



# The other spectrum of the stars

Santiago Henao Castellanos

Director: Dr. Alejandro García

A thesis submitted in partial fulfillment  
for the degree of Bachelor of Physics

Universidad de los Andes  
Science Faculty, Physics Department  
Bogotá, Colombia

December 6th, 2021

# Resumen

# Abstract

*I dedicate the prose to Trinidad,  
for making every morning coffee  
that fueled the writing of this document*

*But all the code is for Miguel  
who will not be able to compile it anymore  
but I think he would have liked it anyway*

# Acknowledgements

There were a lot of people who provided me with their guidance and support on the process of writing this document. It was not purely a research process, but a personal voyage of discovery and motivation.

Thus, I would like to thank first Alejandro García, the director of this monograph, for motivating me since the beginning of my career, and being patient with its eventual delay. The realization of that career would not have been possible without the gargantuan support and love of my parents and my family, who were there on every step along the way. Neither without the aid of the University itself, which staff gave me and othersthe opportunity of continuing our studies on the still ongoing world situation.

I am also grateful with my dearest friends: Miguel, Juan David, Valeria, Camilo, Juanita and the two Nathalias. My time at Bogotá would have been very sorrowful if not for them. Also my dear David, who on recent times of crisis took care of me back in my hometown.

Finally, I write this couple of words to attempt to express the unmeasurable gratitude I have for Laura.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Discovering stellar variability . . . . .	6
1.2	The birth of the Period-Luminosity relation . . . . .	7
1.3	Measuring space and time with variable stars . . . . .	9
1.4	Recent developments . . . . .	10
<b>2</b>	<b>Theoretical Framework</b>	<b>12</b>
2.1	Light curves . . . . .	12
2.1.1	Julian Dates . . . . .	12
2.1.2	Magnitudes . . . . .	16
2.1.3	Phase diagrams . . . . .	23
2.2	Cepheid Variable stars . . . . .	24
2.2.1	Stellar populations . . . . .	27
2.2.2	Radial stellar pulsation . . . . .	27
2.3	The other spectrum, the pulsation spectrum . . . . .	31
2.3.1	Fourier Analysis . . . . .	32
2.3.2	Phase diagram methods . . . . .	37
<b>3</b>	<b>Methodology</b>	<b>41</b>
3.1	OGLE IV classical Cepheids on the Magellanic clouds . . . . .	41
3.1.1	The OGLE project . . . . .	41
3.1.2	Data acquisition and description . . . . .	42
3.2	Implementation details . . . . .	47
3.2.1	Algorithms . . . . .	47

3.2.2	Examples and the sampling problem . . . . .	51
<b>4</b>	<b>Results</b>	<b>55</b>
4.1	Final PL relations . . . . .	55
4.2	Discussion . . . . .	55
<b>5</b>	<b>Conclusions</b>	<b>56</b>
5.1	Future work . . . . .	56
<b>Appendices</b>		
<b>A</b>	<b>Code utilities</b>	<b>58</b>
A.1	C listings . . . . .	58
A.2	Python listings . . . . .	58
<b>B</b>	<b>Complementary figures</b>	<b>64</b>

# List of Figures

1.1	Light curve of $\delta$ Cephei and $\eta$ Aquilæ . . . . .	7
1.2	Leavitt PL relation . . . . .	8
2.1	Conceptual view of the Rømer delay . . . . .	15
2.2	Electromagnetic wave propagation through a sperical surface . . . . .	16
2.3	Johnson-Cousins and OGLE-IV photometric systems . . . . .	20
2.4	Failure of interpolation of a light curve . . . . .	23
2.5	Light curve of OGLE-LMC-CEP-0870 . . . . .	25
2.6	Light curve of OGLE-LMC-CEP-0625 . . . . .	25
2.7	Variable stars in the HR diagram . . . . .	26
2.8	Complex phase diagram: Fourier curl . . . . .	33
2.9	Off-frequency phase diagrams: real and complex . . . . .	35
2.10	Off-frequency phase diagrams: arclength . . . . .	38
2.11	Off-frequency phase diagrams: histograms . . . . .	40
3.1	Distribution of data size for classical Cepheids in the Magellanic clouds . . . . .	44
3.2	Distribution of signal properties for classical Cepheids in the Magellanic clouds . . . . .	45
3.3	Hourly difference between observations for OGLE-LMC-CEP-1234 . . . . .	46
3.4	OGLE-IV observation cadence in the Magellanic system . . . . .	47
3.5	Illustrative benchmark for the algorithms . . . . .	52
3.6	Example of spectra near the true frequency . . . . .	53
3.7	Detail of the failure mode of Fourier and entropy methods . . . . .	53
B.1	Comparison between even and uneven sampled signal past the nyquist limit . . . . .	65



# List of Tables

2.1	Johnson-Cousins effective wavelengths and bandwidths . . . . .	21
3.1	Pulsation mode and filter data distribution for the LMC . . . . .	42
3.2	Pulsation mode and filter data distribution for the LMC . . . . .	43
A.1	Numba type aliases . . . . .	59

# List of Listings

1	<code>eps</code> C function . . . . .	58
2	Library aliases for Python . . . . .	59
3	<code>phase</code> function . . . . .	59
4	Arclength method implementation . . . . .	59
5	2D histogram simple algorithm . . . . .	60
6	Naive entropy method implementation . . . . .	60
7	“Flattened” entropy method implementation . . . . .	61
8	Dispersion method implementation . . . . .	61
9	Non-redundant implementation of the Lomb-Scargle periodogram . . . . .	62
10	Naive implementation of the Fourier periodogram . . . . .	62
11	Incremental iterative implementation of the Fourier Periodogram . . . . .	63

# Chapter 1

## Introduction

### 1.1 Discovering stellar variability

Ancient  
observations

It is undeniable that ancient civilizations looked to the night sky with great interest. Stars were considered innamobile and immutable, with the notable exceptions of planets and Novæ, respectively. But periodic stellar variability was probably known since antiquity too. For instance, Egyptians knew the period of Algol three millennia ago (Jetsu & Porceddu 2015; Jetsu et al. 2013), and the mythology of several cultures seems to have some references to the phenomenon (Wilk 1996). On some cases, periodic variable stars were recorded as Novæ on some ancient sources (Ho Peng Yoke & Ho Ping-Yü 1962).

Western  
rediscovery

The modern rediscovery of stars having periodic changes on their brightness is accredited to Fabricius, who discovered Mira (*o Ceti*) in 1596 (Hoffleit 1997). This phenomenon was in direct contradiction with the classical philosophical view of the Universe, as expressed by Aristotle (350 BCE, book I, part 3): “so far as our inherited records reach, no change appears to have taken place either in the whole scheme of the outermost heaven or in any of its proper parts”. On the following centuries several other stars were marked as candidates for periodic variables. Edward Pigott and his collaborator John Goodricke were verifying those candidates in 1784 (Hoskin 1979). Famously, Pigott (1785) observed  $\eta$  Aquilæ and confirmed its variability, while Goodricke (1786) observed  $\delta$  Cephei, Algol,  $\beta$  Lyræ, and others.

Initial theories  
for stellar  
variability

Pigott and Goodricke results are remarkable, as almost a century would pass until the formalization of the magnitude scale for brightness by Pogson (1856). But despite the lack of truly (instrumental) quantitative measurements of brightness, there were plenty of theories regarding the mechanism

behind stellar variability. Those theories included planetary eclipses, star-star collisions, binary stars, meteor impacts, obfuscation by gas or dust clouds, and sun-like spots coupled with rotation and axial tilts (Hoffleit 1993). And although the eclipse conjecture was later proved to be the correct for Algol (Pickering 1880), the asymmetric behaviour of the brightness of some stars (as those seen in Figure 1.1) could not be accounted for. If the change of brightness were to be the product of an eclipse, the brightness would spend the same amount of time increasing as decreasing, as the transiting object covers and uncovers the star. On the same train of thought, if the star happened to have a side full of spots, or obscured by a dust cloud, no stable movement could make the brightness go from minimum to maximum faster than from maximum to minimum on such a periodic manner.

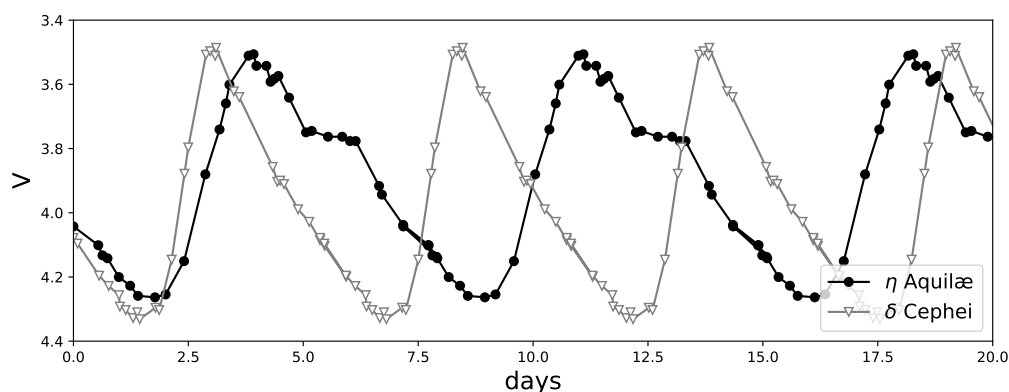


Figure 1.1: Modern light curves for the first two discovered Cepheid variable stars.  $\eta$  Aquilæ has a period of  $\sim 7.17$  days, in contrast with the  $\sim 5.36$  days of  $\delta$  Cephei. Objects are brighter as magnitude ( $V$ ) decreases (see chapter 2 for details). Note how both stars take more time decreasing their brightness than increasing it. Figure reconstructed from Kiss (1998) data.

## 1.2 The birth of the Period-Luminosity relation

The need of  
classification

These variable stars were initially classified by Pickering (1880), only taking into account the shape of the light curve: Novæ, Mira-like, eclipsing, irregular variables, and class for the aforementioned irregular (but highly periodic) case, namely  $\beta$  Lyræ and  $\delta$  Cephei. There was a need for further classification of these stars, with several attempts of break down Pickering classes (Lockyer 1896, 1897), but it would take several important discoveries in astronomy for the classification of variable

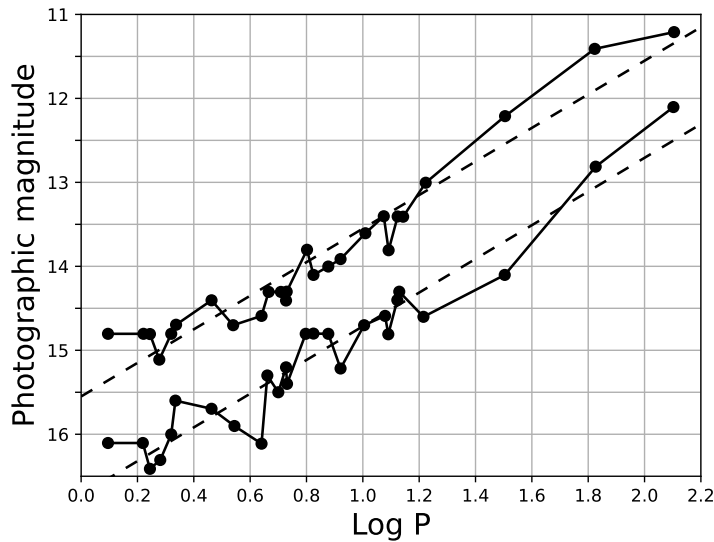


Figure 1.2: [Leavitt & Pickering \(1912\)](#) representation of the relationship between period and luminosity of 25 Cepheid stars in the Small Magellanic Cloud. The two data series refer to the points of maximum and minimum brightness for each star. The period of the stars ( $P$ ) is reported on days, so the abscissa is in units of  $\log(\text{days})$ . The ordinate is given in photographic magnitude.

stars to fully develop, resulting in  $\delta$  Cephei as the prototype of its own class of variable stars.

Work and  
discoveries of  
the Harvard  
Computers

Those next big steps on the field were done by Pickering's "computers" women at Harvard. Although the universities refused to let women study at the time, Pickering (then director of Harvard's observatory) needed people to process and analyze the sheer amount of data being produced in the observatory, on the form of stellar spectra and photographic plaques of the observations. Working there, Williamina Fleming prepared the Draper Catalogue of stellar spectra ([Maury & Pickering 1897](#); [Pickering 1890](#)), which allowed her to classify Mira-type stars and Novæ with only their spectrum. This classification system was later reordered by [Cannon & Pickering \(1901\)](#) to reflect the temperature of the stars. Around that time Henrietta Leavitt joined the Harvard computers, with the task of searching for variable stars in the Magellanic Clouds, satellite Galaxies of the Milky Way known at the time as Nebulæ. There she discovered almost two thousand variables ([Leavitt 1908](#)), the majority of which turned out to be of the  $\delta$  Cephei class. By calculating the periods of 25 of those stars and comparing them to their maximum and minimum magnitudes,

Leavitt found a linear relationship (Figure 1.2).

The Period-  
Luminosity  
relation

Leavitt's discovery was of immense importance as a possible tool for measuring the distance to the farthest cepheid stars. She noted: "Since the variables are probably at nearly the same distance from the Earth, their periods are apparently associated with their actual emission of light" (Leavitt & Pickering 1912, page 3). Leavitt knew she was seeing the *apparent* brightness of the stars, as brightness decreases with distance. If distances to near Cepheids could be found by another method, this Period-Luminosity (PL) relation could be reversed to find the *absolute* brightness of the Cepheids in the Small Magellanic Cloud (SMC), finding the distance to the Nebula. Leavitt correctly proposed the method of parallaxes to solve this task, but his work on Harvard did not let her pursue this investigation.

fine-tuning the  
PR relation and  
the Magellanic  
clouds

The parallaxes for the nearest Cepheids were examined by (Hertzsprung 1913). He used Leavitt's PL relation to calculate the distance to the SMC, giving (after a "pen error", see Fernie 1969) 9.2 kilo parsecs (kpc) (equivalent to 30 kly), which is a vast underestimation. Shapley (1918) improved further the precision of Leavitt's original PL relation, but its zero point (its accuracy) was assumed to be the correct one, despite having several systematic and selection errors (Fernie 1969). Shapley used his results to calculate the distance to the Magellanic Clouds, obtaining 32.5 kpc (106 kly) for the Small Cloud (Shapley 1924a) and 34.4 kpc (112 kly) for the Large Cloud (Shapley 1924b).

### 1.3 Measuring space and time with variable stars

The great  
debate, and  
Hubble  
measuring the  
size of the  
universe

Around this time, there was a big debate in the astronomical community around the nature of the so-called "spiral Nebulae". Were those Nebulae part of the Milky Way, or were they their own separate, distant galaxies? E. Hubble, using Shapley's version of the PL relation, settled the matter when he found a distance of 285 kpc (929 kly) for M31 and M33 (now called the Andromeda and Triangulum galaxies, Hubble 1925a) and 214 (697 kly) for NGC 6822 (Barnard's galaxy, Hubble 1925b), surpassing Shapley's hypothesis that the galaxy (and the whole universe) was only 92 kpc (300 kly) wide (Shapley & Curtis 1921).

Space means  
time: the  
Cosmic age  
problem

With this results, Hubble (1929) measured the redshift of some extra-galactic Nebulae, and encountered a linear relationship between distance as velocity. It was the first experimental evidence of the general relativity prediction of an expanding universe (Friedmann 1922; Lemaître 1927). Hubble estimate of this expanding rate was  $H_0 \approx 550(\text{km/s})/\text{Mpc}$  (the Hubble's constant), which implied the universe must have an age of  $2 \times 10^9$  years. Meanwhile, geologists estimated the age of the earth as  $4 \times 10^9$  years (Dalrymple 1994, for an historical account see). How could be the Earth

older than the universe? There was a big problem in the figures, somewhere.

Baade  
correction

Hubble, again, had half the answer. He had some concerns about the zero point of Shapley’s PL relation, an opinion shared with Baade (1944) after he divided the stars on two populations based on their metallicities. Baade realized that population I Cepheids were 1.5 magnitudes brighter than population II (for the same period) so there were *two* different PL relations, shifted on their zero point by 1.5 magnitudes. He proved his hypothesis at Palomar Observatory, observing the Andromeda nebula (Arp 1955; Baade 1956). As a consequence of the distance-modulus equation, measurements made with population I Cepheids should be scaled by a factor of  $10^{3/10} \approx 2$ . The Hubble distances were based on population I Cepheids, and therefore his distances doubled<sup>1</sup>. This resulted on a halved Hubble constant and consequently doubled the observed age of the universe,  $\sim 1.7 \times 10^9$  years. Patterson et al. (1955) measured the age of the Earth as  $4.55 \times 10^9$  years, so the problem was not solved yet.

Sandage  
correction

A second correction came from the work of Sandage. First reviewing Hubble’s work with more data (Humason et al. 1956), and then correcting another problem: Hubble had made the calibration of the distance for the farthest nebulae with the brightest resolvable stars he had, possibly incurring in an identification mistakes with ionized Hydrogen regions or multiple nearby stars (Sandage 1958). This considerations amounted a total correction of 4.1 magnitudes, and produced a Hubble’s constant of  $H_0 \approx 75(\text{km/s})/\text{Mpc}$ , or a timescale for the universe of  $\sim 1.3 \times 10^{10}$  years, comparable to modern estimates (Freedman et al. 2001).

## 1.4 Recent developments

Present status  
of the distance  
ladder

In the present day, mainly two astronomical projects are entitled to the search of variable stars and the refinement of cosmic distances estimation: the OGLE project<sup>2</sup> and the Araucaria project<sup>3</sup>. While the Gaia space observatory measures parallaxes, the first step on the distance ladder, the OGLE project surveys the galaxy and the Magellanic Clouds in search of variable stars and gravitational microlensing. They have nearly completed Leavitt initial task of catalogue the Cepheid stars on the Magellanic Clouds (Soszyński et al. 2017), and their survey of Cepheids on the Milky Way has allowed to further study its structure (Skowron et al. 2019) and its dynamics (Mróz et al. 2019). On the other side, the Araucaria project focuses on the calibration of standard candles, comparing

---

<sup>1</sup>Shapley’s distance to the SMC was a little more than its actual value of  $\sim 60$  kpc, so this doubling actually made his results very accurate for his time.

