factors, such as the differences in the radial velocities between the provided template and the observed spectra initially proving to be too great for the limits within which the HGRV method can properly work. Therefore, we had to work around those limitations to be able to obtain valid results, like Doppler shifting the template spectra closer to their corresponding observed spectra, after which we took the obtained RV variations and added back the value used for the templates' shifting, which could have resulted in a loss of accuracy. Another factor that could have contributed to the differences between the obtained values and the CCF values is the high number of detected features with very minimal line depth, which could have simply been noises caused by noise and could have been ignored in favour of more prominent features in the spectrum, rather than removing weight from the contribution of these features to the results.

Furthermore, we sadly were not able to experiment with enough data nor for long enough to attempt to perform the necessary analysis in order to obtain the features of the orbiting exoplanets of Tau Ceti. That would have required an amount of data that could span over months of observations, instead of the two observation sessions of half an hour that we used. However, we hope that our work on this method and implementation of this algorithm can serve as an incentive for future development on the observations done and on the algorithm that was already constructed, to further develop the approaches already taken and perhaps even try out new approaches to the HGRV method. Some of the possible improvements on the feature detection and treatment could include some of the features we initially considered adding but ended up unused, like the detection of triple feature blends, due to their overall rarity in comparison to the double feature blends, or the continuum fits of variable degree for the template spectrum, which was only used for the observed spectra. We hope that, in the future, improvements could be made to our algorithm, to then utilize it to not only obtain the radial velocity variations of stars but also obtain the attributes of their orbiting exoplanets.

Bibliography

A. Wolszczan and D. A. Frail, "A planetary system around the millisecond pulsar psr1257 + 12," Nature, vol. 355, pp. 145–147, 1992. [Online]. Available: https://doi.org/10.1038/355145a0 [Cited on page 1.]

M. Mayor and D. Queroz, "A jupiter-mass companion to a solar-type star," Nature, vol. 378, pp. 355–359, 1995. [Online]. Available: https://doi.org/10.1038/378355a0 [Cited on page 1.]

H. J. Deeg and R. Alonso, "Transit photometry as an exoplanet discovery method," in Handbook of Exoplanets. Springer International Publishing, 2018, pp. 633–657. [Cited on page 1.]

F. Rodler, M. Lopez-Morales, and I. Ribas, "WEIGHING THE NON-TRANSITING HOT JUPITER τ boo b," The Astrophysical Journal, vol. 753, no. 1, p. L25, jun 2012. [Cited on page 3.]

S. Zucker, "Cross-correlation and maximum-likelihood analysis: a new approach to combining cross-correlation functions," Monthly Notices of the Royal Astronomical Society, vol. 342, no. 4, pp. 1291–1298, 07 2003. [Online]. Available: https://doi.org/10.1046/j.1365-8711.2003.06633.x [Cited on page 3.]

C. A. Prieto, "Velocities from cross-correlation: A guide for self-improvement," The Astronomical Journal, vol. 134, no. 5, pp. 1843–1848, sep 2007. [Cited on page 3.]

B. Campbell and G. A. H. Walker, "Precision radial velocities with an absorption cell." Publications of the Astronomical Society of the Pacific, vol. 91, no. 542, p. 540, aug 1979. [Cited on page 3.]

G. W. Marcy and R. P. Butler, “Precision radial velocities with an iodine absorption cell,” Publications of the Astronomical Society of the Pacific, vol. 104, no. 674, p. 270, apr 1992. [Cited on page 3.]

R. P. Butler, G. W. Marcy, E. Williams, C. McCarthy, P. Dosanjh, and S. S. Vogt, “Attaining doppler precision of 3 m s-1,” Publications of the Astronomical Society of the Pacific, vol. 108, no. 724, p. 500, jun 1996. [Cited on page 3.]

C. Lovis and D. A. Fischer, “Radial velocity,” 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:39135538 [Cited on page 4.]

J. T. Wright, “Radial velocities as an exoplanet discovery method,” in Handbook of Exoplanets. Springer International Publishing, 2018, pp. 619–631. [Cited on page 4.]

R. Butler, G. Marcy, E. Williams, H. Borland, and P. Shirts, “Three new ”51 peg-type” planets,” Astrophysical Journal, vol. 474, 07 1999. [Cited on page 4.]

M. Mayor, F. Pepe, D. Queloz, F. Bouchy, G. Rupprecht, G. Lo Curto, G. Avila, W. Benz, J. L. Bertaux, X. Bonfils, T. Dall, H. Dekker, B. Delabre, W. Eckert, M. Fleury, A. Gilliotte, D. Gojak, J. C. Guzman, D. Kohler, J. L. Lizon, A. Longinotti, C. Lovis, D. Megevand, L. Pasquini, J. Reyes, J. P. Sivan, D. Sosnowska, R. Soto, S. Udry, A. van Kesteren, L. Weber, and U. Weilenmann, “Setting New Standards with HARPS,” The Messenger, vol. 114, pp. 20–24, Dec. 2003. [Cited on page 4.]

D. A. Fischer, G. Anglada-Escude, P. Arriagada, R. V. Baluev, J. L. Bean, F. Bouchy, L. A. Buchhave, T. Carroll, A. Chakraborty, J. R. Crepp, R. I. Dawson, S. A. Diddams, X. Dumusque, J. D. Eastman, M. Endl, P. Figueira, E. B. Ford, D. Foreman-Mackey, P. Fournier, G. Fű rész, B. S. Gaudi, P. C. Gregory, F. Grundahl, A. P. Hatzes, G. Hébrard, E. Herrero, D. W. Hogg, A. W. Howard, J. A. Johnson, P. Jorden, C. A. Jurgenson, D. W. Latham, G. Laughlin, T. J. Loredo, C. Lovis, S. Mahadevan, T. M. McCracken, F. Pepe, M. Perez, D. F. Phillips, P. P. Plavchan, L. Prato, A. Quirrenbach, A. Reiners, P. Robertson, N. C. Santos, D. Sawyer, D. Segransan, A. Sozzetti, T. Steinmetz, A. Szentgyorgyi, S. Udry, J. A. Valenti, S. X. Wang, R. A. Wittenmyer, and J. T. Wright, “State of the field: Extreme precision radial velocities,” Publications of the Astronomical Society of the Pacific, vol. 128, no. 964, p. 066001, may 2016. [Cited on page 4.]

