UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

UNCERTAINTY CHARACTERIZATION OF ORBITAL PARAMETERS IN CONTEXTS OF PARTIAL INFORMATION

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

RUBÉN MATÍAS CLAVERÍA VEGA

PROFESOR GUÍA:
MARCOS ORCHARD CONCHA

PROFESOR CO-GUÍA:
RENÉ MÉNDEZ BUSSARD

MIEMBROS DE LA COMISIÓN:
JORGE SILVA SÁNCHEZ
PABLO ZEGERS FERNÁNDEZ

SANTIAGO DE CHILE
2017

RESUMEN DE LA TESIS PARA OPTAR AL
GRADO DE MAGÍSTER EN CIENCIAS DE
LA INGENIERÍA, MENCIÓN ELÉCTRICA
POR: RUBÉN MATÍAS CLAVERÍA VEGA
FECHA: 2017
PROF. GUÍA: MARCOS ORCHARD CONCHA

UNCERTAINTY CHARACTERIZATION OF ORBITAL PARAMETERS IN CONTEXTS OF PARTIAL INFORMATION

Mass is arguably the most important property of a star: along with the initial distribution of chemical elements, stellar mass is responsible for the evolution and structure of these objects. The only means for calculating the mass of a star is the study of its orbital movement, for Kepler's Third Law establishes a strict mathematical relation between mass and orbital parameters. Since single stars do not follow orbital trajectories, the observational base of the stellar mass catalog consists mainly of binary and multiple stars; hence the importance of studying their orbits.

This dissertation addresses the problem of characterizing the orbits of binary stars from a Bayesian standpoint, approximating the posterior probability density functions of orbital elements by means of the technique known as Markov Chain Monte Carlo. A selection of 18 visual and 2 spectroscopic binary stars observed by the SOAR telescope are analyzed with the proposed technique, obtaining not only orbit estimates of each object (maximum *a posteriori* for visual binaries, expected value for spectroscopic binaries), but also a characterization of their uncertainty in the form of posterior distributions of orbital elements. By using a mathematical formalism developed as a part of this work, the dimension of the target parameter vector is reduced from 7 to 3 in the case of visual binaries, and from 10 to 7 for spectroscopic binaries, enhancing the efficiency of the estimation routines. The potential of combining astrometric and spectroscopic sources for estimating hypothesis-free parameters is explored, concluding that good estimations can be obtained if both the apparent orbit and the radial velocity profile of the star can be independently characterized.

This work also introduces a strategy to incorporate partial astrometric observations (measurements where either angular separation or position angle is missing) into the proposed Bayesian framework. This strategy combines the MCMC-based estimation of orbital parameters with the multiple imputation approach: instead of being discarded as an input for the parameter estimation routines, incomplete observations are replaced by a set of plausible values, incorporating this partial knowledge into the analysis. This methodology is tested on both synthetic and real data, obtaining not only the distribution of parameters given all observations available (complete and partial ones), but also an estimation of the spatial localization of partial measurements. Results suggest that the incorporation of partial knowledge can lead to a decrease in the uncertainty associated to target parameters (dramatic in some cases); however, partial measurements can also be redundant in some scenarios.

Resumen

Puede argumentarse que masa es la propiedad más determinante de una estrella: en conjunto con la distribución inicial de elements químicos, la masa estelar define la evolución y estructura de estos objetos. La única manera de estimar la masa de una estrella es a través del estudio de su trayectoria orbital, pues la Tercera Ley de Kepler establece una estricta relación entre masa estelar y parámetros orbitales. Dado que los sistemas estelares unitarios no describen trayectorias orbitales, la base observacional del catálogo de masas estelares está constituida esencialmente por sistemas binarios y múltiples; de ahí la importancia de estudiar sus órbitas.

El presente estudio aborda el problema de caracterizar órbitas de estrellas binarias desde una perspectiva Bayesiana, aproximando las densidades de probabilidad *a posteriori* de los parámetros orbitales a través de la técnica denominada Markov Chain Monte Carlo. Una selección de 18 estrellas binarias visuales –algunas de ellas sin órbitas publicadas previamente– y 2 binarias espectroscópicas de línea doble observadas por el telescopio SOAR son analizadas con la técnica desarrollada, obteniendo no sólo estimadores de las órbitas respectivas (máximo *a posteriori* en el caso de binarias visuales, valor esperado en el caso de espectroscópicas) sino también una caracterización de la incertidumbre asociada, entregada por la distribución *a posteriori* de parámetros. Mediante un formalismo matemático desarrollado en este trabajo, se logra reducir la dimensionalidad de vector de parámetros de 7 a 3 en el caso de binarias visuales, y de 10 a 7 en el caso de binarias espectroscópicas, aumentando la eficiencia de los algoritmos de estimación. En el caso de binarias espectroscópicas, se evalúa el potencial de la combinación de fuentes de información –astrometría y velocidad radial– para estimar paralajes, concluyendo que el éxito de este enfoque está sujeto a que la señal de velocidad radial y la órbita aparente puedan ser bien caracterizados de manera independiente.

Se presenta, adicionalmente, una estrategia para incorporar observaciones astrométricas parciales –mediciones donde la separación angular o el ángulo de posición están indeterminados– al método de estimación propuesto. La estrategia combina el método de estimación de parámetros orbitales basado en MCMC con el enfoque de imputación múltiple: en vez de ser descartadas en la rutina de estimación, las observaciones incompletas son reemplazadas por conjuntos de valores plausibles. El esquema desarrollado es puesto a prueba a través de su aplicación a datos sintéticos y a datos reales, obteniendo no sólo una estimación de parámetros orbitales, sino también de la localización de las observaciones parciales. Los resultados obtenidos sugieren que la incorporación de conocimiento parcial puede contribuir a disminuir la incertidumbre en la estimación de parámetros; sin embargo, la información parcial también puede ser redundante en algunos casos.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Mass is arguably the most important property of a star, since it determines, to a great extent, its structure and evolution. Specifically, the well-known Vogt-Russell theorem states that, provided that a star is in hydrostatic and thermal equilibrium, its structure and its evolutionary path are uniquely determined by its initial mass content and the distribution of chemical elements throughout its interior [Kippenhahn et al., 1990].

Binary stars, defined as gravitationally bound pairs of stars, are probably the main source of data on stellar masses, since Kepler's Third Law establishes a strict relation between orbital parameters and the system's total mass:

$$\frac{a^3}{P^2} = \frac{G}{4\pi^2}(m_1 + m_2), \tag{1.1}$$

where P, a and G are the orbital period, orbit's semi-major axis[1] and the gravitational constant, respectively. Individual masses are denoted m_1 (primary star) and m_2 (companion star). Studying the motion of gravitationally bound stars is the only direct method to calculate stellar masses –that is what makes so challenging to estimate the mass of single stars. Multiple star systems can also be a source of stellar masses, however, their analysis is somewhat more complicated, since Kepler's Third Law does not apply when the number of interacting bodies exceeds two –analysis of multiple star systems must rely on numerical methods rather than in an analytic expression. The interest in characterizing the orbit of stellar systems stems, then, from the relevance of building a catalog of stellar masses through which the current theories of stellar evolution, formation and structure can be tested and improved. For example, the discovery of one of the most fundamental relationships depicting the dependency of the star's properties, the mass-luminosity relation[2] (MLR), relied on the availability of estimates of the stellar masses, and even today the refinement of this relation requires the addition of new mass-luminosity points.

Theoretically, few observations may be enough to determine the orbital parameters of a double star. Docobo's analytic method requires as few as three observations of the apparent orbit [Docobo, 1985]; however, those must be high quality observations in well-chosen points of the orbit. When data is obtained from real sources, astronomers have to deal with a range of problems: noisy observations, diversity of sources, poor orbit coverage, incomplete measurements –for example, position observations having one component missing, in either (X, Y) or (ρ, θ) representation. As most algorithms require a full set of measurements as an input, these partial observations are often discarded. Aside from observational difficulties, a common issue when analyzing star motion data is poor characterization of parameter uncertainty: methods –specially optimization-oriented ones– tend to give a single solution of the target parameter set, but fail to offer a measure of confidence about those values.

As a means to cope with the aforementioned difficulties, this work addresses the task of estimating the orbital parameters and characterizing their uncertainty from a Bayesian point of view. Bayesian techniques allow for the incorporation of empirical, theoretical and statistical knowledge (represented, in this case, by the observations, the dynamical model of binary stars and the prior knowledge of the orbital parameters) and, most importantly, they provide naturally what optimization methods do not: uncertainty characterization in the form of a probability density function. The following sections detail the hypotheses, objectives and structure of this work.

[1] Half the longest diameter of an ellipse. This work assumes elliptical paths for all the objects under study. However, some gravitationally bound stars follow hyperbolic trajectories –gravitational interaction between components of that kind of systems will eventually disappear.

[2] First discovered empirically in the early 20th century, and later on explained by the theory of stellar structure [Eddington, 1924].

1.1 Hypotheses

This work focuses on the study of binary stars through astrometric and spectroscopic measurements, and aims at testing the following hypotheses:

1. Markov Chain Monte Carlo is a means to characterize the posterior probability density function of orbital parameters in two different observational settings: (i) data sets with astrometric measurements of relative position between a primary and a companion star[3]; (ii) data sets containing both astrometric observations of relative position and radial velocity measurements [4].

2. The sample-based approximation of the probability density functions provided by Markov Chain Monte Carlo can be used to calculate estimates such as expected value and maximum a posteriori, and those estimates lead to physically feasible orbits of the stars under study.

3. The availability of both astrometric and spectroscopic observations may allow for the estimation of hypothesis-free parallaxes of the stellar system being analyzed.

4. Partial observations can be incorporated into the analysis through a multiple imputation scheme, and may help decrease uncertainty about the parameters.

1.2 Objectives

1.2.1 Main objective:

Formulating a framework for the Bayesian analysis of data of visual and spectroscopic binary stars. This framework must be capable of generating sample-based approximations of the probability density function of orbital parameters as an output, and incorporating partial measurements in the inference process. The approximated PDFs must be used to propose estimates of those parameters, provide uncertainty measures of them, and calculate derived quantities such as stellar masses.

1.2.2 Specific objectives

The main objective can be separated into the following specific objectives:

- Implement Keplerian model to calculate relative position and radial velocity values according to a set of orbital parameters and a specific instant of time.
- Formulate analytic expressions of non-informative prior distributions and likelihood functions for the two settings described in Section 1.1, obtaining a expression for the posterior probability of the parameters.
- Design and implement a Markov Chain Monte Carlo routine to sample the posterior distribution formulated previously.
- Develop a multiple imputation-based scheme to incorporate partial observations into the Bayesian routine.
- Analyze a selection of binary stars observed by the SOAR telescope with the developed tools, aiming at checking the hypotheses presented in Section 1.1.

1.3 Structure

The remainder of this document consists of five chapters, which are described as follows:

Chapter 2 presents the theoretical subjects necessary for the understanding of this work: Section 2.1 describes the Keplerian model of motion of celestial bodies; Section 2.2 provides an introduction to the concept of Bayesian inference and a review of the sampling techniques used for Monte Carlo integration; finally, Section 2.3 explains the theory of Multiple Imputation used to cope with incomplete data.

[3] These objects are referred to as "visual binaries" in this document

[4] In this work, these objects are referred to as "spectroscopic binaries" for the sake of simplicity, even if astrometric/visual data is also available for them

Chapter 3 is a survey of the different approaches from which the problem of orbital fitting has been addressed. Literature on both stars and exoplanets is reviewed, focusing on recent publications that adopt a Bayesian approach.

Chapter 4 details the methodology developed to estimate orbits of binary stars and their uncertainty. The techniques developed are applied on a selection of 18visual binaries and 2 spectroscopic binaries observed by the SOAR telescope. Results are presented and discussed.

Chapter 5 extends the methodology presented in Chapter 4 with Multiple Imputation, allowing for the incorporation of partial data into the analysis. Results based on both synthetic and real data are presented, discussing the effects of including incomplete measurements. In particular, one of the objects analyzed in the previous chapter –visual binary **HU 177**, for which a partial measurement is available– is studied.

Chapter 6 contains the conclusions of this thesis and gives an insight on prospective research lines stemming from this work.

The last pages of this document contain the bibliography and a series of appendices.

Chapter 2

Theoretical framework

2.1 Binary star dynamics

Under some basic assumptions, the dynamics of a binary stellar system is reduced to the two-body problem, whose solution is the well-known Keplerian orbit. Keplerian orbits describe the motion of orbiting celestial bodies in terms of an ellipse, parabola or hyperbola. The specific geometry of the orbit depends on the underlying physical properties (e.g., energy) of the system under study. In this work, we are interested in systems with periodic elliptical orbits.

An ellipse is fully described by its eccentricity, e, and its semi-major axis, a. It can be demonstrated that both the individual bodies of the binary system and the relative position vector follow elliptical paths with the same eccentricity and period (P), but different phase angles (companion star is always 180° ahead of primary star). With m_1, m_2 denoting the masses of the primary and secondary star, individual semi-major axes comply the following relations:

$$a = a_1 + a_2 \tag{2.1}$$

$$a_1 \cdot m_1 = a_2 \cdot m_2 \tag{2.2}$$

where a_1, a_2 and a are the semi-major axis of the primary star, the companion star and the relative orbit, respectively. Thus, determination of the parameters of the relative orbit between the primary and companion star is enough to calculate the total mass of the system.

For a given epoch τ, the value of the relative position between stars is computed as follows. Let be T the epoch when separation between primary and companion star reaches its minimum value. This value is known as *time of periastron passage*. Then, the expression known as Kepler's equation can be written:

$$M = 2\pi(\tau - T)/P = E - e\sin E. \tag{2.3}$$

where terms M and E are called *mean anomaly* and *eccentric anomaly*, respectively[1]. Equation 2.3 does not have analytic solution, and must be solved through numerical methods (in this work we use Newton-Raphson algorithm). Once $E(\tau)$ is obtained, the term known as *true anomaly* is directly determined by:

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}. \tag{2.4}$$

True anomaly ν corresponds to the angle between the main focus of the ellipse and the companion star, provided that the periastron is aligned with the X axis and the primary star occupies the main focus of the ellipse (Figure 2.1). Evaluating $\nu(\tau)$ in the following expression:

$$r(\nu) = \frac{a(1-e^2)}{1+e\cos\nu}. \tag{2.5}$$

[1] The geometrical meaning of this terms, although important in the context of Astronomy or Physics, is not relevant in this article, and therefore won't be explained in detail.

yields the position (r, ν) (polar coordinates) at a given instant τ.

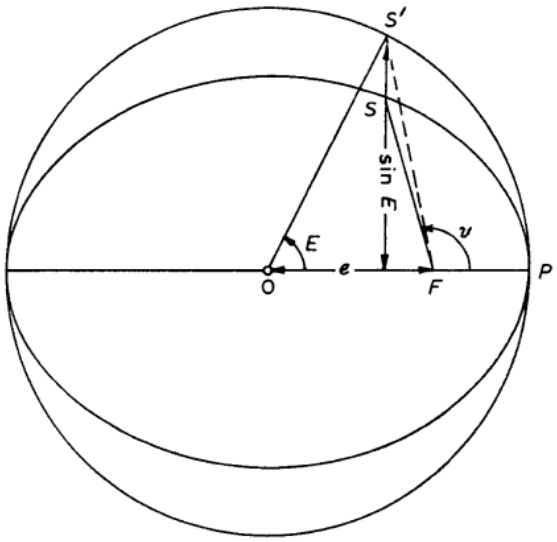

Figure 2.1 Diagram with mean and eccentric anomaly

2.1.1 Observational aspects

In the previous subsection, the reference system has been chosen so that the plane of motion of the binary system (orbital plane) coincides with the XY plane. However, the orbit of a real system may have an arbitrary orientation in the space. What telescopes show is the projection of the real orbit in a reference plane (plane of the sky), known as the *apparent orbit*. In order to explain the mathematical relation between real and apparent orbit, the following quantities must be defined:

- i or *Orbital inclination*: The angle between the reference plane and the orbital plane.
- ω or *Argument of the periapsis*: The angle from the body's ascending node to its periapsis, measured in the direction of motion.
- Ω or *Longitude of the ascending node*: The angle from a reference direction (usually the north celestial pole), called the origin of longitude, to the direction of the ascending node, measured in the reference plane.

Let $\mathcal{O} = \{T, P, e, a, \omega, \Omega, i\}$ be an arbitrary orbital configuration. Then, the apparent orbit is computed as follows:

- Calculate the Thiele-Innes constants:

$$\begin{aligned} A &= a(\cos\omega\ \cos\Omega - \sin\omega\ \sin\Omega\ \cos i) \\ B &= a(\cos\omega\ \sin\Omega + \sin\omega\ \cos\Omega\ \cos i) \\ F &= a(-\sin\omega\ \cos\Omega - \cos\omega\ \sin\Omega\ \cos i) \\ G &= a(-\sin\omega\ \sin\Omega + \cos\omega\ \cos\Omega\ \cos i). \end{aligned} \tag{2.6}$$

- For any given epoch τ, determine the eccentric anomaly E. With this value, calculate the auxiliary variables (x, y).

$$\begin{aligned} x(E) &= \cos E - e \\ y(E) &= \sqrt{1-e^2}\sin E. \end{aligned} \tag{2.7}$$

- Finally, obtain the position in the plane of the sky by evaluating the following expressions:

$$\begin{aligned} X &= Bx + Gy \\ Y &= Ax + Fy \end{aligned} \tag{2.8}$$

  which define a point in the apparent orbit.

Thiele-Innes constants provide not only a straightforward way to calculate the apparent from a set of orbital parameters, but also an alternative mathematical characterization of a binary system. Indeed, parameters $\{a, \omega, \Omega, i\}$ can be substituted by $\{A, B, F, G\}$ since they are mutually mapped by the following expressions:

$$\begin{aligned} tan(\omega + \Omega) &= \frac{B - F}{A + G} \\ tan(\omega - \Omega) &= \frac{-B - F}{A - G} \\ a^2(1 + \cos^2 i) &= A^2 + B^2 + F^2 + G^2 \\ a^2 \cos^2 i &= AG - BF \end{aligned} \tag{2.9}$$

However, this representation is not free of ambiguity –given a value of a, different angles ω, Ω, i may yield the same Thiele-Innes constants. In spite of this, Thiele-Innes representation is widely used by astronomers, since some algorithms can take advantage of the linear dependence of the constants with respect to the values of (x, y).

2.1.2 Spectroscopic data

Astrometric observations of binary or multiple stellar systems can potentially provide enough information to estimate the parameters of the relative orbit[2]. If parallax ϖ is known, semi-major axis a can be converted from arcseconds to Astronomical units, thus allowing for the calculation of total mass $M = m_1 + m_2$ by means of Equation 1.1. The determination of individual masses, however, is not feasible if a set of positions in the plane of the sky is the only source of information available. In order to obtain individual masses, analysis must incorporate spectroscopic data. Nonetheless, spectroscopic data must not be regarded merely as a complement of astrometric measurements, since it is, in fact, the only source of knowledge for certain stellar systems (just a fraction of the binary stars known have been resolved visually). A brief presentation of what spectroscopy is and how it is applied to the study of binary stars is provided as follows.

In general terms, spectroscopy can be defined as the branch of physics that studies the spectra produced when matter interacts with or emits electromagnetic radiation. In astronomy, spectroscopic techniques can be used to derive a number of properties of celestial objects (galaxies, stars, etc.), such as chemical composition, temperature, luminosity, among others. Although many of the properties mentioned previously are obtained from certain permanent features of the spectrum, the study of temporal variations in spectra is also of great importance, since it allows the measurement of relative motion between astronomical objects. This technique relies on Doppler effect and is called Doppler spectroscopy. Basically, Doppler effect is the frequency variation of a perceived signal when there exists relative motion between the observer and the source.

As illustrated in Figure 2.2[3], when the wave source is moving away from the observer, the whole spectrum shifts to lower frequencies (red-shift); when the source moves towards the observer, the spectrum shifts to higher frequencies (blue-shift). Doppler spectroscopy makes use of this effect to calculate radial velocity, that is, the component of source's velocity that points in direction of the radius connecting the observed object and the observer. The *size* of the shift determines the magnitude of radial velocity, whereas the sign is determined by whether the spectrum is blue or red-shifted. Although mostly applied in exo-planet detection, this technique is equally appropriate to measure radial velocity of binary stars.

The rest of this subsection presents the equations of radial velocity equation used in this work. In Figure 2.3, companion star occupies a point D of the orbit, having a separation r with respect to the primary star. Let z be the radial component of the companion star's position, which is perpendicular to the plane of the sky (where the apparent orbit lies). Segment AB is the projection of the real position on the line of nodes (which contains points A, B and F); thus, angle CBD equals the orbit's inclination i. Since true anomaly ν is measured from

[2] As long as there is only one set of parameters that fits the available observations.
[3] ESO Press photo 22e/07, April 2007

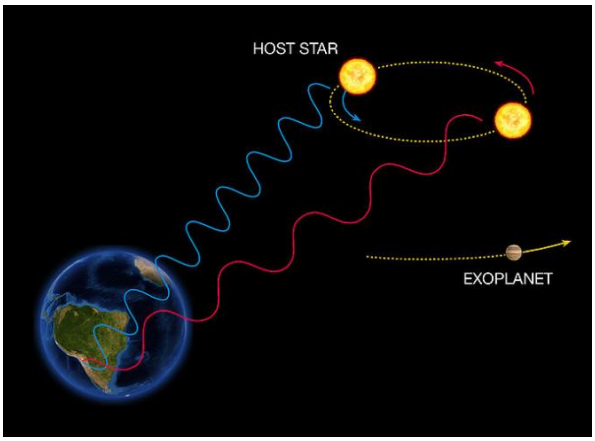

Figure 2.2 Radial velocity variations

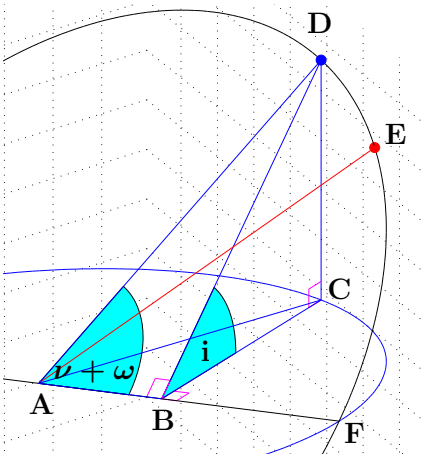

Figure 2.3 Diagram for determination of z

the periastron (indicated by a red mark), and periastron measured from the line of nodes, points D, A, B define the angle $\omega + \nu$. With this, the following equations apply:

$$\sin(\omega + \nu) = \frac{s}{r}, \tag{2.10}$$

$$\sin i = \frac{z}{s}, \tag{2.11}$$

allowing to isolate variable z as:

$$z = r \sin\,(\omega + \nu)\,\sin i. \tag{2.12}$$

In order to obtain an equation for radial velocity, derivative of variable z must be calculated. A closed-form expression for radial velocity is presented in Equation 2.13 (the detailed procedure to derive this formula is found in Appendix A).

$$V_r = \dot{z} = \frac{2\pi a \sin i}{P\sqrt{1-e^2}} \left[\cos(\omega + \nu) + e\cos\omega\right]. \tag{2.13}$$

The expression above was derived for relative position vector $\vec{r} = \vec{r_2} - \vec{r_1}$. However, similar equations hold for individual bodies m_1 (primary star) and m_2 (companion star):

$$V_{r\ (primary)} = V_0 + \frac{2\pi a_1 \sin i}{P\sqrt{1-e^2}} \left[\cos(\omega + \nu) + e\cos\omega\right], \tag{2.14}$$

$$V_{r\ (companion)} = V_0 - \frac{2\pi a_2 \sin i}{P\sqrt{1-e^2}} \left[\cos(\omega + \nu) + e\cos\omega\right], \tag{2.15}$$

where a_1 and a_2 are the semi-major axes of the individual ellipses described by bodies m_1 and m_2, respectively; and V_0 corresponds to the radial velocity of the system's center of mass. The terms outside the brackets can be identified as the amplitude of the radial velocity and are denoted by K_1 and K_2:

$$K_1 = \frac{2\pi a_1 \sin i}{P\sqrt{1-e^2}},\ K_2 = \frac{2\pi a_2 \sin i}{P\sqrt{1-e^2}}. \tag{2.16}$$

One of the main motivations to include spectroscopic data in the study of binary stars is the information they provide about individual masses. In the following paragraphs, the equations that relate orbital parameters, radial velocity and individual masses are derived.

Semi-axis a can be expressed in terms of the masses and the period P by means of Kepler's Third Law (equation 1.1):

$$a = a_1 + a_2 = \left[G\,\frac{P^2(m_1 + m_2)}{4\pi^2}\right]^{1/3}. \tag{2.17}$$

Furthermore, equations 2.1, 2.2 allow the formulation of a_1 and a_2 in terms of a, m_1 and m_2. Replacing a for the

expression above, we obtain:

$$a_1 = a\frac{m_2}{m_1+m_2} \quad \rightarrow \; a_1 = \left[G\,\frac{P^2}{4\pi^2(m_1+m_2)^2}\right]^{1/3} m_2, \tag{2.18}$$

$$a_2 = a\frac{m_1}{m_1+m_2} \quad \rightarrow \; a_2 = \left[G\,\frac{P^2}{4\pi^2(m_1+m_2)^2}\right]^{1/3} m_1, \tag{2.19}$$

which can be replaced in K_1 and K_2, yielding:

$$K_1 = \quad \left[G\,\frac{2\pi}{P}\frac{1}{(m_1+m_2)^2}\right]^{1/3} \frac{m_2 \sin i}{\sqrt{1-e^2}}, \tag{2.20}$$

$$K_2 = \quad \left[G\,\frac{2\pi}{P}\frac{1}{(m_1+m_2)^2}\right]^{1/3} \frac{m_1 \sin i}{\sqrt{1-e^2}}. \tag{2.21}$$

Note that when $m_1 \gg m_2$, the approximation $m_1 + m_2 \sim m_1$ is a valid assumption and the classic expression for the radial velocity of a star orbited by a planet can be recovered:

$$K_1 = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{m_2 \sin i}{m_1^{2/3}} \frac{1}{\sqrt{1-e^2}}. \tag{2.22}$$

The advantage of expressing the amplitudes as shown in equations 2.20, 2.21, 2.22 is that masses are explicitly included within the formulae, thus allowing for the estimation of their individual values.

2.2 Sampling techniques

2.2.1 Bayesian Inference

Mathematical models of physical phenomena enable us to make predictions: given a complete description of a system –that is, a model and the value of its current state–, it is possible to know its evolution over time. The task of calculating the future state of a system given a model and its current value is often referred to as *simulation* or *forward problem*. On the other hand, the problem of translating measurements into actual knowledge of the system –such as a model proposal or a set of parameters– is called *inverse problem*. This kind of problem is commonly found in science and engineering, being oceanography [Wunsch, 1996], seismology [Lailly, 1983] or EEG/ECG techniques in medicine[4] [Pascual-Marqui, 1999, Rudy, 1987] some relevant examples.

Although forward problems usually have a unique solution –at least in deterministic models–, scientists cannot expect the same for inverse problems –frequently, they are not well-posed in the sense of Hadamard[5]. Thus, multiple solutions may exist for them [Tarantola, 2005]. That is the reason why, when faced with an inverse problem, the analysis must include as much information as possible, as well as take into account the uncertainty associated to both the mathematical model used and the measurements taken.