<sup>2</sup><http://ogle.astrouw.edu.pl/>

<sup>3</sup><https://araucaria.camk.edu.pl/>

different methods: they have calibrated the distance to the SMC within a 3% of precision (Graczyk et al. 2014) and to the LMC within a 1% (Pietrzyński et al. 2019).

The other  
spectrum, the  
Fourier  
spectrum

Just one more particular problem has surfaced on the PL relation. As a consequence of the pulsation mechanism of these stars, the underlying physical machinery that makes their brightness oscillate, Cepheid variable stars can pulsate on several *overtones* of their natural frequency. This overtones displaces the PL relation, as is multiplicative operation on the period, which amounts on a different position on the  $\log P$  axis. Therefore, stars with different modes of pulsation must be separated before attempting to calibrate a PL relation (Zabolotskikh et al. 2005). The most natural way of attacking this problem is using the other spectrum of the star: not the light-energy usual spectrum, but the Fourier spectrum, which allow us to see the frequencies and phase properties of the light curve.

We have seen that as experimental techniques evolve and become more precise, the theoretical aspects of the PL relation became increasingly important for the accurate determination of distances on the universe. Even so, with the amount of observational data these projects are producing, the use of efficient and robust methods for analyzing it became imperative too.

Although of capital importance, the methods of image reduction and magnitude determination lie outside the scope of this work. The aim of this monograph is to examine and implement the different methods for finding the period of a Cepheid variable star. Each method will be tested on real Cepheid data, in order to select the one better suited for the task of producing reliable PL relations. The resulting method (or combination of methods) will be used to replicate the PL relations of the Cepheids on the Magellanic Clouds given by Soszyński et al. (2015).



## Chapter 2

# Theoretical Framework

### 2.1 Light curves

The main concept present on the photometric analysis of variable stars is the light curve: a diagram representing the **temporal** evolution of the star's **brightness**. Light curves are used to describe all types of variable stars, both periodic and non-periodic. In the case of periodic variable stars, a **phased** light curve is often preferred.

In order to properly study variable star, we need to have solid definitions for “temporal”, “brightness” and “phase” measurements.

#### 2.1.1 Julian Dates

Common  
calendars as a  
mixed-radix  
complicated  
unit

Calendars have existed for as long as humans have been able to keep track of time; since the dawn of astronomy. But similarly to the astronomical and physical sciences, calendars change with culture and also with time itself. The daylight, lunar and seasonal cycles do not align nicely; their periods are not a simple fraction of each others. This causes all calendars that try to align the day period with any other solar system based period to be a mixed radix unit full of exceptions. If your calendar is moon-based, your months will have unequal number of days. If your calendar is season-based, your years will have unequal number of days. Either way, computing day differences will be a nightmare. You cannot have an integer calendar without leap days every few years.

Troubled days

So, the only option is to do neither and just count the number of days. But as discussed before, the year to day ratio is far from a simple fraction; by the time the earth makes a full revolution

around the sun (with respect to the other stars), a non-integer number of days have passed. But this seems fine, then. If you accept to deal with decimals, you can just take your base unit as a solar day, noon to noon, easy enough to measure and compare.

Atomic clocks  
confidence

Except, the solar day is not exactly constant. The time from terrestrial noon to noon is not a constant number of seconds (McCarthy & Babcock 1986). Now what, measure everything in seconds? will the same argument hold here, and the second will turn out to be variable? It turns out we have a very constant, precise definition for a second. According to the 2019 redefinition of the International System of Units (SI), one second is exactly (BIPM 2019):

$$1 \text{ s} = \frac{9\,192\,631\,770}{\Delta\nu_{\text{Cs}}} \quad (2.1)$$

Where  $\Delta\nu_{\text{Cs}}$  is the unperturbed ground-state hyperfine transition frequency of  $^{133}\text{Cs}$ , which is a fundamental physical constant. Then, supported by hundreds of years of experiments in the natural sciences, we measure everything in seconds.

The Unix Time  
and the Julian  
Date

But for astronomical events, seconds are a tiny unit. Measuring temporal scales of days and years in seconds is cumbersome and unpopular. Besides, historical astronomical records—even Before the Common Era—were and are still in use (Morrison & Stephenson 2004), and astronomers and historians liked to have a unit of time that did not extend to the negative numbers. Therefore astronomical dates were historically to be measured in days since a date early enough to capture historical records. The original idea for this zero point came from Joseph Scaliger in 1583 (Carroll & Ostlie 2017). He considered three time periods used in his day:

- The solar cycle: in the Julian calendar, the weekday of a given date on the year would repeat every 28 years<sup>1</sup>. The solar number of a (Julian) calendar year  $y$  in this cycle is given by

$$\text{SolarNumber}(y) = \text{mod}(y + 8, 28) + 1 \quad (2.2)$$

- The lunar cycle: 19 years, the number of years in which a moon phase will occur in the same day of the year<sup>2</sup>. The so called “Golden Number” of a Julian year  $y$  is given by

$$\text{GoldenNumber}(y) = \text{mod}(y, 19) + 1 \quad (2.3)$$

---

<sup>1</sup>With exactly  $365.25 = 1461/4$  days in a year and 7 days a week, the week days cycle every  $\text{LCM}(1461, 7) = 10227$  days, = 28 years.

<sup>2</sup>A mean synodic month (time between two full moons) is 29.53059 days. 235 synodic months was sufficiently close to 19 years for the Julian calendar.

- the Indiction: a 15 year period for tax census used by the Roman Empire and later by the Holy Roman Empire. Its calculation for a Julian year is:

$$\text{Indiction}(y) = \text{mod}(y + 2, 15) + 1 \quad (2.4)$$

The phases on the modulus operations are selected such that historical records make sense. The solution, then to the integer equation

$$\text{SolarNumber}(y) = \text{GoldenNumber}(y) = \text{Indiction}(y) = 1 \quad (2.5)$$

is, using the Chinese remainder theorem,

$$y = 3268 + 7980n \quad (2.6)$$

where  $n$  is any whole number. Scaliger selected  $n = -1$  as the zero point for the Julian Day system, which would be Julian year  $-4712$  CE, and as there was no year 0, that would be Julian 4713 BCE, or in the Gregorian modern calendar, noon of November 24, 4714 BCE (Epoch).

Final definition  
and the Römer  
Delay

Those are the ingredients needed to define the modern Julian Day system. If an Ephemeris Day is defined as  $24 \times 60 \times 60 = 86400$  SI seconds, the Julian Day (JD) is the fractional number of Ephemeris days that have passed since noon of November 24, 4714 BCE on the current calendar. There are at least 4 principal corrections to the Julian Day due to astronomical, geological or relativistic effects (Eastman et al. 2010), but here we will only consider one of them, as it's the only strictly needed to understand the data used on this work. As earth moves around the sun, the light from a distant object may reach its surface with different time delays, consequence of the finite speed of light. In order to be able to accurately compare time series taken around all positions of earth's orbit, the signal is corrected with a delay as if the measurement was taken on the position of the sun. This is called **Heliocentric Julian Day** (HJD), and that correction is given by (see Figure 2.1):

$$\Delta t = \frac{\vec{r} \cdot \hat{n}}{c} = -\frac{r}{c} (\cos \delta \cos \delta_{\odot} \cos(\alpha - \alpha_{\odot}) + \sin \delta \sin \delta_{\odot}) \quad (2.7)$$

where  $\vec{r}$  is the vector sun-to-earth,  $\hat{n}$  is the direction of the measurement on earth's sky, and  $(\alpha, \delta)$ ,  $(\alpha_{\odot}, \delta_{\odot})$  are the equatorial coordinates of the observed object and the sun respectively, as seen from earth, *i.e.*  $\vec{r} = (r, \phi = \frac{\pi}{2} - \delta_{\odot}, \theta = \alpha_{\odot})$ ,  $\hat{n} = (1, \phi = \frac{\pi}{2} - \delta, \alpha)$  The maximum possible value for this correction is  $\sim 1 \text{ AU}/c = 16.63 \text{ min} = 0.00577$  Ephemeris days.

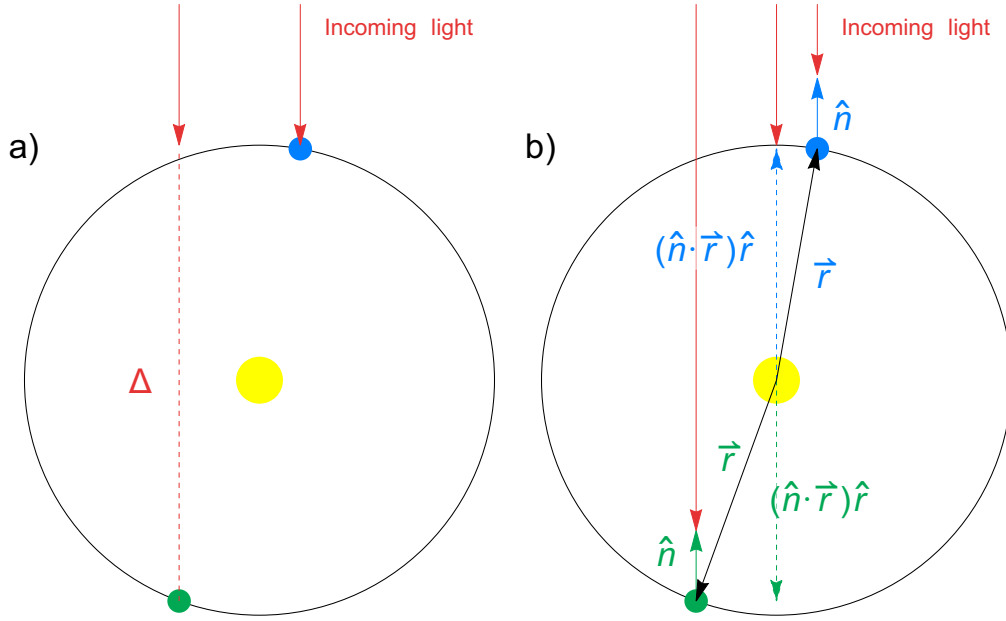


Figure 2.1: Conceptual view of the Rømer delay for a signal coming from above (in red). The comparison is made between the earth at two points of its orbit (blue and green disks) around the sun (yellow disk). **a)** The distance  $\Delta$  that the incoming light has to travel if the earth is in the opposite side of its orbit (green earth) will mean a time delay of  $\Delta/c$  relative to the nearest part of the orbit (blue earth). The limit value for  $\Delta$  is  $\sim 2$  AU, so the time delay would be around 16 minutes. **b)** This time the excess distance is calculated from the sun, and its value is the projection of the sun-to-earth vector to the vector normal to the incoming signal.

### Remark about floating point representations

Unit in last  
place of  
floating point  
data

Typically, decimal numbers are stored on a computer as a double-precision floating point number, which is basically “fixed point” scientific notation in base 2, with 52 binary digits for the mantissa and 11 for the exponent (IEEE 2019). Floating point numbers are of course imprecise, and in fact more as the number they represent grows. To quantify this loss of precision exists the notion of the machine epsilon: the difference from a given number to the next representable number on the system. This epsilon (also called “unit in last place”) is greater for greater numbers, and there is the question of whether we will lose or not our anhelated temporal scale precision.

typical  
precision of the  
numbers used

As a random example from the OGLE IV database (see [subsection 3.1.2](#)), let's check the precision at  $\text{HJD} = 2455262.5065$ . In [Listing 1](#) there is a simple function to estimate the precision. If we use it with our number we'll get  $\sim 4.6 \times 10^{-10}$  days, which is  $40 \mu\text{s}$ . If instead of the HJD we choose to store the number of seconds since the Julian Epoch ( $\sim 2.12 \times 10^{11}$  s) the precision would be  $\sim 30 \mu\text{s}$ . But this HJD in particular was stored as a *reduced* HJD. In this case, the reported number was  $\text{RHJD} = \text{HJD} - 2.45 \times 10^6 = 5262.5065$ . The machine precision at 5262.5065 is  $\sim 9.09 \times 10^{-13}$  days, or 78.5 ns, so we conclude that floating point precision is not a problem for the storage of this reduced time unit in this case.

### 2.1.2 Magnitudes

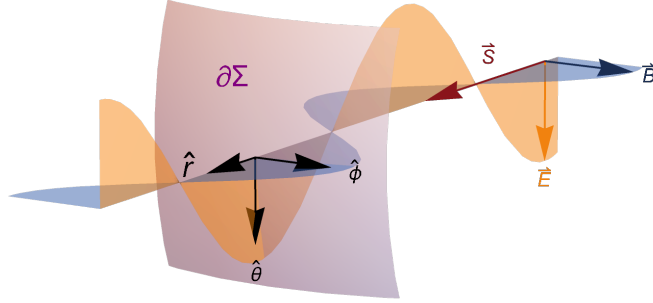


Figure 2.2: An electromagnetic wave ( $\vec{E}, \vec{B}$ ) propagating outwards ( $\vec{S}$ ) through a spherical surface  $\partial\Sigma$  in a vacuum space. The correspondence of the directions of  $\vec{S}, \vec{B}$  and  $\vec{E}$  with the spherical directions  $\hat{r}, \hat{\phi}, \hat{\theta}$  is to be noted.

#### Energy conservation and the inverse square law

The magnitude system, vertical axis of a light curve, can be understood from the principles of electromagnetic waves. We would first want to ascertain the distance dependence of the energy flux for an isotropic light source. Here, isotropic means that the magnitude of the electric and magnetic fields of the wave would not depend on any angle. In spherical coordinates, for an outgoing wave of frequency  $\omega$  and wave number  $k$ , we have (see [Figure 2.2](#))

$$\vec{E} = E(r) \cos(kr - \omega t) \hat{\theta} \quad \vec{B} = B(r) \cos(kr - \omega t) \hat{\phi} \quad (2.8)$$

where  $B(r) = E(r)/c$ . Using Poynting's theorem (Griffiths 2013, §8.1.2) we can calculate the total energy flux<sup>3</sup> through an entire sphere centered at the source. First, we calculate the Poynting vector

$$\vec{S} = \frac{\vec{E} \times \vec{B}}{\mu_0} = \frac{\cos^2(kr - \omega t)E(r)^2}{c\mu_0} \hat{r} \quad (2.9)$$

and its temporal mean

$$\langle \vec{S} \rangle = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \frac{\cos^2(kr - \omega t)E(r)^2}{c\mu_0} \hat{r} dt = \frac{E(r)^2}{2c\mu_0} \hat{r} \quad (2.10)$$

Its flux is then the energy density flux through the whole spherical surface  $\Sigma$  of radius  $r$ , also called Luminosity

$$L = \oint_{\partial\Sigma} \langle \vec{S} \rangle \cdot d\vec{A} = \oint_{\partial\Sigma} \langle \vec{S} \rangle \cdot \hat{r} r^2 d\Omega = \oint_{\partial\Sigma} \frac{E(r)^2 r^2}{2c\mu_0} d\Omega = \frac{2\pi r^2 E(r)^2}{c\mu_0} \quad (2.11)$$

But the luminosity should *not* be a function of the radius. The electromagnetic energy that passes through a  $r_0$ -radius sphere should be the same that later passes through a  $2r_0$ -radius sphere. It's the same electromagnetic wave after all; there are no charges or currents. Therefore, the magnitude of the electric field  $E(r)$  should be inversely proportional to  $r$ , and that way  $L$  is constant and energy is conserved.

Knowing  $E(r)$  we can then do the same process the other way around: calculate the energy flux through a properly aligned detector of known area  $A$ , simply known as Flux

$$F = \int_S \langle \vec{S} \rangle \cdot \hat{r} dA = \frac{E(r)^2 A}{2c\mu_0} \quad (2.12)$$

where the integration occurs along the surface  $S$  of the detector, which typically would be flat surface, but considered here as a section of a sphere with a radius so immense that the difference is meaningless. As  $E(r) \propto 1/r$ ,  $F \propto 1/r^2$ , or equivalently,

$$F(r) = \frac{LA}{4\pi r^2} \quad (2.13)$$

## Logarithmic scales

Working with the same example as before, if we put our detector at a distance  $r = r_0$  of a source of light  $a$ , and measure a flux  $F_a(r_0)$ , the flux at a distance  $2r_0$  will be  $F_a(r_0)/4$ , according to Equation 2.13. In general, if we change the distance by a factor  $d$ , the change in flux would be  $1/d^2$ .

---

<sup>3</sup>All the "energy fluxes" referred in this section are per unit of time.

But what if we wanted a brightness scale where *multiplicative* changes on the distance became *additive* changes of brightness? If we are confident in our measurement of  $F_a(r_0)$ , we could measure the fluxes relative to  $F_a(r_0)$ , and define a new logarithmic unit for the brightness of any object  $b$  as

$$\mathcal{F}_b(r) = \alpha \ln \left( \frac{F_b(r)}{F_a(r_0)} \right) \quad (2.14)$$

where  $\alpha$  is a free parameter of our unit system. That way, the “brightness” difference between a measure of object  $b$  at distances  $r_1$  and  $r_2$  would be

$$\Delta \mathcal{F}_b = \mathcal{F}_b(r_2) - \mathcal{F}_b(r_1) = \alpha \ln \left( \frac{F_b(r_2)}{F_b(r_1)} \right) = -2\alpha \ln \left( \frac{r_2}{r_1} \right) \quad (2.15)$$

where we used Equation 2.13 as before. One immediate consequence of this definition of brightness is that the brightness of the reference object  $a$  at distance  $r_0$  would be exactly 0.