P. Holzer, J. Cisewski-Kehe, D. Fischer, and L. Zhao, “A hermite-gaussian based radial velocity estimation method,” 2020. [Cited on pages 5, 12, 13, 14, 17, 18, 25, 31, 35, and 43.]

F. Pepe, S. Cristiani, R. Rebolo, N. C. Santos, H. Dekker, A. Cabral, P. Di Marcantonio, P. Figueira, G. Lo Curto, C. Lovis, M. Mayor, D. Mégevand, P. Molaro, M. Riva, M. R. Zapatero Osorio, M. Amate, A. Manescau, L. Pasquini, F. M. Zerbi, V. Adibekyan, M. Abreu, M. Affolter, Y. Alibert, M. Aliverti, R. Allart, C. Allende Prieto, D. Álvarez, D. Alves, G. Avila, V. Baldini, T. Bandy, S. C. C. Barros, W. Benz, A. Bianco, F. Borsa, V. Bourrier, F. Bouchy, C. Broeg, G. Calderone, R. Cirami, J. Coelho, P. Conconi, I. Coretti, C. Cumani, G. Cupani, V. D’Odorico, M. Damasso, S. Deiries, B. Delabre, O. D. S. Demangeon, X. Dumusque, D. Ehrenreich, J. P. Faria, A. Fragoso, L. Genolet, M. Genoni, R. Génova Santos, J. I. González Hernández, I. Hughes, O. Iwert, F. Kerber, J. Knudstrup, M. Landoni, B. Lavie, J. Lillo-Box, J. L. Lizon, C. Maire, C. J. A. P. Martins, A. Mehner, G. Micela, A. Modigliani, M. A. Monteiro, M. J. P. F. G. Monteiro, M. Moschetti, M. T. Murphy, N. Nunes, L. Oggioni, A. Oliveira, M. Oshagh, E. Pallé, G. Pariani, E. Poretti, J. L. Rasilla, J. Rebordão, E. M. Redaelli, S. Santana Tschudi, P. Santin, P. Santos, D. Ségransan, T. M. Schmidt, A. Segovia, D. Sosnowska, A. Sozzetti, S. G. Sousa, P. Spanò, A. Suárez Mascareño, H. Tabernero, F. Tenegi, S. Udry, and A. Zanutta, “ESPRESSO at VLT. On-sky performance and first results,” Astronomy & Astrophysics, vol. 645, p. A96, jan 2021. [Cited on page 5.]

Tuomi, M., Jones, H. R. A., Jenkins, J. S., Tinney, C. G., Butler, R. P., Vogt, S. S., Barnes, J. R., Wittenmyer, R. A., O´Toole, S., Horner, J., Bailey, J., Carter, B. D., Wright, D. J., Salter, G. S., and Pinfield, D., “Signals embedded in the radial velocity noise - periodic variations in the i velocities,” A&A, vol. 551, p. A79, 2013. [Online]. Available: https://doi.org/10.1051/0004-6361/201220509 [Cited on page 5.]

F. Feng, M. Tuomi, H. R. A. Jones, J. Barnes, G. Anglada-Escudé , S. S. Vogt, and R. P. Butler, “Color difference makes a difference: Four planet candidates around tau ceti,” The Astronomical Journal, vol. 154, no. 4, p. 135, sep 2017. [Cited on page 5.]

M. Planck, “On the law of distribution of energy in the normal spectrum,” Annalen der Physik, vol. 4, pp. 553–562, 1901. [Cited on page 9.]

D. F. Gray, The Observation and Analysis of Stellar Photospheres. Cambridge University Press, 2005. [Cited on page 10.]

J. T. Wright and J. D. Eastman, “Barycentric corrections at 1 cm s^{-1} for precise doppler velocities,” Publications of the Astronomical Society of the Pacific, vol. 126, no. 943, pp. 838–852, sep 2014. [Cited on page 10.]

R. T. Blackman, A. E. Szymkowiak, D. A. Fischer, and C. A. Jurgenson, “Accounting for chromatic atmospheric effects on barycentric corrections,” The Astrophysical Journal, vol. 837, no. 1, p. 18, feb 2017. [Cited on page 10.]

R. T. Blackman, D. A. Fischer, C. A. Jurgenson, D. Sawyer, T. M. McCracken, A. E. Szymkowiak, R. R. Petersburg, J. M. J. Ong, J. M. Brewer, L. L. Zhao, C. Leet, L. A. Buchhave, R. Tronsgaard, J. Llama, T. Sawyer, A. B. Davis, S. H. C. Cabot, M. Shao, R. Trahan, B. Nemati, M. Genoni, G. Pariani, M. Riva, P. Fournier, and R. Pawluczyk, “Performance verification of the extreme precision spectrograph,” The Astronomical Journal, vol. 159, no. 5, p. 238, apr 2020. [Cited on page 10.]

W. Johnston, “The weighted hermite polynomials form a basis for l2($\mathbb{R}$),” The American Mathematical Monthly, vol. 121, no. 3, pp. 249–253, 2014. [Online]. Available: https://www.tandfonline.com/doi/abs/10.4169/amer.math.monthly.121.03.249 [Cited on page 11.]