The idea pointed above is what encourages the use of the Bayesian approach, since it provides a means to incorporate prior information into the analysis. Term *Bayesian* refers to a series of concepts and techniques in Probability and Statistics that are based on Bayes Theorem (Equation 2.23), which reformulates conditional probability between two statements X and Y. In particular, term *Bayesian Inference* designates the kind of statistical inference in which measurements are used to upgrade or infer the probability of certain hypothesis.

$$p(X|Y) = \frac{p(Y|X)\ p(X)}{p(Y)} \tag{2.23}$$

Notation used in Equation 2.23 is not casual: although the theorem holds for any statements X and Y, it is most used when the variable of interest, typically called X, is studied through an observation or *evidence*, usually designated by Y. The main advantage of this theorem is that, while $p(X|Y)$ –posterior probability of X– may not be computable in a direct manner, terms appearing in expression $\frac{p(Y|X)\ p(X)}{p(Y)}$ are more likely to have a known (possibly closed) form. Expression $p(Y|X)$ is referred to as *likelihood*, whereas $p(X)$ is known as *prior probability*. In a wide range of problems, term $p(Y)$, called *marginal probability* or *model evidence*, does not hold the same importance as $p(Y|X)$, $p(X)$, as its value is the same for all the hypotheses being considered, thus not taking any role in determining their relative probabilities. Value of $p(Y)$ must be equal to the normalization factor needed for $p(X|Y)$ in order to satisfy the second probability axiom, $\int p(X|Y)dX = 1$.

2.2.2 Markov Chain Monte Carlo: an introduction

Markov Chain Monte Carlo, usually shortened to MCMC, designates a wide class of sampling techniques that rely on constructing a Markov Chain whose equilibrium distribution is the same as the one we desire to sample from. The chain is designed to explore the domain of the target p.d.f. in such a way that it spends most of the time in areas of high probability [Andrieu et al., 2003]. Since the implementation of the algorithm is essentially independent of the target distribution, MCMC provides a means to efficiently draw samples from distributions with complex analytic formulae and/or multidimensional domain. Furthermore, MCMC does not require complete knowledge of target distribution $p(X)$, but just being able to evaluate $p(X)$ up to a normalizing constant. Thus, it is particularly suitable for Bayesian inference problems, where posteriors have the form:

$$p(X|Y) \propto p(Y|X) \cdot p(X). \tag{2.24}$$

First used to perform statistical mechanics simulations in the context of the Manhattan project in Los Alamos Laboratory, California, the technique later known as MCMC was publicly introduced in [Metropolis and Ulam, 1949], [Metropolis et al., 1953]. Although Monte Carlo integration (see **The Monte Carlo principle** in Section 2.2.3) was not unknown by that time, the sampling scheme presented in the latter work represents a major turn in numerical integration methods. In that paper, authors describe a procedure to compute the equations of state

[4] Electroencephalogram and electrocardiogram, respectively

[5] A detailed definition of this concept can be found in [Kabanikhin, 2008]

of a system of interacting molecules by sampling single-molecule configurations[6] with probability $\exp(-E/kT)$ (the Boltzmann factor) and weighting them evenly, rather than choosing them randomly and weighting them by the Boltzmann factor, as would be done in traditional Monte Carlo. Thus, most computer time is dedicated to highly probable configurations, instead of wasting resources in sampling states with little to none weight in the integral. The way in which that sampling regime is achieved, as well as other theoretical aspects of this technique, is explained in Section 2.2.3.

After the pioneering publications of Metropolis around 1950, several works on Monte Carlo simulation were published in the physics literature; however, the potential of this approach was not fully exploited in the statistics community until several decades later. [Hastings, 1970] generalized the Metropolis method to a larger family of algorithms and formulated the Metropolis-Hastings algorithm, arguably the basic building block on which a large part of modern inference techniques rely on. In the 1980s and early 1990s, a bunch of milestone papers were published as well: in a work aimed at the computer vision community, [Geman and Geman, 1984] presented what later would be known as Gibbs sampler; [Tanner and Wong, 1987] proposed data augmentation scheme for Monte Carlo; [Gelfand and Smith, 1990] introduced Markov Chain Monte Carlo into the Bayesian statistics community; [Gelman and Rubin, 1992] addressed the issue of convergence of the MCMC algorithm, introducing the idea of running multiple chains and proposing diagnostics indicators that are used until today. Over the last decades, MCMC has been established as a key tool within the Bayesian community, allowing to calculate posteriors, marginals and integrals that would be otherwise intractable.

The remainder of this section presents the essential theory of the statistical tools used in this work: Subsection 2.2.3 presents the foundations of MCMC and some of its most used formulations; Subsection 2.3 introduces multiple imputation theory.

2.2.3 MCMC theory

2.2.3.1 The Monte Carlo principle

Monte Carlo designates a wide class or algorithms based on the idea of using random sampling to obtain numerical solutions. In Monte Carlo simulations, one draws samples from a target density $\pi(\cdot)$ and uses them to approximate that density as a discrete collection of points:

$$\hat{\pi}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^{(i)}}(x). \tag{2.25}$$

Thus, integrals that depend on density $\pi(x)$ (e.g., the expected value of any function $f(\cdot)$ on the domain of x) can be approximated as the discrete sum of N terms:

$$\hat{I}(f) = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)}), \tag{2.26}$$

which converges almost surely to the value of the integral $I(f) = \int f(x)\pi(x)dx$ as $N \to \infty$ [Andrieu et al., 2003]. If variance of $f(x)$ meets the condition $\sigma_f^2 = \mathbb{E}_\pi(f^2(x)) - I^2(f) < \infty$, then the variance of the approximation $\hat{I}(f)$ satisfies that $var(\hat{I}(f)) = \sigma_f^2/N$. As a consequence of central limit theorem, the following statement holds true:

$$\sqrt{N}(\hat{I}(f) - I(f)) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2), \tag{2.27}$$

that is, the error term $\sqrt{N}(\hat{I}(f) - I(f))$ converges in distribution to a Gaussian. The importance of this statement is that, in addition to guaranteeing convergence, it drops some practical advice: if one desires to increase the precision of the estimate $\hat{I}(f)$ by a factor of n, then sample size must increase as n^2.

[6] In classical statistical mechanics, a system's property Q at equilibrium is calculated as an integral over the possible individual configurations of its constituent particles:

$$Q = \int Q(\vec{X}) \cdot p(\vec{X}) d\vec{X},$$

where $\vec{X}$ is a vector describing the microstate of a single-particle (e.g., geometrical configuration) and $p(\vec{X})$ is the Boltzmann probability, defined as $p(\vec{X}) = \exp(-E(\vec{X})/kT) / \int \exp(-E(\vec{X}')/kT) d\vec{X}'$. In a multiple particles system, Boltzmann factor represents the fraction of particles found in the configuration $\vec{X}$.

2.2.3.2 Preceding ideas: rejection sampling, importance sampling

Aside from The Monte Carlo principle in itself, there are other concepts that have somehow influenced the conception of Markov Chain Monte Carlo. The techniques presented next, introduced before the seminal work of [Metropolis et al., 1953], also deal with the problem of sampling arbitrary distributions, and may be appropriate to facilitate the understanding of MCMC:

- **Rejection sampling**: Arguably the simplest among the methods reviewed in this work, rejection sampling was introduced in [Von Neumann, 1951]. If target distribution $\pi(\cdot)$ is known up to a proportionality constant, it can be sampled by using a proposal distribution $q(x)$ whose support contains the support of π. After finding a constant M that meets the condition $\pi(x) \leq Mq(x)\ \forall x,\ M < \infty$, the sampling scheme can be carried out in the manner shown in Algorithm 1. The samples thus generated can be proven to distribute as $\pi(\cdot)$ [Robert and Casella, 2004]. Although appealing due to its simplicity, rejection sampling suffers from serious drawbacks: as the dimension of x increases, it is increasingly difficult to find a M constant that meets $M > \pi(x)/q(x)\ \forall x$. Moreover, if the constant M is chosen to be too large, samples are rejected with excessively high probability, leading to poor sampling (if the algorithm is programmed for a fixed number of iterations, regardless of the acceptance or rejection of the samples) or long execution times (if algorithm is programmed to draw fixed number of samples). The notion of a proposal distribution, however, is key to the development of more advanced sampling techniques.

Algorithm 1 Rejection sampling

Rejection Sampling

$\bullet k = 0$
while $k < N_{samples}$ **do**
$\rightarrow x' = x$, with $x \sim q(x)$
$\rightarrow u' = u$, with $u \sim \mathcal{U}(0,1)$.
$\rightarrow r = \dfrac{\pi(x')}{Mq(x')}$
if $u' < r$ **then**
• Accept sample (if not, sample is said to be *rejected*)
$\rightarrow X[k] = x'$
$\rightarrow k = k + 1$
end if
end while

- **Importance sampling** is another sampling scheme proposed before the formulation of MCMC. According to [Geweke, 1989, Rubinstein and Kroese, 2016], the idea of importance sampling can be traced back to 1940s. If $q(x)$ is a proposal distribution whose support includes that of the target distribution, integrals involving $\pi(x)$ can be rewritten as follows:

$$I(f) = \int f(x)w(x)q(x)dx, \tag{2.28}$$

where $w(x)$ is defined as $w(x) = \pi(x)/q(x)$ and is referred to as *importance weight*. Accordingly, if N samples are drawn from distribution $q(x)$, integral $I(f)$ can be approximated as:

$$\hat{I}(f) = \sum_{i=1}^{N} f(x^{(i)})w(x^{(i)}), \tag{2.29}$$

provided that one can evaluate $w(x)$. The estimate $\hat{I}(f)$ can be proven, under weak assumptions, to satisfy almost sure convergence to the integral $I(f)$. Under this approach, estimate $\hat{I}(f)$ can be interpreted as the integral of function $f(x)$ with respect to the discrete approximation of π:

$$\hat{\pi}(x) = \sum_{i=1}^{N} w(x^{(i)})\ \delta_{x^{(i)}}(x). \tag{2.30}$$

It remains unclear, however, how to choose the proposal distribution $q(x)$. A possible criterion is to choose a distribution that minimizes the variance of the estimate $\hat{I}(f)$:

$$var_{q(x)}(f(x)w(x)) = \mathbb{E}_{q(x)}(f^2(x)w^2(x)) - I^2(f). \tag{2.31}$$

Since term $I^2(f)$ does not depend on the proposal distribution, efforts must be focused on $\mathbb{E}_{q(x)}(f^2(x)w^2(x))$. Following Jensen's inequality[7], that term satisfies:

$$\mathbb{E}_{q(x)}(f^2(x)w^2(x)) \geq \Big(\mathbb{E}_{q(x)}(f(x)w(x))\Big)^2, \tag{2.32}$$

which reaches equality by adopting the proposal distribution $q^*(x) = \frac{|f(x)|w(x)}{\int |f(x)|w(x)dx}$. Although the result may not be of straightforward practical usefulness –sampling from $|f(x)|w(x)$ is virtually impossible in any non-trivial case–, it gives an insight about how proposal distributions should be chosen: higher sampling efficiency is achieved when samples concentrate in regions where $|f(x)|p(x)$ (rather than $p(x)$ alone) has high values. That is what justifies the name of *importance sampling*. A telling example of this concept comes from communication networks [Smith et al., 1997]: being the event of interest the bit error rate, authors focus on sampling the region where bit errors occur rather than on the whole domain of events.

2.2.3.3 A discrete analogy of MCMC

The main innovation introduced by MCMC with respect to previous numerical integration schemes is its focus on the sampling in itself rather than in weighting the samples (as in Equation 2.30). Proposed as a solution to a practical problem (namely, how to compute the equation of state of a system of interacting particles), and probably unaware of the theoretical implications of their ideas, authors in [Metropolis et al., 1953] state: "instead of choosing configurations randomly, then weighting them with $\exp(-E/kT)$, we choose configurations with a probability $\exp(-E/kT)$ and weight them evenly."[8]

In order to understand the theoretical basis of MCMC, a discrete analogy may be useful. Let $\mathcal{X}$ be a discrete space of size S: $\mathcal{X} = \{x_1, x_2, \ldots, x_S\}$. A stochastic process $x^{(i)}$ is a Markov Chain if:

$$Pr(x^{(i)}|x^{(i-1)}, \ldots, x^{(1)}) = T(x^{(i)}|x^{(i-1)}), \tag{2.33}$$

that is, the probability of state $x^{(i)}$ is completely determined by the previous state $x^{(i-1)}$. This definition allows for the representation of the stochastic process in terms of a probability vector p and a transition matrix $\mathbf{T}$:

$$p^{(i|i-1)} = \begin{bmatrix} p_{(1,i)} \\ \vdots \\ p_{(k,i)} \\ \vdots \\ p_{(S,i)} \end{bmatrix} = \begin{bmatrix} T_{(1,1)} & \cdots & T_{(1,S)} \\ \vdots & \vdots & \vdots \\ T_{(k,1)} & \cdots & T_{(k,S)} \\ \vdots & \vdots & \vdots \\ T_{(S,1)} & \cdots & T_{(S,S)} \end{bmatrix} \begin{bmatrix} p_{(1,i-1)} \\ \vdots \\ p_{(k,i-1)} \\ \vdots \\ p_{(S,i-1)} \end{bmatrix}, \tag{2.34}$$

where $p^{(i|i-1)}$ is a vector containing the probability mass function (p.m.f. hereinafter) of state $x^{(i)}$ given the p.m.f. of the previous state, $x^{(i-1)}$. Thus, terms in $p^{(i|i-1)}$ represent the probability of a particular state: $p_{(k,i)} = Pr(x^{(i)} = x_k)$. Terms in matrix $\mathbf{T}$ represent individual transition probabilities: $T_{(k_1,k_2)} = T(x_{k_1}|x_{k_2})$. In this formulation, the sum of terms in the same column must equal 1.

A chain is said to be *homogeneous* if transition probabilities remain invariant throughout the stochastic process (that is, $\mathbf{T}$ is independent of i); thus, $x^{(i)}$ depends only on a fixed transition matrix and previous state $x^{(i-1)}$. Some chains stabilize at a fixed probability vector after a number of transition steps: let μ be any probability vector for the initial state; then, after n iterations:

$$p^{(n)} = \mathbf{T}^n \mu = p^*, \tag{2.35}$$

[7] In the context of probability theory, $\phi(\mathbb{E}(x)) \leq \mathbb{E}(\phi(x))$, being $\phi(\cdot)$ any convex function.

[8] Physical meaning of term $\exp(-E/kT)$ was explained in the Subsection 2.2.2.

and the probability vector remains at p^* even after a new transition step is applied:

$$p^{(n+1)} = \mathbf{T}p^* = p^*. \tag{2.36}$$

This stability behaviour plays a fundamental role in MCMC simulation, particularly in the choice of transition probabilities: if properly designed, the chain will converge to the target distribution $\pi(\cdot)$ for any starting point. In discrete Markov processes, the chain converges to a stable distribution p^* as long as matrix $\mathbf{T}$ satisfies the following conditions:

- **Irreducibility**: the probability of visiting all other states is always positive, regardless of the current state. From an algebraic point of view, this is equivalent to state that matrix $\mathbf{T}$ cannot be decomposed as a set of separate smaller matrices. From a graph-theoretic standpoint, irreducibility is equivalent to the connectivity property of the transition graph.
- **Aperiodicity**: the property of not getting trapped in loops.

If these properties hold, then p^* is the invariant distribution. If $\mathbf{T}$ does not meet the irreducibility condition, the chain may still converge to a probability vector, but that vector would depend on the initial probabilities μ. If aperiodicity does not apply, the chain may get trapped in cycles, so that it does not converge.

One way to ensure that the properties mentioned above apply is to study the reversibility condition. A chain is *reversible* if it satisfies the *detailed balance* condition:

$$Pr(x_{k_1})T(x_{k_2}|x_{k_1}) = Pr(x_{k_2})T(x_{k_1}|x_{k_2}). \tag{2.37}$$

Summing over x_{k_2} yields:

$$Pr(x_{k_1}) = \sum_{x_{k_2}} Pr(x_{k_2})T(x_{k_1}|x_{k_2}). \tag{2.38}$$

Thus, if Pr is found to meet the detailed balance condition, then Pr is the stationary distribution p^*. An interpretation of detailed balance is that, regardless of the other states, the probability of moving from x_{k_1} to x_{k_2} is the same that that of moving from x_{k_2} to x_{k_2}. Reversibility guarantees stationarity (i.e., it is a sufficient condition), but stationarity does not require detailed balance. Reversibility is stronger than stationarity in the sense that it implies that a record of the successive transitions looks the same when viewed forwards or backwards.

From equations 2.36, 2.38 it can be noticed that vector p^* is the right eigenvector of matrix $\mathbf{T}$ with corresponding eigenvalue 1. By means of the Perron-Frobenius theorem from linear algebra, it can be proven that all the remaining eigenvalues are smaller in absolute value than 1. Second largest eigenvalue determines the rate of convergence of the chain. In MCMC samplers, a major part of the efforts are focused on designing chains that converge as quickly as possible.

Although discrete Markov chains are of little use in the context of sampling techniques, the concepts presented to this point give an insight of how MCMC works. The leap to the continuous is made through the definition of the transition probabilities in those spaces: instead of a discrete collection of probabilities of moving from one state to another $T(x_{k_1}|x_{k_2})$, an integral kernel K representing the conditional density of $x^{(i+1)}$ given $x^{(i)}$ is used:

$$\int p(x^{(i)})K(x^{(i+1)}|x^{(i)})dx^{(i)} = p(x^{(i+1)}), \tag{2.39}$$

whereas the probability vector is replaced by a probability density function.

2.2.3.4 Metropolis/Metropolis-Hastings algorithm

Proposed in [Metropolis et al., 1953] as a practical solution, MCMC algorithm took time to be formalized –at first, concepts such as stationarity, proposal distribution, transition kernel were implied rather than explicitly stated. In [Hastings, 1970], Metropolis work is analyzed mathematically and generalized to a wider family of algorithms, the so-called Metropolis-Hastings (MH) method. A major part of the sampling could be considered particular cases or extensions of the MH algorithm, hence the importance of studying it in detail.

Given a target distribution $\pi(\cdot)$ (known up to a constant), MH performs sampling by drawing a new value x' from a proposal distribution $q(x'|x)$ (which depends on a previous sample x), and accepting it with probability

$\mathcal{A} = \min\{1, \frac{\pi(x') \cdot q(x^{(i-1)}|x')}{\pi(x^{(i-1)}) \cdot q(x'|x^{(i-1)})}\}$. If x' is rejected, the chain remains at x. This procedure is summarized in Algorithm 2.

Algorithm 2 Metropolis-Hastings algorithm

Metropolis-Hastings sampler

- Initialization

$\rightarrow x^{(0)} = x$, with $x \sim PriorDist$

for $i = 1, \ldots, N_{steps}$ **do**

$\rightarrow x' = x$, with $x \sim q(x|x^{(i-1)})$

$\rightarrow u' = u$, with $u \sim \mathcal{U}(0, 1)$.

$\rightarrow \mathcal{A} = \min\{1, \frac{\pi(x') \cdot q(x^{(i-1)}|x')}{\pi(x^{(i-1)}) \cdot q(x'|x^{(i-1)})}\}$

if $u' < \mathcal{A}$ **then**

- Accept sample

$\rightarrow x^{(i)} = x'$

else

$\rightarrow x^{(i)} = x^{(i-1)}$

end if

end for

Analogous to the discrete approach presented previously, in MCMC the chain runs following a transition probability (here defined by an integral kernel K) that achieves a stationary regime that generate samples as if they were drawn directly from π. That is, instead of the stationary probability vector p^*, MCMC has π as its stationary distribution. As long as the initial samples of the chain are representative of target distribution, the routine is guaranteed to mimic the action of sampling from π. However, most times the chain takes a number of iterations before reaching an stationary regime; that set of samples is called *burn-in period* and is often discarded.

The kernel that represents the transition probability defined by Algorithm 2 is:

$$K_{MH}(x^{(i+1)}|x^{(i)}) = q(x^{(i+1)}|x^{(i)})\mathcal{A}(x^{(i)}, x^{(i+1)}) + \delta_{x^{(i)}}(x^{(i)})r(x^{(i)}), \tag{2.40}$$

from which it can be noticed that transition probability does not depend solely on proposal distribution $q(\cdot)$, but also on the acceptance and rejection probabilities. Second term $\delta_{x^{(i)}}(x^{(i)})r(x^{(i)})$ represents the probability of remaining at the previous sample $x^{(i)}$:

$$r(x^{(i)}) = \int_{\mathcal{X}} q(x'|x^{(i)})(1 - \mathcal{A}(x^{(i)}, x'))dx'. \tag{2.41}$$

By construction, the MH kernel defined in Equation 2.40 meets the detailed balance condition:

$$\pi(x^{(i)})K_{MH}(x^{(i+1)}|x^{(i)}) = \pi(x^{(i+1)})K_{MH}(x^{(i)}|x^{(i+1)}), \tag{2.42}$$

and this proves that the chain defined by K_{MH} admits π as a stationary distribution. MH algorithm satisfies irreducibility (since q can be chosen to contain the support of π) and aperiodicity (there is always a probability of rejection, therefore the chain cannot get trapped in loops). Under this conditions, asymptotic convergence is guaranteed [Tierney, 1994].

MH encompasses a wide range of algorithms, since depending on the choice of proposal distribution q one can recover algorithms that were proposed before the publication of [Hastings, 1970]. When the next sample does not depend on the current state, $q(x'|x) = q(x')$ and then the acceptance probability takes the form:

$$\mathcal{A}(x^{(i)}, x') = \min\left\{1, \frac{\pi(x') \cdot q(x^{(i)})}{\pi(x^{(i)}) \cdot q(x')}\right\} = \min\left\{1, \frac{w(x')}{w(x^{(i)})}\right\}, \tag{2.43}$$

which is an *independent sampler* close to the idea of important sampling (but with correlated samples). If proposal distribution is symmetric (i.e., $q(x'|x) = q(x|x')$), then $q(\cdot)$ terms in $\mathcal{A}$ cancel out, recovering the Metropolis algorithm:

$$\mathcal{A}(x^{(i)}, x') = \min\left\{1, \frac{\pi(x')}{\pi(x^{(i)})}\right\}. \tag{2.44}$$

In summary, the key of MH –and MCMC in general– is that kernels are constructed in such a way that the chains they induce always lead to stationary distributions that correspond with $\pi(\cdot)$. However, in practice the choice of proposal distributions affects to a great extent the chances of convergence of the chain –both the distribution q in itself (for example, a Gaussian) and its parameters (e.g., the value pf σ if q is a Gaussian) must be chosen according to the target distribution.

2.2.3.5 Kernel composition and Gibbs sampler

As long as two different kernels K_1, K_2 admit the same distribution $\pi(\cdot)$ as their stationary distribution, they can be combined to generate new samplers. According to [Tierney, 1994], both the *cycle hybrid kernel* ($K_1 \cdot K_2$) and the *mixture hybrid kernel* ($\alpha K_1 + (1-\alpha)K_2$) are transition kernels with invariant distribution π. Algorithmically, cycle hybrid kernel corresponds to the sequential application of two or more kernels, whereas mixture hybrid kernel is equivalent to applying K_1 with probability α and K_2 with probability $1-\alpha$ (this scheme is extensible to more than two kernels).

Kernel composition can be used to explore the feature space in a sensible way, depending on the particular aims of the problem. For example, a kernel K_1 may provide a means to explore the domain of $\pi(\cdot)$ globally, whereas K_2 has a local scope intended to discover finer details (e.g., narrow peaks).

An important case of cycle hybrid kernel is the well-known Gibbs sampler [Geman and Geman, 1984], which relies on the sequential sampling of conditional probabilities. On the long run, such scheme is equivalent to drawing samples from the joint target distribution. Provided that samples are defined as $\vec{x} = [x_1, \ldots, x_d]$, Gibbs algorithm operates as follows:

Algorithm 3 Gibbs sampler

- Initialize $x^{(1)}$

for $i = 2, \ldots, N_{steps}$ **do**
  for $j = 1, \ldots, d$ **do**
    $\rightarrow x_j^{(i)} \sim \pi(x_j | x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \ldots, x_d^{(i-1)})$
  end for
end for

Formally, the proposal distribution of Gibbs sampler is defined for each component $j \in \{1, \ldots, d\}$:

$$q(x'|x^{(i)}) = \begin{cases} \pi(x'_j | x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \ldots, x_d^{(i-1)}) & \text{If components other than } j \text{ are equal in } x^{(i)} \text{ and } x' \\ 0 & \text{Otherwise} \end{cases} \tag{2.45}$$

Thus, acceptance probability is always 1:

$$\begin{aligned} \mathcal{A}(x^{(i)}, x') &= \min\left\{1, \frac{\pi(x') \cdot q(x^{(i)}|x')}{\pi(x^{(i)}) \cdot q(x'|x^{(i)})}\right\} \\ &= \min\left\{1, \frac{\pi(x') \cdot \pi(x_j^{(i)} | x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i)}, \ldots, x_d^{(i)})}{\pi(x^{(i)}) \cdot \pi(x_j | x'_1, \ldots, x'_{j-1}, x'_{j+1}, \ldots, x'_d)}\right\} \\ &= \min\left\{1, \frac{\pi(x'_j | x'_1, \ldots, x'_{j-1}, x'_{j+1}, \ldots, x'_d)}{\pi(x_j^{(i)} | x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i)}, \ldots, x_d^{(i)})}\right\} = 1 \quad (x' \text{ and } x^{(i)} \text{ are equal except for } x_j), \end{aligned}$$

where the rewriting of $\pi(x)$ in terms of conditional probability was used from step 2 to step 3: $\pi(x) = \pi(x_j | x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d) \cdot \pi(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)$. Although explained in terms of one-dimensional components, Gibbs sampling admits component grouping as well. If conveniently blocked, sub-components may have conditional probabilities with a standard form (Gaussian, Poisson, Gamma, etc.), thus being easier to sample than the target joint probability. If conditional probabilities do not follow a standard form, one can draw "conditional" samples by means of other sampling schemes. In particular, being Gibbs algorithm a special case of Metropolis-Hastings (namely, that with Equation 2.45 as proposal distribution), MH steps can be included within the Gibbs sampler. The resulting procedure is summarized in Algorithm 4, where an auxiliary proposal distribution q'_j is defined: let $x_{-j}^{(i)}$ be $[x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i)}, \ldots, x_d^{(i)}]$ (all components other than j), then $q'_j(x|x^{(i)})$ is a distribution

that induces leaps only on the j-th component (i.e., x_{-j} remaining equal to $x_{-j}^{(i)}$ and x_j being the result of a random variation on $x_j^{(i)}$). This scheme is appropriate to sample multivariate distributions of large dimension, since sequential sampling of parameters prevents the algorithm from falling in zones of near-zero probability.

Algorithm 4 MH-within-Gibbs

- Initialize $x^{(1)}$.

for $i = 2, \ldots, N_{steps}$ **do**

- Retrieve value from the last iteration:

$\rightarrow x^{(i)} = x^{(i-1)}$

for $j = 1, \ldots, d$ **do**

$\rightarrow x' = x$, with $x \sim q'_j(x|x^{(i)})$

$\rightarrow u' = u$, with $u \sim \mathcal{U}(0,1)$.