For our  $\mathcal{F}$  to become the standard magnitude system, we have to choose  $\alpha$ , and  $F_a(r_0)$

### Historical choice for $\alpha$

In order to formalize the magnitude scale, and still maintain some sense with the historical records, Pogson (1856) proposed that an increment of 5 magnitudes would correspond to a flux 100 times *smaller*. That means that a magnitude 5 object would have a flux  $F_b = F_a(r_0)/100$ . Replacing that in Equation 2.14 and solving for  $\alpha$  results in

$$5 = \alpha \ln \frac{1}{100} \quad \Rightarrow \quad \alpha = \alpha = -\frac{5}{\ln 100} = -\frac{5}{\log_{10}(100) \ln(10)} = -\frac{2.5}{\ln 10}$$

which allows us to rewrite Equation 2.14 as

$$\mathcal{M}_b(r) = -2.5 \log_{10} \left( \frac{F_b(r)}{F_a(r_0)} \right) \quad (2.16)$$

and Equation 2.15 as

$$\Delta \mathcal{M}_b = \mathcal{M}_b(r_2) - \mathcal{M}_b(r_1) = -2.5 \log_{10} \left( \frac{F_b(r_2)}{F_b(r_1)} \right) = 5 \log_{10} \left( \frac{r_2}{r_1} \right) \quad (2.17)$$

which implies that a positive difference in magnitude of 1 corresponds to a flux ratio  $f$  of

$$1 = -2.5 \log_{10}(f) \quad \Rightarrow \quad f = 10^{-1/2.5} = 0.3981$$

and correspondingly, a negative difference in one magnitude corresponds to  $10^{1/2.5} = 2.51188$  times the flux.

As is standard in astronomy, the 10 in  $\log_{10}$  would be obviated from now on.

### The reference flux, absolute and apparent magnitudes

The standard flux for the magnitude is defined by the International Astronomical Union as  $F_a(r_0) = F_0 = 2.518021002 \times 10^{-8} \text{ W/m}^2$  at  $r_0 = 10 \text{ pc}$ <sup>4</sup>, which corresponds to a luminosity  $L_0 = 3.0128 \times 10^{28} \text{ W}$  (Mamajek et al. 2015, IAU, B2).

With the reference flux, we can measure the flux of any light source and calculate its **apparent magnitude**  $m$ , which for an object  $b$  is at a distance  $r_b$  is defined using Equation 2.16 as

$$m = -2.5 \log \left( \frac{F_b(r_b)}{F_0} \right) \quad (2.18)$$

By the convention  $r_0 = 10 \text{ pc}$  the **absolute magnitude** is defined as

$$M = -2.5 \log \left( \frac{F_b(10 \text{ pc})}{F_0} \right) \quad (2.19)$$

If you know the distance  $r_b$  and the apparent magnitude, you can calculate the absolute magnitude using Equation 2.17:

$$m - M = 5 \log \left( \frac{r_b}{10 \text{ pc}} \right) \quad (2.20)$$

which is known as the distance modulus.

Before the standard flux was given in base SI units, it was taken as that of Vega ( $\alpha \text{ Lyr}\alpha$ ).

### Wavelength dependence on the magnitude

Until now there have been no discussion about the wavelength of the light on the energy flux. But of course the energy of an electromagnetic wave depends on its wavelength. For a black body, this is governed by the Planck's Law (Planck 1901), stating that the spectral radiance

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \left( e^{\frac{hc}{\lambda k_B T}} - 1 \right)^{-1} \quad (2.21)$$

This spectral radiance is conceptually equivalent to the term  $\langle \vec{S} \rangle \cdot \hat{r}$  used in Equation 2.10<sup>5</sup>. With this approximation and with Equation 2.11 we can define the monochromatic Flux and Luminosity (Carroll & Ostlie 2017, Chapter 3, Section 5):

$$F_\lambda d\lambda = \frac{L_\lambda A}{4\pi r^2} d\lambda = B_\lambda \left( \frac{R}{r} \right)^2 d\lambda \quad (2.22)$$

---

<sup>4</sup>A parsec is defined as  $\text{pc} = 648000/\pi \text{ AU}$

<sup>5</sup>The spectral radiance  $B_\lambda$  is defined as the energy flux per hertz per second, and is in fact connected with the statistics of the magnitude of the electric field  $E(r)$  considered before, because the different occupation levels of some energies that arise from the thermodynamical considerations. Again, refer to the original work of Planck (1901) for details.



where  $R$  is the radius of the black body (the emitting sphere) and  $r$  is the distance where the flux is measured. One can integrate that to obtain the Stephan-Boltzmann law:

$$L = 4\pi R^2 \sigma T^4 \quad (2.23)$$

With this equation we can deduce a way to calculate  $F_{0\lambda}$ , the standard flux at any given wavelength; we have the standard luminosity:  $L_0 = 3.0128 \times 10^{28}$  W, and the Stefan-Boltzmann constant is  $\sigma = 5.6703744 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4}$ . Additionally, as the standard had to be numerically based on the previous standard, the emitting radius must be the radius of Vega (Yoon et al. 2010),  $R \approx 1.789 \times 10^9 \text{m}$ . From that we deduce an effective temperature  $T_0 = 10720$  K

### The Johnson-Cousins photometric system

Albeit one *could* study starlight in a monochromatic fashion, it would be instrumentally and phenomenologically inconvenient to work with individual wavelengths. Rather, astronomers use filters to focus their attention on a portion of the spectrum, usually based on the spectral properties of the target object or star.

Each filter is an optical device that let the light pass through more or less depending of its wavelength. This function is called the transmittance of the filter.

The most commonly used photometric system of filters is called the Johnson-Cousins UBVRI system. Transmittance curves for these five filters, one for each letter, are given in Figure 2.3, along with the filters used in the OGLE-IV project. But in place of a full function, it is common to give an effective wavelength and bandwidth for the filter. This is presented on Table 2.1.

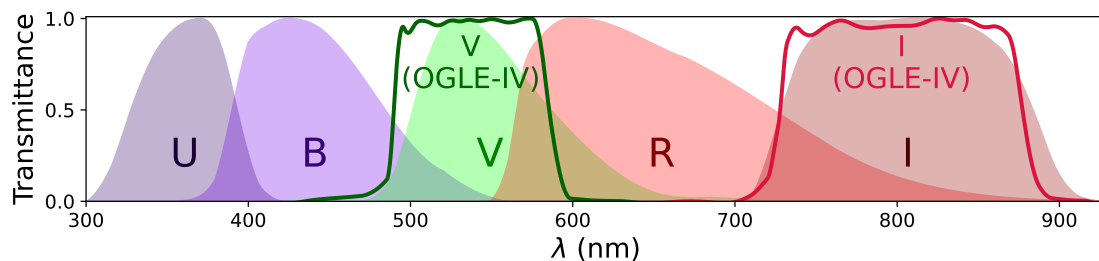


Figure 2.3: Johnson-Cousins (filled) and OGLE-IV equivalent (solid) filters transmittance curves. Colors are meant to be representative, not exact. UBVRI curves were adapted from Bessell (2005), and OGLE-IV curves from Udalski et al. (2015). It is to note that OGLE-IV curves are experimental measurements of custom made filters.

Filter	description	$\lambda_{eff}$ (nm)	$\Delta_\lambda$ (nm)
U	Ultraviolet	366.3	65
B	Blue	436.1	89
V	Visual	544.8	84
R	Red	640.7	158
I	Infrared	798	154

Table 2.1: Effective wavelengths and bandwidths for the Johnson-Cousins photometric system. Data from [Bessell \(2005\)](#).

Absolute magnitudes for an object in these standard filters are presented as a sub-index:  $M_B$  would be the blue absolute magnitude. Relative magnitudes are presented as the filter letters:  $B$  would be the blue apparent magnitude.

If the transmittance function of a filter “X” is denoted  $T_X(\lambda)$ , the apparent magnitude  $X$  is physically obtained by the process

$$X = \int_0^\infty T_X(\lambda) R(\lambda) F(\lambda) d\lambda \quad (2.24)$$

where  $F(\lambda)$  would be the monochromatic flux (the light spectrum) of the star, and  $R(\lambda)$  would be the optical response of the instrumental system used to take the measure, without the filter <sup>6</sup>. This process is rarely dealt with in a theoretical manner; rather it serves here as a conceptual tool to understand what is going on where the measure is taken.

### Color, extinction, and the Wesenheit index

The dimming of light intensity by effects of the distance has been discussed, and reached conclusion on [Equation 2.20](#). But that equation is only valid if the light from the distant object is completely unperturbed. Even ignoring atmospheric or instrumental effects on that light, the interstellar medium will have a considerable impact on the amount of light that reaches the detector, as the dust and gas absorbs and scatters the starlight.

This absorption would have a multiplicative effect of the flux, making it dimmer, and therefore an *additive* effect on the magnitude, because magnitudes are an inverse scale ([Karttunen et al. 2017](#)).

---

<sup>6</sup>this would include things like the optical response of the telescope and mirrors, and in the case of CCD detectors, the quantum efficiency, and a term  $\frac{\lambda}{hc}$  for the photon counting process ([Bessell 2005](#)).

Continuing with the filter X example, if we denote extinction-free quantities by the subscript 0, and this additive extinction on the filter X by  $A_X$ , we have the full extent of [Equation 2.20](#):

$$X = X_0 + A_X = M_X + 5 \log \left( \frac{r}{10 \text{ pc}} \right) + A_X \quad (2.25)$$

Astronomers define color indexes between pairs of filters as the difference on apparent magnitudes. Using [Equation 2.25](#)

$$X - Y = (M_X - M_Y) + (A_X - A_Y) = (X_0 - Y_0) + E(X - Y) \quad (2.26)$$

where  $X_0 - Y_0$  is called the extinction-free color index, sometimes denoted  $(X - Y)_0$ , and  $E(X - Y) = A_X - A_Y$  is called the color excess.

The ratio of total-to-selective extinction is defined for the filter  $V$  as  $R_V = A_V/E(B - V)$ . Experimentally it is known that its value can vary from 2.6 to 5.5 ([Clayton & Cardelli 1988](#)), but it is widely taken as its mean value of 3.1 in the interstellar medium, which is valid for the Magellanic Clouds ([Cardelli et al. 1989](#); [Górski et al. 2020](#)).

If for a color index  $X - Y$  and for a certain star,  $R_{X-Y} = A_Y/E(X - Y)$  is known, the so called extinction-free Wesenheit<sup>7</sup> index ([Madore 1982](#)) is defined as

$$W_{X-Y} = Y - R_{X-Y}(X - Y) \quad (2.27)$$

which upon expanding turns into

$$\begin{aligned} W_{X-Y} &= Y_0 + A_Y - \frac{A_Y}{A_X - A_Y} [(M_X - M_Y) + (A_X - A_Y)] \\ &= Y_0 - \frac{A_Y}{A_X - A_Y} (M_X - M_Y) \\ &= Y_0 + R_{X-Y}(X - Y)_0 \end{aligned}$$

and thus is indeed free of extinction effects. Through this work, we will use the  $V - I$  color index, and the Wesenheit index:

$$W_I = I - R_I(V - I)$$

Where  $R_I = \frac{A_I}{E(V-I)} = \frac{A_I}{A_V - A_I}$ . Similarly to the  $R_V$  case,  $R_I$  has a wide range of values on the literature, even with  $R_V = 3.1$  fixed (for a discussion see [Nataf 2015](#)). It was measured in the Galactic bulge as  $1.080 \pm 0.007$  by [Pietrukowicz et al. \(2012\)](#). Using the tables from [Schlegel et al. \(1998\)](#) and [Schlafly & Finkbeiner \(2011\)](#) yields 1.411 and 1.217 respectively, for the Landolt filters.

---

<sup>7</sup>German word for “essentiality, essential being”, philosophical term for the fundamental identity of things ([Waibl & Herdina 1997](#), page 341).

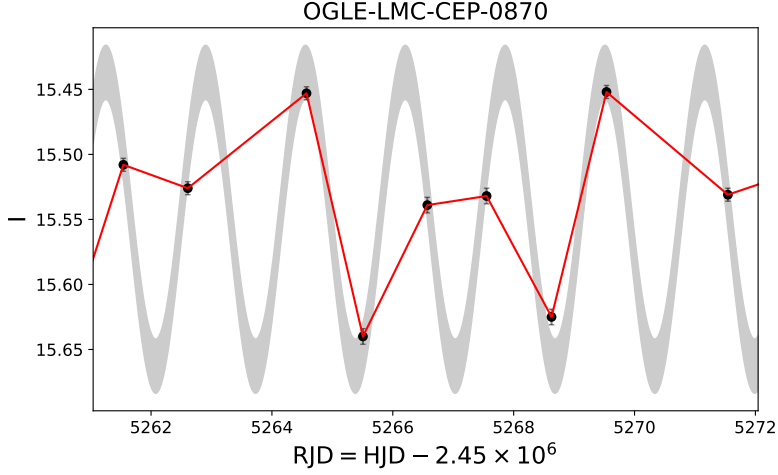


Figure 2.4: First points of the I band time series for a randomly selected star with low period from the OGLE database (Soszyński et al. 2015). Data points are in black, and the actual light curve (Fourier series fit to all data) is shown as a gray band. Simple interpolation is shown in red, showing that there is not enough data for a point-by-point interpolation to represent the underlying function.

The first direct OGLE measurement of  $A_I$  and  $E(V - I)$  on the Magellanic system has a star-number weighted mean of  $1.57 \pm 0.33$ . Through the OGLE literature a value of 1.55 has been widely used (see for instance Udalski et al. 1999, 2015; Ulaczyk et al. 2013), and thus we will use that value, but from the latest reddening maps, Skowron et al. (2021) derived  $R_{I;SMC} = 1.74$  and  $R_{I;LMC} = 1.67$ . Actually, this  $R_I$  should be calculated on a star-by-star basis, using those reddening maps, but that lies outside the scope of this work.

### 2.1.3 Phase diagrams

With time and magnitude scales covered, the last definition needed is that of the phase. Astronomical observations taken from the earth surface, ignoring those of the sun, can obviously only taken at night. Atmospheric related distortions are less prominent near zenith, so measurements are ideally taken when the star is in its highest point in the sky, but telescope times are heavily scheduled, so the measure will be taken whenever possible. Additionally, because of the rotation of the earth around the sun, there will be months of the year when the object will not be visible during night.

These irregularities in the measuring times, apart from having consequences on the signal analysis

process of the data, can obfuscate the phenomenology of the star’s light curve, specially if its variability period is on the order of days. Continuous measurements  $I_{t=0}, I_{t=1}$  (approximately one day apart) are not necessarily in the same period, and thus there is no way to grant that the value at an intermediate time (say  $I_{t=1/2}$ ) is even inside the  $[I_0, I_1]$  range; any interpolation would eventually fail, as it is seen in [Figure 2.4](#).

But as already stated, the shape of the light curve carries phenomenologic details about the star variation. In order to recuperate that shape without having to fit the data, and if the period is known, one can define the *phase*  $\phi$  of a certain time  $t$  as

$$\phi(t) = \frac{\text{mod}(t - t_0, P)}{P} = \text{mod}_1 \left( \frac{t - t_0}{P} \right) \quad (2.28)$$

where  $t_0$  is called the ephemeris time, and is commonly taken as the point with maximum brightness (minimum magnitude). The first definition is just the time that has passed since the last maximum divided by the period, and the second is the fractional part of the time since *any maximum*<sup>8</sup> in units of  $P$ .

A phase diagram is then a plot of the phase against the magnitude. Examples can be found on [Figure 2.5](#) and [Figure 2.6](#).

The condition that the period has to be known to properly make a phase diagram can seem ridiculous, since the purpose of this work is to find periods. But this ends up playing in our favor: it is precisely the fact that phase diagrams are only correct in the period that will allows us to find the period using phase diagrams. The discussion about how the phase diagram behaves on a wrong period and how that can be used to find the correct one is presented on [subsection 2.3.2](#).

## 2.2 Cepheid Variable stars

Classical Cepheid are radial pulsating variable stars, named after the prototypical  $\delta$  Cephei as discussed in [section 1.1](#). But that can be a bit misleading, as the term actually refers to an evolutionary stage that some stars present even multiple times in their lifespan. Some stars (depending on their mass) leave the main sequence as their Hydrogen fuel gets consumed, and evolve into what is known as the instability strip (see [Figure 2.7](#)). Cepheids are yellow giants or supergiants at this point ([Catelan & Smith 2015](#)), and began to experience radial pulsation, typically with a period of 1 to 50 days, changing their brightness with amplitudes ranging from 0.1 to 2.5 magnitudes, and going from spectral type F to G or even K ([Karttunen et al. 2017](#)). Their masses can vary from  $\sim 1M_{\odot}$  to  $20M_{\odot}$ , and their radius from  $\sim 10R_{\odot}$  to  $200R_{\odot}$

---

<sup>8</sup>By properties of the mod function, the phase remains the same if  $t_0$  is shifted by any whole multiple of  $P$ .

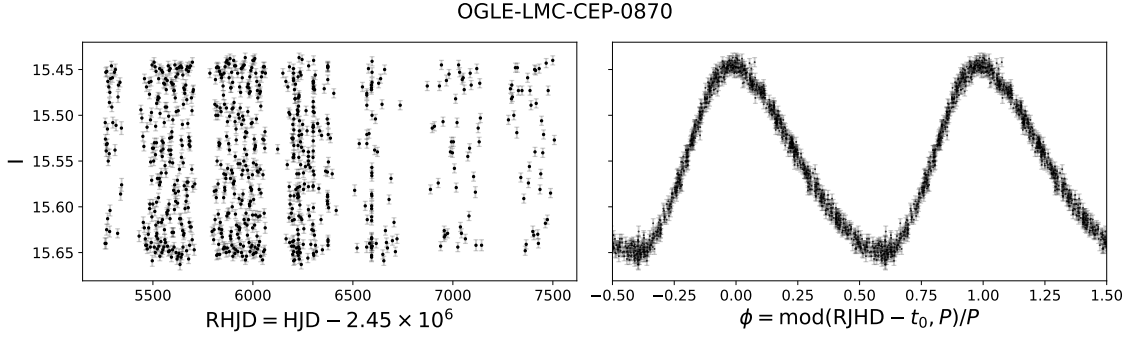


Figure 2.5: I band time series and phase diagram for a randomly selected star with low period from the OGLE database (Soszyński et al. 2015). Uncertainties for the magnitudes are shown as light gray error bars. Temporal units are ephemeris days, and ephemeris time  $t_0$  is taken as the time with minimum magnitude. A period of  $P = 1.6518$  days was used for the phase. The yearly windows when the object cannot be observed from the ground observatory are clearly present in the time series.