$\rightarrow \mathcal{A} = \min\left\{1, \frac{f(x') \cdot q'_j(x^{(i)}|x')}{f(x^{(i)}) \cdot q'_j(x'|x^{(i)})}\right\}$

if $u' < \mathcal{A}$ **then**

- Accept sample (if not, $x^{(i)}$ remains at the current value):

$\rightarrow x^{(i)} = x'$

end if

end for

end for

Up to this point, the basic MCMC techniques have been presented. Other important MCMC and MCMC-based methods, such as Slice sampler [Damien et al., 1999], Hybrid Monte Carlo [Duane et al., 1987], Reversible jump MCMC [Green, 1995] have been left out from this review. However, Chapter 3 examines some of the MCMC techniques that have been used in exoplanet research and binary/multiple star analysis.

2.3 Multiple imputation theory

The presence of incomplete or missing data occurs in a wide range of statistical situations. Since standard inference methods are conceived to be used with complete data sets, missing values imply a challenge for data analysts. Several approaches can be taken to deal with partial or missing data. The most straightforward strategy is discarding data entries with partial measurements and apply a standard technique on the subset of complete measurements. However, the validity and efficiency of complete-data based methods cannot be guaranteed when data are incomplete [Rubin, 1976].

A number of techniques specifically aimed at addressing the problem of missing or partial measurements have been proposed. For example, [Dempster et al., 1977] uses Expectation-Maximization (EM hereinafter) algorithm to calculate maximum likelihood estimate from incomplete data. Another approach is to fill in the blank spaces due to missing data with plausible values[9] –referred to as imputations–, thus generating a set of observations on which complete-data based methods can be applied. Regardless the specific manner in which imputations are generated, methods that proceed like that can be labeled as "single imputation techniques". Both Expectation-Maximization and single imputation techniques have certain drawbacks: EM is not suitable when the object of interest is the likelihood or posterior distribution rather than just a maximizer (and the curvature at that point, at most); on the other hand, single imputation techniques omit the sources of uncertainty associated to the imputation process, which may lead to biased results. These sources of uncertainty are enumerated as follows: One is the uncertainty associated to the modeling of the joint distribution of the response variables Y (observed and unobserved) and missingness indicators R (Equation 2.47); the second is the uncertainty of the imputation model, assuming that values of the observed data and the model parameters are known (associated with the term later identified as *conditional predictive* distribution); the third is the uncertainty about the model parameters (θ) themselves, i.e., $p(\theta|Y)$. Further details are explained throughout this section.

Introduced in [Rubin, 1987], the approach known as *multiple imputation* is aimed at performing inference from incomplete data while taking into account the uncertainty sources that single imputation techniques ignore. Multiple imputation relies on replacing missing values with not one, but multiple plausible values, thus generating several complete data sets which differ from each other only in the imputed values –entries with complete data remain the same. These data sets are analyzed individually with standard techniques and the final inference is performed by combining the individual results (e.g., by calculating an average).

Certain definitions must be introduced in order to formalize the problem of missing data and the techniques to address it. Let Y be an observation matrix, with each row y_i being a single multivariate observation of dimension p drawn from a probability distribution $p(y|\theta)$ governed by the model parameter vector θ. Components of missingness indicator R are defined as:

$$r_{ij} = \begin{cases} 0 \text{ if } y_{ij} \text{ is observed,} \\ 1 \text{ if } y_{ij} \text{ is missing.} \end{cases} \tag{2.46}$$

Defining $Pr(r_{ij} = 0|y_{ij}) = p_{ij}$ (and $Pr(r_{ij} = 1|y_{ij}) = 1 - p_{ij}$), the value of R is subject to a probability distribution $p(R|\xi, Y)$ which depends on parameters ξ. Thus, by virtue of the chain rule of of probability, joint distribution of Y and R can be expressed as:

$$p(Y, R|\theta, \xi) = p(Y|\theta) \cdot p(R|\xi, Y), \tag{2.47}$$

where term $p(Y|\theta)$ is the conditional distribution of the observations given the model parameters. It has been assumed that $p(Y|\theta, \xi) = p(Y|\theta)$ and $p(R|\xi, \theta, Y) = p(R|\xi, \theta)$. Values within Y can be separated into Y_{mis} (values in Y such that $r_{ij} = 1$) and Y_{obs} (values in Y such that $r_{ij} = 0$). Since conclusions about the target parameters θ must be based on the joint probability model (Equation 2.47), the manner in which the missingness depends on Y must be taken into account when performing inference. [Little and Rubin, 1987] identifies three missingness mechanisms:

- Missing completely at random (MCAR): $p(R|\xi, (Y_{mis}, Y_{obs})) = p(R|\xi)$. For example, in a clinical trial participants would flip a coin to decide whether they fill in a depression survey.
- Missing at random (MAR): $p(R|\xi, (Y_{mis}, Y_{obs})) = p(R|\xi, Y_{obs})$, i.e., occurrence of data loss depends only

[9] The way in which these "plausible values" can be generated ranges from replacing blank spaces with average values, to problem-specific probability models.

on the observed values. Following the example above, male participants could be more likely to refuse to complete the depression survey, regardless their individual levels of depression ("they tend to skip the survey just because they are male").

- Missing not at random (MNAR): $p(R|\xi,(Y_{mis},Y_{obs})) \neq p(R|\xi,Y_{obs})$, that is, occurrence of data loss may depend on unobserved values. In the clinical trial mentioned above, male participants might be more reluctant to complete the survey as their level of depression is higher ("they tend to skip the survey because they are depressed").

Terms θ and ξ are said to be *distinct* if their joint probability can be expressed as a product of independent marginal PDFs [Rubin, 1976]. If that condition holds, and if either MCAR or MAR applies, inferences based on observed-data likelihood function $L(\theta,\xi|Y_{obs},R)$ will be the same as those based on $L(\theta|Y_{obs})$. In those cases, the missingness mechanism is said to be *ignorable*. However, the precision of the inference thus performed is reduced if a large portion of the information is missing. That is what motivates the multiple imputation approach.

Conditional probability of Y_{mis} given Y_{obs} can be derived by integrating over the parameter space of θ, that is:

$$p(Y_{mis}|Y_{obs}) = \int p(Y_{mis}|Y_{obs},\theta)p(\theta|Y_{obs})d\theta. \tag{2.48}$$

Term $p(Y_{mis}|Y_{obs},\theta)$ is identified as *conditional predictive* distribution of Y_{mis} given Y_{obs} and θ. Term $p(Y_{mis}|Y_{obs})$ is identified as the *posterior predictive distribution* of Y_{mis} given Y_{obs}, and must be understood as the condition predictive distribution averaged over the observed-data posterior distribution of θ, $p(\theta|Y_{obs})$. Although $p(Y_{mis}|Y_{obs})$ is seldom found in a closed form, expressions for both $p(Y_{mis}|Y_{obs},\theta)$ and $p(\theta|Y_{obs})$ may be found in certain situations. On the other hand, $p(Y_{mis}|Y_{obs},\theta)$ depends on the imputation model adopted. Examples of standard imputation models existing in the literature are the *predictive model method* [Little and Rubin, 1987] and the *propensity score method* [Lavori et al., 1995], which are only applicable when data follows a monotone missing pattern[10].

Let Q be any quantity of interest to be estimated from the observed data. Then, the core of the multiple imputation approach is the estimation of Q by averaging the completed-data posterior, $p(Q|Y_{obs},Y_{mis})$, over the feasible values of Y_{mis} given Y_{obs} (which are represented by the distribution $p(Y_{mis}|Y_{obs})$ and depend, ultimately, on the chosen imputation model):

$$p(Q|Y_{obs}) = \int p(Q|Y_{obs},Y_{mis})p(Y_{mis}|Y_{obs})dY_{mis}. \tag{2.49}$$

This integral can be approximated as the discrete average of the values of Q obtained from a finite, possibly small, number of data sets filled in with imputations. [Zhang, 2003] presents a discussion on how the number of imputations affects the estimation variance.

2.3.1 MCMC for multiple imputation

When data loss is not governed by a monotone pattern, imputation models such as the *predictive model method* and the *propensity score method* cannot be applied. To imput the missing values in cases with arbitrary missing patterns, more advanced techniques must be used. The data augmentation algorithm introduced in [Tanner and Wong, 1987], which is formally a Markov Chain Monte Carlo method, appears as a useful tool for those cases.

The data augmentation algorithm is motivated by the representation of the desired posterior distribution found in Equation 2.49, where the quantity of interest Q is typically a parameter vector θ. Term $p(\theta|Y_{obs})$ depends, in turn, on $p(Y_{mis}|Y_{obs})$. In [Tanner and Wong, 1987], authors prove, by a fixed point-based argument, that under mild conditions the scheme presented next converges to $p(\theta|Y_{obs})$:

1. Generate $Y_{mis}^{(i,1)},\ldots,Y_{mis}^{(i,m)}$ from the current approximation of the predictive posterior distribution, $p_i(Y_{mis}|Y_{obs})$. This can be accomplished by: (i) drawing a sample θ_i from $p_i(\theta|Y_{obs})$; (ii) sampling m values of Y_{mis} from $p(Y_{mis}|\theta_i,Y_{obs})$.

[10] Monotone missing pattern implies that if certain datum y_{ij} in the data matrix Y is missing, then all subsequent y_{ik} $(k>j)$ are also missing.

2. Update the approximation of the desired posterior, $p_i(\theta|Y_{obs})$, by averaging the results obtained from the completed data sets:

$$p_{i+1}(\theta|Y_{obs}) = \frac{1}{m}\sum_{j=1}^{m} p(\theta|Y_{mis}^{(i,j)}, Y_{obs}) \tag{2.50}$$

3. If a stopping criterion has been reached, stop. If not, go to step 1.

Steps 1 and 2 are referred to as I-step (Imputation step) and P-step (Posterior step), respectively, in analogy to the Expectation and Maximization steps of the EM algorithm. The evaluation of function $p(\theta|Y_{mis}, Y_{obs})$ in step 2 is replaced by a sampling procedure in certain situations (e.g., if $p(\theta|Y_{mis}, Y_{obs})$ is known up to constant, if the distribution is not known in a closed form but it is possible to sample from, etc.).

Authors in [Tanner and Wong, 1987] suggest that even a value of m as small as $m = 1$ leads to a correct approximation, in the sense that the average of the posterior distribution across the iterations –that obtained in the step 2– will converge to $p(\theta|Y_{obs})$. This motivates the MCMC scheme actually used in later publications on missing data [Zhang, 2003] and commercial software implementations [Yuan, 2010]. That scheme is described in Algorithm 5.

Algorithm 5 MCMC for multiple imputation

- Initialize $\theta^{(0)}$.

for $i = 1, \ldots, N_{steps}$ **do**

- Imputation step:

$\rightarrow Y_{mis}^{(i)} = Y'_{mis}$, with $Y'_{mis} \sim p(Y_{mis}|\theta^{(i-1)}, Y_{obs})$

- Posterior step:

$\rightarrow \theta^{(i)} = \theta'$, with $\theta' \sim p(\theta|Y_{mis}^{(i)}, Y_{obs})$

end for

By following the scheme presented above during a sufficiently large number of iterations, one obtains a sequence $\{\theta^{(i)}, Y_{mis}^{(i)}\}_{i=0,\cdots,N_{steps}}$ whose stationary distribution is $p(\theta, Y_{mis}|Y_{obs})$. Marginalizing out Y_{mis} yields the desired posterior distribution $p(\theta|Y_{obs})$, whereas by marginalizing out θ one obtains the posterior predictive distribution $p(Y_{mis}|Y_{obs})$.

Chapter 3

State of the art

This section presents a general review of the techniques that have been used in the astronomical community to address the problem of fitting stellar orbits and radial velocity curves. Although classic techniques such as Thiele-Innes-Van den Bos [Thiele, 1883] have not been excluded from the survey, this review is focused on recent works that adopt a Bayesian approach. Since both apparent orbits and radial velocity curves are manifestations of the same underlying physical phenomenon –the motion of celestial bodies that are gravitationally bound–, the tools suitable for analyzing one kind of data are usually also suitable for the other. Moreover, from a purely kinematic point of view there are no major differences between studying binary/multiple stars and sub-stellar objects such as brown dwarfs with multiple components or planetary systems –all of them behave according to Newton's laws of motion[1]. That is why this section covers certain advances made in exoplanet research and brown dwarfs analysis in addition to the available tools for the study of visual stellar systems.

3.1 Classic orbit fitting techniques

The method proposed in [Thiele, 1883] is the first to provide a solution to the problem of finding the elliptic orbit of a visual binary star given a set of observations of relative position. After further refinements made by other authors, that technique came to be known as Thiele-Innes-Van den Bos method during the 20th century. It relies on the selection of three complete observations of the form (τ, ρ, θ) and an areal constant c (obtained from additional data), which are used to formulate an auxiliary system of equations whose solution is a set of orbital parameters that matches the chosen measurements. A similar approach is taken in [Cid Palacios, 1958], where three complete observations of the form (τ, ρ, θ) and one partial observation of the form (τ, θ) are required as input. In line with the aforementioned works, the technique presented in [Docobo, 1985] also uses three complete observations, but explores the dependency of eccentricity e as a function of an independent auxiliary angular variable V instead of the areal constant c. That method yields different orbital solutions depending on the chosen value of V. This class of methods depends to a great extent on the quality of the selected observations –the output orbit passes through those points–, which does not facilitate the incorporation of observational uncertainty into the model. Furthermore, even though these techniques have a physical basis –all of them are based on the invariability of areal velocity–, their approach is heuristic rather than formal (at least from a statistical point of view).

3.2 Orbital fitting as an optimization problem

The consistent progress in computing power during the last decades led to the popularization of computationally intensive algorithms to solve optimization, estimation and simulation problems that were previously intractable. The availability of both computer power and theoretical knowledge enabled researchers to address the task of fitting stellar orbits as an optimization problem. Rather than the unique solution of an algebraic problem derived from physical equations, orbital fitting can be seen as the search of a configuration that maximizes (or minimizes) certain merit function. In [Tokovinin, 1992], orbital parameters are found by minimizing the weighted sum of the

[1] The incorporation of additional phenomena such as mass transfer and relativistic effects into the model is also applicable to both stellar and sub-stellar systems to some extent.

squared differences between observations and model-generated values. To do so, the author uses the Levenberg-Marquardt algorithm, which interpolates between gradient-descent and Newton-Gauss methods [Levenberg, 1944], [Marquardt, 1963]. A similar approach is found in [MacKnight and Horch, 2004], where authors develop a downhill simplex minimization technique to estimate orbital parameters. This method has been extensively used by Horch and his collaborators (e.g., [Horch et al., 2015]).

Other methods rely on randomness to find optimal or sub-optimal solutions to the problem. One of the pioneers in the application of such kind of algorithms to orbital fitting is [Pourbaix, 1994], where author uses the Simulated Annealing method [Kirkpatrick et al., 1983] to find a set of orbital elements that minimizes differences between measurements and model-generated values (the same merit function used in [Tokovinin, 1992]). This *trial-and-error* approach, as described by its author, is an adaptation of the Metropolis algorithm previously explained in this work. Whereas Metropolis and MCMC methods in general are focused on sampling (for either integration or p.d.f. characterization purposes), Simulated Annealing is aimed at optimization. Like Metropolis algorithm, Simulated Annealing is a physically-inspired method[2]: let s be a vector encoding the configuration of a system, with an energy level $E(s)$ associated. A transition from s_n to s_{n+1} is always allowed if it leads to a lower energy level (that is, $E(s_{n+1}) \leq E(s_n)$), but can also be accepted with probability $\exp\left(-\frac{E(s_{n+1})-E(s_n)}{kT}\right)$ if $E(s_{n+1}) > E(s_n)$. As in the real physical systems that inspire this algorithm[3], temperature T decreases progressively over time. At high temperatures, configurations of higher energy are more likely to be accepted, whereas at lower temperatures the system undergoes smaller configuration leaps –it has reached a point of minimum energy. That scheme, which it outlined in Algorithm 6, can be used for general optimization purposes –it just requires configuration s to be replaced by the set of parameters the user desires to optimize and energy function $E(\cdot)$ to be replaced by an adequate merit function.

Algorithm 6 Simulated annealing

- Pick a random initial state:

$\rightarrow s = s_0$

for $n = 1, \ldots, N_{steps}$ **do**

- Choose a temperature according to the cooling schedule:

$\rightarrow T = temperature(n)$

- Pick a random neighbor:

$\rightarrow s' = neighbor(s)$

- Perform the Boltzmann probability test:

$\rightarrow u' = u$, with $u \sim \mathcal{U}(0,1)$.

if $u' < \exp\left(-\frac{E(s')-E(s)}{kT}\right)$ **then**

- Move to new state:

$\rightarrow s = s'$

end if

end for

In Algorithm 6, $temperature(\cdot)$ is a function that assigns a value of T to each step n according to a predetermined cooling schedule, and $neighbor(\cdot)$ is a function that outputs a new configuration in the vicinity of input value s. In [Pourbaix, 1994], s is a vector with the seven orbital elements (Thiele-Innes representation): $s = [P, T, e, A, B, F, G]$, whereas merit function has the form of a weighted mean squared error:

$$E(s) = \sum_{j=1}^{N} w_i \left((x_{obs}(i) - x_{model}(i;s))^2 + (y_{obs}(i) - y_{model}(i;s))^2 \right). \tag{3.1}$$

Later developments by the same author extend this scheme to binary stars with visual and spectroscopic observations [Pourbaix, 1998]. Both works refine the results obtained by Simulated Annealing by adding a second optimization stage: Powell's method in [Pourbaix, 1994] and Broyden–Fletcher–Goldfarb–Shanno local search

[2]Boltzmann constant k is kept just in order to show the physical analogy explicitly, but from an algorithmic point of view, the only value that matters is temperature T. Term kT can be replaced for a single value C that varies over time.

[3]Annealing is technique used in metallurgy to modify certain physical and chemical properties of a metal. It involves heating the material and then cooling it in a controlled way, thus altering the structure of its crystal lattice.

method in [Pourbaix, 1998].

A major disadvantage of Simulated Annealing is the need of a cooling schedule –formulating a sensible cooling schedule requires some degree of trial-and-error, since not all schemes to decrease the value of T generate good results. Another drawback has relation with uncertainty characterization: Simulated Annealing gives a single solution to the optimization problem, but does not provide a means to evaluate how solutions are distributed –required for establishing a confidence intervals, for example. This problem may be alleviated to some extent by computing an analytic estimate of the covariance matrix during a later stage (i.e., once Simulated Annealing algorithm has given an output). That estimate relies on the calculation of Fisher information matrix **J** (provided that merit function $E(\cdot)$ is differentiable) and is used by Levenberg-Marquardt algorithm and [Pourbaix, 1994]. In turn, this **J**-based estimate of covariance matrix has some limitations: it depends on the solution found (where **J** function is evaluated) and may be a good approximation of the covariance matrix in that vicinity, but not from a global point of view (for example, if merit function has several peaks); furthermore, it does not drop a hint on the shape of the merit function –if merit function admits a statistical interpretation, it may make a difference whether the solution is located at the top of a narrow peak or in a plateau-shaped surface.

3.3 Bayesian estimation of orbital parameters

In a Bayesian setting, a p.d.f. encodes the current state of knowledge of a variable of interest (parameter, state) rather than a probability of occurrence. In other words, parameters do not distribute over the range in which the p.d.f. is defined –they have a definite value–, but certain points in the parameter space are more likely to be the actual value. Thus, the Bayesian approach is at its core a tool to accomplish what optimization methods do not naturally provide: uncertainty characterization.

During the last years, the Bayesian approach has become somewhat of a standard in exoplanet research –MCMC and MCMC-inspired techniques have been profusely applied to estimate posterior distributions of parameters of multi-planet systems. However, that kind of analysis has not been systematically applied to visual binary stars. Endeavors made in the Bayesian analysis of exoplanets, brown dwarfs and binary stars are covered in the remainder of this section.

3.3.1 Bayesian analysis of exoplanet parameters

The breakthrough discovery of the first known planet orbiting a main-sequence star other than the Sun [Mayor and Queloz, 1995] paved the way for a new era of astronomical research. Up to this day, thousands of extrasolar planets have been discovered, mainly through radial velocity and –more recently– transit methods [IPAC Caltech, 2017]. However, it was not until mid 2000s that MCMC established as a mainstream technique for analyzing exoplanet radial velocity data. The pioneering works of Ford [Ford, 2005] and Gregory [Gregory, 2005] established guidelines for the application of MCMC to exoplanet research.

In [Ford, 2005], author introduces the concept of uncertainty quantification of orbits of extrasolar planets, being one of the first scholars to address orbital fitting as a p.d.f. estimation rather than as an optimization problem. To do so, the posterior distribution of the parameter vector of a single-planet system[4] (Equation 3.2) is estimated by means of a Metropolis-within-Gibbs scheme.

$$x = [P, K, e, \omega, T, V_{CoM}]. \tag{3.2}$$

The scheme is extended to multiple-planet systems by using a multi-Keplerian model for the radial velocity signal of the host star:

$$V(t) = \sum_{i=1}^{N_p} K_i(\cos(\omega_i + \nu(t; P_i, T_i, e_i)) + e_i \cos\omega_i) + V_{CoM}, \tag{3.3}$$

[4] Actually, author uses an alternative parameter vector that accelerates MCMC convergence:

$$x = [\log P, \log K, e\cos\omega, e\sin\omega, M_0, V_{CoM}],$$

where: (i) parameters P (orbital period) and K (amplitude) are explored in the logarithmic space in accordance to the scaling arguments suggested in [Gelman et al., 2003]; (ii) elements e and ω are reparametrized as $e\cos\omega$ and $e\sin\omega$ in order to avoid parameter correlation; (iii) term M_0 is the mean anomaly at time of periastron passage T and V_{CoM} the velocity offset of the spectroscopic data.

where $\nu(\cdot)$ is the true anomaly and N_p the number of planets. As the number of planets increases, so does the dimension of parameter vector. Although MCMC is not free from the so-called *curse of dimensionality*, it is more efficient than other algorithms in high-dimensional scenarios [Ford, 2005], hence the popularity it has achieved in exoplanet research.

In [Gregory, 2005], the author uses Parallel Tempering Markov Chain Monte Carlo (PT-MCMC hereinafter) to analyze radial velocity data of the yellow dwarf HD 73526, and compares the results with those previously obtained in [Tinney et al., 2003]. Both works detect the presence of a planet orbiting HD 73526: by performing a nonlinear least-squares analysis, [Tinney et al., 2003] finds a planet with an orbital period around 190 days; [Gregory, 2005], on the other hand, uses PT-MCMC to estimate the posterior distribution of the parameters of a single-planet model, finding three possible orbits –three modes in the p.d.f.– that fit the data, one of them corresponding to that found in [Tinney et al., 2003]. However, that peak is not the one with the most significant probability –another orbit, with a period about 376 days, is found to be formally more probable, for its vicinity contains a volume of parameter space with larger probability density. This is an example of the sort of analyses that Bayesian approach allows for.

Introduced by [Swendsen and Wang, 1986], Parallel Tempering (also known as Replica Exchange algorithm) may not be part of the "classical" arsenal of MCMC techniques, but it provides a means to cope with distributions with widely separated peaks –a problem the raw Metropolis-Hastings algorithm might struggle with. Like Metropolis sampler and Simulated Annealing, PT-MCMC is a thermodynamics-inspired method. A formal description of a basic version of this technique is presented as follows:

Let $\pi(x)$ be the posterior distribution of interest, $\pi(x) = p(x|y) \propto p(x) \cdot p(y|x)$, where x is a parameter or state vector and y a observation vector. Flattened versions of π are obtained by raising the likelihood term to the reciprocal of temperature T:

$$p'(x|y) \propto p(x) \cdot p(y|x)^{1/T}. \tag{3.4}$$

Although not physically meaningful, reciprocal $\beta = 1/T$ is preferred over T for practical reasons ($T \in (0, \infty)$, whereas $\beta \in (0, 1]$). A collection of values for β is defined: $\vec{\beta} = [\beta_1, \ldots, \beta_{N_{chains}}]$, with $\beta_1 = 1$, $\beta_2 > \beta_3 > \cdots > \beta_{N_{chains}} > 0$. This array of values induces a collection of probability density functions: let $p_j(\cdot)$ be $p_j(x|y) = p(x) \cdot p(y|x)^{\beta_j}$ (the function to evaluate target distributions is not necessarily normalized). Then, a chain –also called "replica"– is run for each β, and state swaps between replicas with adjacent[5] values of β occur with fixed probability $P_{exchange}$ –hence the original name of the algorithm. A proposal distribution $q_j(\cdot)$ must be defined for each chain, taking into account that "hotter" chains may require larger jumps to be properly explored. The routine is summarized in Algorithm 7.

The underlying idea is that in chains with flatter, "hotter" versions of π, significantly different configurations can arise from one step to the next, whereas in low temperature chains the current state is given the chance of refine itself. Thus, if a dramatically different configuration –possibly another mode– is discovered at the bottom of the ladder (where "hotter" versions of π lie), it can climb to the upper, "colder" chains through successive state swaps. The inference is usually performed at the chain with temperature $T = 1$ –that is, the chain with the π itself as its target distribution.

Additionally, [Gregory, 2005] presents a control system to automatically tune the σ-values of Gaussian proposal distributions $q_j(\cdot)$. Subsequent works by Gregory extend the methodology presented in [Gregory, 2005] with enhancements in the control system, incorporation of crossover operations inspired by genetic algorithms [Gregory and Fischer, 2010], and the formulation of an alternative parametrization of Keplerian elements, aimed at improving the performance of MCMC in presence of target distributions with highly correlated parameters [Gregory, 2011]. This methodology has been consistently used by Gregory and other researchers to analyze exoplanet spectroscopic data: in [Gregory, 2007a], [Gregory, 2007b], for example, author uses PT-MCMC to study objects HD 208487 and HD 11964, finding evidence of additional bodies orbiting the main star.

Although not the focus of exoplanet research, methodological aspects of MCMC have been a matter of concern for certain authors in the area: the choice of priors, proposal distributions and the specific class of MCMC method to be used is a non-trivial task –there is no technique suitable for all cases. Akin to the control system introduced in [Gregory, 2005], other scholars have proposed a number of different schemes to automatically tune the mechanism to propose new samples. An interesting case –and one that certainly influenced this work– is the use of Differential Evolution Markov Chain (DE-MC hereinafter) [Nelson et al., 2014]. That work presents a updated study of the planets orbiting star 55 Cancri A using DE-MC as a tool of analysis, arguing that this technique helps accelerate the burn-in phase in presence of a covariant structure among model parameters. As shown next, a particularly

[5] That is, only pairs of the form (β_k, β_{k+1}) are eligible.

Algorithm 7 Parallel tempering MCMC

• Pick a random initial state for each chain:
for $j = 1, \ldots, N_{chains}$ **do**
  $\rightarrow x_j(0) = x$, with $x \sim \ Prior_j$
end for
• Run chains in parallel:
for $i = 1, \ldots, N_{steps}$ **do**
  for $j = 1, \ldots, N_{chains}$ **do**
    • Make a proposal for $x_j^{(i)}$:
    $\rightarrow x'_j = x$, with $x \sim q_j(x|x_j(i-1))$
    • Accept or reject proposal:
    $\rightarrow u' = u$, with $u \sim \mathcal{U}(0,1)$.
    $\rightarrow \mathcal{A} = \min\{1, \frac{p_j(x'_j|y) \cdot q_j(x_j^{(i-1)}|x'_j)}{p_j(x_j^{(i-1)}|y) \cdot q_j(x'_j|x_j^{(i-1)})}\}$
    if $u' < \mathcal{A}$ **then**
      $\rightarrow x_j^{(i)} = x'_j$
    else
      $\rightarrow x_j^{(i)} = x_j^{(i-1)}$
    end if
  end for
  • Perform state swap with probability $P_{exchange}$:
  $\rightarrow s' = s$, with $s \sim \mathcal{U}(0,1)$.
  if $s' < P_{exchange}$ **then**
    • Choose a pair of adjacent chains at random:
    $\rightarrow$ Draw j' from $\{1, \ldots, N_{chains} - 1\}$ with uniform probability.
    • Accept or reject swap with probability $\mathcal{A}'$:
    $\rightarrow \mathcal{A}' = \min\{1, \frac{p_{j'}(x_{j'+1}^{(i)}|y) \cdot p_{j'+1}(x_{j'}^{(i)}|y)}{p_{j'}(x_{j'}^{(i)}|y) \cdot p_{j'+1}(x_{j'+1}^{(i)}|y)}\}$
    $\rightarrow a' = a$, with $a \sim \mathcal{U}(0,1)$.
    if $a' < \mathcal{A}'$ **then**
      $\rightarrow x_{j'}^{(i)} = x_{j'+1}^{(i)}$
      $\rightarrow x_{j'+1}^{(i)} = x_{j'}^{(i)}$
    end if
  end if
end for

appealing property of DE-MC is that it exempts the user from determining an appropriate proposal distribution –relying on the mutual learning between chains run in parallel, the algorithm adjusts itself to the particular structure of the distribution being sampled.

Introduced in [Braak, 2006] and refined later in [Vrugt et al., 2009], DE-MC is a population-based MCMC algorithm that integrates MCMC and Differential Evolution [Storn and Price, 1997], "a simple and efficient heuristic for global optimization over continuous spaces" as described by its authors. Mutual learning between chains is accomplished by using a proposal distribution based on the DE jumping step considered in [Storn and Price, 1995]. In DE-MC, the proposal sample $x_{proposal}$ of each chain is obtained by adding to the previous sample ($x_{previous}$) the difference between the current values of two other randomly chosen chains (say $R1$, $R2$):

$$x_{proposal} = x_{previous} + \gamma(x_{R1} - x_{R2}) + w, \tag{3.5}$$

where coefficient γ is a term to modulate the difference vector (its optimal value depends on the dimension of the feature space, d), and w is an additional perturbation drawn from a distribution with unbounded support (e.g., a normal) and small variance with respect to that of the objective distribution π. Term w is aimed at guaranteeing the irreducibility condition of MCMC. In practice, the additional noise term is useful to explore the

feature space at a local level, whereas term $\gamma(x_{R1} - x_{R2})$ contributes to make larger leaps without falling in zones of low likelihood. By using this scheme, jumps between consecutive steps of the algorithm are given in accordance to the scale and orientation for the target distribution. Algorithm 8 details how DE-MC works. In [Braak, 2006], the technique is proven to meet the reversibility, aperiodicity and irreducibility conditions required for MCMC.

Algorithm 8 Differential Evolution Markov Chain

• Initialize $x^{(1,j)}$ for $j = 1, \ldots, N_{chains}$.
for $i = 2, \ldots, N_{steps}$ **do**
  for $j = 1, \ldots, N_{chains}$ **do**
    • Randomly select two chains:
    $\rightarrow$ Draw j_1 from $\{1, \ldots, N_{chains}\}\backslash\{j\}$ with uniform probability.
    $\rightarrow$ Draw j_2 from $\{1, \ldots, N_{chains}\}\backslash\{j, j_1\}$ with uniform probability.
    • Propose a new sample:
    $\rightarrow w^{(i,j)} = w$, with $w \sim \mathcal{S}$.
    $\rightarrow x' = x^{(i-1,j)} + \gamma(x^{(i-1,j_1)} - x^{(i-1,j_2)}) + w^{(i,j)}$
    • Calculate acceptance probability through Metropolis-Hastings ratio:
    $\rightarrow \mathcal{A} = \min\left\{1, \frac{\pi(x')}{\pi(x^{(i-1,j)})}\right\}$
    • Accept or reject the proposed sample:
    $\rightarrow u' = u$, with $u \sim \mathcal{U}(0, 1)$
    if $u' < \mathcal{A}$ **then**
      $\rightarrow x^{(i,j)} = x'$
    else
      $\rightarrow x^{(i,j)} = x^{(i-1,j)}$
    end if
  end for
end for

As a final remark, the relevance of MCMC as a tool for analyzing exoplanet data extends to these days: besides making up an important part of the Kepler Mission data processing pipeline (see [Rowe et al., 2014], [Gautier III et al., 2012] for example), recent findings such as the celebrated discovery of seven temperate terrestrial planets orbiting ultra-cool dwarf star TRAPPIST-1 [Gillon et al., 2017] relied to some extent on MCMC (although applied to planet transit data rather than radial velocity measurements). Of course, much of the work done in exoplanet data analysis has been left out, since this survey does not intend by any means to be extensive, but rather focused in: (i) publications that have shaped the development of this area; (ii) publications that have directly influenced this work. In [Loredo et al., 2010], several lines of research regarding extrasolar planet data analysis are listed, such as the incorporation information-theoretic criteria into the sampling routines, the aforementioned Differential Evolution approach, and other adaptive MCMC methods. However, most part of the research is actually done based on the available tools, rather than introducing methodological innovations.

3.3.2 Bayesian approach applied to astrometric data

The application of Bayesian techniques to perform inference from astrometric observations has not been as extensive as in extrasolar planet research. However, a few relevant examples can be enumerated. In [Sahlmann et al., 2013], authors study a series of position measurements of ultra-cool dwarf DENIS-P J082303.1-491201 over two years, detecting a signal of orbital motion after parallactic, proper motion and chromatic refraction effects were corrected. This additional signal is explained by the presence of a low-mass companion –an unseen planet orbiting the main star. Authors perform a joint estimation of the orbital elements of the companion (i.e., P, T, e, a, ω, Ω, i) and parameters related to parallactic, proper motion and chromatic refraction effects. To do so, they separate the estimation into two stages: first, an optimal (or sub-optimal) set of parameters is found by means of Genetic Algorithms; then, the results of the previous stage are used to seed MCMC, in order to characterize the uncertainty of the parameters being estimated.

In [Lucy, 2014], the problem of estimating orbital parameters of visual binary stars is addressed from a Bayesian point of view. By virtue of a dimensionality reduction in the parameter space[6], the posterior mean of orbital

[6]The author takes advantage of the Thiele-Innes representation of orbital elements, in which the linearity of positions X, Y with

elements is calculated as a triple integral of the likelihood function[7] over period P, time of periastron passage T and eccentricity e. Numerical integration is carried out with a grid-based method. Besides introducing the Bayesian approach in the study of visual binaries in a formal way, a relevant aspect of [Lucy, 2014] is the support of the so-called "Eggen's effect": a good estimate of the total mass of a binary star ($\propto a^3/P^2$) can be found even if parameters P and a are not well-constrained. Numerical results obtained in [Lucy, 2014] suggest that if orbital coverage exceeds 40%, the total mass of the system can be reasonably estimated.

An example of joint estimation of orbital (the aforementioned P, T, e, a, ω, Ω, i) and radial velocity-related parameters (amplitudes K_1, K_2 and velocity of the center of mass, V_{CoM}) is found in [Burgasser et al., 2015]. That works extends the MCMC-based methodology introduced in [Burgasser et al., 2012], where authors use it to analyze radial velocity data of a brown dwarf, to include astrometric observations, thus allowing the estimation of a, Ω and i. Although applied to a brown dwarf, the approach presented is applicable to binary stars as well, since they are similar to brown dwarfs from the point of view of dynamics —aspects such as mechanisms of formation or chemical composition play no role in orbital motion.

respect to Thiele-Innes constants A, B, G, F allows for the analytic calculation of the least-squares solution of these parameters.

[7] Uniform priors were assumed, therefore they do not play any role in the integration.

Chapter 4

MCMC for orbit calculation

4.1 Motivation

As suggested by [Eggen, 1967], the quotient a^3/P^2 can often be well-determined even if the individual values of a and P are not accurately known –this is often referred to as the *Eggen's effect* [Lucy, 2014]. Thus, if the estimate of P suffers a dramatic change after new observations are incorporated into the analysis, a may undergo a shift such that compensates that variation, yielding a a^3/P^2 that falls not far from the previous value. The numerical results presented in [Lucy, 2014] strongly support this conjecture, as they suggest that, if orbital coverage exceeds 40%, a reasonable estimate of a^3/P^2 can be obtained. Of course, there are additional factors to take into account, such as the quality of the observations, the specific orbital section being covered (points near the periastron are significantly more informative that those that are far from it) and even the particular orbital configuration being observed –some orbits may be intrinsically more challenging to examine than others; think of cases with inclination (i) very close to 90°, for example.

As long as the observations provide a minimal orbital coverage of the object under study, the so-called Eggen's effect opens the possibility of estimating the mass by identifying the set of feasible orbital configurations, even if they involve a wide range of values for a and P. The basic idea is that, rather than calculating the mass based on a single estimate, one can characterize mass (and its uncertainty) on the base of this set of feasible values.

Since our aim is not only to find a single feasible estimate for the object under study, but rather to characterize the uncertainty of its orbital elements, we adopt a Bayesian approach for this problem. From a Bayesian standpoint, the set of feasible values mentioned in the paragraph above takes the form of a posterior PDF. In this work, we describe the posterior PDF a set of samples, which are drawn by means of the technique known as Markov Chain Monte Carlo (MCMC hereinafter). The remainder of this section is structured as follows: Section 4.2 describe the model used to describe the orbits and the dimensionality reduction carried out in both visual and spectroscopic stars; Section 4.3 briefly discusses how the selection of the objects under study was performed; sections 4.4 and 4.5 contain the methodology, results and discussion associated to the selected visual and spectroscopic binaries, respectively.

4.2 Model description

Assuming that phenomena such as mass transfer, relativistic effects or even the presence of non visible additional bodies do not affect the observed objects to a significant degree, the Keplerian model introduced in Section 2.1 is used to describe the orbits of the analyzed binary stars. This model requires seven parameters to fully characterize the trajectory of a visual binary star (Equation 4.1), that is, to compute the ephemeris for any given epoch τ.

$$\theta = \{P, T, e, a, \omega, \Omega, i\}. \tag{4.1}$$

Section 2.1 presents a brief explanation on the the physical and geometrical meaning of each parameter in Equation 4.1, and Subsection 2.1.1 details how to calculate the orbit (i.e., the trajectory over time) of a binary star given certain values for those elements. As to systems for which both astrometric and radial velocity measurements are available, one can perform a joint analysis by extending the parameter vector presented in Equation 4.1 in the

manner shown in Equation 4.2:

$$\theta = \{P, T, e, a, \omega, \Omega, i, V_{CoM}, \varpi, q\}, \tag{4.2}$$

where $\{P, T, e, a, \omega, \Omega, i\}$ are the well-known Campbell elements and V_{CoM}, ϖ and q denote the velocity of the center of mass, the parallax and the mass ratio m_2/m_1, respectively. The procedure to calculate the radial velocity signal as a function of parameters P, T, e, a, ω, i, V_{CoM}, ϖ, q can be found in Subsection 2.1.2.

The representation presented in Equation 4.2, used previously in [Burgasser et al., 2015], has some distinct characteristics. First, it includes parallax ϖ as one of the parameters to be estimated rather than a value known in advance, thus putting into practice the somewhat unexplored possibility of utilizing combined data (i.e., astrometry and radial velocity) to estimate hypothesis-free parallaxes [Pourbaix, 2000]. Secondly, it exploits all the restrictions imposed by the formulae of orbit position (equations 2.6, 2.8) and radial velocity (equations 2.14, 2.15), possibly leading to more precise inferences about the parameters. In contrast, some methods, such as the well-known ORBIT routine [Tokovinin, 1992], estimate amplitudes K_1, K_2 as free parameters. Those methods have the advantage of not requiring a parallax value to perform the estimation, but omit the dependency of K_1, K_2 on a, i and ω.

4.2.1 On the dimensionality of $\vec{\theta}$

Since the set of objects studied in this work makes up a relatively long list, it seems reasonable to devote some effort to reduce the computational costs involved in the analysis. In exploration-based methods such as the MCMC technique, the computer time required to obtain *good* results (in terms of convergence, precision and accuracy of the estimates, etc.) grows as the dimension of the feature space increases. For that reason, and at the expense of not exploring the whole seven-dimensional feature space of orbital parameters (10-dimensional space in the case of spectroscopic binaries), we propose a dimensionality reduction based on the separation of the parameter vector into two lower dimension vectors: one containing components whose least-squares solution cannot be determined analytically, ($\vec{\theta}_1$); and the other containing the components whose linear dependency[1] makes possible to calculate their least-squares solution with simple matrix algebra, ($\vec{\theta}_2$).

In the case of binaries with astrometric measurements only, one exploits the linear dependency of the well-known Thiele-Innes constants (A, B, G, F) with respect to the *normalized coordinates* x, y (which in turn depend on P, T, e and the collection of epochs of observation, $\{\tau_i\}_{i=1,\ldots,N}$). The procedure to obtain the least-squares solution of Thiele-Innes is detailed in Appendix B.1. Thus, instead of exploring the whole 7D space, the search is focused on $\vec{\theta}^{\,1} = [P, T, e]$, with $\vec{\theta}^{\,2} = [A, B, F, G]$ determined individually from each combination of free parameters. Campbell elements a, ω, Ω, i are recovered by using equations 2.9 (detailed procedure in Appendix B.2). This work follows the convention of choosing solutions with $\Omega \in (0°, 180°)$ in absence of information about the real orientation of the orbit. This approach has been previously adopted in [Hartkopf et al., 1989], [Lucy, 2014], among others.

Some definitions must be delivered before describing of the approach taken for binaries with spectroscopic data. In addition to parameters A, B, F, G, Thiele-Innes representation uses parameters C to H to compute coordinates in the Z-axis:

$$Z = Cx + Hy. \tag{4.3}$$

Those quantities are defined as follows:

$$\begin{aligned} C &= a\ \sin\omega\ \sin i, \\ H &= a\ \cos\omega\ \sin i. \end{aligned} \tag{4.4}$$

In [Wright and Howard, 2009], the authors take advantage of this representation to propose an efficient method to fit multi-Keplerian models to purely spectroscopic, purely astrometric and combined data sets. The core of their approach is the reformulation of the equations 2.14, 2.15 in a manner such that V_1 and V_2 are linear in the parameters, allowing for analytic calculation of least-square solutions. Making use of some trigonometric identities, radial velocity equations can be expressed as:

$$V(\tau) = h\cos\nu(\tau) + c\sin\nu(\tau) + \gamma, \tag{4.5}$$

where $h = K\cos\omega$, $c = -K\sin\omega$, $\gamma = V_{CoM} + K \cdot e \cdot \cos\omega$. Thus, $h = H/\varpi \cdot 2\pi/(P\sqrt{1-e^2})$, $c = -C/\varpi \cdot$

[1] With respect to quantities determined by $\vec{\theta}_1$.

$2\pi/(P\sqrt{1-e^2})$.

Nonetheless, as that paper is targeted at exoplanet research, each body involved is modeled with an independent Keplerian orbit, omitting the influence each system's component exerts on the others. Since that influence is not negligible when analyzing objects with masses of similar order of magnitude, that approach is not directly applicable to binary stars. Concretely, when analyzing binary stars the conditions shown below must be met, making the orbital parameters of the primary and those of the secondary interdependent.

$$\begin{aligned} a &= a_1 + a_2, \\ \frac{a_1}{a_2} &= \frac{m_2}{m_1} = q, \end{aligned} \tag{4.6}$$

The equalities above impose constraints on the parameters being estimated: if we reformulate equations 2.14, 2.15 according to the parametrization presented in Equation 4.5, then $h_1 = H_1/\varpi \cdot 2\pi/(P\sqrt{1-e^2})$, $c_1 = -C_1/\varpi \cdot 2\pi/(P\sqrt{1-e^2})$, being $H_1 = a_1 \cos\omega \sin i = \frac{q}{1+q}H$, $C_1 = a_1 \sin\omega \sin i = \frac{q}{1+q}C$ (analogous equations for V_2: specifically, $H_2 = \frac{1}{1+q}H$, $C_2 = \frac{1}{1+q}C$). The strict relations that (H_1, C_1) and (H_2, C_2) must comply (namely, $H = H_1 + H_2$, $C = C_1 + C_2$, $H_1/H_2 = C_1/C_2 = q$) do not stem naturally when calculating these quantities them as free parameters. Therefore, those conditions must be enforced as an additional mathematical restriction.

Although one could address that problem using Lagrange multipliers, there is no guarantee that the resulting set of non-linear equations will be analytically tractable (it may even have no unique solution). However, one can manipulate the formulae in such a manner that both orbital and radial velocity values of the Keplerian model are expressed as a linear combination of parameters, and meet the restrictions mentioned in paragraph above at the same time:

- In an approach similar to that used in [Wright and Howard, 2009], the first step is to use a combination of H, C –which are simpler expressions– to reconstruct parameters A, B, F, G (this requires the aim of trigonometric functions of Ω, i):

$$X = \underbrace{[H \sin\Omega \csc i + C \cos\Omega \cot i]}_{\mathrm{B}} x + \underbrace{[-C \sin\Omega \csc i + H \cos\Omega \cot i]}_{\mathrm{G}} y$$

$$Y = \underbrace{[H \cos\Omega \csc i - C \sin\Omega \cot i]}_{\mathrm{A}} x + \underbrace{[-C \cos\Omega \csc i - H \sin\Omega \cot i]}_{\mathrm{F}} y$$

- Grouping terms multiplied by C and H yields:

$$\begin{aligned} X &= \underbrace{(\sin\Omega \csc i\ x + \cos\Omega \cot i\ y)}_{S_X} H \quad + \underbrace{(\cos\Omega \cot i\ x - \sin\Omega \csc i\ y)}_{T_X} C \\ Y &= \underbrace{(\cos\Omega \csc i\ x - \sin\Omega \cot i\ y)}_{S_Y} H \quad + \underbrace{(-\sin\Omega \cot i\ x - \cos\Omega \csc i\ y)}_{T_Y} C \end{aligned}$$

  Thus, coordinates X, Y can be written as the linear combination of terms S_X, T_X, S_Y, T_Y (which can be easily computed from x, y, Ω and i), being H, C their accompanying constants.

- Finally, by using $\lambda_1 = \frac{q}{1+q} \cdot \frac{2\pi}{\varpi P\sqrt{1-e^2}}$, $\lambda_2 = \frac{1}{1+q} \cdot \frac{2\pi}{\varpi P\sqrt{1-e^2}}$ to transform H, C into h_1, h_2, c_1, c_2, one can express both the astrometric coordinates (Equation 2.8) and radial velocity values (equations 2.14, 2.15) in terms of a vector of parameters $\vec{\theta}_2 = [H, C, V_{CoM}]$:

$$\vec{\theta}_2 \cdot \mathbf{F} = [\vec{X}^{model}, \vec{Y}^{model}, \vec{V_1}^{model}, \vec{V_2}^{model}], \tag{4.7}$$

  where $\mathbf{F}$ is:

$$\mathbf{F} = \begin{bmatrix} S_x(1) & \dots & S_x(N_x) & S_y(1) & \dots & S_y(N_y) & \lambda_1(\cos\nu_1(1)+e) & \dots & \lambda_1(\cos\nu_1(N_1)+e) & \lambda_2(\cos\nu_2(1)+e) & \dots & \lambda_2(\cos\nu_2(N_2)+e) \\ T_x(1) & \dots & T_x(N_x) & T_y(1) & \dots & T_y(N_y) & -\lambda_1\sin\nu_1(1) & \dots & -\lambda_1\sin\nu_1(N_1) & -\lambda_2\sin\nu_2(1) & \dots & -\lambda_2\sin\nu_2(N_2) \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \end{bmatrix}. \tag{4.8}$$

  This allows for the calculation of the least-squares solution of $\vec{\theta}_2$ as:

$$\vec{\theta}_2 = \vec{x}\ \mathbf{W}\mathbf{F}^T(\mathbf{F}\mathbf{W}\mathbf{F}^T)^{-1}, \tag{4.9}$$

  being $\vec{x} = [\vec{X}^{obs}, \vec{Y}^{obs}, \vec{V_1}^{obs}, \vec{V_2}^{obs}]$ the data vector and $\mathbf{W}$ a diagonal matrix with the weight of each

observation. From the resulting $\hat{H}$, $\hat{C}$ values, parameters a and ω can be recovered as:

$$\hat{a} = \sqrt{\frac{\hat{C}^2 + \hat{H}^2}{\sin^2 i}}, \tag{4.10}$$

$$\hat{\omega} = \tan^{-1}\left(\frac{\hat{C}}{\hat{H}}\right), \tag{4.11}$$

The third component of $\vec{\theta}_2$ (V_{CoM}) has direct physical meaning and does not need to be transformed. Under this scheme, only seven parameters (P, T, e, Ω, i, q, ϖ) must be explored, whereas a, ω and V_{CoM} are calculated analytically. Although MCMC is the technique used in this work, the presented representation –and the dimensionality reduction that it involves– can be applied to other sorts of methods as well, even if they are not strictly exploration-based, such as Levenberg–Marquardt algorithm.

4.3 Sample selection

The primary sample of this work consists of Hipparcos' set of 1112 double visual star discoveries and 325 spectroscopic binaries from the Geneva-Copenhagen spectroscopic survey. During the years after the end of the Hipparcos Mission, visual binaries stars discovered by Hipparcos Satellite have been systematically observed in the Northern Hemisphere at the WIYN Telescope by Horch, van Altena and their collaborators [Horch et al., 2011, Horch et al., 2017]. On the other hand, Tokovinin has shown the capabilities of the instrument HRCam at the SOAR 4m telescope in northern Chile for binary-star research, producing significant results [Tokovinin et al., 2015]. Since 2014, the two primary samples (Hipparcos and Geneva-Copenhagen survey) have been observed by the SOAR telescope in order to confirm the binary nature of Hipparcos "suspected binaries" and to add observational data to the confirmed and spectroscopic binaries, with the aim of eventually computing their orbits and masses.

The final sample is largely derived from the published speckle data by Tokovinin and collaborators [Tokovinin et al., 2010, Tokovinin et al., 2014, Tokovinin et al., 2015, Tokovinin et al., 2016], which includes objects from the samples indicated in the previous paragraph, plus objects from Tokovinin's own sample of nearby F- and G-type dwarf-stars within 67 pc of the Sun ([Tokovinin et al., 2014]). The selection considered those orbital pairs for which their [O-C][2] in either angular separation ($[\text{O-C}]_\rho$) or position angle ($[\text{O-C}]_\theta$) evaluated at the epoch of SOAR Speckle data was too large in comparison with the internal precision of SOAR telescope, thus indicating that their orbits should be revised or improved with the addition of these new data points. Certain binaries whose first-time orbits could be computed using the SOAR observations were added. The final list is summarized in Table 4.1 (visual binaries) and Table 4.4 (spectroscopic binaries), which include an orbit estimate of each object.

4.4 Visual Binaries

Our list comprehends 18 visual and two spectroscopic binary systems, with periods ranging from few months in the case of spectroscopic stars to possibly thousands of years in the case of some visual binaries. Although parametrization of T as $T' = (T - \tau_0)/P$ (an approach used in this work and previously in [Lucy, 2014]) is useful to restrict the search range of the time of periastron passage to $[0, 1)$, both the initial distribution and the search range of the period P remain a difficult guess. The dimensionality reduction presented in Section 4.2 alleviates to some extent to need of choosing an initial guess –as it suppresses parameters a, ω, Ω, i from the analysis–, however, the variability of the feasible ranges of P among the studied objects imposes a diversity of scales among the posterior distributions. Moreover, the shape and orientation of each posterior PDF is unique. These factors make difficult to choose a single set of algorithm-related parameters. At the same time, the list is long enough to make individual case analysis undesirable.

4.4.1 Method description

In an effort to study the visual binaries in the list under a single unified framework, rather than choosing algorithm-related parameters for each star separately –a task usually involving a lot of trial and error iterations–, we adopt

[2] Observation minus computed value.

the Differential Evolution MCMC approach [Braak, 2006] introduced in Subsection 3.3.1. As a reminder, DE-MC stems as the combination of a genetic algorithm called Differential Evolution [Storn and Price, 1997], "a simple and efficient heuristic for global optimization over continuous spaces" as described by its authors, and MCMC. The algorithm is based on the idea of running several Markov chains in parallel but, instead of letting them run independently as in MCMC convergence tests, it lets the chains learn from each other. This aims at dealing with the problem of choosing an appropriate scale and orientation for the proposal distribution.

In DE-MC, the multiple chains "collaborate" with each other by using a proposal distribution based on the DE jumping step considered in [Storn and Price, 1995]. In that scheme, the proposal sample $x_{proposal}$ of each chain[3] is obtained by adding to the previous sample ($x_{previous}$) the difference between the current values of two other randomly chosen samples of the current generation (say $R1$, $R2$):

$$x_{proposal} = x_{previous} + \gamma(x_{R1} - x_{R2}) + w, \tag{4.12}$$

where γ is a term to modulate the difference vector and w is an additional perturbation drawn from a distribution with unbounded support. As mentioned in Subsection 3.3.1, term w is aimed at guaranteeing the irreducibility condition of MCMC and, from practical point of view, this additional noise is useful to explore the feature space at a local level, whereas term $\gamma(x_{R1} - x_{R2})$ contributes to make larger leaps without falling in zones of low likelihood.

Then, the only algorithm-related[4] parameters of DE-MC are: the number of chains N_{chains}; the coefficient γ; the parameters of the distribution of w. The actual implementation of this technique is summarized as follows:

- The parameters of interest are the orbital elements contained in Equation 4.1. Thiele-Innes representation is used: $x = [P, T', e, B, A, F, G]$, with vector x separated into non-linear and linear elements: $x_{NL} = [P, T', e]$ (those on which MCMC perform its "exploration") and $x_L = [B, A, F, G]$ (calculated by least-squares in the manner described in Appendix B.1).

- According to [Ford, 2005], it is a common practice to choose a non-informative prior that is uniform in the logarithm space for positive definite magnitudes (for example, period P and semi-major axis a), following certain scaling arguments presented in [Gelman et al., 2003]. In the context of orbital estimation, this approach has also been used in, for example, [Gregory, 2005], [Sahlmann et al., 2013], [Lucy, 2014]. From a practical point of view, such representation increases the rate of convergence in systems for which period is not well constrained. In exploration-based methods such as MCMC, the use of a Gaussian in the logarithmic of P as proposal distribution has proven to be useful [Gregory, 2005], and that is the approach adopted in this work.

- Following the guidelines presented in [Braak, 2006], algorithm-related parameters were fixed to the following values: $N_{chains} = 10$ (the recommendation is to choose $N_{chains} > 2 \cdot d$), $\gamma = 2.38/\sqrt{2 \cdot d}$. Values for w are drawn from $\mathcal{N}(0, \Sigma)$, where Σ is a diagonal matrix with: $\sigma^2_{T'} = 0.01$, $\sigma^2_{\log P} = 0.01 \cdot (\log P_{upper} - \log P_{lower})$ (using $\log P$ instead of P in accordance to the arguments presented in the previous point) and $\sigma^2_e = 0.01$.