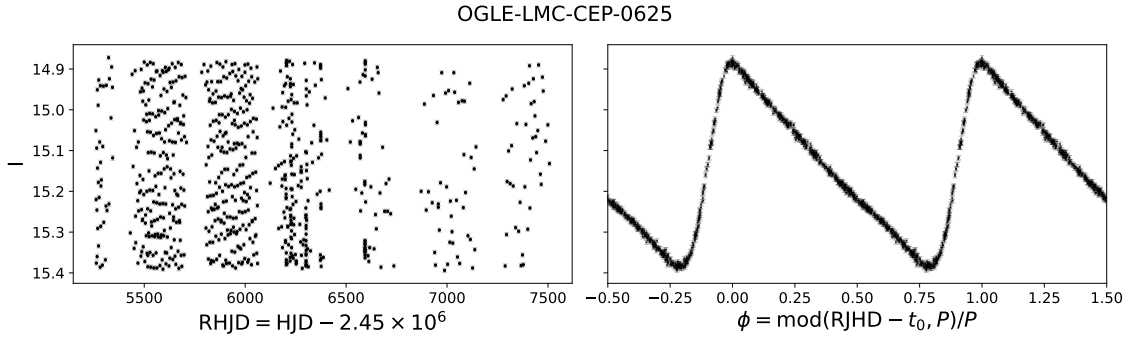


Figure 2.6: Same as in Figure 2.5, but for a typical Cepheid in the LMC. A period of  $P = 3.7503$  days was used for the phase. Phases are padded cyclically for values outside the  $[0, 1]$  range for clarity, to better observe the behavior of all parts of the light curve. Data from Soszyński et al. (2015). A notable difference of this figure and Figure 2.5 is the amount of “dispersion” or width of the light curve. This dispersion is much larger than the experimental uncertainty, and should be intrinsic to the star, a phenomenon which deserves further study.

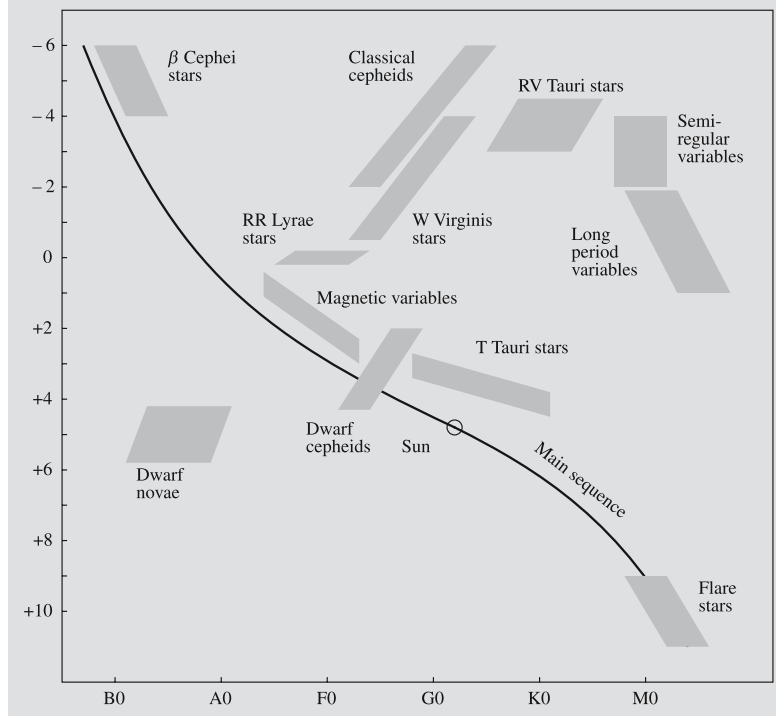


Figure 2.7: Position of several types of variable stars in a Hertzsprung-Russell diagram. The abscissa is given as spectral class, proportional to temperature and color: red (cooler, 3750 K at M0) on the right and blue-white (hotter, 27000 K at B0) to the left. The ordinate is given in absolute visual magnitude  $M_V$ . The main sequence is represented as a solid line, and the Sun as a circle. Classical Cepheids can be seen on the red giant branch. Taken from [Karttunen et al. \(2017\)](#), Figure 14.2.

The former are the characteristics of galactic Cepheids, but it is known that there are extra-galactic Cepheids with shorter and larger periods, specially in the Magellanic system ([Cox 1980](#); [Payne-Gaposchkin & Gaposchkin 1967](#)). The data taken by the OGLE project is in agreement with that ranges; see [subsection 3.1.2](#) for more details.

### 2.2.1 Stellar populations

Stars contents are generally classified percentually as Hydrogen ( $X$ ), Helium ( $Y$ ) and “metals” ( $Z$ ), that constitute really any other element present in the star<sup>9</sup>. As the nuclear fusion in the star core turns Hydrogen into Helium, and Helium into heavier elements up to Iron, the balance of the mass fractions over the stars moves from  $X$  to  $Y$  and eventually to  $Z$  as the time passes. Moreover, massive stars die in cataclysmic events known as novæ, which have the energy to produce metals heavier than Iron. As new stars form from the remnants of the old ones, they have a greater amount of metals from the beginning, but early stars on the Universe would have been mostly Hydrogen, the simplest element.

The idea of the stars coming in generations is attributed to Oort, according to [Baade \(1944\)](#). Baade classified as population I the high metallicity ones (modern value  $Z \approx 2\%$ ), the youngest, and population II the oldest (modern value  $Z \approx 0.1\%$ ) ([Carroll & Ostlie 2017](#)). A population III have been theorized since then, those of the first generation of stars, but none have been observed ([Heger & Woosley 2002](#)). There is a discussion around the timescales needed for the violent death of the stars to occur (specifically type Ia supernovæ), and if it is sufficient to produce the metallicity difference among populations ([Carroll & Ostlie 2017](#)). Regardless, the PL relations are dependent on population, and in this work only Cepheids of population I will be considered.

### 2.2.2 Radial stellar pulsation

#### The period-density relation

The most intuitive way of interpreting the radial pulsations of a star is by considering sound-like waves traveling from its center to its surface. As a crude approximation, we could consider the equilibrium at the last layer of a star of total mass  $M$  and radius  $R$ , say, a spherical shell of width  $dr$  and mass  $m \ll M$  at the limit  $r \rightarrow R$ . Assuming that the pressure outside of the star is zero, and the pressure just below our outer shell is  $P$ , there would be an outward force  $P A = P 2\pi r^2$  opposed by the gravitational pull  $GMm/r^2$ . This naive equation of motion would then be

$$m\ddot{r} = P4\pi r^2 - \frac{GMm}{r^2} \quad (2.29)$$

---

<sup>9</sup>For reference, the current surface values for the sun are  $X = 73.81\%$ ,  $Y = 24.85\%$ ,  $Z = 1.34\%$  ([Asplund et al. 2009](#)).



After a painful process of linealization, we will get an equivalent to the Linear Adiabatic Wave Equation (LAW):

$$-\frac{1}{r^4 \rho} \frac{d}{dr} \left( \Gamma_1 P r^4 \frac{d\eta}{dr} \right) - \frac{1}{\rho r} \left( \frac{d}{dr} [(3\Gamma_1 - 4) P] \right) \eta = \omega^2 \eta \quad (2.30)$$

where perturbations on  $r$  of the form  $\xi = \eta e^{i\omega t}$  have been considered,  $\rho$  is the mean density and  $\Gamma_1$  is the first adiabatic exponent<sup>10</sup>. This is equation 5.90 of [Catelan & Smith \(2015\)](#), where the whole linealization process can be consulted. [Equation 2.30](#) is a more general form of that first derived by [Eddington \(1918\)](#). As already stated, the perturbations on the radius are assumed as outward waves with amplitude  $\eta$  and angular frequency  $\omega$ . In our simplistic model, the amplitude does not depend on the position of the spherical shell, so  $\frac{d\eta}{dr} = 0$ , and the first term vanishes. We shall take away all the radial dependencies, except of course for that of the pressure, as we are considering sound waves:

$$\left( \frac{3\Gamma_1 - 4}{R\rho} \right) \left( -\frac{dP}{dr} \right) = \omega^2 \quad (2.31)$$

That pressure dependence comes from the pressure gradient equation of stellar evolution (in its hydrodynamical form) ([Catelan & Smith 2015](#), equation 4.18):

$$\ddot{r} = -\frac{1}{\rho} \frac{dP}{dr} - \frac{GM}{r^3} \quad (2.32)$$

which we can use in equilibrium ( $\ddot{r} = 0$ ) on [Equation 2.31](#) to get

$$\omega^2 = (3\Gamma_1 - 4) \frac{GM}{R^3}$$

And in terms of the period  $\Pi$  and the mean density  $\rho = M/(\frac{4}{3}\pi R^3)$

$$\Pi = \frac{2\pi}{\sqrt{(3\Gamma_1 - 4) \frac{4}{3}\pi G \rho}} \quad (2.33)$$

This is usually just a rough estimation of the order of magnitude of the period. For instance, the parameters for  $\delta$  Cephei would be  $M = 4.5 \pm 0.3 M_\odot$ ,  $R = 44.5 R_\odot$  ([Matthews et al. 2012](#)), which assuming ideal gas for  $\Gamma_1$  gives  $\Pi = 16.2 \pm 0.5$  days. The real period is 5.366249 days ([Samus' et al. 2017](#)). As seen below these oscillations are not necessarily adiabatic, and the amplitude should actually decay.

---

<sup>10</sup> $\Gamma_1 = 5/3$  for an ideal gas, and it is the value recommended by [Cox \(1980\)](#). Interestingly enough,  $\Gamma_1 = 4/3$  for radiation, and the period actually diverges if one blindly uses [Equation 2.33](#).

## The Eddington valve and the $\epsilon$ , $\kappa$ and $\gamma$ mechanisms

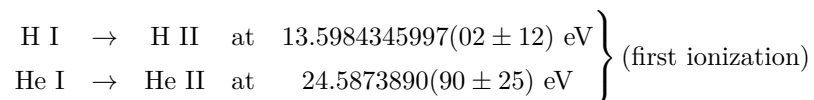
Through the 20th century Eddington (1918, 1926, 1941) proposed a thermodynamic cycle to be responsible for the radial pulsation of stars. He devised two possible physical causes to drive this “valve”, known today as the  $\epsilon$  and  $\kappa$  mechanisms.

The  $\epsilon$  mechanism is based on the idea that the nuclear processes at the core of the star are more intense in the contraction than in the expansion phase of the valve. This would increase the pressure of the stellar material in the contraction phase, and decrease it in the expansion, working in a similar fashion as a Diesel engine (Zhevakin 1963). Although it is true that nuclear reaction rates increase with temperature and pressure, this effect only becomes dominant at higher masses ( $\sim 100M_{\odot}$ ) and is in fact destabilizing (Carroll & Ostlie 2017; Catelan & Smith 2015; Zhevakin 1963). Thus, at least for Cepheids and related variable stars types, the  $\epsilon$  mechanisms is not the primary source of radial pulsation.

The second mechanism proposed by Eddington depends on the star stopping the energy leakage from the nucleus when the star is compressed, and allowing that energy to leak when the star expands. If the energy is *somehow* retained in the core when the star is contracted, the temperature and the pressure will raise, and the star will be forced to expand. If then, at the peak of the expansion, the star *somehow* allows that nuclear energy to escape at a higher rate than normal, the temperature and pressure in the inside will drop, and the gravitational pull would then win over the pressure gradient, and force the system to began compression again, closing the valve’s cycle. This would allow the star to pulse without depending on violent variations on the nuclear reaction rates.

But those “somehow” are quite problematic. Kramers opacity law states that the optical opacity of the stellar material behaves as  $\kappa \propto \rho T^{-3.5}$  (Carroll & Ostlie 2017). As the dependence is stronger with the temperature, the opacity in the compression phase would be lower, not higher. The opacity is directly proportional to the energy (radiation) absorption, and thus the system would not work as Sir Eddington expected.

But there is a way for the temperature changes to be damped at compression and expansion on certain layers of the star. These are called partial ionization zones, and are cause by the increase of degrees of freedom on Hydrogen and Helium gases undergoing ionization (Cox 1963), which effectively rises the heat capacities  $C_V$  and  $C_P$  of the layers (Carroll & Ostlie 2017). There are two such layers, one where Hydrogen and Helium experience their first ionization:



and one deeper layer where the Helium experiences its second ionization:

$$\text{He II} \rightarrow \text{He III} \quad \text{at} \quad 54.4177654(86 \pm 25) \text{ eV} \quad (\text{second ionization})$$

Those energies (Kramida et al. 2021) put all those transitions in the ultraviolet region of the electromagnetic spectrum. The ionizations “leak” a fraction of the energy that otherwise would have been spent on increasing the temperature. As the temperature increase is less prominent, the density term in Kramers law dominates, and therefore the opacity increases.

Conversely, in the expansion phase, the ions interact with free electrons, releasing energy that is absorbed by the gas, increasing the temperature. As the temperature drop for the expansion is less than expected, the density term dominates again, and the opacity decreases. This two processes are known as the  $\kappa$  mechanism, proposed by Zhevakin (1963) and verified by Cox (1963). Moreover, the changes in the heat capacities and the damped temperature gradients causes more heat to flow into those layers, positively reinforcing the cycle in which is known as the  $\gamma$  mechanism.

The position and behavior of this layers is critically dependent of the temperature, and hence the instability strip discussed before must be almost vertical, as Cox (1963) derived theoretically. This can be seen in the position of the Cepheid branch on Figure 2.7.

This effect also —and finally— explains the asymmetrical variability of the brightness on pulsating stars. The amount of luminosity retained by the partial ionization layers depends on their position, and it releases when the layers are propelled outwards and its opacity diminishes (Carroll & Ostlie 2017). The more energy trapped on the compression phase, the greater the force that would make the star expand, and as the opacity has just decreased, the increase of brightness happens faster. On the other side, the gravitational pull and the pressure gradient that closes the cycle and causes compression again are slower, since all of that energy has escaped in the form of light.

The consensus is that all of those mechanisms do require exceptional conditions to actually work, and that could explain why stellar pulsators are indeed rare in the bulk population of stars in the universe, being approximately one for each  $10^5$  stars (Carroll & Ostlie 2017).

## Pulsation modes

Returning to the intuition of the sound waves of the beginning of the section, one could be surprised to know that Equation 2.30 can actually be solve exactly. In fact, it constitutes an special case of what is called a Sturm–Liouville problem. As any other physics equation with similar generality, there is an infinite family of solutions, and we will not go into the details here; it can be consulted on §5.6 of Catelan & Smith (2015), and more generally on §9.3 of Butkov (1968).

But there are important conclusions derived from the solution of Equation 2.30. In the first place, it is always the case that  $\omega^2$  is a real number, but it can be negative, so  $\omega$  is not always real. As the waves are defined with  $e^{i\omega t}$ , imaginary frequencies correspond usually to a cataclysmic behavior of the star. We will be concerned with the real frequency case, where the wave act as in oscillatory motion. This frequency is called the eigenvalue of the system, say  $\omega_n$ , and it is associated with an eigenfunction  $\eta_n$ .

Those eigenfunctions act as basis functions for all the waveforms than could be present in the pulsation of the star, in a similar way as the sinusoidal functions act as a basis (the Fourier basis) for the sounds produced in a pipe or a flute. Moreover, in the same way a pipe can present harmonics with higher frequencies than its lowest standing wave, each of the  $\omega_n$ ,  $n > 0$ , is a harmonic standing wave pulsating radially onto the star.

Therefore it can (and will) be the case that a star is pulsating primarily in, say, the first harmonic ( $\omega_1$ ), or a combination of the fundamental ( $\omega_0$ ) and other harmonics. This is denoted as **n0**, where **n** is the harmonic (or Overtone, **0**) number. The “zeroth” overtone is denoted **F**, for Fundamental. For instance, a star pulsating in its fundamental and second overtone would be denoted **F20**. A first and second and third overtone pulsator would be **102030**.

The direct solutions of Equation 2.30 would be of little to no use for this work, as the observational data only includes the magnitude variation, and it is difficult to translate that into radial motion. Nevertheless, in the following section will be described how to analyze the magnitude variability using the Fourier theory, and some additional methods.

## 2.3 The other spectrum, the pulsation spectrum

If we wanted to extract information from a periodic signal, the most natural thing to do is to check how the signal “correlates” with certain frequencies. The pure, ideal, harmonic oscillator would correlate only with a single frequency, but more complex phenomena could correlate with several frequencies with different intensities, as if one were to describe several coupled oscillators in terms of their normal modes.

This “correlation” is not precisely defined at this point, but left as an intuitive concept. This is done purposefully, as we will present several ways to measure it. Any function that express the intensity of the correlation of a given signal with a frequency as independent variable would be called a power spectrum. The terms spectrogram, Fourierogram and periodogram are also used through the literature.

### 2.3.1 Fourier Analysis

When we say frequency, we mean frequency  $\nu$  of uniform rotation in some space, as described by a simple waveform  $e^{2i\pi\nu t}$ . The decomposition of a function into a linear combination of regular rotations must be based in the theory of Fourier analysis, if interrogated well enough.