- Since the aim of this work is the characterization of the posterior distribution, fitness function f is defined as the posterior PDF, which has the canonical form of *prior* $\times$ *likelihood*. Terms from the prior PDF can be dropped, as uniform priors were used for T' (range $(0, 1)$), $\log P$ (range $(\log 10, \log 5000)$ and e (range $(0, 0.99)$). Thus, likelihood function can be directly utilized to evaluate the Metropolis-Hastings ratio. Assuming individual Gaussian errors for each observation, the likelihood function for the i'-th sample of the MCMC routine[5] is computed as:

$$f(x^{i'}) \propto \exp\left(-\frac{1}{2}\left(\sum_{k=1}^{N_x} \frac{1}{\sigma_x^2(k)}[X(k) - X^{model}(k, i')]^2 + \sum_{k=1}^{N_y} \frac{1}{\sigma_y^2(k)}[Y(k) - Y^{model}(k, i')]^2\right)\right), \tag{4.13}$$

  where positions $X^{model}(k, i)$ and $Y^{model}(k, i)$ are calculated according to the instant of observation τ_k and the orbital parameters contained in $x^{i'}$, following the formulae introduced in Section 2.1.

[3] Term "chain" is used in order to keep a degree of familiarity between Differential Evolution and MCMC, but the samples of the current generation are actually called "agents" or simply "individuals" in the Differential Evolution literature.

[4] This term is used to avoid confusion with the target parameters being estimated: the orbital elements.

[5] Index i' encodes (i, j), that is, the iteration, i, and the chain j of DE-MC algorithm.

- By arguments of symmetry, the proposal distribution induced by DE-MC, which depends on the random selection of two chains and may be difficult to express in an analytic form, is canceled out in the calculation of the acceptance probability $\mathcal{A}$.
- The essential steps of this procedure are described in Algorithm 9:

Algorithm 9 Differential Evolution Markov Chain

• Initialize $x^{(1,j)}$ for $j = 1, \ldots, N_{chains}$ drawing values from prior distribution.
for $i = 2, \ldots, N_{steps}$ **do**
  for $j = 1, \ldots, N_{chains}$ **do**
    • Randomly select two chains:
    → Draw j_1 from $\{1, \ldots, N_{chains}\}\backslash\{j\}$ with uniform probability.
    → Draw j_2 from $\{1, \ldots, N_{chains}\}\backslash\{j, j_1\}$ with uniform probability.
    • Propose a new sample:
    → $w^{(i,j)} = w$, with $w \sim \mathcal{N}(0, \Sigma)$.
    → $x'_{NL} = x_{NL}^{(i-1,j)} + \gamma(x_{NL}^{(i-1,j_1)} - x_{NL}^{(i-1,j_2)}) + w^{(i,j)}$
    • Calculate the least-squares estimate of Thiele-Innes elements:
    → $x'_L = [\vec{X}_{obs}\vec{Y}_{obs}]\ \mathbf{WF}^T(\mathbf{FWF}^T)^{-1}$
    • Calculate acceptance probability through Metropolis-Hastings ratio:
    → $\mathcal{A} = \min\left\{1, \frac{f(x')}{f(x^{(i-1,j)})}\right\}$
    • Accept or reject the proposed sample:
    → $u' = u$, with $u \sim \mathcal{U}(0, 1)$
    if $u' < \mathcal{A}$ **then**
      → $x^{(i,j)} = x'$
    else
      → $x^{(i,j)} = x^{(i-1,j)}$
    end if
  end for
end for

4.4.2 Results

The methodology described in Subsection 4.4.1 is applied to a set of 18 visual binaries, in order to obtain an approximation of the posterior distribution of their orbital parameters. By virtue of the arguments presented in Appendix C, observation weights[6] were set to $\sigma_x = \sigma_y = \sigma_\rho$. The DE-MC algorithm was run with $N_{steps} = 10^6$. To evaluate whether the algorithm has reached a stationary regime, the classic convergence criterion introduced in [Gelman and Rubin, 1992] was used. The so-called "Gelman–Rubin R statistic" is a tool to evaluate convergence by comparing multiple independent Markov chains. To apply the Gelman–Rubin R statistic, several chains must be run in parallel, each of them seeded by a point drawn from an over-dispersed distribution (i.e., not exactly the target distribution but not too far from it either). Then, after calculating quantities B (inter-chain variance, Equation 4.14) and W (intra-chain variance, Equation 4.15), R statistic is computed as in 4.16.

$$B = \frac{N_{steps}}{N_{chains} - 1} \sum_{j=1}^{N_{chains}} (\hat{\theta}_j - \hat{\theta})^2. \tag{4.14}$$

$$W = \frac{1}{N_{chains}} \sum_{j=1}^{N_{chains}} \hat{\sigma}_j^2. \tag{4.15}$$

$$R = 1 + \frac{1}{N_{steps}} \cdot \left(\frac{B}{W} - 1\right). \tag{4.16}$$

[6] Typically, it is the value of σ_ρ that can be retrieved from previous works and from the own interferometric measurements, as (ρ, θ) is the prevailing representation of astrometric data.

In equations 4.14, 4.15, 4.15, θ denotes one of the parameters being estimated (i.e., a scalar component of the parameter vector), with $\hat{\theta}$ the average value calculated over all the chains and $\hat{\theta}_j$ the average of a single chain. The basic assumption of the R statistic is that, as in theory the stationary distribution is the same for all the chains (they are just a different instance of the same algorithm), if all of them behave similarly, then they all have reached the stationary regime. A convention among statisticians states that R values less than 1.05 are "good numbers" (that is, one can consider the chains have converged).

A burn-in period of 10^5 was considered for all objects. It was found that, according to the Gelman–Rubin R statistic, the stationary distribution was reached in no more than 10^5 iterations (actually significantly fewer iterations were needed in the general case). During the burn-in period, chains exhibit irregular behaviour, for they are in a stage of "accomodation", merely moving towards a stationary regime but not having reached it yet. Thus, considering burn-in periods with less than 10^5 samples induces higher values of the R statistic in certain cases (the algorithm does not have the same convergence rate for all the binaries analyzed). On the other hand, discarding more than 10^5 samples at the beginning of the chain does not affect the R statistic to a great extent, but does affect the quality of estimation, since fewer samples are used to approximate the target distribution (the ideal is to use as many samples as possible).

As proven in [Braak, 2006], each chain distributes as the stationary desired distribution $\pi(\cdot)$, so that: (i) the set of all chains can be treated as a single collection of samples once the samples in the burn-in period are dropped; (ii) the convergence diagnostics tools can be applied to the chains of a single run of DE-MC. This yields, after discarding the samples corresponding to he burn-in period, a total of $9 \cdot 10^6$ samples per object.

For each binary star of the list, R statistic happened to be less than 1.05 for all the parameters. The only exception was the first run of DE-MC for **STT 358AB**, because one of the chains got stuck in a vicinity of the initial point (just one chain showing a significantly different behaviour "ruins" the results of the convergence diagnostics). However, after a re-run of the algorithm for the aforementioned object (from which the values shown at Table 4.1 were taken), R values in the acceptable range were obtained. The reported anomaly can be explained, thus, simply as a consequence of the probabilistic nature of the method –priors may place the initial point of the chain in an inconvenient zone–, rather than a peculiarity of the object in itself.

The estimated orbital parameters obtained with the procedure previously described are presented in Table 4.1, whose structure is explained as follows: for each object there are two rows, the first one displays the Washington Double Star Catalog denomination of the object (first column) and the maximum a posteriori estimate[7] (columns 3 to 9); the second one shows the Hipparcos denomination (first column) and a quartile representation of the marginal PDF of each parameter: the main term corresponds to the median, whereas the lower (first) quartile (Q25) and the upper (third) quartile (Q75) are shown in the form of a subscript and a superscript, respectively. The reason why maximum likelihood/maximum a posteriori is preferred over expected value is that, for most of the cases studied in this work, PDFs are rather disperse and asymmetrical, thus yielding average values that are not in good agreement with the observations. The only exceptions are those orbits with good orbital coverage: **STF 2729AB**, **HU177**, **I 669AB** (to mention a few examples), for which the expected value approximately coincides with MAP/ML estimate.

Figure 4.2 displays some examples of orbital solutions and PDFs for objects in the sample. In order to visualize the scope and possibilities of the proposed method, examples of both "good" and "poor" orbits were selected: double stars **I 669AB STF 2729AB** being the "'good" cases (i.e., almost final orbital parameters, tight and possibly Gaussian-like PDFs); and **STF2434BC**, **STT358AB** being examples of orbits with poor orbital coverage and large uncertainty about their parameters. Right column shows some marginal PDFs of orbital parameters, as well as the total stellar mass. Orbital solutions displayed in Figure 4.2 (left column) must be interpreted in the following manner: the orbit induced by the maximum likelihood estimate appears in blue; observations appear in grey (linked to the proposed orbit by a red line); the black line indicates the apastron (point of maximum separation between the two stars); yellow line indicates the periastron (point of minimum separation); straight blue line indicates the line of nodes (intersection between the real orbit and the plane of the sky).

By looking at Table 4.1 and Figure 4.2 one can say that, in general, well determined orbits show a MAP value that approximately coincides with the median of the PDF, the inter-quartile range is relatively well constrained, and the PDFs show a Gaussian-like shape. On the other hand, orbits with insufficient orbital coverage or very uncertain measurements have PDFs with very long tails (and therefore larger interquartile ranges) on which the

[7] Due to the use of uniform priors, and the definition of likelihood function as in Equation 4.13, this estimate coincides in value, although not in meaning, with both the least-squares and maximum likelihood solution. In MCMC, the maximum a posteriori is simply approximated as the sample that maximizes the posterior PDF; however, the real MAP may not have been sampled.

MAP value usually exceeds by much the median value (it may be located somewhere along the "tail" of the PDF rather than in zones with a high density of samples, see **OL 18** in Table 4.1 for example).

The orbital parameters presented in Table 4.1 allow for the estimation of the stellar mass of each object. Table 4.2 indicates, for each of the visual binaries in the list, their spectral type from the literature (third column), trigonometric parallax (fourth column, on the second line the uncertainty of the published parallax is indicated) and the astrometric mass sum (in solar masses, last column of the table) obtained from the published parallax and the period and semi-major axis from Table 4.1. For each object, the upper line indicates the mass sum from the maximum a posteriori (MAP) solution, whereas the second line reports the median, with the first quartile as a subscript and the third quartile as a superscript (in the same manner found in Table 4.1). The quartiles reported for the mass sum comes from the MCMC orbital results alone, i.e., they are based on the published trigonometric parallax (fourth column, upper line) and do not take into account the uncertainty associated (fourth column, second line). Dynamic parallax (fifth column) and the mass values derived from it (sixth, seventh and eighth column) are computed as a reliability test for the solutions found with DE-MC. Dynamic parallax may be interpreted as the solution for the parallax value given the period P, the semi-major axis a and the apparent magnitude of both primary and secondary star, assuming that the analyzed object is a Main Sequence star. As dynamic parallax is an approximation based on a quite restrictive assumptions (namely, that the object belongs to the Main Sequence, and that luminosity and mass are related by a polynomial), the resulting sets of mass values (which are induced PDFs, indeed), tend to be more concentrated than those calculated based on a single parallax value and the whole set of solutions for P and a (which, except for the assumed value for the parallax, are hypothesis-free calculations).

To compute all these quantities, the photometry values for the primary (V_P) and secondary (V_S) from Table D.2 have been adopted[8], with the values of P and a taken from Table 4.1. The Mass-Luminosity Relation from [Henry and McCarthy Jr, 1993] provides an easy-to-evaluate mass vs. visual magnitude (M_v) polynomial relationship[9] for objects below $1M_\circ$. The uncertainty values for the dynamical parallax reported in Table 4.2 come exclusively from the range of solutions of our MCMC simulations, and not from uncertainties on either the photometry or the width of the Mass-Luminosity Relation.

Figure 4.1 shows a comparison between the orbits recently reported in Miles & Mason in the IAU Double Stars circular #191[10] and the solutions obtained in this work by means of MCMC, for objects **B1352** and **HU177**. The values reported by Miles & Mason are: $P = 123.84$ $[yr]$, $T = 2039.29$ $[yr]$, $e = 0.45$, $\Omega = 240.0°$, $a = 0.257''$, $i = 51.4°$, $\omega = 81.9°$ for **B1352**; $P = 220.41$ $[yr]$, $T = 1987.21$ $[yr]$, $e = 0.541$, $a = 0.293''$, $\omega = 252.6°$, $\Omega = 179.5°$, $i = 156.3°$ for **HU177**. The mean square error values obtained are: for **B1352**, 0.81 (Miles & Mason) and 0.098 (this work); for **HU177**, 8.3 (Miles & Mason) and 2.5 (this work). However, the weights assigned to each observation in this work may be different to those used by Miles & Mason, leading to very different results regardless of the estimation tool utilized. Moreover, the disagreement between both orbits may find an explanation in the fact that WDS have not included the SOAR observations yet.

4.4.3 Analysis and conclusions

A collection of 18 visual binary stars has been analyzed in this section. As can be seen in both Table 4.1 and Figure 4.2, the orbital coverage, as well as the reliability of the fitted orbital parameters, ranges from what one could consider as almost definitive orbits (e.g., **HU 177**, **A 2192**, **STF 2729AB**, **I 669AB**, **CHR 187**) to those with very poor coverage and uncertain orbital parameters (e.g., **COU 499**, **A 1374AB**, **OL 18**). The proposed method is not exempt from the degeneracy problems that most orbit-fitting techniques undergo (see **FIN 354**, for example, where low values of eccentricity provoke large uncertainty in parameters T and ω); however, it achieves the objective of characterizing the posterior probability of the target parameters as a set of samples. As to the estimators, this sections does not provide any numeric evidence of the suitability of MCMC for approximating integrals of the orbital parameters, since expected value is discarded in favour of maximum likelihood estimator[11] (reasons discussed briefly in Subsection 4.4.2). However, it proves useful to identify zones of high probability density and find, approximately, maximum points (see Figure 4.2). To support the last point, MCMC results were compared with the output of other orbital estimation tools, such as Tokovinin's ORBIT routine, as well as published orbits, obtaining good agreement.

[8] Included in Appendix D, since photometric analysis is out of the scope of this work.

[9] Several of the objects in Table 4.2 have masses above $1M_\circ$, but it is assumed that the polynomial fits of MLR are gentle enough to allow some degree of extrapolation.

[10] Available at http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/wds/dsl.

[11] Examples of the use of expected value, however, are presented in Section 4.5.

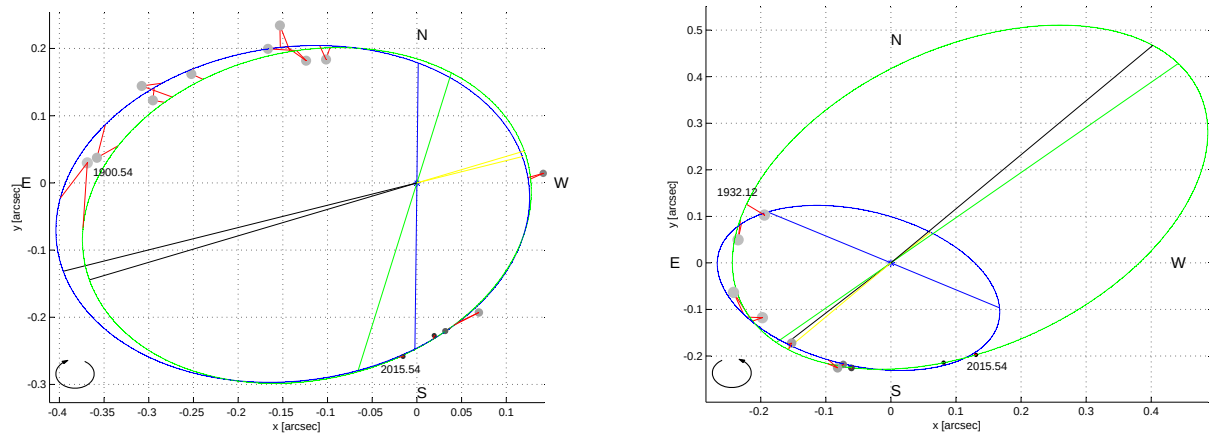

Figure 4.1 Comparison with published orbits: HU177 (left), B1352 (right). Blue line is used to draw the published orbit, whereas green is used to draw solutions from Table 4.1.

Table 4.1 Estimated parameters (visual binaries)

WDS HIP	Discoverer designation	P (yr)	T_0 (yr)	e	a ($''$)	ω ($^\circ$)	Ω ($^\circ$)	i ($^\circ$)
16115+0943	FIN 354	60.597	1999.655	0.049	0.128	92.266	83.361	91.126
79337		$59.314^{60.226}_{58.635}$	$1993.835^{1999.707}_{1966.473}$	$0.016^{0.040}_{0.007}$	$0.1285^{0.1287}_{0.1282}$	$120.622^{235.403}_{90.084}$	$83.352^{83.361}_{83.344}$	$91.113^{91.124}_{91.100}$
17305−1446	HU 177	202.320	1986.527	0.506	0.286	238.587	166.765	151.719
85679		$200.928^{214.525}_{189.235}$	$1986.461^{1986.897}_{1986.007}$	$0.503^{0.530}_{0.475}$	$0.286^{0.294}_{0.278}$	$237.488^{243.101}_{231.014}$	$165.817^{170.233}_{160.666}$	$150.896^{153.929}_{148.097}$
17313+1901	COU 499	2302.935	1988.173	0.868	1.317	6.828	59.211	109.264
85740		$136.088^{183.459}_{105.064}$	$2025.041^{2048.626}_{1983.290}$	$0.263^{0.428}_{0.128}$	$0.191^{0.241}_{0.173}$	$246.873^{288.625}_{129.610}$	$69.880^{75.733}_{66.0518}$	$116.749^{120.605}_{112.615}$
17533−3444	BU 1123	894.881	1961.744	0.750	0.641	300.303	36.712	36.265
87567		$753.043^{1003.634}_{599.282}$	$1961.517^{1962.248}_{1960.744}$	$0.718^{0.769}_{0.669}$	$0.571^{0.694}_{0.491}$	$296.034^{302.902}_{288.208}$	$40.352^{46.977}_{34.368}$	$36.116^{37.911}_{34.313}$
18003+2154	A 1374AB	3431.339	1976.577	0.844	3.308	218.824	153.313	125.795
1566-1708-1[a]		$496.492^{790.310}_{346.522}$	$1978.048^{1981.482}_{1975.345}$	$0.409^{0.574}_{0.234}$	$0.976^{1.285}_{0.808}$	$232.998^{243.150}_{225.578}$	$166.037^{171.236}_{160.982}$	$124.106^{126.038}_{122.061}$
18108−3529	B 1352	427.302	1977.431	0.433	0.432	6.873	134.114	41.939
89076		$229.101^{292.841}_{192.696}$	$1990.432^{2053.526}_{1959.921}$	$0.207^{0.361}_{0.097}$	$0.311^{0.369}_{0.279}$	$250.477^{309.862}_{104.260}$	$99.666^{115.055}_{88.126}$	$43.361^{51.724}_{35.058}$
18191−3509	OL 18	1118.536	2000.075	0.609	3.051	113.984	119.375	84.809
89766		$309.014^{422.653}_{231.059}$	$2048.950^{2057.546}_{2034.143}$	$0.383^{0.559}_{0.292}$	$1.960^{2.232}_{1.827}$	$209.372^{242.455}_{159.636}$	$119.437^{119.599}_{119.286}$	$86.731^{87.520}_{86.102}$
18359+1659	STT 358AB	532.063	2364.301	0.709	2.802	95.629	34.322	110.068
91159		$594.527^{783.753}_{495.219}$	$2379.588^{2508.472}_{2289.974}$	$0.686^{0.739}_{0.642}$	$2.937^{3.226}_{2.770}$	$99.603^{111.129}_{93.195}$	$32.184^{35.925}_{28.142}$	$110.542^{111.670}_{109.313}$
18537−0533	A 93	763.800	1914.744	0.746	1.202	342.186	29.623	35.341
92726		$549.097^{797.700}_{414.001}$	$1911.607^{1915.047}_{1907.944}$	$0.702^{0.753}_{0.660}$	$1.032^{1.238}_{0.927}$	$338.430^{349.231}_{330.184}$	$16.896^{28.119}_{9.438}$	$42.872^{49.355}_{35.941}$
18558+0327	A 2192	154.719	2008.234	0.221	0.295	62.839	80.138	129.307
92909		$154.221^{164.899}_{145.574}$	$2008.339^{2011.338}_{2005.542}$	$0.222^{0.248}_{0.201}$	$0.295^{0.310}_{0.281}$	$62.593^{73.114}_{52.990}$	$80.189^{81.284}_{79.081}$	$129.328^{131.142}_{127.580}$
19027−0043	STF 2434BC	1122.607	1992.129	0.647	1.972	95.071	47.870	151.316
93519		$975.012^{1238.389}_{807.202}$	$1993.480^{1995.979}_{1991.172}$	$0.612^{0.668}_{0.562}$	$1.821^{2.095}_{1.642}$	$97.225^{106.975}_{83.156}$	$49.298^{55.496}_{39.191}$	$148.979^{151.510}_{146.350}$
19350+2328	A 162	2650.764	1992.726	0.829	1.384	194.787	76.882	70.669
96317		$278.516^{404.744}_{222.688}$	$2014.700^{2028.929}_{2000.837}$	$0.330^{0.448}_{0.229}$	$0.294^{0.383}_{0.250}$	$234.338^{271.239}_{209.589}$	$77.924^{82.490}_{73.635}$	$62.821^{65.866}_{59.633}$
20073−5127	RST 1059	52.318	2017.245	0.265	0.166	159.239	65.697	22.426
99114		$154.489^{182.392}_{135.950}$	$2016.762^{2022.565}_{2009.591}$	$0.294^{0.347}_{0.251}$	$0.175^{0.188}_{0.167}$	$154.001^{194.197}_{75.893}$	$77.169^{112.011}_{58.562}$	$32.565^{38.979}_{25.982}$
20514−0538	STF 2729AB	200.676	1896.815	0.535	0.816	45.832	174.316	64.058
102945		$200.742^{201.870}_{199.623}$	$1896.803^{1897.198}_{1896.401}$	$0.535^{0.541}_{0.530}$	$0.816^{0.822}_{0.810}$	$45.861^{47.014}_{44.698}$	$174.306^{174.687}_{173.919}$	$64.061^{64.326}_{63.795}$
20597−5211	I 669AB	113.710	2010.542	0.611	0.682	207.169	63.068	93.692
103620		$113.546^{114.863}_{112.303}$	$2010.623^{2011.104}_{2010.161}$	$0.613^{0.625}_{0.602}$	$0.684^{0.691}_{0.678}$	$207.618^{210.430}_{205.032}$	$63.015^{63.321}_{62.701}$	$93.663^{93.819}_{93.478}$
21504−5818	HDS 3109	32.772	2014.133	0.158	0.224	98.178	127.217	87.836
107806		$32.915^{35.849}_{30.704}$	$2013.858^{2015.062}_{2012.248}$	$0.164^{0.185}_{0.143}$	$0.226^{0.241}_{0.215}$	$96.765^{116.391}_{74.771}$	$127.234^{127.333}_{127.143}$	$87.881^{88.019}_{87.764}$
22156−4121	CHR 187	19.089	1996.390	0.562	0.169	92.483	105.038	65.699
109908		$19.042^{19.199}_{18.860}$	$1996.350^{1996.516}_{1996.165}$	$0.561^{0.586}_{0.541}$	$0.168^{0.171}_{0.166}$	$92.608^{93.283}_{92.065}$	$104.191^{105.772}_{102.352}$	$65.693^{66.135}_{65.161}$
23171−1349	BU 182AB	380.704	1927.822	0.470	0.942	92.049	44.252	86.890
114962		$388.368^{422.365}_{363.982}$	$1928.592^{1930.696}_{1927.099}$	$0.464^{0.504}_{0.423}$	$0.957^{1.010}_{0.916}$	$93.700^{99.923}_{89.361}$	$44.335^{44.541}_{44.139}$	$86.930^{86.999}_{86.870}$

a. Tycho number.

The dimensionality reduction of the parameter vector, the particular implementation of MCMC chosen for analyzing visual binary stars (DE-MC) and the $\log P$ representation of orbital period in the feature space have been of great help for analyzing objects as dissimilar as **CHR 111**, with an orbital period as short as 20 [yr] and **STF 2434BC**, with a period around 1000 [yr], under the same framework. DE-MC bypasses the need of

Table 4.2 Masses and parallaxes (visual binaries)

WDS	HIP	Sp. Type	Trig. Parallax (mas)	Dyn. parallax (mas)	$\mathrm{Mass}_{\mathrm{P}}^{\mathrm{dyn}}$ ($M_\odot$)	$\mathrm{Mass}_{\mathrm{S}}^{\mathrm{dyn}}$ ($M_\odot$)	$\mathrm{Mass}_{\mathrm{T}}^{\mathrm{dyn}}$ ($M_\odot$)	$\mathrm{Mass}_{\mathrm{T}}$[a] ($M_\odot$)
16115+0943	79337	F0IV	5.13[b]	5.02	2.35	2.19	4.53	4.24
			$\pm$0.70	$5.13^{5.17}_{5.06}$	$2.32^{2.34}_{2.31}$	$2.16^{2.18}_{2.16}$	$4.49^{4.52}_{4.47}$	$4.48^{4.58}_{4.32}$
17305−1446	85679	F0V	5.06	5.72	1.69	1.37	3.05	4.41
			$\pm$0.97	$5.74^{5.85}_{5.64}$	$1.68^{1.70}_{1.67}$	$1.36^{1.37}_{1.35}$	$3.05^{3.07}_{3.03}$	$4.45^{4.68}_{4.25}$
17313+1901	85740	A5	2.24	5.00	1.72	1.72	3.44	38.29
			$\pm$1.32	$4.91^{5.47}_{4.53}$	$1.73^{1.79}_{1.65}$	$1.73^{1.79}_{1.65}$	$3.46^{3.59}_{3.31}$	$36.54^{48.24}_{29.74}$
17533−3444	87567	B3/5III[c]	3.68	3.87	2.86	2.82	5.68	6.61
			$\pm$0.54	$3.89^{3.93}_{3.84}$	$2.85^{2.87}_{2.84}$	$2.81^{2.83}_{2.80}$	$5.67^{5.70}_{5.63}$	$6.67^{6.87}_{6.48}$
18003+2154	1566−1708−1	G0	–	11.49	1.18	0.85	2.03	–
			–	$12.46^{13.20}_{11.93}$	$1.14^{1.16}_{1.12}$	$0.83^{0.84}_{0.81}$	$1.97^{2.00}_{1.93}$	–
18108−3529	89076	G3V	9.88	5.51	1.32	1.32	2.64	0.46
			$\pm$1.43	$5.90^{6.54}_{5.54}$	$1.28^{1.32}_{1.23}$	$1.28^{1.31}_{1.23}$	$2.57^{2.63}_{2.47}$	$0.55^{0.72}_{0.46}$
18191−3509	89766	K3+Vk[d]	31.35	23.89	0.87	0.80	1.66	0.74
			$\pm$1.25	$35.00^{45.38}_{30.07}$	$0.77^{0.80}_{0.71}$	$0.71^{0.74}_{0.65}$	$1.47^{1.54}_{1.36}$	$2.05^{4.13}_{1.36}$
18359+1659	91159	G2V[e]	29.63[f]	32.78	1.12	1.09	2.21	2.99
			$\pm$0.83	$30.49^{34.30}_{27.42}$	$1.15^{1.19}_{1.10}$	$1.12^{1.17}_{1.07}$	$2.27^{2.36}_{2.17}$	$2.47^{3.36}_{1.87}$
18537−0533	92726	G5V	12.99	11.23	1.14	0.96	2.10	1.36
			$\pm$1.65	$12.15^{13.38}_{11.16}$	$1.10^{1.14}_{1.07}$	$0.94^{0.97}_{0.91}$	$2.04^{2.11}_{1.97}$	$1.67^{2.16}_{1.33}$
18558+0327	92909	A3IV[g]	6.99	6.70	1.84	1.74	3.58	3.15
			$\pm$0.83	$6.69^{6.79}_{6.60}$	$1.84^{1.85}_{1.83}$	$1.74^{1.75}_{1.73}$	$3.58^{3.60}_{3.56}$	$3.14^{3.26}_{3.03}$
19027−0043	93519	G3/5V	9.48[h]	13.86	1.19	1.09	2.28	7.14
			$\pm$0.25	$14.11^{14.47}_{13.79}$	$1.18^{1.19}_{1.17}$	$1.09^{1.10}_{1.08}$	$2.27^{2.29}_{2.25}$	$7.47^{7.99}_{7.05}$
19350+2328	96317	A0	6.42	4.77	1.74	1.73	3.47	1.43
			$\pm$1.33	$4.51^{4.67}_{4.37}$	$1.79^{1.81}_{1.76}$	$1.77^{1.80}_{1.75}$	$3.56^{3.61}_{3.51}$	$1.23^{1.35}_{1.14}$
20073−5127	99114	F2IV	3.90[i]	3.77	1.87	1.82	3.69	3.34
			$\pm$0.65	$3.90^{4.15}_{3.76}$	$1.84^{1.87}_{1.79}$	$1.79^{1.82}_{1.74}$	$3.64^{3.70}_{3.54}$	$3.63^{4.26}_{3.30}$
20514−0538	102945	F6V[j]	16.47	16.59	1.62	1.33	2.95	3.02
			$\pm$0.59	$16.59^{16.68}_{16.50}$	$1.62^{1.62}_{1.61}$	$1.33^{1.33}_{1.33}$	$2.95^{2.96}_{2.94}$	$3.02^{3.06}_{2.97}$
20597−5211	103620	K0Vq	23.56[k]	24.34	0.88	0.82	1.70	1.88
			$\pm$0.31	$24.39^{24.61}_{24.26}$	$0.88^{0.88}_{0.88}$	$0.82^{0.82}_{0.82}$	$1.70^{1.70}_{1.70}$	$1.89^{1.93}_{1.86}$
21504−5818	107806	G6V	24.09	17.18	1.08	0.99	2.06	0.75
			$\pm$1.03	$17.35^{17.60}_{17.13}$	$1.07^{1.08}_{1.07}$	$0.98^{0.99}_{0.98}$	$2.06^{2.07}_{2.04}$	$0.77^{0.80}_{0.74}$
22156−4121	109908	G8III+G	11.87	15.22	2.14	1.59	3.73	7.88
			$\pm$0.43	$15.20^{15.42}_{14.94}$	$2.15^{2.16}_{2.13}$	$1.59^{1.60}_{1.58}$	$3.74^{3.77}_{3.71}$	$7.84^{8.14}_{7.50}$
23171−1349	114962	F(8)w[l]	15.08	13.80	1.13	1.07	2.19	1.68
			$\pm$1.80	$13.80^{13.95}_{13.66}$	$1.13^{1.13}_{1.12}$	$1.07^{1.07}_{1.06}$	$2.19^{2.20}_{2.18}$	$1.68^{1.73}_{1.64}$

a. Using the solution from Table 4.1, and the published trigonometric parallax from this table.
b. This is the revised parallax from Gaia DR1. The Hipparcos value was 6.16 ± 0.57 [mas].
c. B8V according to WDS
d. K3V according to WDS.
e. F8V according to WDS.
f. This is the revised parallax from Gaia DR1. The Hipparcos value was 30.41 ± 0.90 [mas].
g. A5V according to WDS.
h. This is the revised parallax from Gaia DR1. The Hipparcos value was 14.95 ± 3.80 [mas] - note the large difference!.
i. This is the revised parallax from Gaia DR1. The Hipparcos value was 3.64 ± 1.02 [mas].
j. F5IV-V according to WDS.
k. This is the revised parallax from Gaia DR1. The Hipparcos value was 24.59 ± 1.14 [mas].
l. F8IV according to WDS.

tuning the proposal distribution for each object individually (except for the additional noise term w, which was, in fact, fixed to the same value for all the objects in the list); moreover, since it accelerates the convergence of the algorithm [Nelson et al., 2014], as few as 10^6 iterations were sufficient to get good estimates and PDFs. Convergence, however, is a matter of concern, as with the majority of iterative methods. The mutual learning between chains, although useful for exploring the feature space in an adaptive way, does not guarantee that all the chains will reach the stationary distribution; hence the importance of performing convergence diagnostics.

Although some of the assumptions adopted in this work might be a topic of further discussion, such as the Gaussianity of the observational error or even the Keplerian model of the orbits, the proposed MCMC-based methodology is expected to be flexible enough to adapt to different scenarios, since none of the physics-based premises are directly linked with the main structure of the sampling routine. In fact, refining the results by utilizing a finer modelization of the phenomena associated with this problem might be a matter of future research. Another aspect that is not directly related to the sampling routine in itself, but has substantial impact on the estimation, is the assignation of weights. The case presented in Figure 4.1 suggests that mathematical correctness of estimation techniques may be misleading, since the results (either PDFs obtained with MCMC or point-estimates obtained with optimization techniques) are largely determined by the weights of each observation. In that sense, weight assignation may be more important than the technique used, and may require a specific discussion in an astronomical context (in this work, weights are merely part of the input).

Most of the mass estimates and mass uncertainties reported in Table 4.2 show good agreement with the conjecture referred to as *Eggen's effect*: even if orbital parameters are poorly determined, the PDFs of the mass sum is (relatively) tightly constrained; the object identified as **STF 2434BC/19027−0043** is a good example of this. The most noticeable exception is the star identified as **COU 499** or **17313+1901** (WDS denomination), where both orbital parameters and mass sum show disperse values, with the mass sum being, additionally, too large. The only explanation for this is the poor orbital coverage –although Eggen's effect suggests that reasonable mass estimates can be obtained even with few measurements, there exists a minimum threshold of orbital coverage that must be fulfilled. As to the large value of the mass sum, the explanation is equally simple: the reported trigonometric parallax is small and has a large uncertainty associated, so that small changes in its value induce dramatic changes in the mass sum estimate. For example, if the trigonometric parallax reported in Table 4.2 is increased by just 2σ, the mass sum obtained with the best-fit orbit indicated in Table 4.1 would be equal to 3.5 $[M_\circ]$ –a significantly more feasible mass according to the spectral type of this object, and largely different from the 38.29 $[M_\circ]$ value reported in Table 4.2. In-depth analysis of the stars listed in Table 4.1 is provided in [Mendez et al., 2017], a forthcoming publication largely based on this work, where the individual results of orbital parameters and mass are addressed from an astrophysical standpoint.

In certain cases, the mass sum reported falls out of the feasible range[12] (see **17533−3444**, for example), but has a small value of interquartile range. As with **COU 499**, the explanation for this stems from the published trigonometric parallax: since it has not already been precisely determined, small corrections in ϖ would render the mass sum estimate within a feasible range.

As seen in Chapter 3, over the last years MCMC has established as a standard tool for analyzing a quite related, similar problem to that addressed in this work –exoplanet detection and characterization. The trend, of course, is not accidental: it stems from a general tendency, in recent years, to favor Bayesian over parametric approach in estimation problems; and from the suitability of MCMC for processing data in Bayesian settings. In this sense, the results presented in this section are coherent with what astronomical literature has already suggested: MCMC is a useful tool for estimating orbits.

[12] Because of the spectral types reported in Table D.2, most of the stars are expected to have mass sums lower than 3 $[M_\circ]$.

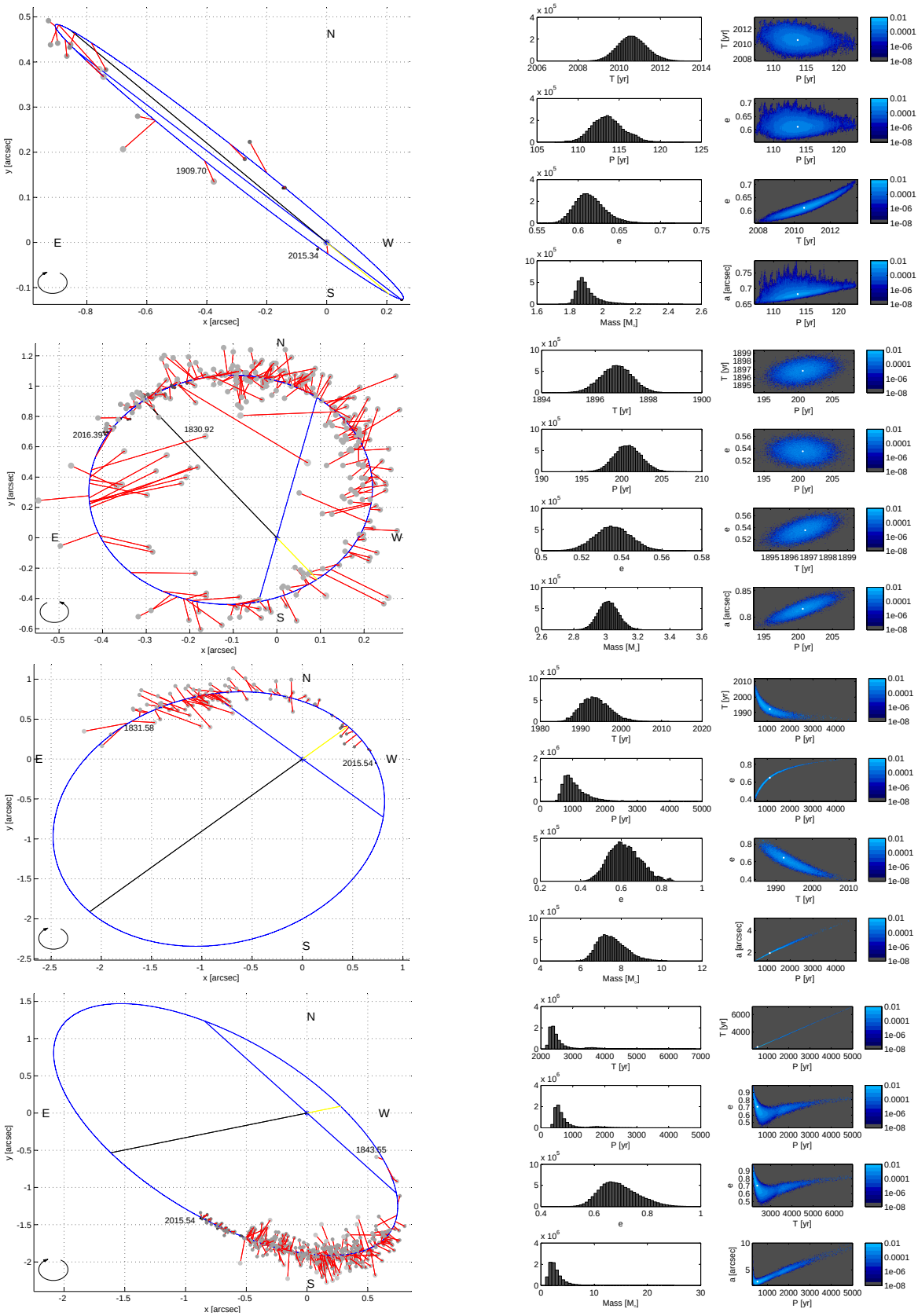

Figure 4.2 Best orbit and PDFs of (descending order): I699AB, STF2729AB, STF2434BC, STT358AB.

4.5 Spectroscopic binaries

The spectroscopic binaries are analyzed by means of the more traditional Gibbs sampler, introduced in [Geman and Geman, 1984] and used in [Ford, 2005] for the analysis of radial velocity data for exoplanet characterization. A number of practical reasons support this decision: first, as the dimensionality of the feature space is larger than that of the visual binary problem –10D, or 7D with the dimensionality reduction–, a larger number of chains must be run within the DE-MC algorithm (at least two times the size of the feature space, and preferably more), increasing the computational costs; secondly, unlike visual binaries in the list, a reasonable search range for the period can be proposed from visual inspection of the observations. Finally, although a raw Metropolis-Hastings MCMC would be a simpler approach, the dimension of the problem is large enough, and the location of the solutions concentrated enough, to make highly probable to fall in zones of low likelihood after a multidimensional random jump is applied on a sample, possibly reaching pathologically low levels of acceptance probability.

As explained in Subsubsection 2.2.3.5, Gibbs sampler relies on sequentially sampling each component of the feature space according to the conditional distributions. On the long run, that scheme is equivalent to drawing samples from the joint posterior distribution, $\pi(\cdot)$. Components can be grouped in blocks if it offers some advantage (for example, when a sub-set of the parameter vector has a closed-form and easy to sample distribution, or if two or more parameters have a known correlation). As the conditional posteriors do not have a standard form in this particular problem, the approach adopted is the scheme known as Metropolis-within-Gibbs, that is, generating a new sample according to a proposal distribution (modifying one component or block of components at once) and rejecting or accepting it according to the Metropolis-Hastings ratio. Denoting the parameter vector as $x = [x_1, \ldots, x_d] = [P, T', e, \Omega, i, q, \varpi]$, the implementation can be summarized as shown in Algorithm 10. As in Algorithm 9, subscript L indicates linearity of the parameters, thus x_L denotes $[a, \omega, V_{CoM}]$, and its least-squares estimate (given the parameters $P, T', e, \Omega, i, q, \varpi$ and the available measurements, both astrometric and spectroscopic) can be calculated in the manner described in Equation 4.9. The actual implementation uses the $\log P$ representation of orbital period, for the same reasons presented in Section 4.4.

Algorithm 10 Gibbs sampler for double star orbits

- Initialize $x^{(1)}$.

for $i = 2, \ldots, N_{steps}$ **do**
- Retrieve value from the last iteration:

$\rightarrow x^{(i)} = x^{(i-1)}$

for $j = 1, \ldots, d$ **do**
- Propose a new sample:

$\rightarrow x' = x^{(i)}$ (Copying the current value of $x^{(i)}$)

$\rightarrow x'_j = x_j$, with $x_j \sim \mathcal{N}(x'_j, \sigma_j^2)$ (Applying an additive Gaussian perturbation on component x_j)
- Calculate the least-squares estimate of elements a, ω, V_{CoM}:

$\rightarrow x'_L = [\vec{X}^{obs}, \vec{Y}^{obs}, \vec{V_1}^{obs}, \vec{V_2}^{obs}]$ $\mathbf{W}\mathbf{F}^T(\mathbf{F}\mathbf{W}\mathbf{F}^T)^{-1}$
- Acceptance/Rejection step:

$\rightarrow u' = u$, with $u \sim \mathcal{U}(0, 1)$.

$\rightarrow \mathcal{A} = \min\left\{1, \frac{f(x')}{f(x^{(i)})}\right\}$

if $u' < \mathcal{A}$ **then**
- Accept sample (if not, $x^{(i)}$ remains at the current value):

$\rightarrow x^{(i)} = x'$

end if

end for

end for

Under the assumption that individual errors of both astrometric and radial velocity observations are Gaussian, likelihood function has the form of:

$$\mathcal{L}(x^{(i)}) \propto \exp\Bigg(-\frac{1}{2}\Big(\sum_{k=1}^{N_x}\frac{1}{\sigma_x^2(k)}[X(k) - X^{model}(k,i)]^2 + \sum_{k=1}^{N_y}\frac{1}{\sigma_y^2(k)}[Y(k) - Y^{model}(k,i)]^2 + \sum_{k=1}^{N_{V1}}\frac{1}{\sigma_p^2(k)}[V_1(k) - V_1^{model}(k,i)]^2 + \sum_{k=1}^{N_{V2}}\frac{1}{\sigma_s^2(k)}[V_2(k) - V_2^{model}(k,i)]^2\Big)\Bigg). \tag{4.17}$$

Unlike the case of visual binaries, the priors chosen for the spectroscopic binaries studied in this work are not necessarily uniform. In the case of **YSC132Aa,Ab**, a Gaussian prior for the parallax was included in the fitness function f, since unfeasible values (parallaxes less than zero, for example) were explored if ϖ was set free (this simply means that the available data is not yet informative enough to give an estimate for ϖ); thus, the fitness function used for this object is the product of the likelihood $\mathcal{L}(\cdot)$ (defined in Equation 4.17) and the Gaussian PDF with the parameters indicated in Table 4.3. For **CHR111**, in absence non-uniform priors, ϖ consistently converges to a value close to the dynamic parallax and not far from the published trigonometric parallax; thus, the fitness function used for the final results is exactly the likelihood defined Equation $\mathcal{L}$. In the latter case, priors only determine the value of the initial sample of the MCMC routine, but do not play any role in the posterior probability function. Since the posterior PDFs obtained in this section are quite tight and Gaussian-like, the estimates of orbital parameters are calculated by averaging the sampled values –an approximation of the expected value.

On each Metropolis-Hastings step, an additive Gaussian perturbation applied to the previous value is used to propose new samples. Since that kind of proposal function meets the condition $q(x|x') = q(x'|x)$, those terms cancel out in the Metropolis-Hastings ratio. Parameters of the proposal distributions, as well as the boundaries of the initial uniform distributions used for the two objects are detailed in Table 4.3. The Gibbs sampler was run with $N_{steps} = 2\ 10^6$ and burn-in periods of $2\ 10^5$ for both **YSC132Aa,Ab** and **CHR111**, yielding a final sequence of $1.9\ 10^6$ samples for each. The burn-in period was determined by visual inspection of the evolution of the parameter values over the successive iterations of MCMC: initially placed at any point in the ranges defined by the priors, parameters move towards a zone of greater density of points, with values in the "transient zone" (the first 10^5 initial samples at most) not being visited again.

Table 4.3 Algorithm-related parameters for spectroscopic binaries

Object		P (yr)	T'	e	Ω $(^\circ)$	i $(^\circ)$	ϖ (mas)	q
CHR111	σ	0.5[a]	0.01	0.01	$\pi/45$	$\pi/45$	5	0.01
	Range	$(0.2, 3.0)$	$(0, 1)$	$(0, 0.99)$	$(0, 2\pi)$	$(0, \pi)$	$(32.35, 46.35)$	$(0, 1)$
YSC132Aa,Ab	σ	0.1[a]	0.01	0.01	$\pi/45$	$\pi/45$	5	0.01
	Range	$(0.1, 1.0)$	$(0, 1)$	$(0, 0.99)$	$(0, 2\pi)$	$(0, \pi)$	$-$[b]	$(0, 1)$

a. In the $\log P$ domain.
b. Gaussian prior with mean and standard deviation values indicated in Table 4.5.

4.5.1 Results and discussion

For the spectroscopic binaries analyzed in this section, a series of factors contribute to a more precise estimation of orbital parameters (Table 4.4) and masses (Table 4.5), in comparison with the visual binaries studied in Section 4.4. With orbital periods significantly shorter than those of the the visual binaries, these objects require smaller periods of observation to generate an accurate and precise estimation of this parameter. Specifically, the large amount spectroscopic data collected over the years allows for an almost definitive characterization of the radial velocity profiles (Figure 4.3), which translates into a very precise estimation of some of the target parameters (where period P is one of them). The astrometric observations, on the contrary, are poorer in comparison with both spectroscopic measurements and astrometric data of certain binaries of Section 4.4, mainly for two reasons: (i) the apparent orbit occupies a rather small "patch" of the plane of the sky (in the order of $0.01''$ approximately), therefore more precise imaging devices are required to obtain better measurements of the movement of these objects on the plane of the sky (for "larger" objects, the current degree of precision would be enough); (ii) especially in the case of **YSC132AaAb**, astrometric observations are scarce, leaving certain zones of the orbit uncovered or covered with very imprecise measurements.

As a consequence of the poor astrometric data, parameters that rely on astrometric measurements[13], such as ϖ or i, are not as well estimated as one would wish; in fact, for **YSC132AaAb** a prior for the parallax had to be included in the target probability in order to avoid sampling unfeasible values. This can be interpreted, however, in a more positive light: although the routine is not estimating hypothesis-free parallaxes, it reveals the potential of MCMC to incorporate knowledge from external sources (in this case, the trigonometric measurements of

[13]It is more accurate to say that parallax depends on the combination of sources (spectroscopic and astrometric) to be estimated, but the spectroscopic contribution is already good enough.

parallax). The results of **CHR111**, in turn, shed some light on the combination of astrometric and spectroscopic sources as a powerful tool for estimating parallaxes –the posterior PDF of ϖ for **CHR111**, displayed in Figure 4.4, is completely hypothesis-free–, although better astrometric measurements would be required to get tighter estimates.

From a methodological point of view, the results presented in this section reinforce the conclusions of Section 4.4, in the sense that the MCMC-based technique developed in this work proves to be a reliable tool for estimating orbital parameters and their uncertainty. Moreover, certain aspects not evaluated when analyzing visual binaries, such as the use of expected value as an estimator, are presented here with good results (tables 4.4, 4.5). The dimensionality reduction scheme developed in Subsection 4.2.1 for the case of objects with both spectroscopic and astrometric data makes easier, to some extent, the task of estimating the ten target parameters of this problem. However, more could be done in this regard, since the particular MCMC implementation used (Gibbs sampler) is not particularly efficient: although it successfully avoids exploring low-likelihood points (which would be rejected in the Metropolis-Hastings step) by modifying just one component at a time, on each iteration the Kepler's equation has to be solved as many times as the number of non-linear components.

It is worth noting the good agreement for the triad (V_{CoM}, K_1, K_2)[14] reported by the Catalogue of Spectroscopic Binary Orbits and the calculation performed here, namely: (-14.12 ± 0.04, 14.20 ± 0.07, 14.80 ± 0.07) *vs.* (-14.13 ± 0.08, 14.158 ± 0.026, 14.806 ± 0.015) km s^{-1} for **YSC131AaAb** and (-9.716 ± 0.097, 11.44 ± 0.16, 20.96 ± 0.61) *vs.* (-9.57 ± 0.01, 11.114 ± 0.080, 20.68 ± 0.55) km s^{-1} for **CHR111** respectively. This is particularly interesting, since it validates both the use of MCMC as a tool for orbital estimation of binary stars, and the mathematical formalism developed in this Subsection 4.2.1 for dimensionality reduction.

Table 4.4 Estimated parameters (spectroscopic binaries). On each row, upper lines: expected value; lower lines: quartiles of marginal PDFs.

WDS HIP	Discoverer designation	P (yr)	T_0 (yr)	e	a ($''$)	ω (°)	Ω (°)	i (°)	V_{cm} ($km\ s^{-1}$)	m_S/m_P	ϖ ($''$)	$Mass_T$ (M_{odot})
18099+0307	YSC 132Aa,Ab	0.54643	1990.675	0.302	0.019	86.650	51.189	146.167	−14.131	0.956	21.308	2.522
89000		$0.54643^{0.54649}_{0.54637}$	$1990.675^{1990.676}_{1990.674}$	$0.302^{0.304}_{0.300}$	$0.019^{0.020}_{0.019}$	$86.650^{86.922}_{86.377}$	$51.176^{53.587}_{48.777}$	$146.255^{147.433}_{144.996}$	$-14.131^{-14.124}_{-14.139}$	$0.956^{0.959}_{0.954}$	$21.308^{21.517}_{21.099}$	$2.539^{2.791}_{2.306}$
22313−0633	CHR 111	1.731	1965.475	0.367	0.066	172.100	261.393	67.141	−9.573	0.538	35.542	2.167
111170		$1.731^{1.731}_{1.730}$	$1965.475^{1965.484}_{1965.467}$	$0.367^{0.374}_{0.360}$	$0.066^{0.067}_{0.066}$	$172.092^{173.253}_{170.932}$	$261.341^{262.282}_{260.445}$	$67.229^{69.920}_{64.493}$	$-9.574^{-9.564}_{-9.584}$	$0.537^{0.553}_{0.522}$	$35.569^{36.629}_{34.497}$	$2.170^{2.353}_{2.011}$

Table 4.5 Trigonometric and dynamic parallaxes (spectroscopic binaries). On each row, upper lines: expected value; lower lines: quartiles of marginal PDFs.

WDS	HIP	Sp. Type	Trig. Parallax (mas)	Dyn. parallax (mas)	$Mass_P^{dyn}$ ($M_\odot$)	$Mass_S^{dyn}$ ($M_\odot$)	$Mass_T^{dyn}$ ($M_\odot$)	$Mass_T$ [a] ($M_\odot$)	$Mass_P^{comb}$ ($M_\odot$)	$Mass_S^{comb}$ ($M_\odot$)	$Mass_T^{comb}$ ($M_\odot$)
18099+0307	89000	F6V[b]	21.31	20.40	1.57	1.30	2.88	2.52	1.29	1.23	2.52
			±0.31	$20.45^{21.18}_{19.73}$	$1.57^{1.59}_{1.55}$	$1.30^{1.32}_{1.28}$	$2.87^{2.91}_{2.83}$	$2.54^{2.78}_{2.31}$	$1.30^{1.43}_{1.18}$	$1.24^{1.36}_{1.13}$	$2.54^{2.79}_{2.31}$
22313−0633	111170	F8V[c]	39.35	36.37	1.20	0.82	2.02	1.60	1.41	0.76	2.17
			±0.70	$36.42^{36.79}_{36.06}$	$1.20^{1.20}_{1.20}$	$0.82^{0.83}_{0.82}$	$2.02^{2.03}_{2.01}$	$1.60^{1.65}_{1.56}$	$1.41^{1.54}_{1.30}$	$0.76^{0.81}_{0.71}$	$2.17^{2.35}_{2.01}$

a. Using the solution from Table 4.4, and the published trigonometric parallax on the 4th column of this table.
b. F7V+F7.5V according to WDS.
c. F7V according to WDS.

[14] As a remainder, K_1 and K_2 are the amplitudes of the radial velocity signals.