Then, we will define the basics of the theory. If we have a theoretical signal  $f(t)$ , that is mathematically well-behaved, its representation in frequency space would be given by the function

$$F(\nu) = \int_{-\infty}^{\infty} f(t) e^{2i\pi\nu t} dt \quad (2.34)$$

and its called the Fourier transform. The inverse transformation is given by

$$f(t) = \int_{-\infty}^{\infty} F(\nu) e^{-2i\pi\nu t} d\nu \quad (2.35)$$

where we follow the conventions from [Deeming \(1975\)](#). In general,  $F(\nu)$  would be a complex number. The canonical power spectrum is defined as its magnitude:

$$P(\nu) = F^*(\nu) F(\nu) \quad (2.36)$$

#### Discrete non-uniform Fourier transform

One could be purely practical and force the discretization of [Equation 2.34](#) to a signal with data  $\{t_k, f_k\}$ , defining  $f(t) = \sum_k f_k \delta(t - t_k)$ , but I propose to deduce it by a visual construction, building it up from the phase diagram.

In [Equation 2.28](#), the phase was defined as a dimensionless number; the fraction of the period elapsed since the last maximum. As an amount of time relative to the period, it goes from 0 to 1. We can translate that to a rotational motion interpretation, by making that temporal phase  $\phi$  into an angular one  $\psi = 2\pi\phi$ , as if that phase were not seen as the fraction of a period, but as the fraction of a rotation. This is made to eliminate the need for a modulus operation. For each data point, our definitions would be

$$\phi_k = \frac{t_k - t_0}{P} = \frac{\omega}{2\pi}(t_k - t_0) \quad \psi_k = \omega(t_k - t_0)$$

That way, the angular phase  $\psi_k$  corresponds to the argument of a typical sinusoidal function  $\sin(\omega(t_k - t_0))$ . The phase diagram was a (temporal) phase against magnitude plot, but we can construct a polar plot with these  $\phi_k$  as angles and the magnitudes ( $f_k$ ) as radii. Such comparison of linear-to-angular representations can be seen in [Figure 2.8](#). One could call this polar plot a “Fourier curl”, or also a “complex phase diagram”; as it is composed of the points  $\{f_k \cos(\omega(t_k -$

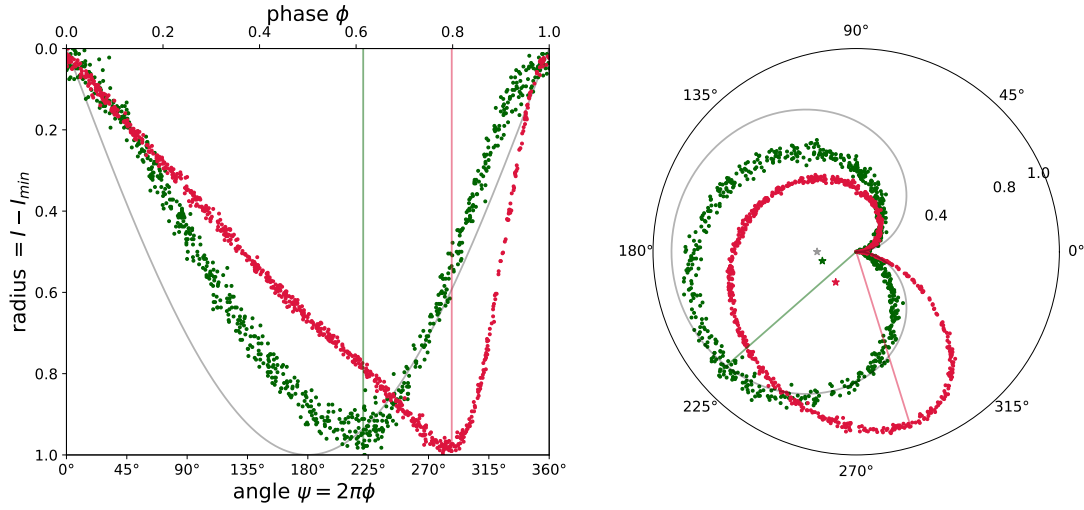


Figure 2.8: Normal and polar phase diagrams for the sample stars in Figure 2.5 and 2.6, in green and magenta respectively. A simple sine curve is shown in light grey as reference. The magnitude was normalized to the  $[0, 1]$  range for visibility. In the phase diagram, brighter is up. In the polar diagram, brighter is towards the center, and dimmer the larger the radius. A transparent line was drawn from the zero of each axis to the point of minimum brightness, for reference. In the polar plot, a  $\star$  sign represents the centroid of each of the curled data. It is of note how the symmetric sinusoid curls up to a cardioid.

$t_0)), f_k \sin(\omega(t_k - t_0))\}$ , anyone would be tempted to just define it using Euler's formula in the complex plane as  $f_k e^{i\omega(t_k - t_0)}$ .

If the data were to be flat, or just random, one would expect the points in the polar plot to be distributed on an uniform fashion. On the opposite case, when the signal is just a sine wave, the polar figure would be a cardioid. The middle case of just a periodic function with relatively low noise is pictured in [Figure 2.8](#). There is a definite clump of points in one direction. The most natural way to discern between noise and patterns in this case would be to look at the centroid of the curled figure. In the complex plane this is just the mean of the points:

$$C = \frac{1}{N} \sum_{k=1}^N f_k e^{i\omega(t_k - t_0)} \quad (2.37)$$

If we drop the  $1/N$  factor and the ephemeris  $t_0$ , and express it in normal frequency terms  $\nu = \omega/(2\pi)$ , we get the general discrete Fourier transform:

$$F(\nu) = \sum_{k=1}^N f_k e^{2\pi i \nu t_k} \quad (2.38)$$

as defined by [Deeming \(1975\)](#); [Schuster \(1898\)](#); [Thompson \(1971\)](#). The power spectrum in this case would be proportional to the magnitude of the centroid of the Fourier curl, and therefore the independence from  $t_0$  is justified.

Provided that there are enough points, and perhaps more importantly, sufficiently wide in time, if  $\nu$  were not to be the correct frequency for the signal, the polar plot would be misaligned. That is, it would have several “petals” along all angles, like a rose curve, which means that the centroid would be displaced towards the center, its magnitude becoming smaller. The more “correlation” a signal has with a given frequency, the better the alignment of the figure in the polar plot; the points with larger radii will clump to one side, and the magnitude of the centroid will be maximum, as can be observed in [Figure 2.9](#).

A power spectrum calculated this way would have a peak at the principal frequency, and smaller peaks in the harmonics of the signal. In theory, these peaks of the power spectrum are invariant over scale changes, but in practice their widths and heights would depend on the particularities of the data.

As this method involve just a sum over the data, it is expected to be  $O(n)$ ,<sup>11</sup> with an overhead due to the complex exponential (or trigonometric) calculations.

---

<sup>11</sup>All time complexities are given per single frequency iteration. Calculating an spectrum with  $f$  frequencies, the complexity of a single iteration have to be multiplied by  $f$ .

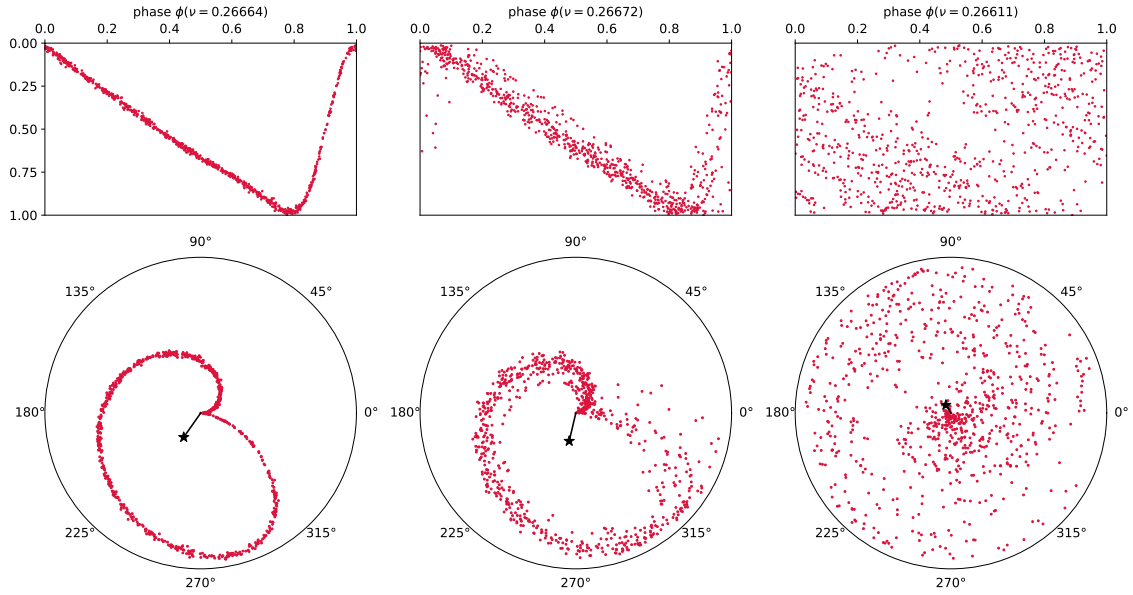


Figure 2.9: Several phase diagrams (Fourier curls) calculated with different sample frequencies for the star in Figure 2.6. Each column has the normal phase diagram and the polar one. A black line and a star ( $\star$ ) symbol represents the centroid of the curled figure. The first column has the correct frequency, and thus the points are clumped together. The second column has a frequency only slightly higher; the figure begins to misalign, but the magnitude of the centroid is mostly stable: this periodogram is not as sensitive as the others. The last column has a “completely wrong” phase, despite having two correct decimals. The data is distributed more or less uniformly along the angles, and therefore the centroid is almost at zero.



But the most important properties of this way of defining the Fourier power spectrum, in contrast with the fast Fourier transform (FFT), is the ability to be evaluated in any frequency  $\nu$  with some freedom on the sample times  $t_k$ <sup>12</sup>. The FFT requires the data to be evenly spaced on time, and it calculates the power spectrum on an evenly spaced frequency grid (Brigham 1974). As we will see in subsection 3.1.2, astronomical terrestrial data is far from being evenly sampled, and Figure 2.4 already stated that interpolation is not an option either.

### The Lomb-Scargle periodogram

Apart from the very theoretical definition of the correlation between a signal and a frequency that we just presented, there is an even simpler one. We could define the correlation of the signal  $\{t_k, f_k\}$  with a frequency  $\nu$  as how well does the signal is fitted by the model

$$f_k + \epsilon_k = A \cos(2\pi\nu t_k + \phi)$$

where  $\epsilon_k$  are independent, normally distributed error terms, and  $A$  and  $\phi$  are the model parameters. In the literature the model is often presented as  $A \sin(\omega t_k) + B \cos(\omega t_k)$ , which is completely equivalent, as both are forms of a first order Fourier series. After the least squares minimization, the coefficient of determination of the fit is given by (Lomb 1976; Scargle 1982)

$$R^2(\omega) \propto \frac{\left(\sum_j f_j \cos \omega(t_j - \tau_\omega)\right)^2}{\sum_j \cos^2 \omega(t_j - \tau_\omega)} + \frac{\left(\sum_j f_j \sin \omega(t_j - \tau_\omega)\right)^2}{\sum_j \sin^2 \omega(t_j - \tau_\omega)} \quad (2.39)$$

with

$$\tau_\omega = \frac{1}{2\omega} \arctan \left( \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \right) \quad (2.40)$$

This  $R^2$  can be thought as a first order approximation of the Fourier spectrum, as is lacking the high order terms, but it is good enough to discern the peaks of the spectrum.

Cepheid light curves, though, are typically not that sinusoidal, as we have established with emphasis. But the maximum of the light curve is periodic, so despite a single sinusoid would not be a good fit for the signal, the least bad fit would indeed have the correct period.

For a more detailed explanation on how this definition of the power spectrum works, I suggest reading VanderPlas (2018). The original derivation from Lomb (1976) is clean and instructive too, and provides approximations for Equation 2.39 and Equation 2.40.

---

<sup>12</sup>There are some limitations in the reliability of the spectrum depending on the sampling frequencies (Marvasti 2001), but nonuniform sampling seems to mitigate the problem, as will be discussed in section 3.1.2.

With this definition it becomes more clear that, provided there is enough data, the model would go progressively out of phase with the signal if the frequency is off. The overall quality of the fit  $R^2$  would then decrease significantly, and again the power spectrum will present peaks at the principal frequency and its harmonics.

Although the complexity on this algorithm seems  $O(n)$ , the redundant trigonometric calculations would present a significant overhead.

## 2.3.2 Phase diagram methods

### Minimum arclength

Leaving the Fourier side for the moment, there is a curious technique devised by [Burke et al. \(1970\)](#) and analyzed more in depth by [Dworetsky \(1983\)](#) that uses the phase directly.

If the period used to construct the diagram is slightly incorrect, the phases would shift more and more. If two points were very near in the good phase diagram, their distance would increase as the period changes. Therefore, one way to directly measure the dispersion of the phase diagram is to join all the adjacent points with a line, and calculate its length. Mathematically, the arclength  $L$  of the diagram would be:

$$L = \sum_{k=0}^{N-1} \sqrt{(f'_k - f'_{k+1})^2 + (\phi'_k - \phi'_{k+1})^2} \quad (2.41)$$

where *the phases have to be sorted*, hence the primes. This is important:  $\phi'_k$  is *not* the phase of  $t_k$ , but the  $k$ -th lowest phase, and  $f'_k$  its corresponding magnitude, not the magnitude at time  $t_k$ . Contrary to the Fourier method,  $L$  will be maximum with the wrong period and minimum with the correct one. A graphic depiction of this arclength is given in [Figure 2.10](#).

The square root could be obviated, as any maximum of  $\sum_k \sqrt{x_k}$  would be a maximum of  $\sum_k x_k$ . The need for sorting the whole phase array each time the test frequency is changed will make this algorithm less efficient compared to the trigonometric operations of the Fourier and Lomb-Scargle periodograms; usually the sorting process would be  $O(n \ln n)$ .

There is an additional peculiarity of this method: the magnitude scale is numerically different than the phase scale, and therefore magnitude differences would contribute differently than phase ones. That is solved by normalizing the magnitude to the interval  $[0,1]$ , as in [Figure 2.8a](#), but typically the highest contributions to  $L$  would be vertical, not horizontal, as can be seen in [Figure 2.10](#).

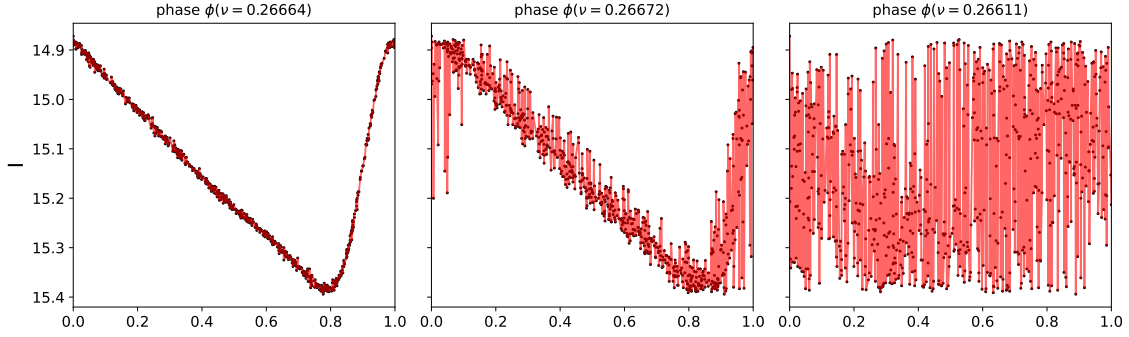


Figure 2.10: Different phase diagrams for tests frequencies on the star from figure 2.6. The red lines are connecting each pair of consecutive dots on the phase diagram, and its total length is minimum at the correct phase, and higher otherwise. Contrary to Figure 2.9, even the slightest deviation from the correct frequency causes the arclength to rise; this method is extremely sensible to phase changes.

### Phase diagram statistics

As an alternative to sorting all the phases each time, we could just do it partially, dividing the  $[0,1]$  phase axis into  $M$  equal parts. Then we could use the arclength method with the centroid of the point on those parts, but this does not turn out to be very useful. An alternative is the method proposed by Stellingwerf (1978), which defines the global variance of the phase diagram as

$$\sigma^2 = \frac{\sum_k^N (f_k - \bar{f})^2}{N - 1} \quad (2.42)$$

and a sort of moving variance

$$s_j^2 = \frac{\sum_k^{n_j} (f_k - \bar{f}_j)^2}{n_j - 1} \quad (2.43)$$

where the sum is performed over the  $n_j$  points in the  $j$ -th interval,  $j \in [1, M]$ . The pooled variance for the  $M$  samples would be

$$s^2 = \frac{\sum_j^M (n_j - 1) s_j^2}{\sum_j^M n_j - M} \quad (2.44)$$

and the statistic to consider will be defined as  $\Theta = s^2 / \sigma^2$ . If the frequency used to calculate the phases is wrong,  $\Theta \approx 1$ , and if its right or nearly right,  $\Theta$  would reach a local minimum.

This method technically fits the phase curve with the moving mean of the phase curve, and is thus another approximation of the power spectrum. The coefficient of determination  $R^2 = 1 - \Theta$  could

be used directly to compare it with a Lomb-Scargle periodogram, but this method will probably be slower, as the tally of a length  $n$  array into  $M$  sub-intervals is  $O(nM)$ .

The same phase diagrams as in [Figure 2.10](#), but partially sorted (binned as bidimensional histograms), can be seen in [Figure 2.11](#). Note that this method only requires discretization of the data in the phase axis.

### Phase diagram entropy

If instead of calculating the variance of each column of this histograms, we calculate the overall entropy of the data, we would be using the minimum entropy method proposed by [Cincotta et al. \(1995\)](#). If the phase diagram is partitioned in  $m = M^2$  zones of the same area, the probability  $\mu_j$  for a point to be inside any given zone is just the points inside that zone divided by the total number of points. The entropy of this configuration in the [Shannon \(1948\)](#) sense is

$$S = - \sum_j^m \mu_j \ln(\mu_j) \quad (2.45)$$

As already discussed, if the test period is off, we expect the phase diagram to be populated everywhere; therefore those probabilities would be uniform, and the entropy would be maximum. In the correct period the probabilities for each zone are heavily unbalanced towards the occupied spaces, and then the histogram is “ordered” and the information entropy minimum.

On a more mathematical level, if the histogram is populated everywhere, as in the noise case, most of the  $\mu_j$  would be nonzero. As probabilities, they lie in the  $[0, 1]$  interval, making their logarithms negative, adding up to a positive and maximal sum.

On the other side, on the ocrrect period there would be many unoccupied zones (see [Figure 2.11](#)), and therefore most of the  $\mu_j$  would be zero. As  $\lim_{\mu \rightarrow 0} \mu \ln \mu = 0$ . The growth of the other probabilities cannot compensate for this loss, so the sum is minimum. This has been rigorously proved by [Cincotta et al. \(1999\)](#).

As the obvious implementation involve the full bidimensional binning of the phase diagram, this would be  $O(nM^2)$ , making it technically slower than the other methods. There are faster methods to compute the entropy of a signal ([Cohen & Procaccia 1985](#)), but for our purposes it is possible to implement an entropy method that is faster that both the  $O(n \ln n)$  arclength and the  $O(nM)$  method, by transforming the bidimensional histogram into a one dimensional one; see [Listing 7](#) for details.

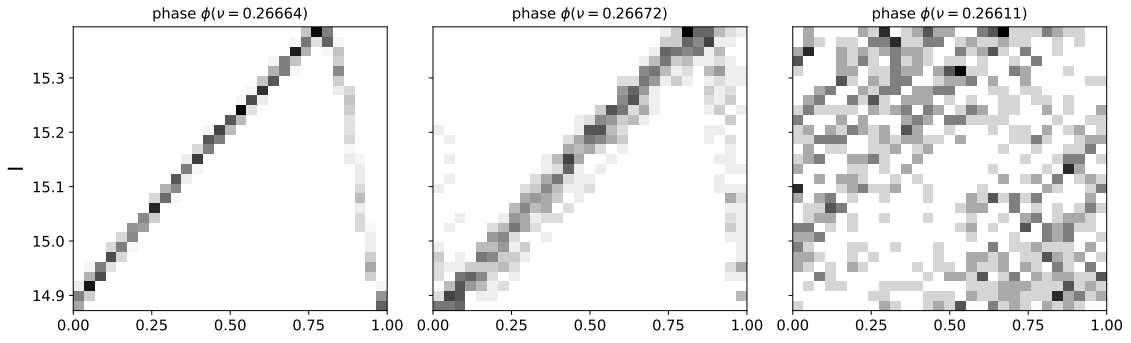


Figure 2.11: Same as Figure 2.10 but “pixelated” phase diagrams, *i.e.* bidimensional histograms.  $M = 30$  bins were taken on each axis, for a total of  $m = 900$  zones. The greater columnwise dispersion of the off-phase diagrams is notable, and when the frequency is completely wrong, the image is just noise. On the applications discussed here, the number of bins on each axis are typically between 2 and 10.

## Chapter 3

# Methodology

### 3.1 OGLE IV classical Cepheids on the Magellanic clouds

#### 3.1.1 The OGLE project

The Optical Gravitational Lensing Experiment (OGLE) project is an earth-based observational astronomical project with the objective of study dark matter and microlensing phenomena in the Magellanic clouds and the Galactic bulge. This project was originated by an idea of Bohdan Paczyński (1986). Its first phase began in 1992, with very limited time in the 1m Swope telescope, in Las Campanas Observatory, Chile, operated by the Carnegie Institution of Washington.

Those initial observation continued until 1995, and produced many interesting results: apart from its initial goal of observing microlensing events (Udalski et al. 1993), they published extinction maps of the galactic bulge (Stanek 1996), and the first version of their catalog of variable stars (Udalski et al. 1997b).

The whole idea of Paczyński was a great success despite its difficulties, and therefore the Warsaw University decided to further fund OGLE. In 1995 began the construction of the 1.3m Warsaw telescope and observing site at the same observatory, due to its optimal position. The construction, assembly and tests were concluded by late 1996, and in 1997 began the second phase of the project.

Third and fourth OGLE phases began in 2001 and 2010 respectively. The last phase, OGLE-IV, is still on pause at the time of writing this document (late 2021) due to the COVID-19 pandemic. As of 2019, their main site<sup>1</sup> reported ~800 papers published from the main OGLE results, and

---

<sup>1</sup><http://ogle.astrouw.edu.pl/>

mode	stars	with I	only I	with V	only V	both	none
F	2477	2428	151	2287	10	2277	39
1O	1776	1766	116	1650	0	1650	10
2O	26	26	0	26	0	26	0
F1O	95	95	6	89	0	89	0
1O2O	322	321	16	305	0	305	1
F1O2O	1	1	0	1	0	1	0
1O2O3O	7	7	0	7	0	7	0
1O3O	1	1	0	1	0	1	0
2O3O	1	1	0	1	0	1	0
total	4706	4646	289	4367	10	4357	50

Table 3.1: Distribution of the number of stars in the database for the LMC according with their reported pulsation mode. The first column give the total of stars in that mode. The next four columns give how many of those stars have data in *I* band, *only I* band, and the same for *V* band. The last two columns enumerates the stars that have data in both bands and the stars of which no data were found in the database. As can be seen, there are so few stars with pulsation modes F1O2O, 1O2O3O, 1O3O, and 2O3O, that their inclusion in the following analysis would be less than satisfactory.

~1300 OGLE-related results. All of that makes the OGLE project one of the oldest observational projects that have been focused on the same objects, making it an invaluable resource for studying the Magellanic system and the galactic bulge.

### 3.1.2 Data acquisition and description

As previously mentioned, the OGLE project uses its own 1.3m telescope at Las Campanas Observatory, Chile, fully described by Udalski et al. (1997a). The OGLE-IV phase uses this telescope, and since 2009 they use a mosaic camera of 32 thin E2V44-82 2048×4096 CCD chips, with 0.26 arcsec/pixel scale and 1.4 square degrees total field of view. The whole data obtention process, the specific mosaic arrangement and the sky coverage are described by Udalski et al. (2015).

This work uses a subset of the The Ogle Catalog of Variable Stars (OCVS), publicly available to download for the astronomic community at <http://www.astrouw.edu.pl/ogle/ogle4/OCVS/>. Specifically, we will use the photometric data from Classical Cepheids in the Small and Large

mode	stars	with I	only I	with V	only V	both	none
F	2754	2739	125	2617	3	2614	12
1O	1791	1783	86	1697	0	1697	8
2O	91	91	2	89	0	89	0
F1O	68	68	2	66	0	66	0
1O2O	239	239	9	230	0	230	0
1O2O3O	1	1	0	1	0	1	0
total	4944	4921	224	4700	3	4697	20

Table 3.2: Same as Table 3.1 but for the SMC. This time, there is a single star with pulsation mode 1O2O3O, and therefore it will be excluded from the analysis. This choice leaves the LMC and the SMC with the same pulsation categories.

Magellanic clouds (SMC, LMC). This section of the catalog is composed of 9583 individual stars, with data on the  $I$  and  $V$  magnitude, taken with the filters shown in Figure 2.3, but converted to the standard Johnson-Cousins photometric system (Udalski et al. 2015). Each data file is composed of three columns, the date of the observation in  $RHJD = HJD - 2.45 \times 10^6$ , the measured magnitude, and its uncertainty. Uncertainties though the data varies from 0.004 mag to 0.1 mag in the most extreme cases. Those cases are extremely rare, however, and the mean stays at around 0.006 mag.

There are some stars that, perhaps by observation schedule, perhaps by reclassification, have no data files in some filters, or have data in only one. This information, separated by reported pulsation mode, is detailed in Table 3.1 for the LMC and in Table 3.2 for the SMC. The stars with no data will be ignored, as well as the rarest pulsation modes. From those tables is clear that stars with  $I$  magnitude data are slightly more prevalent. In fact, even for a star that has data in both filters, there is usually a lot more  $I$  magnitude observation than  $V$  ones. The distribution of data points per star per filter can be seen in Figure 3.1. Probably the  $V$  data was only taken to measure its mean, in order to construct the Wesenheit index, while the  $I$  data was taken both for the mean  $I$  magnitude and to calculate the period.

In order to define our frequency grid for the period search, the OGLE reported periods can be examined to deduce a reasonable range. The distributions of  $I$  mean magnitude and amplitude can also be examined, as they are all presented in Figure 3.2. There one can see some bimodal behaviors with small tails in the long-period high-luminosity end of the data, and sometimes a distinction between the pulsation modes can be made.



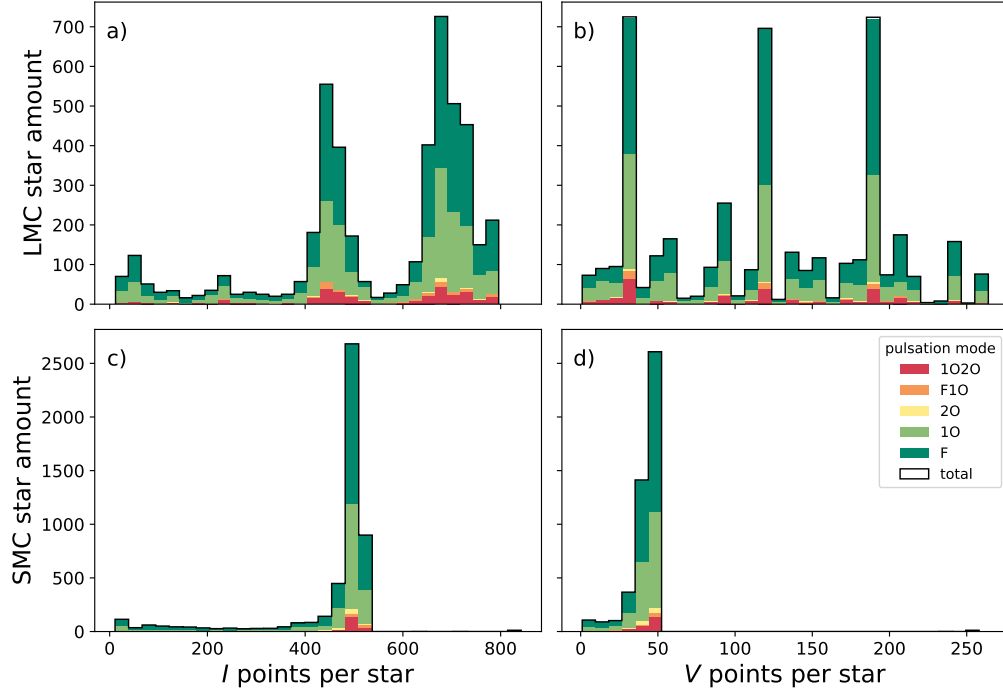


Figure 3.1: Distribution of data points per star in both *I* and *V* filters for the LMC and the SMC. Histograms are cumulative, and each color represents the contribution fraction of each pulsation mode on each bin. Total is outlined in black. As expected, there are much more data points in the LMC compared to the SMC due to their size difference. There is also much more data in *I* magnitude, as outlined in the text. The threshold for minimum number of points required by the period finding algorithms is around 50, which made the *V* data unusable for finding periods on the SMC case (d). If only *I* data is to be used for this purpose, the lower tails on both clouds can be cut at around 400 points without losing too many stars.

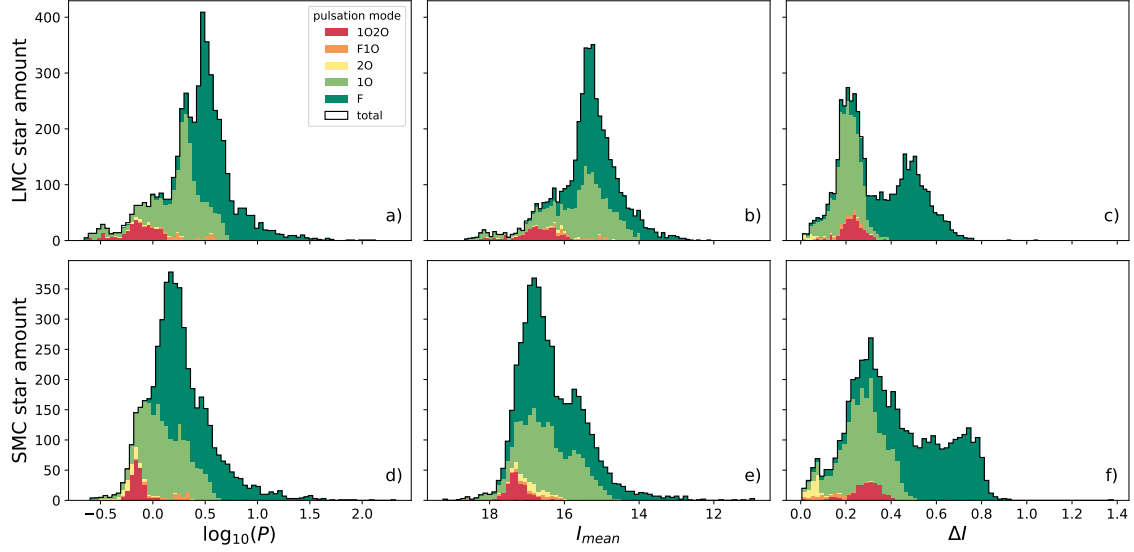


Figure 3.2: Distribution of periods, mean  $I$  magnitudes and  $I$  amplitudes for the OGLE classical Cepheids in the Magellanic system. Histograms are cumulative, and each color represents the contribution fraction of each pulsation mode on each bin. Total is outlined in black. In the case of multi-period stars, only the lower one is considered, as reported in the data headers `*.cep`. The tails on the longer periods and brightest and widest stars are kept in the plot range, despite having so few stars, to represent the difficulty of studying the bright side of the PL relation.

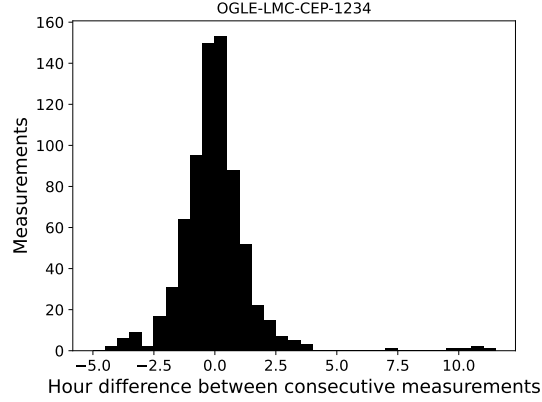


Figure 3.3: Histogram of the *hour* difference between consecutive observations for a sample star in the LMC. This hour difference is defined as the time difference modulo one day; for example, if one observation was made on Monday at 1:00am, and the next on Tuesday at 3:00am, the hour difference would be 3. The small clump on  $\sim 10$  hours is due to the yearly pauses on observations, as can be seen on Figures 2.5 and 2.6. If the data were to be evenly sampled, this histogram would look like a delta function.

### Sampling and observation cadence

All of these observations are made on earth, and as such, they are limited to the sky conditions, tight scheduling and the earth rotation. Therefore, the cadence of observations is far from being evenly sampled. An example of this uneven nature of the data is presented on Figure 3.3 for a single star, and on Figure 3.4 for the whole of OGLE-IV measurements.

As we are not interested in signal reconstruction, but only in finding the true period, we can bypass the mean sampling frequency limit given by Marvasti (2001). In theory, when a signal is evenly sampled, one could only calculate its periodogram up to half that frequency (the Nyquist frequency); all the subsequent peaks are false peaks, reflections of the spectrum. An example of this can be seen on Figure B.1.

The is shown how uneven sampling might be even beneficial, as one can freely search for a frequency much higher that the mean Nyquist limit for the data. How much this effect lasts and hwy many points does it need to be useful remains to bee seen.

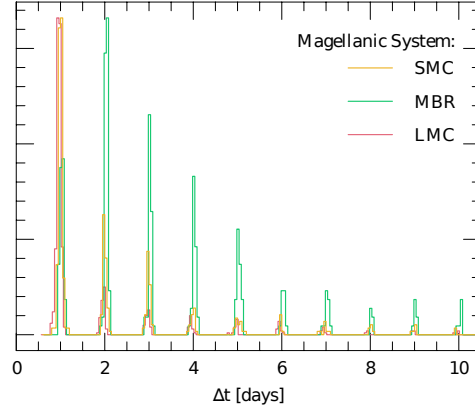


Figure 3.4: Histogram of the time difference between consecutive observations on OGLE-IV, for the clouds (LMC, SMC) and the Magellanic Bridge (MBR). Figure 3.3 is basically a phase diagram of the LMC histogram, but only considering the observations made to one star. The vertical scale is normalized to the highest point of each curve. Taken from Udalski et al. (2015), Figure 19.

## 3.2 Implementation details

All of the algorithms presented here are programmed in Python (Van Rossum & Drake 2009) for simplicity, but with extensive use of the NumPy (van der Walt et al. 2011) and Numba (Lam et al. 2015) libraries to achieve greater performance.

No claim is made that these algorithms are optimal, but effort was putted on implement the optimizations commonly used in the literature.

All of the algorithms take a time array, a magnitude array, and a frequency array as parameters. The minimum entropy and minimum dispersion algorithms also take an integer number of bins.