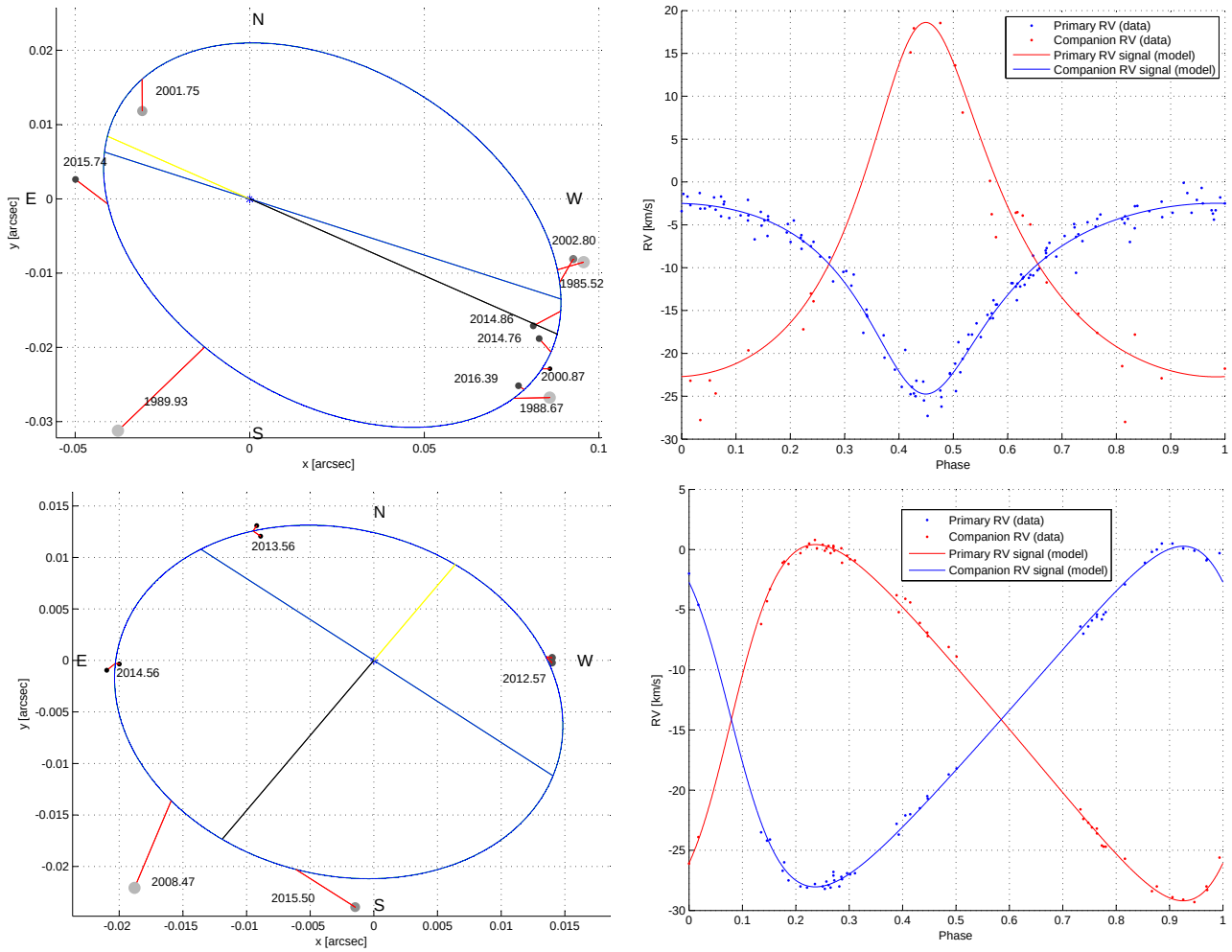

Figure 4.3 Orbit estimates and radial velocity profiles of: **CHR111** (top) and **YSC132AaAb** (bottom). In left column, yellow and black lines indicate periastron and apastron, respectively; straight blue line indicates line of nodes.

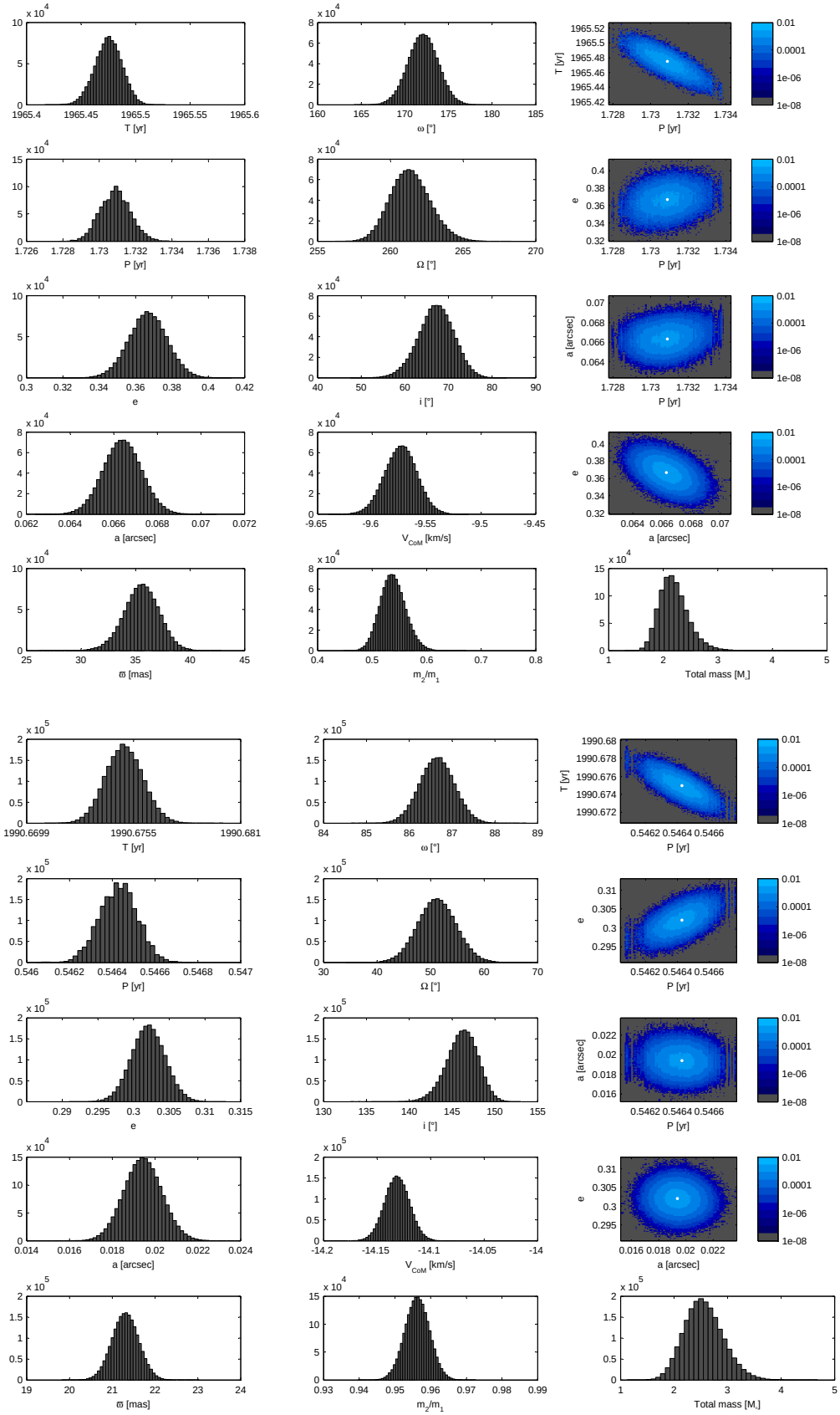

Figure 4.4 Posterior distributions of spectroscopic binaries: **CHR111** (top), **YSC132AaAb** (bottom).

Chapter 5

Orbit calculation with incomplete data

Partial data is a relatively common phenomenon in astrometric observations of binary stars. Usually, it does not derive from random data loss (as occurs in communication systems, for example), but rather from the impossibility of resolving the relative position between components on an image or sequence of images. Images of visual pairs, obtained by means of optical telescopes or interferometric techniques, are expected to display each component as a sharp point over a fainter background, thus enabling the viewer to identify their position in the plane of the sky. However, a number of factors may hinder the process of resolving a binary star: excessive or insufficient exposure time of the imaging devices, presence of additional noise due to atmospheric conditions, varying degrees of brightness of the objects being observed, angular separations beyond the angular resolution of the telescope. The latter case is arguably the most common source of partial information: in the vicinity of the periastron, the angular separation ρ between primary and secondary star reaches its minimum values; if those values fall below the resolution threshold of the imaging device, then the angular separation is confined to a range (i.e., $\rho \in (0, \rho_{max})$) instead of being reduced to a single value (as occurs with regular, successfully resolved measurements), and usually no information about position angle ϑ can be inferred either. In certain cases, databases of visual binary observations report partial data of the form $\rho = (0, \infty)$, $\vartheta = \vartheta^*$, that is, angular separation is missing but position angle ϑ is well-defined.

In general, incomplete entries are discarded when estimating the orbits of binary stars, with the information contained in them remaining unused. The incorporation of incomplete data, however, may have the potential to reduce the uncertainty about orbital parameters. A hint of that can be appreciated in the fact that some astronomers use partial data in a qualitative way: incomplete entries are not included as an input of the estimation routine (whatever it may be: least-squares methods, MCMC, etc.); however, within a set of orbit proposals (obtained by means of complete data), the solutions that violate the ranges imposed by partial information can be rejected (for example, if ρ is known to be in $(0, \rho_{max})$ for certain epoch, but the model predicts an angular separation significantly larger than ρ_{max}). This Chapter extends the routine presented in Chapter 4 with the incorporation of partial data. The cases addressed are those mentioned in the preceding paragraph, namely, $\rho \in (0, \rho_{max}), \vartheta \in (0, 2\pi)$ and $\rho \in (0, \infty)$, $\vartheta = \vartheta^*$. Entries where a Cartesian coordinate is missing, as well as cases where the feasible zone of the missing observation has an arbitrary geometry, are theoretical possibilities; however, they are not addressed in this work, since they do not occur in real situations (as far as the author's knowledge goes).

5.1 Methodology

The concept of multiple imputation is introduced in Section 2.3. At its core, it involves filling missing measurements with a set of plausible values, in order to generate data sets that are suitable for being analyzed with complete-data techniques. Subsection 2.3.1 details how MCMC can be applied to generate those plausible values and perform inference based on data sets with some degree of "missingness". That mechanism was introduced in the statistics literature by [Tanner and Wong, 1987], where authors detail how their *data-augmentation algorithm* works. Although a larger number of imputations per iteration (m) may lead to a better approximation of the integral in Equation 2.49, authors indicate that running the algorithm with an m value as small $m = 1$ leads to a correct approximation, in the sense that the average of the posterior distribution across the iterations will converge to $p(\theta|\mathcal{Y}_{obs})$ on the long run (that is, with a sufficient number of iterations). Subsequent works on

MCMC with partial data take advantage of that property, proposing MCMC routines that use one imputation at a time, but generate correct approximations of $p(\theta|\mathcal{Y}_{obs})$, $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs})$ across the successive MCMC iterations [Zhang, 2003], [Yuan, 2010]. That approach, summarized in Algorithm 5, is the scheme adopted in this work. Algorithm 11 describes the actual implementation used in this work, where the generic algorithm presented in Subsection 2.3.1 is adapted to the problem of orbital parameter estimation.

A fundamental aspect of imputation-based techniques is determining, based on reasonable assumptions, how missing values ($\mathcal{Y}_{mis}$), observed values ($\mathcal{Y}_{obs}$) and the target parameters (θ) depend on each other. This involves defining probability distributions relating each set of values (i.e., how $\mathcal{Y}_{mis}$ depends on $\mathcal{Y}_{obs}$, how observations $\mathcal{Y} = \mathcal{Y}_{obs} \cup \mathcal{Y}_{mis}$ depend on θ, etc.). The following list details the form that the probability distributions involved in the estimation routine (Algorithm 11) adopt in this particular problem:

- Imputations must be drawn from $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs})$, but this distribution is rarely sampled directly. As explained in Subsection 2.3.1, samples from $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs})$ can be obtained by a two-step procedure that involves drawing values from easier-to-sample distributions: first, getting θ from $p(\theta|\mathcal{Y}_{obs})$; then, conditional to the obtained value, generating a sample of $\mathcal{Y}_{mis}$ from $p(\mathcal{Y}_{mis}|\theta, \mathcal{Y}_{obs})$. Repeating that procedure a large number of times is equivalent to integrating $p(\mathcal{Y}_{mis}|\theta, \mathcal{Y}_{obs}) \cdot p(\theta|\mathcal{Y}_{obs})$ over θ, which yields $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs})$. In the context of orbital fitting, measurements $\mathcal{Y}$ correspond to observations of relative position: $\{\tau_i, \rho_i, \vartheta_i\}_{i=1,\dots,N_{obs}}$, where τ indicates an epoch of observation, ρ the angular separation and ϑ the position angle. Measurement matrix $\mathcal{Y}$ can be separated into $\mathcal{Y}_{obs}$ (observed values) and $\mathcal{Y}_{mis}$ (missing values). Thus, entries in $\mathcal{Y}_{obs}$ have a well-defined value assigned to each field (τ, ρ, ϑ, plus information of the observational error, σ_θ), whereas entries in $\mathcal{Y}_{mis}$ have one of the forms described before: at a given epoch τ, either $\rho \in (0, \rho_{max}), \vartheta \in (0, 2\pi)$ or $\rho \in (0, \infty)$, $\vartheta = \vartheta^*$.

- Term $p(\theta|\mathcal{Y}_{mis}, \mathcal{Y}_{obs})$, referred to as *posterior predictive distribution* in the context of multiple imputation, is simply the posterior PDF given the complete set of observations, $p(\theta|\mathcal{Y})$. Thus, its analytic form is given by Equation 4.13, that is, the posterior PDF of orbital parameters in a setting with complete astrometric measurements (Subsection 4.4). Its use in presence of partial information is slightly different, since, in order to generate an input with no partial data ($\mathcal{Y}$), the unobserved values $\mathcal{Y}_{mis}$ are filled with samples $\mathcal{Y}'_{mis}$ extracted from a predictive model $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs}, \theta)$. The set of orbital parameters , referred to as θ in this Chapter, has the same meaning that vector x' has in Equation 4.13.

- Expression $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs}, \theta)$, identified as *conditional predictive* distribution of $\mathcal{Y}_{mis}$ given $\mathcal{Y}_{obs}$ and θ, is a key term in the scheme, since it is the distribution from which the values to fill the missing observations are finally drawn. The localization of a missing observation in the plane of the sky depends on three factors:

  - Geometric restrictions, which are indicated as an input of the procedure. As a reminder, this kind of information may take the form of either $\rho \in (0, \rho_{max}), \vartheta \in (0, 2\pi)$ or $\rho \in (0, \infty)$, $\vartheta = \vartheta^*$. The former involves a circular shape with radius ρ_{max} and the primary as its center; the latter corresponds to a line that forms an angle of ϑ^* with the north celestial pole. A clear example of this can be observed in Figure 5.1.
  - A given value of orbital parameters θ (which appears explicitly in the expression for *conditional predictive* distribution). These orbital parameters determine the "real" orbit, but observations are affected by observational noise.
  - A characterization of the of observational error. In this work, it is assumed that that the difference between the real position and the observed position is Gaussian, therefore it can be described by its standard deviation σ_ρ. A value for σ_ρ must be provided for each epoch of observation: not only it is fundamental to assign a weight to each datum, but also to characterize where missing observations may fall, given that the "real position" is determined by θ.
  - Complete measurements of relative position, $\mathcal{Y}_{mis}$, are indirectly involved, since they limit the possibilities of the values that θ can adopt, for θ is previously drawn from $p(\theta|\mathcal{Y}_{mis}, \mathcal{Y}_{obs})$ (where $\mathcal{Y}_{mis}$ was, in turn, drawn from $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs}, \theta)$ in the last iteration).

This chapter addresses the problem of orbital estimation in a purely astrometric setting, i.e., when the only observations available are those of relative position –the same situation addressed in Section 4.4, but in presence of partial information. However, the same approach, which is summarized in Algorithm 11, can be adapted to the case of spectroscopic binaries. As in Section 4.4, the dimensionality reduction based on Thiele-Innes representation

(briefly documented in Subsection 4.2.1) is applied. Since the geometric constraints of the partial observations are not restrictive enough, it is necessary to have a reasonable knowledge of θ before running the algorithm –if the prior guess of θ is too broad, the imputations generated will cover a wide zone of the plane of the sky, thus being non-informative. Because of this, it is recommended to run a estimation routine based on complete-data (for example, that presented in Subsection 4.4.1) before executing Algorithm 11. The PDF thus obtained is then used as a prior distribution for the MCMC routine with multiple imputation (see the initialization step in Algorithm 11).

In order to maintain the notation used in Chapter 4, in Algorithm 11 the target parameter vector θ, which contains the orbital parameters, is named x. Thus, $x = [P, T', e, a, \omega, \Omega, i]$ and is separated into its non-linear (x_{NL}) and its linear (x_L) components. Additionally, plausible values of the missing observations, $\mathcal{Y}_{mis}$, are calculated over the successive iterations. The number of partial observations is denoted by N_{mis}. Let $\tau_{mis}^{(j)}$ denote an epoch where a partial observation occurs, with $j = 1, \ldots, N_{mis}$. Then, the imputation corresponding to epoch $\tau_{mis}^{(j)}$ in the i-th iteration of the MCMC algorithm is denoted $\mathcal{Y}_{mis}^{(i,j)}$, and has the format $\mathcal{Y}_{mis}^{(i,j)} = (\rho^{(i,j)}, \vartheta^{(i,j)})$ (polar coordinates) or $\mathcal{Y}_{mis}^{(i,j)} = (X^{(i,j)}, Y^{(i,j)})$ (Cartesian coordinates). Term $\mathcal{Y}_{gen}^{(i)}$ denotes the set with complete data generated in iteration i, that is, the union of successfully resolved measurements of relative position (whose value is fixed) and the particular values that have been imputed on the i-th iteration: $\mathcal{Y}_{gen}^{(i)} = \mathcal{Y}_{obs} \cup \mathcal{Y}_{mis}^{(i)} = \{\vec{X}_{obs}, \vec{Y}_{obs}\} \cup \{\vec{X}_{mis}^{(i)}, \vec{Y}_{mis}^{(i)}\}$.

By similar reasons to those adduced in Section 4.5 for the case of spectroscopic binaries, a Metropolis-within-Gibbs approach is adopted. However, other MCMC formulations (as DE-MC) may be suitable as well –the aim is to draw samples from distributions $p(\theta|\mathcal{Y}_{mis}, \mathcal{Y}_{obs})$ and $p(\mathcal{Y}_{mis}|\mathcal{Y}_{obs}, \theta)$, regardless of the manner in which it is performed.

Algorithm 11 MCMC with multiple imputation for orbital fitting

- Initialize $x^{(0)}$:

$\rightarrow x^{(0)} = x$, where x is drawn from a prior distribution

for $i = 1, \ldots, N_{steps}$ **do**

- Retrieve value from the last iteration:

$\rightarrow x^{(i)} = x^{(i-1)}$

- Imputation step (drawing $\mathcal{Y}'_{mis}$ from $p(\mathcal{Y}_{mis}|x^{(i-1)}, \mathcal{Y}_{obs})$):

for $j = 1, \ldots, N_{mis}$ **do**

$\rightarrow \mathcal{Y}_{mis}^{(i,j)} = generateImputation(\tau_{mis}^{(j)}, x^{(i-1)}, \sigma_\rho^{(j)})$

end for

- Posterior step (drawing x' from $p(x|\mathcal{Y}_{mis}^{(i)}, \mathcal{Y}_{obs})$)

for $j = 1, \ldots, d$ **do**

- Propose a new sample:

$\rightarrow x' = x^{(i)}$ (Copying the current value of $x^{(i)}$)

$\rightarrow x'_j = x_j$, with $x_j \sim \mathcal{N}(x'_j, \sigma_j^2)$ (Applying an additive Gaussian perturbation on component x_j)

- Calculate the least-squares estimate of elements a, ω, Ω, i:

$\rightarrow x'_L = [\vec{X}_{gen}^{(i)}, \vec{Y}_{gen}^{(i)}]\ \mathbf{WF}(\mathbf{FWF}^T)^{-1}$

- Acceptance/Rejection step:

$\rightarrow u' = u$, with $u \sim \mathcal{U}(0, 1)$.

$\rightarrow \mathcal{A} = \min\left\{1, \dfrac{f(x', \mathcal{Y}_{gen}^{(i)})}{f(x^{(i)}, \mathcal{Y}_{gen}^{(i)})}\right\}$

if $u' < \mathcal{A}$ **then**

- Accept sample (if not, $x^{(i)}$ remains at the current value):

$\rightarrow x^{(i)} = x'$

end if

end for

end for

For the sake of briefness and clarity, the Imputation step in Algorithm 11 is not explicitly explained, but shortened to an instruction called $generateImputation(\cdot)$. The reason for this is that the manner in which imputations are generated, intended to replicate the action of drawing samples from distribution $p(\mathcal{Y}_{mis}|x^{(i-1)}, \mathcal{Y}_{obs})$,

requires a separate explanation. The procedure is summarized in the points presented next:

- Partial observation is of the type $\rho \in (0, \rho_{max}), \vartheta \in (0, 2\pi)$: first, the predicted relative position of the system at epoch $\tau_{mis}^{(j)}$ is calculated according to the orbital parameters contained in $x^{(i-1)}$; then, an additive Gaussian perturbation with standard deviation $\sigma_\rho^{(j)}$ (an approximation of the observational error for that particular measurement) is applied. If the result meets the geometric constraint of being in $(0, \rho_{max})$, then it is accepted; if not, another realization of a Gaussian distribution is performed and added to the originally predicted position. The process is repeated until the geometric restriction is complied. In the tests performed in this work, the number of repetitions required to obtain a value within the radius ρ_{max} is not larger than 2.

- Partial observation is of the type $\rho \in (0, \infty)$, $\vartheta = \vartheta^*$: as in the case presented previously, the predicted relative position of the system at epoch $\tau_{mis}^{(j)}$ is calculated according to the orbital parameters contained in $x^{(i-1)}$ and a Gaussian perturbation with standard deviation $\sigma_\rho^{(j)}$ is applied on the model-based value. Then, instead of evaluating if the obtained value meets the geometric restriction –which means, in this case, falling on the line defined by $\vartheta = \vartheta^*$–, that value is *projected* on that line. This is equivalent to sample the bivariate Gaussian centered at the model-based position and $\sigma_X = \sigma_Y = \sigma_\rho^{(j)}$, $\sigma_{XY} = 0$, *given* that the observed value lies on $\vartheta = \vartheta^*$

Fitness function $f(\cdot)$, used to calculate the Metropolis-Hastings ratio, receives two arguments: an orbit encoded in the parameter vector x, and a collection of observations. The reason behind this is that the set of complete measurements $\mathcal{Y}$ changes on each iteration because of the imputations, unlike the data set in Equation 4.13, which has fixed values.

5.2 Simulation-based tests

The methodology proposed in Section 5 is evaluated on both synthetic and real data. This section presents the experimental design and the results obtained with the application of Algorithm 11 to a synthetic data set of astrometric observations of a binary star. The aim of this experiment is giving a glance on what kind of visual results this method can provide (see Figure 5.1), and how the uncertainty about orbital parameters changes once partial information is incorporated.

Evaluating the effect of incorporating new measurements (whether complete or partial) to the problem of orbital fitting has an intrinsic complexity: not all observations are equally informative, and as long as there are no theoretically-backed tools to quantify the information contained in each measurement, it is difficult to obtain results that hold for all cases[1]. For that reason, this study does not intend to be exhaustive.

The generation of artificial observations of relative position are loosely based on the orbit of the binary star Sirius, which is the brightest star observed from the Earth. The list presented next details how the synthetic observations are generated:

- Orbital parameters have fixed values: $P = 50.090$ $[yr]$, $T = 1944.220$ $[yr]$, $e = 0.5923$, $a = 0.750/3''$ (Sirius has a semi-major axis of $a = 7.50''$), $\omega = 147.27°$, $\Omega = 44.57°$, $i = 147.27°$. The first epoch of observation, τ_1, is set to occur on 1914.00, then $T' = (1944.220 - 1914.00)/50.090 = 30.22/50.090$.

- Complete measurements (i.e., those with known scalar values for ρ and ϑ) are positioned in such a way that they cover the orbit section near the apoastron. There are two partial observations: one of the type $\rho \in (0, \rho_{max}), \vartheta \in (0, 2\pi)$ and the other of the type $\rho \in (0, \infty)$, $\vartheta = \vartheta^*$; both partial measurements are located near the periastron. The idea is to mimic the fact that, in real settings, the occurrence of partial measurements is more probable in zones close to the periastron due to the resolution threshold of the imaging devices. Information of each measurement is reported Table 5.1 (epoch, angular separation, position angle and observational precision). Synthetic observations were generated as a particular realization of a Gaussian noise applied on the model-predicted positions (with the parameter values indicated previously). Partial measurements were obtained by dropping one of the components (ρ or ϑ).

[1] An example of "general result" is that stated in [Lucy, 2014]: "an orbital coverage of 40% leads to non-biased estimates of the mass sum". However, the cited paper fails to mention that it may not apply when one of the conditions present in that work is not met (for example, the fact that the periastron is within or out of the range of the observations makes a huge difference in terms of the quality of estimation, regardless of the fulfillment of the "40% rule"). Then, the value of this kind of synthetic data-based experiments, even if they lead to more general results than the analyses of real objects, is rather qualitative.

- Observational error follows a Gaussian distribution with $\sigma_x = \sigma_y = 0.008''$ and $\sigma_x = \sigma_y = 0.004''$ depending on the observation, as indicated in Table 5.1.
- Parallax is set to $\varpi = 37.9210/3$ $[mas]$ (Sirius has a parallax of $\varpi = 379.210$ $[mas]$). Both a and ϖ were modified in order to keep the same mass that Sirius has, but at the same time admitting more realistic values of observational noise. Since Sirius is a bright and close star, observations obtained with instruments with $\sigma \sim 0.004''$ would be excessively small in comparison with the apparent size of the orbit. To compensate, significantly larger values of σ would have had to be imposed.

Table 5.1 Ephemerides of a visual binary (synthetic data)

Epoch (yr)	ρ ($''$)	ϑ ($^\circ$)	σ_ρ ($''$)
1914.00	0.3519	10.17	0.008
1916.50	0.3624	15.07	0.008
1919.01	0.3748	20.69	0.008
1921.51	0.3716	24.45	0.008
1924.02	0.3657	29.69	0.008
1926.52	0.3651	33.57	0.008
1929.03	0.3617	39.13	0.008
1931.53	0.3270	44.68	0.008
1944.00[a]	0.0976[b] or $(0, 0.0976)$[c]	195.42[b] or $(0, 360)$[c]	0.004
1945.00[a]	0.1047[b] or $(0, \infty)$[c]	217.14	0.004
1956.58	0.2460	346.73	0.004
1959.08	0.2805	356.38	0.004

a. Partial observation.
b. Used in complete information scenario.
c. Used in partial information scenario.

The experiment comprehends three scenarios where the algorithms developed in this work are applied. In the first scenario, partial observations are discarded and orbital parameters are estimated by applying the Gibbs sampler[2] to the sub-set of complete observations. In the second scenario, partial observations are incorporated by means of Algorithm 11, thus sampling orbital parameters θ and "missing" observations $\mathcal{Y}_{mis}$ simultaneously. Note that, over the successive iterations of that method, sampled values of θ and Y_{mis} influence each other (that is how the additional information is incorporated). In the third scenario, partial measurements of the second scenario are completely known (see Table 5.1). This setting has the role of a ground-truth, in the sense that it yields the "best estimation possible" given the available data (it answers the question of how the estimation would improve if certain values of the observations associated to epochs 1944.00 and 1945.00 had not been dropped). As in the first case, estimation is carried out by using the Gibbs sampler. The MCMC algorithms were run with the following parameters: $N_{steps} = 4 \cdot 10^6$; Gaussian proposal distributions with parameters $\sigma_{\log P} = 0.4$, $\sigma_{T'} = 0.01$, $\sigma_e = 0.01$. A burn-out period of 10^5 samples was determined based on visual inspection of the parameter values over the successive iterations of the algorithm. In the case of MCMC with multiple imputation, a Gaussian proposal distribution with $\sigma_\rho = 0.004''$ (the same value "reported" as observational error in Table 5.1) was used to generate samples of $\mathcal{Y}_{mis}$ the in Imputation step.