### 3.2.1 Algorithms

#### Arclength

For each frequency in the search array:

- Listing 3 is used to calculate the phase with the time and magnitude arrays.
- The phase array is sorted, and the magnitude array is rearranged accordingly.

- The differences between contiguous phases and magnitudes are calculated.
- Those differences are squared and then added, according to [Equation 2.41](#)
- Optionally, the square root is taken. This does not affect the peaks except by a scaling.
- The total sum is returned as the arclength.

The best guess for the correct frequency would be where the arclength is minimum. A possible implementation of a single iteration of the method is given in [Listing 4](#).

## Entropy

For each frequency in the search array:

- [Listing 3](#) is used to calculate the phase.
- A binning algorithm, like [Listing 5](#), is used to calculate a 2D histogram of the phase diagram.
- The histogram counts are transformed in the probabilities  $\mu_j$  of [Equation 2.45](#).
- The zero probability bins are deleted, to avoid a numerical error taking the logarithm.
- The sum on [Equation 2.45](#) is returned as the entropy.

The best guess for the correct frequency would be where the entropy is minimum. A possible implementation of an iteration is given in [Listing 6](#).

A big improvement can be made by translating the phase-magnitude plane  $\{\phi_k, m_k\}$  into a line  $\lfloor m_k n_{bins} \rfloor + \phi_k$ , analogous at making the phase “decens” and the phase “units”. Over that transformed line, a 1D histogram algorithm can be used with  $n_{bins}^2$  number of bins, hence the phase diagram is “flattened”. That implementation is given in [Listing 7](#). This flattening allows the use of very efficient 1D binning algorithms.

## Dispersion

For each frequency in the search array:

- [Listing 3](#) is used to calculate the phase.
- For each uniform zone (partition of the phase  $[0, 1]$  range), calculate the occupation number and the variance of the magnitudes corresponding to the phases on the zone.

- If the zone is empty, the variance is defined as zero. This makes some sense because the (biased) variance  $\frac{\sum (x-\bar{x})^2}{n}$  would be an empty sum with a second order zero in the numerator and a first order zero in the denominator.
- The pooled variance  $s^2$  of all the zones is calculated with [Equation 2.44](#). The overall variance  $\sigma^2$  is just the variance of all the magnitudes.
- The coefficient of determination  $R^2 = 1 - s^2/\sigma^2$  is returned as the dispersion.

The best guess for the correct frequency would be where this  $R^2$  is maximum. This  $R^2$  approximates really well the Fourier spectrum, as can be seen in [Figure 3.6](#). The implementation is in [Listing 8](#)

### Lomb-Scargle

The original formulas for the Lomb-Scargle periodogram (Equations [2.39](#) and [2.40](#)) involve several redundant trigonometric calculations. Here we implement the optimizations given by [Townsend \(2010\)](#), where no trigonometric function is repeated.

For each frequency  $\nu$  in the search array, being  $t$  the time array and  $m$  the magnitude array:

- Calculate:

- $\omega t = 2\pi\nu t$
- $C = \cos \omega t$
- $S = \sin \omega t$
- $mC = m \cdot C$
- $mS = m \cdot S$
- $CC = C \cdot C$
- $SS = S \cdot S$
- $CS = C \cdot S$
- $\omega\tau = \frac{1}{2} \arctan \left( \frac{2CS}{CC-SS} \right)$
- $c_\tau = \cos \omega\tau$
- $s_\tau = \sin \omega\tau$

- return the coefficient of determination

$$R^2 = \frac{1}{2} \left( \frac{(c_\tau mC + s_\tau mS)^2}{c_\tau^2 CC + 2c_\tau s_\tau CS + s_\tau^2 SS} + \frac{(c_\tau mS - s_\tau mC)^2}{c_\tau^2 SS + 2c_\tau s_\tau CS + s_\tau^2 CC} \right)$$

As before, the correct frequency is at maximum  $R^2$ , and this is also a good approximation of the Fourier spectrum.

## Fourier

A straightforward calculation of Equation 2.38 would actually be enough to outperform some (if not all) of the aforementioned methods, but there is an iterative optimization provided by Kurtz (1985), that calculates the transform on the next frequency modifying the last calculation, saving a lot of time.

Thus, contrary to all the other methods, this implementation cannot be putted into a single-iteration function. Also contrary to the other methods, it requires the frequency grid to be evenly spaced. This is not a problem, as the only problem were the *times* to be unevenly spaced. Additionally, if we have no idea of the correct frequency in the first place, there is no worse nor better way to search for it than in an evenly spaced frequency grid.

The idea behind this optimization is that if the frequencies are of the form  $\nu_j = \nu_{j-1} + \Delta\nu$ , the centroid from Equation 2.37 can be written (element-wise) as

$$C(\nu_j)_k = f_k e^{2i\pi(\nu_{j-1} + \Delta\nu)t_k} = f_k e^{2i\pi\nu_{j-1}t_k} e^{2i\pi\Delta\nu t_k} = C(\nu_{j-1})_k e^{2i\pi\Delta\nu t_k} \quad (3.1)$$

Where  $j$  goes over the frequencies and  $k$  over times/magnitudes. The array  $D_k = e^{2i\pi\Delta\nu t_k}$  is constant, so the coordinates of the “curled object” in the complex plane that we took the centroid off (see Figure 2.9) for a certain frequency, can be calculated from the last coordinates by just multiplying element-wise by  $D$ .

Given  $\nu_0, \nu_f, \Delta\nu$ , the time array  $t_k$  and the magnitude array  $f_k$ , the algorithm would look like:

1. Calculate  $D_k = e^{2i\pi\Delta\nu t_k}$
2. Allocate an array  $F_k$  of the same size as the frequency range,  $J = \lfloor \frac{\nu_f - \nu_0}{\Delta\nu} \rfloor$ , to store the centroid magnitudes.
3. Initialize  $\nu = \nu_0$ .
4. Calculate the first “curled object” in the complex plane,  $C_k = f_k e^{2i\pi\nu_0 t_k}$ .
5. Initialize  $F_0 = |\sum_k C_k|^2$
6. While  $\nu < \nu_f$ , or equivalently, for  $j \in [1, J]$

- $C_k \rightarrow C_k \star D_k$ , where  $\star$  means element-wise multiplication
- $F_j \rightarrow |\sum_k C_k|^2$

7. Return  $\nu_k = \nu_0 + k\Delta\nu$  where  $F_k$  is maximum.

Perhaps the code on [Listing 11](#) would be clearer.

### 3.2.2 Examples and the sampling problem

A simple benchmark of the algorithms can be seen on [Figure 3.5](#). It is of note that the minimum entropy algorithm (on the naive implementation) was the slowest, and the simple Fourier method was the fastest by far, even without optimizations. The “flattened” implementation of minimum entropy (on red) is faster than the dispersion one, because the  $M^2$  term in the  $O(nM^2)$  complexity of the entropy calculation is generally small, and there is no linear term on counting bins in the histogram. On the other side, despite the  $O(nM)$  complexity of the dispersion method, each iteration contains variances and counting in addition to the tally, so there is linear overhead on  $n$ . The full spectrum iterative Fourier method has complexity  $O(nf)$ , with  $f$  the number of frequencies. It is slower than the  $O(n \ln n)$ , as for our purposes  $f \gg n$ . Even with this, the incremental Fourier method seems to be the fastest on uneven sampled data.

Examples of the spectra produced by each method are in [Figure 3.6](#). As expected, the Lomb-Scargle and dispersion spectra are very good approximations of the pure Fourier one. Of course, there are some random noise and minor peaks that make those three differ, and even more with the arclength spectrum. The entropy spectrum is the most different, considering that the peak is sharper (in the first star) and it presents false peaks at  $\nu \approx 1/\text{day}^2$ . In fact, the Fourier methods also present that peaks, but they can be eliminated.

These peaks are a common occurrence, because of the near-nightly cadence of the data acquisition process (see [Figure 3.4](#)). We can take a closer look at the phase diagram and Fourier curling on one of those false peaks in [Figure 3.7](#).

Then it is suddenly clear why the entropy method fails: the dots are distributed phase-wise in a similar fashion as in [Figure 3.3](#). That leaves a huge gap (in this case in the middle), and if there are many bins with probability zero the entropy would certainly be lower, because it is more ordered that way; there are chunks of the phase diagram histogram that are uniform, and therefore the entropy method gives a false positive.

---

<sup>2</sup>It’s never exactly 1/day, because of the peak on [Figure 3.3](#). The earth moves around the sun, and optimal observing times also move. Then the peak will not be exactly at zero.



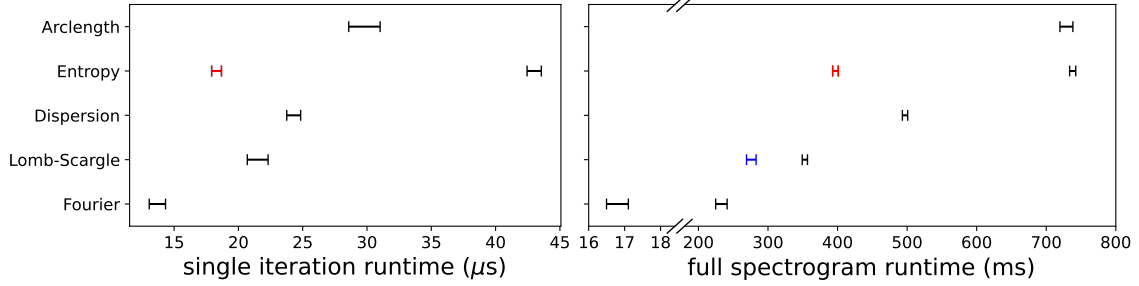


Figure 3.5: Sample benchmark of repeated single-threaded runs over all the algorithms described in this section. The single iteration times were calculated with 50000 repetitions of 7 runs, and the full spectrum times were made on a regular frequency grid of size 14000, with 10 repetitions. The red points of the entropy method refer to the flattened implementation on [Listing 7](#), and the black one to the naive implementation. For the Lomb-Scargle method, a blue point with the times for the `scipy.signal.lombscargle` implementation was given as a reference value. The full spectrum time axis is broken to show the difference between the incremental method ([Listing 11](#)) and the simple method ([Listing 10](#)). The machine used was a domestic computer with Intel® Core™ i7-4702MQ CPU @ 2.20GHz.

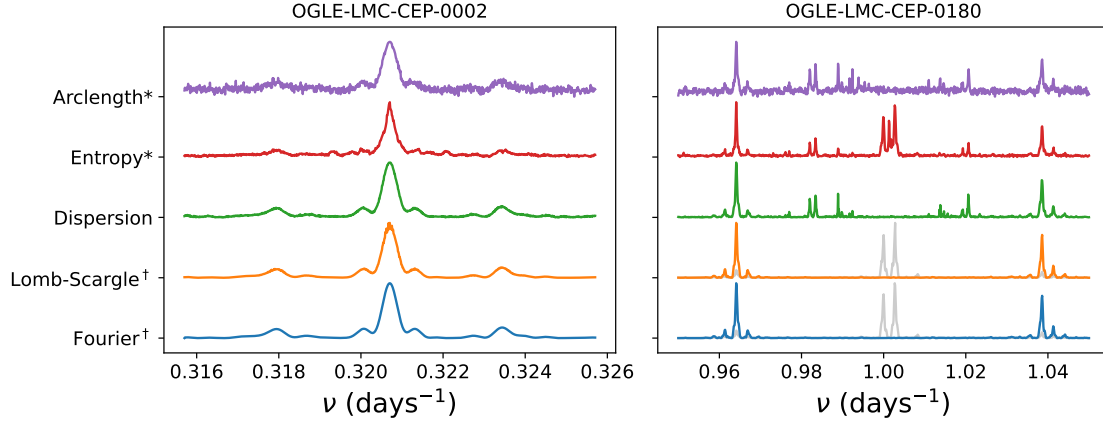


Figure 3.6: Sample spectra for the five methods described in this section, for two different stars, near the true frequency. (\*): Arclength and Entropy spectra were inverted to allow direct comparison with the Fourier approximations. (†): The Fourier, Lomb-Scargle and entropy methods present a false peak (in light gray) when the frequency is very near 1/day; see text for details.

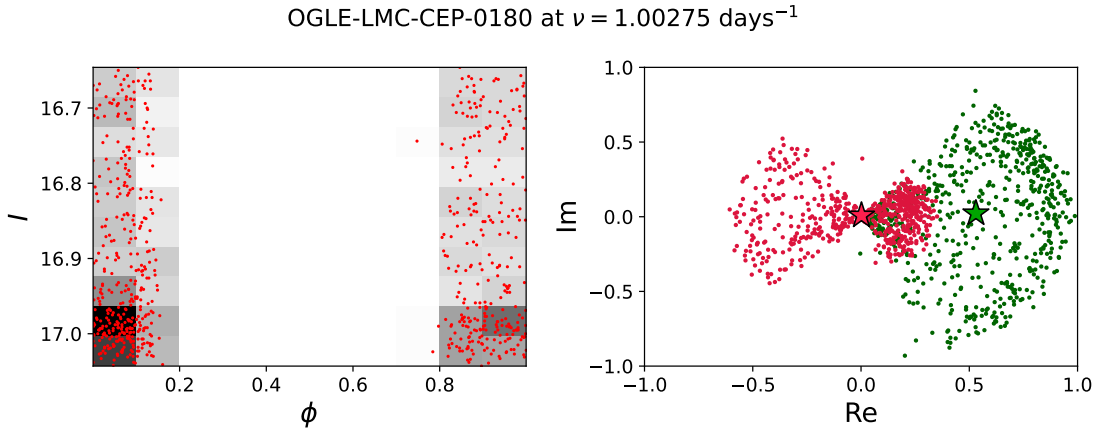


Figure 3.7: Detail of the phase diagram and complex phase curling of the star from Figure 3.6, very near the false peak at 1.00275/day. The histogram was made with 10 bins on each axis. The green and magenta dots represent a Fourier curling where the magnitude has been normalized in the respective intervals  $[0, \max - \min]$  and  $[\min - \text{mean}, \max - \text{mean}]$  (such that the mean becomes zero). The big stars lie on the centroid.

Something similar happens with the curled figure in the complex plane. As each point is rotated almost  $360^\circ$  (again, how much more or less than that is dictated by [Figure 3.3](#)), and if the radii are all positive that means the centroid would be near the positive reals. There is a simple solution: make the mean radius zero. That way, the negative radius would end in the  $180^\circ$  side of the complex plane, and everything would cancel nicely.

[Figure 3.7](#) also allow us to understand how the arclength and dispersion spectra are “immune” to this effect. If we connect the dots of the phase diagram, at least one line should join the two groups, and it will be a big line. So no matter how close get those points on the sides, the big line between them should carry the weight of this strange alignment. That works until everything aligns in one clump in the middle, and suddenly there is no line. This behavior deserves further study. On the other side, we defined the empty variance as zero. Therefore the terms corresponding at the empty gaps would just not contribute, neither in the denominator nor in the numerator of [Equation 2.44](#). It would be as if someone had just cut out the empty columns. The rest is just vertical chaos, having in fact maximal variance, so the dispersion method is truly immune with our definitions.

## Chapter 4

# Results

### 4.1 Final PL relations

### 4.2 Discussion

## Chapter 5

# Conclusions

### 5.1 Future work

# Appendices

# Appendix A

## Code utilities

### A.1 C listings

```
double eps(double x) {  
    long i = *(long*) &x;  
    i++;  
    double x_next = *(double*) &i;  
    return x_next - x;  
}
```

Listing 1: C code of a function that estimates the machine epsilon for a given double precision floating point: the minimum value for which  $x + \epsilon/2 > x$  still holds. This code exploits the fact that consecutive floating point numbers must have consecutive bit representations.

### A.2 Python listings

For all the code fragments in this section, [Listing 2](#) is used to define the names of the NumPy and Numba package namespaces. The Numba notation for function signatures is given on [Table A.1](#). A string using these types in the `njit` decorator is given to allow ahead-of-time compilation. The notation for these strings is `return_type(arg1_type, arg2_type, ...)`.

Type	notation
64 bit float	f8
64 bit signed integer	i8
Contiguous 1D array	f8[:, :1]
Row-contiguous 2D array	f8[:, ::1]

Table A.1: Numba types used in the code for this work. The full list can be consulted on <https://numba.pydata.org/numba-doc/dev/reference/types.html>

```
import numpy as np
from numba import njit
```

Listing 2: Common library imports for all the Python listings.

```
@njit("f8[:, :1](f8[:, :1], f8)")
def phase(t_arr, f_test):
    return np.mod(t_arr*f_test, 1.0)
```

Listing 3: Straightforward implementation of 2.28. The ephemeris is not included, since it does not matter for any of the algorithms, an is used just as a visual aid to present the light curves.

```
@njit("f8(f8[:, :1], f8[:, :1], f8)")
def arclength(t_arg, m_arg, f_test):
    phi = phase(t_arg, f_test)
    index = phi.argsort()
    return sum(np.diff(phi[index])**2 + np.diff(m_arg[index])**2)
```

Listing 4: Here, `numpy.argsort` (which is implemented as a quicksort) is used to simultaneously sort the phase array and rearrange the magnitude array. The phase array is calculated with Listing 3.



```

@njit("f8[:,::1](f8[:,::1],f8[:,::1],i8)")
def hist2d(x_arr,y_arr,bins):
    x = x_arr.copy()*bins
    y = y_arr.copy()*bins
    H = np.zeros((bins,bins))
    for i in range(bins):
        for j in range(bins):
            H[i,j] += sum((i<=x)*(x<=i+1)*(j<=y)*(y<=j+1))
    return H

```

Listing 5: A simple algorithm to calculate a 2D histogram. It returns a matrix of shape `(bins,bins)` showing how many points of the data fall on each bin. Both `x_arr` and `y_arr` are assumed to be normalized on the `[0, 1]` interval, for simplicity.

```

@njit("f8(f8[:,::1],f8[:,::1],f8,i8)")
def entropy(t_arg,m_arg,f_test,bins):
    phi = phase(t_arg,f_test)
    H = hist2d(phi,m_arg,bins)
    mu = H.ravel() / len(phi)
    mu = mu[np.nonzero(mu)]
    return -mu@np.log(mu)

```

Listing 6: Algorithm to calculate the entropy of a phase diagram, avoiding numerical error on large `bins` by deleting the zero probability zones.

```

@njit("f8(f8[::1],f8[::1],f8,i8)")
def entropy_flattened(t_arg,m_arg,f_test,bins):
    phi = phase(t_arg,f_test)
    rounded_mag = np.empty_like(m_arg)
    np.round(bins*m_arg,0,rounded_mag)
    counts = np.histogram(phi+rounded_mag,bins=bins**2)[0]
    mu = counts/len(t_arg)
    mu = mu[np.nonzero(mu)]
    return -mu*np.log(mu)

```

Listing 7: Algorithm to calculate the entropy of a phase diagram, flattening and rounding the magnitude scale in order to use a 1D histogram algorithm.

```

@njit("f8(f8[::1],f8[::1],f8,i8)")
def dispersion(t_arg,m_arg,f_test,bins):
    x = phase(t_arg,f_test)*bins
    ss = np.zeros(bins) # zone variances
    n = np.zeros(bins) # zone occupations
    for i in range(bins):
        mask = (i<=x)&(x<=i+1)
        n[i] = mask.sum()
        if n[i] == 0:
            ss[i] = 0 # empty variance
        else:
            ss[i] = np.var(m_arg[mask])
    s2 = ((n-1)*ss)/(n-1).sum() # pooled variance
    return 1-s2/np.var(m_arg)

```

Listing 8: Algorithm to calculate the dispersion of a phase diagram. There is some resemblance between this code and Listing 5; the `mask` array effectively takes a 1D histogram of the phase, contrary to a 2D one, and thus its faster. Care is taken when taking an empty variance, because it is not defined in NumPy.

```

@njit("f8(f8[::1],f8[::1],f8)")
def lombscargle(t_arg,m_arg,f_test):
    omega_t = f_test*2*np.pi*t_arg
    c = np.cos(omega_t)
    s = np.sin(omega_t)
    mc = sum(m_arg*c)
    ms = sum(m_arg*s)
    cc = sum(c**2)
    ss = sum(s**2)
    cs = sum(c*s)
    omega_tau = np.arctan(2*cs/(cc-ss))/2
    c_tau = np.cos(omega_tau)
    s_tau = np.sin(omega_tau)
    P = 1/2 * (
        (c_tau*mc+s_tau*ms)**2/(c_tau**2*cc+2*c_tau*s_tau*cs+s_tau**2*ss) +
        (c_tau*ms-s_tau*mc)**2/(c_tau**2*ss+2*c_tau*s_tau*cs+s_tau**2*cc)
    )
    return P

```

Listing 9: Non-redundant implementation of the Lomb-Scargle periodogram, according to [Townsend \(2010\)](#). This is similar to the [SciPy implementation](#), but here Numba is used instead of Pythran.

```

@njit("f8(f8[::1],f8[::1],f8)")
def fourier(t_arg,m_arg,f_test):
    return abs(sum(m_arg*np.exp(-2j*np.pi*t_arg*f_test)))**2

```

Listing 10: Simple implementation of an iteration of the Fourier periodogram, without any optimizations.

```

@njit("f8[:, :1](f8[:, :1], f8[:, :1], f8, f8, f8)")
def fourier_incremental(t_arg, m_arg, f_ini, f_end, df):
    dF = np.exp(-2j*np.pi*t_arg*df)
    F = np.zeros_like(np.arange(f_ini, f_end, df))
    curl = m_arg * np.exp(-2j*np.pi*t_arg*f_ini)
    F[0] = np.abs(sum(curl))
    for i in range(1, len(F)):
        curl = curl*dF
        F[i] = np.abs(sum(curl))**2
    return F

```

Listing 11: Incremental iterative implementation of the nonuniform discrete Fourier transform, according to Kurtz (1985). A production-grade implementation of this algorithm can be found in the `fnpeaks` package by Zbigniew Kołaczkowski; this is in fact the software used by OGLE-IV (Soszyński et al. 2015).

## Appendix B

### Complementary figures

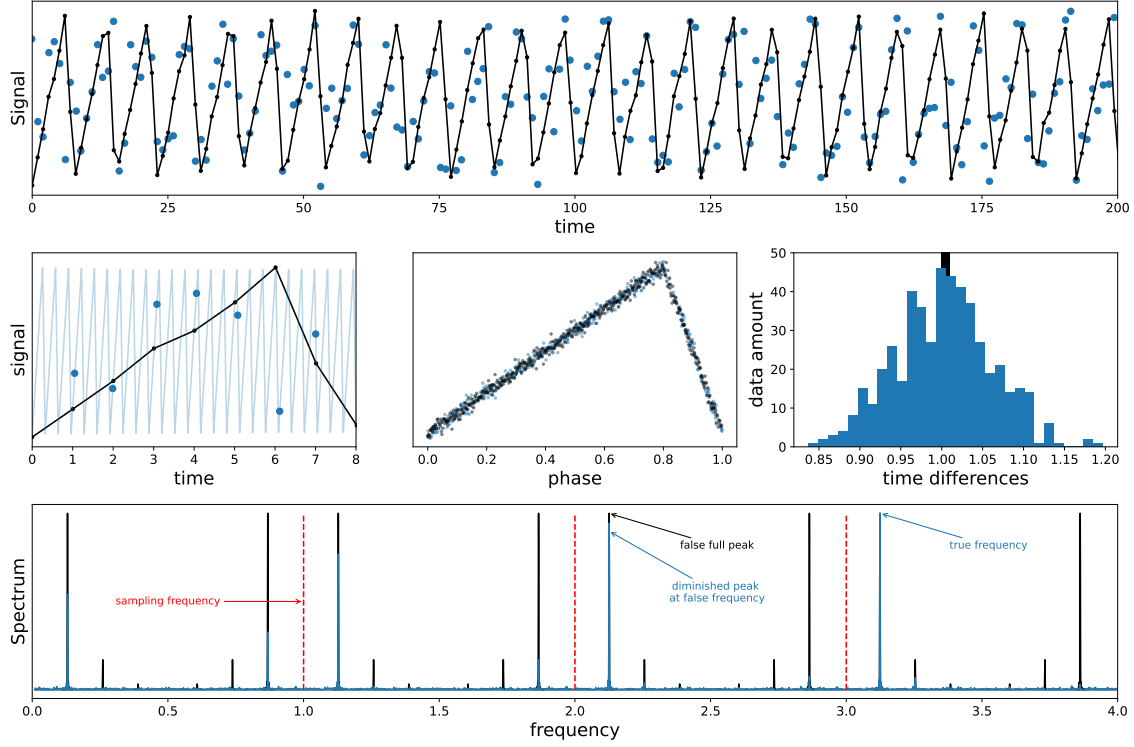


Figure B.1: Sample skewed triangular wave sampled two ways: the black dots are sampled one each unit of time. The sample times of the blue dots are very near one sample each unit of time, but have a little dispersion, as can be seen in the left histogram. The frequency of the generated data is 3.124 time units. It is clear that the spectrum reflects every half of the sampling time (Nyquist frequency). On the uneven sampling case, the spectrum does not reflect exactly; there are peaks greater than others, and the differences last until several times past the sampling frequency. The greatest peak is in the correct frequency, and the rest are partially canceled out. This cancellation auments with the dispersion on the sampling times, and the number of data. The detailed curve on the mid-left shows how exactly this happens: if the frequency is high, the dots with take an entire different path on the curve with the slighests of the temporal deviations.

# Bibliography

Aristotle. 350 BCE, *On the Heavens* (The Clarendon Press, 1992), translated to english by Stocks, J. L.  
[Aviable online.](#)

Arp, H. C. 1955, *AJ*, 60, 1

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481

Baade, W. 1944, *ApJ*, 100, 137

Baade, W. 1956, *PASP*, 68, 5

Bessell, M. S. 2005, *ARA&A*, 43, 293

BIPM. 2019, *Le Système international d’unités*, The International System of Units (‘The SI Brochure’), 9th edn. (Bureau international des poids et mesures)

Brigham, E. 1974, *The Fast Fourier Transform*, Prentice-Hall signal processing series (Prentice-Hall)

Burke, Edward W., J., Rolland, W. W., & Boy, W. R. 1970, *JRASC*, 64, 353

Butkov, E. 1968, *Mathematical Physics* (Addison-Wesley, Reading)

Cannon, A. J. & Pickering, E. C. 1901, *Annals of Harvard College Observatory*, 28, 129

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245

Carroll, B. & Ostlie, D. 2017, *An Introduction to Modern Astrophysics*, 2nd edn. (Cambridge University Press)

Catelan, M. & Smith, H. A., eds. 2015, *Pulsating Stars* (Wiley-VCH Verlag GmbH & Co. KGaA)

Cincotta, P. M., Helmi, A., Mendez, M., Nunez, J. A., & Vucetich, H. 1999, *MNRAS*, 302, 582

- Cincotta, P. M., Mendez, M., & Nunez, J. A. 1995, *ApJ*, 449, 231
- Clayton, G. C. & Cardelli, J. A. 1988, *AJ*, 96, 695
- Cohen, A. & Procaccia, I. 1985, *Phys. Rev. A*, 31, 1872
- Cox, J. P. 1963, *ApJ*, 138, 487
- Cox, J. P. 1980, *Theory of stellar pulsation* (Princeton University Press)
- Dalrymple, G. 1994, *The Age of the Earth* (Stanford University Press)
- Deeming, T. J. 1975, *Ap&SS*, 36, 137
- Dworetzky, M. M. 1983, *MNRAS*, 203, 917
- Eastman, J., Siverd, R., & Gaudi, B. S. 2010, *Publications of the Astronomical Society of the Pacific*, 122, 935
- Eddington, A. S. 1918, *MNRAS*, 79, 2
- Eddington, A. S. 1926, *The Internal Constitution of the Stars* (Cambride University Press)
- Eddington, A. S., S. 1941, *MNRAS*, 101, 182
- Fernie, J. D. 1969, *Publications of the Astronomical Society of the Pacific*, 81, 707
- Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, *ApJ*, 553, 47
- Friedmann, A. 1922, *Zeitschrift für Physik*, 10, 377
- Goodricke, J. 1786, *Philosophical Transactions of the Royal Society of London*, 76, 48
- Górski, M., Zgirski, B., Pietrzyński, G., et al. 2020, *ApJ*, 889, 179
- Graczyk, D., Pietrzyński, G., Thompson, I. B., et al. 2014, *ApJ*, 780, 59
- Griffiths, D. 2013, *Introduction to Electrodynamics*, Always learning (Pearson)
- Heger, A. & Woosley, S. E. 2002, *ApJ*, 567, 532
- Hertzsprung, E. 1913, *Astronomische Nachrichten*, 196, 201
- Ho Peng Yoke & Ho Ping-Yü. 1962, *Vistas in Astronomy*, 5, 127



- Hoffleit, D. 1993, Women in the history of variable star astronomy (American Association of Variable Stars Observers)
- Hoffleit, D. 1997, Journal of the American Association of Variable Star Observers, 25, 115
- Hoskin, M. 1979, Journal for the History of Astronomy, 10, 23
- Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168
- Hubble, E. P. 1925a, Popular Astronomy, 33, 252
- Hubble, E. P. 1925b, ApJ, 62, 409
- Humason, M. L., Mayall, N. U., & Sandage, A. R. 1956, AJ, 61, 97
- IEEE. 2019, IEEE Std 754-2019 (Revision of IEEE 754-2008), 1
- Jetsu, L. & Porceddu, S. 2015, PLoS ONE, 10, 44
- Jetsu, L., Porceddu, S., Lyytinen, J., et al. 2013, ApJ, 773, 1
- Karttunen, H., Kröger, P., Oja, H., Poutanen, M., & Donner, K. J. 2017, Fundamental Astronomy (Springer Berlin Heidelberg)
- Kiss, L. L. 1998, MNRAS, 297, 825
- Kramida, A., Yu. Ralchenko, Reader, J., & and NIST ASD Team. 2021, NIST Atomic Spectra Database (ver. 5.9). National Institute of Standards and Technology, Gaithersburg, MD.
- Kurtz, D. W. 1985, MNRAS, 213, 773
- Lam, S. K., Pitrou, A., & Seibert, S. 2015, in Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC (ACM Press)
- Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87
- Leavitt, H. S. & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1
- Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49
- Lockyer, J. N. 1896, Proceedings of the Royal Society of London, 59, 101
- Lockyer, J. N. 1897, Proceedings of the Royal Society of London, 61, 445

- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- Madore, B. F. 1982, *ApJ*, 253, 575
- Mamajek, E. E., Torres, G., Prsa, A., et al. 2015, arXiv e-prints, arXiv:1510.06262
- Marvasti, F. 2001, *Nonuniform Sampling: Theory and Practice*, Information Technology Series (Springer US)
- Matthews, L. D., Marengo, M., Evans, N. R., & Bono, G. 2012, *ApJ*, 744, 53
- Maury, A. C. & Pickering, E. C. 1897, *Annals of Harvard College Observatory*, 28, 1
- McCarthy, D. D. & Babcock, A. K. 1986, *Physics of the Earth and Planetary Interiors*, 44, 281
- Morrison, L. V. & Stephenson, F. R. 2004, *Journal for the History of Astronomy*, 35, 327
- Mróz, P., Udalski, A., Skowron, D. M., et al. 2019, *ApJ*, 870, L10
- Nataf, D. M. 2015, *MNRAS*, 449, 1171
- Paczýński, B. 1986, *ApJ*, 304, 1
- Patterson, C., Tilton, G., & Inghram, M. 1955, *Science*, 121, 69
- Payne-Gaposchkin, C. & Gaposchkin, S. 1967, *ZAp*, 67, 168
- Pickering, E. C. 1880, *Proceedings of the American Academy of Arts and Sciences*, 16, 1
- Pickering, E. C. 1890, *Annals of Harvard College Observatory*, 27, 1
- Pietrukowicz, P., Udalski, A., Soszyński, I., et al. 2012, *ApJ*, 750, 169
- Pietrzyński, G., Graczyk, D., Gallenne, A., et al. 2019, *Nature*, 567, 200
- Pigott, E. 1785, *Philosophical Transactions of the Royal Society of London*, 75, 127
- Planck, M. 1901, *Annalen der Physik*, 309, 553
- Pogson, N. 1856, *MNRAS*, 17, 12
- Samus', N. N., Kazarovets, E. V., Durlevich, O. V., Kireeva, N. N., & Pastukhova, E. N. 2017, *Astronomy Reports*, 61, 80
- Sandage, A. 1958, *ApJ*, 127, 513

- Scargle, J. D. 1982, *ApJ*, 263, 835
- Schlaflly, E. F. & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schuster, A. 1898, *Terrestrial Magnetism (Journal of Geophysical Research)*, 3, 13
- Shannon, C. E. 1948, *The Bell System Technical Journal*, 27, 379
- Shapley, H. 1918, *Contributions from the Mount Wilson Observatory*, 152, 1
- Shapley, H. 1924a, *Harvard College Observatory Circular*, 255, 1
- Shapley, H. 1924b, *Harvard College Observatory Circular*, 268, 1
- Shapley, H. & Curtis, H. D. 1921, *Bulletin of the National Research Council*, 2, 171
- Skowron, D. M., Skowron, J., Mróz, P., et al. 2019, *Science*, 365, 478
- Skowron, D. M., Skowron, J., Udalski, A., et al. 2021, *ApJS*, 252, 23
- Soszyński, I., Udalski, A., Szymański, M., et al. 2015, *Acta Astron.*, 65, 297
- Soszyński, I., Udalski, A., Szymański, M., et al. 2017, *Acta Astron.*, 67, 103
- Stanek, K. Z. 1996, *ApJ*, 460, L37
- Stellingwerf, R. F. 1978, *ApJ*, 224, 953
- Thompson, R. R. Y. 1971, *IEEE Transactions on Geoscience Electronics*, 9, 107
- Townsend, R. H. D. 2010, *ApJS*, 191, 247
- Udalski, A., Kubiak, M., & Szymański, M. 1997a, *Acta Astron.*, 47, 319
- Udalski, A., Olech, A., Szymański, M., et al. 1997b, *Acta Astron.*, 47, 1
- Udalski, A., Soszyński, I., Szymański, M., et al. 1999, *Acta Astron.*, 49, 437
- Udalski, A., Szymański, M., Kałużny, J., et al. 1993, *Acta Astron.*, 43, 289
- Udalski, A., Szymański, M. K., & Szymański, G. 2015, *Acta Astron.*, 65, 1
- Ulaczyk, K., Szymański, M. K., Udalski, A., et al. 2013, *Acta Astron.*, 63, 159

- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science & Engineering*, 13, 22
- Van Rossum, G. & Drake, F. L. 2009, *Python 3 Reference Manual* (Scotts Valley, CA: CreateSpace)
- VanderPlas, J. T. 2018, *ApJS*, 236, 16
- Waibl, E. & Herdina, P. 1997, *German Dictionary of Philosophical Terms*, Vol. 1 (Routledge)
- Wilk, S. R. 1996, *Journal of the American Association of Variable Star Observers*, 24, 129
- Yoon, J., Peterson, D. M., Kurucz, R. L., & Zagarelli, R. J. 2010, *ApJ*, 708, 71
- Zabolotskikh, M. V., Sachkov, M. E., Berdnikov, L. N., Rastorguev, A. S., & Egorov, I. E. 2005, in *ESA Special Publication*, Vol. 576, *The Three-Dimensional Universe with Gaia*, ed. C. Turon, K. S. O’Flaherty, & M. A. C. Perryman, 723
- Zhevakin, S. A. 1963, *ARA&A*, 1, 367