Figure 5.1 shows the synthetic measurements and the estimated orbit (a maximum likelihood estimated is used). PDFs of partial measurements $\mathcal{Y}_{mis}$ (obtained in the second scenario) are superimposed on the graph that displays the observations and the estimated orbit. Those PDFs indicate where the partial observations may fall on the plane of the sky, given the available information. Depending on the type of partial measurement, the domain of these PDFs can take the form of an area (case $\rho \in (0, \rho_{max})$; lighter shades of blue indicate higher values of probability density) or a line (case $\vartheta = \vartheta^*$). The original values of the partial measurements are displayed as black dots for referential purposes, but are not actually used in the estimation procedure of the second scenario.

Figure 5.2 presents the marginal PDFs of the main orbital parameters (P, T, e, a and total mass) obtained on each setting, and adds a comparison of three scenarios on the same graph. In order to facilitate visual comparison, histograms are replaced by kernel-based densities in the lower right panel, although both convey essentially the same information.

[2]This Gibbs sampler operates like Algorithm 10 adapted to the purely astrometric setting, or like Algorithm 11 without the Imputation step.

5.2.1 Results and analysis

In order to assess the effects of incorporating partial measurements into the estimation routine, the methodology developed in Section 5.1 has been applied to three different settings. In a certain sense, the general technique described in Algorithm 5 (being Algorithm 11 an implementation for the problem of visual binaries) is based on augmenting the parameter vector with "slots" for the missing measurements, thus estimating both of them simultaneously. The simplest, most evident conclusion of the experiment is that the scheme developed in this section achieves the objective of sampling both quantities of interest in parallel, regardless of the quality of the results from an estimation point of view –how the incorporation of additional information affects the estimation is a rather different question. Figure 5.1 gives a demonstration of the capability of this method to characterize the uncertainty about the missing observations; Figure 5.2, on the other hand, reveals how the resulting posterior PDFs of orbital parameters differ from each other on each scenario.

As can be seen in Figure 5.1, this "MCMC plus multiple imputation" scheme manages to characterize the feasible values of the partial measurements on the plane of the sky, integrating the geometric restrictions of the partial measurements, the complete observations and the previous knowledge about the orbital parameters. It is worth noting that, as values of Y_{mis} and θ (denoted x within Algorithm 11) influence each other along the succession of Imputation and Posterior steps, the final result is different from what would arise from each source of knowledge on its own: according to geometric restrictions, feasible values would be evenly distributed within either a circular zone ($\rho \in (0, \rho_{max})$) or an infinite line ($\vartheta = \vartheta^*$); according to complete observations and prior knowledge of the parameters[3], the location of partial observations is defined by an *a priori* set of feasible orbits (which induce "real" positions in the epochs of interest) plus some kind of observational noise, however, that scheme ignores the influence that imputed observations would exert on the orbital parameters being estimated (in the case of the $\rho \in (0, \rho_{max})$ observation, for example, the support of the resulting PDF covers a much larger area within the circle if compared with that displayed in Figure 5.1 and obtained by means of Algorithm 11).

In Figure 5.2, the posterior PDFs of the target parameters obtained for each of the scenarios previously described are displayed for comparison. In general, parameter uncertainty decreases as more information is incorporated: the tightest PDFs are obtained in the third scenario (ideal setting with complete information), whereas the widest, least concentrated PDFs are obtained when partial information is discarded (first scenario). Results of the second setting (partial information incorporated via Algorithm 11) lie somewhere in the middle. The lower right panel of Figure 5.2 is the most explicit graph in this regard, clearly showing the peaks of the PDFs and how concentrated they are. However, not all quantities are equally affected by the incorporation of information (or lack of): while the uncertainties of T and the mass sum decrease dramatically with the incorporation of partial data, the marginal PDF of e obtained in the second scenario resembles more that obtained in the first setting than that obtained with complete information; the PDFs obtained for P, on the other hand, are quite similar in the three cases. Joint marginal PDFs undergo a similar change: distribution e vs. T, for example, shows a bow-shaped profile in the first scenario, turning into a convex shape[4] when partial observations are included, and reaching a Gaussian-like form in the complete information scenario.

In summary, the test performed in this section, aside from presenting a demonstration of the developed methodology, suggests that the incorporation of partial information into the analysis tools has potential of decreasing the estimation uncertainty.

[3] The distribution of orbital parameters obtained by using only complete observations, as described in the first scenario.
[4] In the sense of defining a non-concave contour.

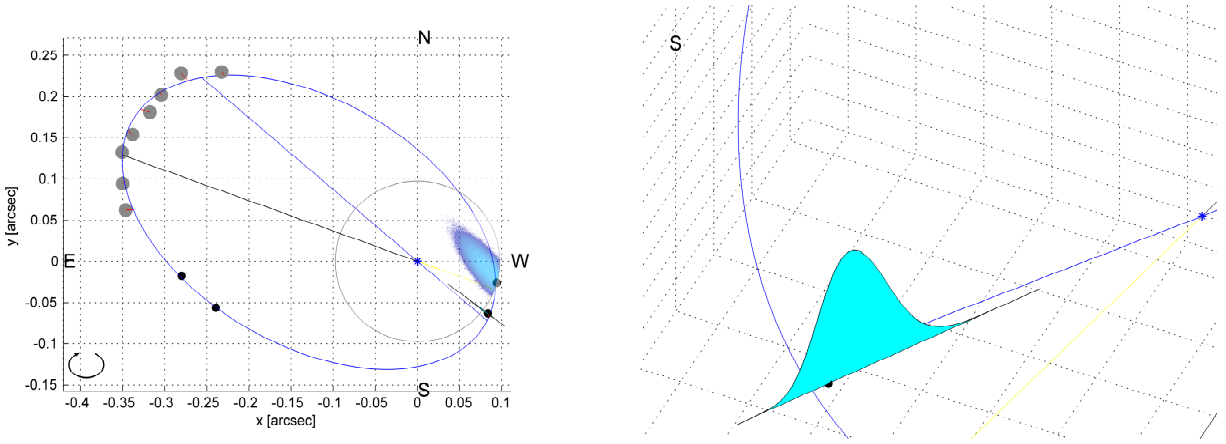

Figure 5.1 Orbit estimate in presence of incomplete information (synthetic data). Left panel shows the original measurements, the ML orbit estimate and the zones were missing observations may fall (calculated with the proposed method); right panel is a view in detail of the PDF along the line $\vartheta = 217.14°$.

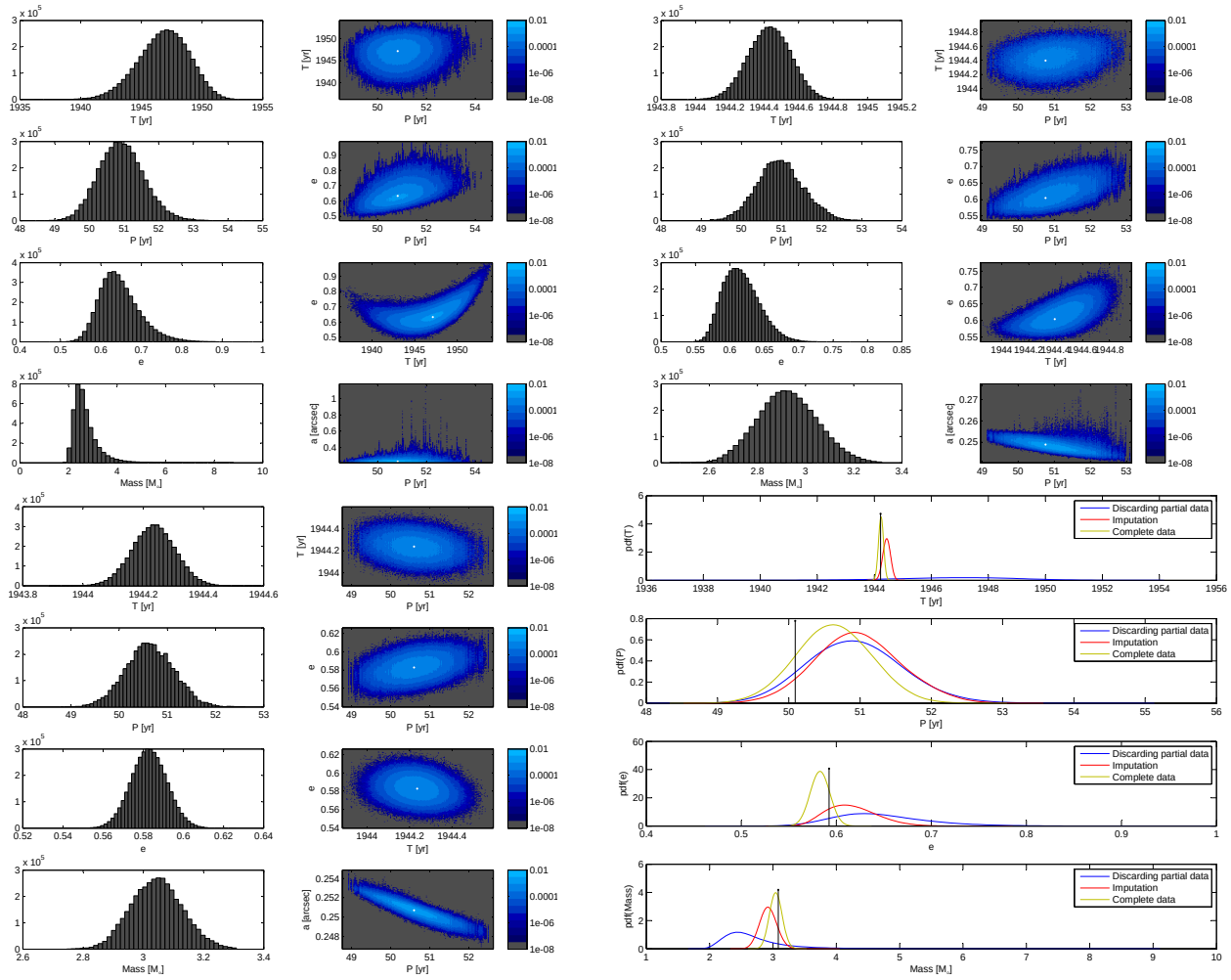

Figure 5.2 Panels with the posterior PDFs (synthetic data). Upper left panel shows the results obtained using only complete observations; upper right panel shows the results obtained when partial information has been incorporated; lower right panel shows the results obtained using complete information; lower left panel compares marginal PDFs of the three cases on the same figure (referential parameter value is indicated with a black vertical bar).

5.3 A case study: binary star HU177

In this Section, the methodology proposed in Section 5.1 and tested in Section 5.2 with synthetic data is applied to a real object: binary star **HU177**. Within the selection presented in Table 4.1, **HU177** is one of the few objects with partial data available. The reason for choosing that object for a case study is that, among objects with partial data, this is one for which the incorporation of partial information may make a difference in terms of estimation. For objects with an almost definite orbit, like **STF 2729AB** (for which a large amount of partial observations are available), the incorporation of partial measurements into the analysis would hardly affect the value of the estimated parameters and the uncertainty associated. Objects with a highly indeterminate orbit, on the other hand, are equally inadequate for testing the incorporation of partial measurements, since such a level of uncertainty would induce spatially disperse imputations (compare left and right panels of Figure 5.3: with less information, the area covered by the support of the PDF of the partial measurement is larger).

HU177 has relatively few complete observations (16) with varying degrees of quality, distributed unevenly from 1900 to 2015. The results presented in Table 4.1 indicate that, with the available data, the orbital parameters of this star are neither tightly constrained nor extremely indeterminate –that is the scenario in which partial measurements may be a contribution. In Figure 4.1, the solution proposed in Section 4.4 is compared with a referential orbit recently published in the astronomic literature. In both solutions, the periastron is located in (roughly) the same zone, so that it can be assumed that at least this particular aspect of the orbit is well estimated. There is a partial observation with $\rho \in (0,1)''$ and no ϑ, occurring in 1991.25 (an epoch very close to the predicted periastron passage). For this reason, the available partial measurement is expected to be, qualitatively speaking, informative.

The test performed in this section comprehends two different data sets: the first is identified by **HU177** and contains all the complete measurements available, plus the partial observation; the second one, identified as **HU177***, contains all the complete measurements except for the one near the periastron (epoch 1989.312), plus the partial observation (the discarded observation can be noticed in Figure 5.3). The "assembly" of two separate data sets is carried out in order to better assess the impact of the partial observation: in presence of a complete observation occurring closely in time and space (as in data set **HU177**), the partial observation may be redundant or even a source of additional uncertainty (recall that Section 2.3 identifies three sources of uncertainty associated to the imputation process). For each data set, two cases are addressed: with and without multiple imputation. In the first case the partial measurement is omitted, then the data set is analyzed with the Gibbs sampler used in the previous section; the second case is analyzed by means of Algorithm 11. N_{steps}, $\sigma_{\log P}$, $\sigma_{T'}$, σ_e are set to the same values used in Section 5.2. When applying the multiple imputation approach, the proposal distribution used to generate samples of $\mathcal{Y}_{mis}$ is set to $\sigma_\rho = 0.015''$, based on the observational error associated to the imaging device that produced the partial measurement. The criteria to fix that value, as well as the weight assigned to the imputed observations in the likelihood function, are topic of further research.

Figure 5.3 displays the orbits associated to the maximum likelihood estimates obtained for data sets **HU177*** and **HU177** with the incorporation of the partial data. The complete measurements of relative position, as well as the PDF of the partial observation, are shown in the same plane. The parameter values corresponding to each of these solutions are reported in 5.2. Figure 5.4 shows the posterior PDFs of the parameters obtained in each scenario. Table 5.3 reports numerically the results obtained for the four cases analyzed: each row has two lines, the first showing the quartiles of the orbital parameters (as in the second lines of Table 4.1), and the second line reporting the interquartile range (i.e., the difference between 75% and 25% percentiles). Interquartile range is used as a means to assess how uncertainty changes when partial information is incorporated.

5.3.1 Results and analysis

From Figure 5.3 and Table 5.2 it is observed that, although similar at first sight, orbit estimates obtained of **HU177*** and **HU177** (with multiple imputation) have considerable, although not large, differences. The most significant one, from a physical rather than numerical standpoint, is P. Estimates of angular elements ω and Ω exhibit a large numerical difference between both scenarios, but are geometrically close due to the circular nature of angular quantities –in fact, both the line of nodes (in blue) and the line that connects periastron and apoastron (yellow and black in the respective ends) share similar orientation and location in the two orbits. More noticeable is the difference between the PDF of the partial measurement obtained in each case: for **HU177***, it comprehends almost half the circle defined by $\rho \in (0, \rho_{max})$ (although zones with higher probability are concentrated towards the "western" border); for **HU177**, the PDF of the partial measurement is confined to a much smaller area.

Despite the larger uncertainty about the partial observation in the case of **HU177***, the incorporation of partial information has a larger impact on the estimation in **HU177*** than in the **HU177** data set: in the first setting, the interquartile range of most parameters undergo a dramatic reduction once the multiple imputation scheme is adopted (Table 5.3, from first to second row), whereas in the second setting the is no statistically significant evidence that the partial observation translates into additional knowledge of the system[5] (Table 5.3, from third to second fourth). Since in **HU177*** the partial measurement is the only information about that particular zone of the orbit (the vicinity of the periastron), it contributes to reduce the uncertainty about orbital parameters. In **HU177**, on the contrary, that information is redundant –it is the observation dropped in **HU177*** that contributes to characterize that sector of the orbit. Furthermore, as imputation process in itself has some uncertainty associated (the three sources enumerated in Section 2.3), the multiple imputation may add uncertainty to the estimation instead of reducing it. In **HU177***, for example, mass sum is more constrained when no imputations are used (Figure 5.4, fourth row of the lower left panel). A question stems from the latter observation: if the interquartile range of both P and a decrease when the multiple imputation scheme is adopted, why (and how) the uncertainty of a derived quantity such as the mass sum increases? The explanation might be found in the fact that interquartile range is a rather naïve tool to measure uncertainty, since it basically ignores the shape of the PDFs (aspects such as heaviness of tails, symmetry, etc.). Thus, assessing uncertainty by means of more sophisticated criteria, such as information-theoretic indicators, appears as the next step in this research line.

Finally, it is worth mentioning that defining more restrictive feasible areas for missing observations (instead of mere circles and lines) would be of great help to the estimation –the more constricted they are, the more they resemble a well-resolved, point-like observation. In other words, it might be beneficial, from the point of view of parameter estimation and uncertainty characterization, to save as much information as possible when resolving relative position values from interferometric measurements; that means that, if complete resolution is not feasible, an effort could be made to confine the plausible values to, for example, an angular section instead of a circle centered at the primary star.

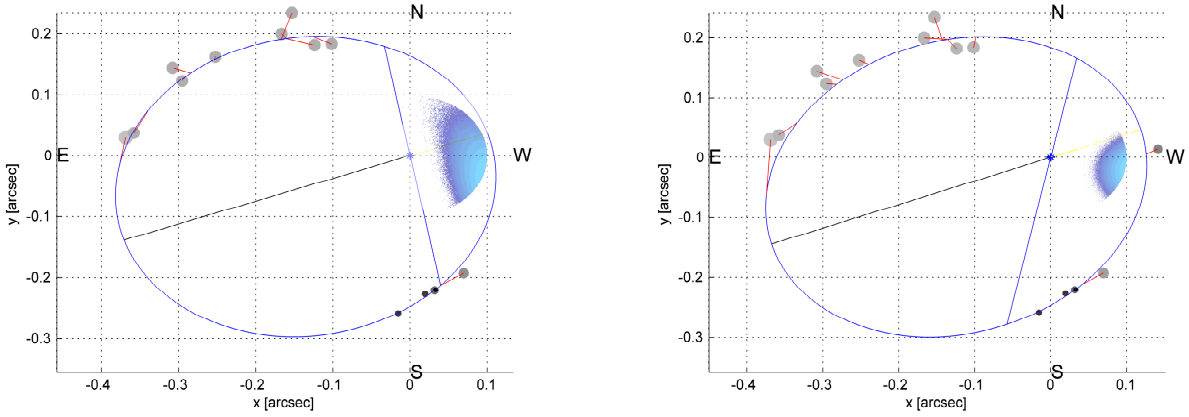

Figure 5.3 Comparison: visualization of orbit estimates of **HU177*** (left) and **HU177** (right).

Table 5.2 Maximum likelihood orbit estimates in presence of incomplete information (**HU177**)

Data set	P (yr)	T (yr)	e	a $''$	ω ($^\circ$)	Ω ($^\circ$)	i ($^\circ$)	Mass_T ($M_\odot$)
HU177*	222.906	1987.135	0.575	0.296	81.572	10.540	147.771	4.022
HU177	204.607	1986.604	0.512	0.287	240.178	168.311	151.376	4.367

[5]Although for some parameters the interquartile range has actually lower values in "**HU177** - MI", those difference are within the variation range that is inherent to MCMC (due to the random nature of the algorithm).

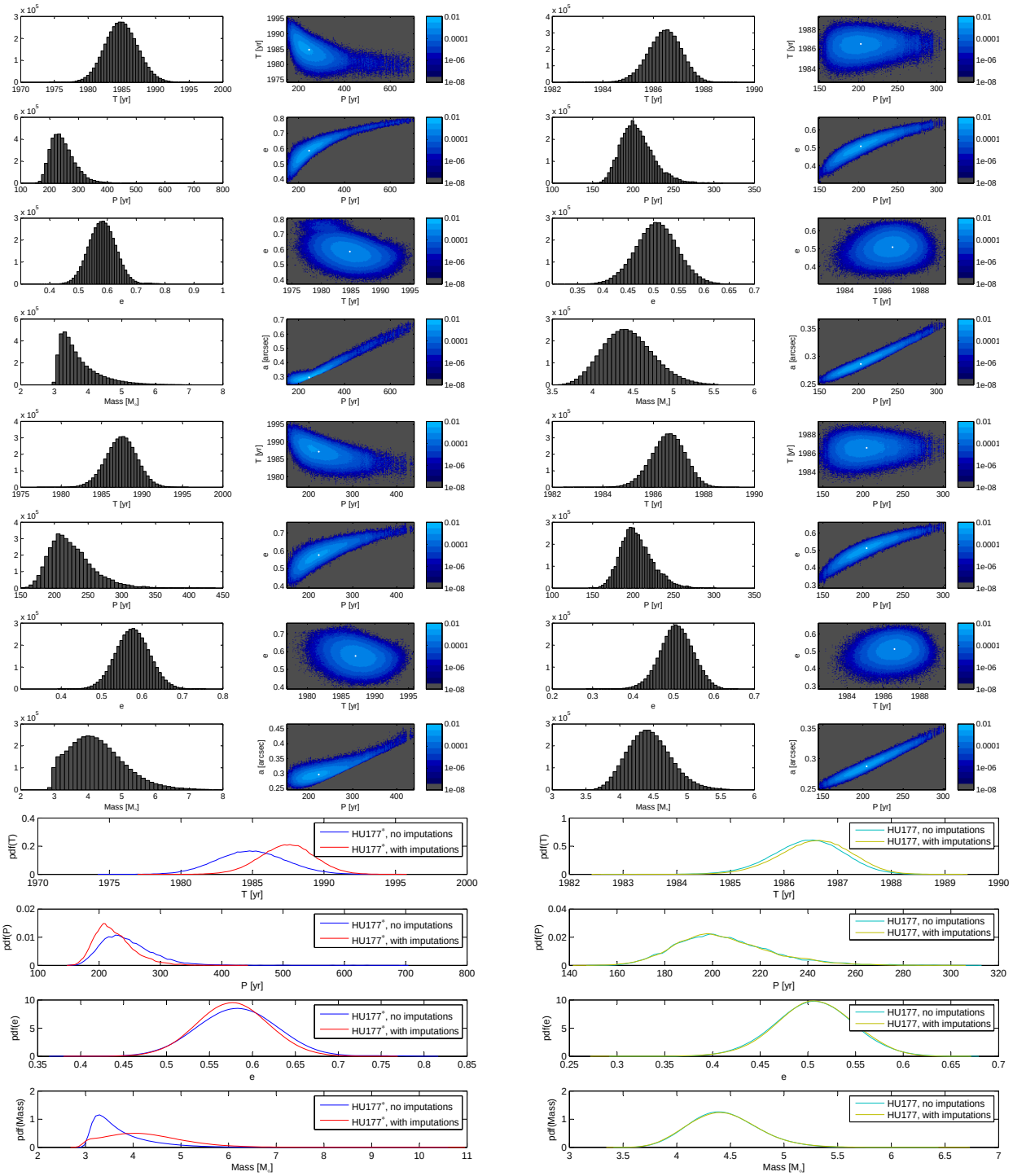

Figure 5.4 PDFs of orbital parameters of the four cases reported in Table 5.3: No MI, **HU177*** (first column, first row); **HU177***, MI (first column, second row); **HU177**, no MI (second column, first row); **HU177**, MI (second column, second row). Lower panels show a comparison between the marginal PDFs obtained for the **HU177*** data set (first column, third row), and those obtained for **HU177** data set (second column, third row).

Table 5.3 Estimation of orbital parameters in presence of incomplete information (**HU177**)

Data set		P (yr)	T (yr)	e	a $''$	ω (°)	Ω (°)	i (°)	Mass_{T} ($M_{\circ}$)
HU177*	No MI[a]	$239.85^{270.13}_{215.61}$	$1984.91^{1986.56}_{1983.30}$	$0.5826^{0.6141}_{0.5513}$	$0.2987^{0.3167}_{0.2863}$	$131.47^{179.43}_{82.47}$	$64.76^{113.81}_{10.70}$	$154.65^{160.65}_{148.03}$	$3.546^{3.976}_{3.300}$
		54.5176	3.2568	0.0628	0.0303	96.9656	103.1041	12.6236	0.6758
HU177*	MI[b]	$220.00^{242.75}_{202.22}$	$1987.47^{1988.72}_{1986.18}$	$0.5769^{0.6050}_{0.5490}$	$0.2986^{0.3099}_{0.2890}$	$84.93^{89.05}_{80.68}$	$12.11^{17.13}_{8.32}$	$145.66^{152.01}_{140.52}$	$4.171^{4.752}_{3.658}$
		40.5331	2.5481	0.0561	0.0209	8.3685	8.8175	11.4864	1.0939
HU177	No MI[a]	$202.79^{216.81}_{190.83}$	$1986.47^{1986.90}_{1986.01}$	$0.5067^{0.5340}_{0.4792}$	$0.2869^{0.2958}_{0.2793}$	$237.18^{242.94}_{230.50}$	$165.52^{170.09}_{160.14}$	$151.24^{154.30}_{148.41}$	$4.424^{4.645}_{4.220}$
		25.9799	0.8856	0.0548	0.0164	12.4439	9.9519	5.8927	0.4246
HU177	MI[b]	$202.28^{216.27}_{190.69}$	$1986.60^{1987.04}_{1986.14}$	$0.5077^{0.5352}_{0.4807}$	$0.2865^{0.2953}_{0.2792}$	$239.53^{245.02}_{233.05}$	$167.77^{172.06}_{162.67}$	$150.43^{153.42}_{147.67}$	$4.426^{4.649}_{4.216}$
		25.5799	0.8947	0.0545	0.0161	11.9774	9.3903	5.7510	0.4331

a. Multiple imputation strategy not applied, incomplete data unused.
b. Incomplete data incorporated through multiple imputation strategy.

Chapter 6

Final remarks

This dissertation presents a Bayesian framework for the estimation and uncertainty characterization of orbital elements of binary stars. Relying on Markov Chain Monte Carlo as a sampling tool, the algorithms proposed in this work achieve the aim of characterizing the posterior PDFs of orbital parameters given a set of measurements. In Chapter 4, both visual (Sect.ion 4.4) and spectroscopic (Section 4.5) observations of stars are analyzed by means of the MCMC routines developed, obtaining a characterization of the posterior distribution of the target parameters in the form of a collection of samples; from these collections, estimators such as maximum a posteriori (MAP) and expected value can be calculated and proposed as candidate orbits. By means of a dimensionality reduction developed in Subsection 4.2.1, the techniques presented take a step ahead in terms of efficiency, since the number of parameters required to be explored decrease from 7 to 3 for visual binaries (similar approaches have been applied before, however, in [Hartkopf et al., 1989], [Lucy, 2014], among others), and from 10 to 7 for spectroscopic binaries (using a formalism entirely developed in this work). Beyond the agreement with the objectives stated in Chapter 1, and the support given to hypotheses 1 and 2 declared in the same chapter, it is worth noting that the contribution of this work is not merely methodological: the framework was successfully applied for analyzing real data of a selection of objects observed by the SOAR telescope, obtaining orbital solutions that range from about 1 yr of orbital period in the case of spectroscopic stars, to more than 500 yr in the case of some visual binaries, providing inputs for actual astrophysical discussion. These orbits and their respective stellar masses–arguably the final product of this work–are presented and analyzed in a forthcoming publication [Mendez et al., 2017].

The potential of combining sources, i.e., astrometric and spectroscopic simultaneously, to estimate hypothesis-free parallaxes is explored in Section 4.5, and is certainly proven useful when both the radial velocity profile and the apparent orbit of the object are well sampled. However, if data is insufficient, a better option is to restrict the value of ϖ by introducing a prior distribution of it based on the measured trigonometric parallax.

Chapter 5 introduces a scheme for coping with partial measurements in the Bayesian framework proposed in Chapter 4. The scheme, based on filling missing or partial measurements with a set of plausible values–the so-called multiple imputation approach–, is tested with both synthetic and real measurements of visual binaries. Although results suggest that incorporation of partial information can lead to a decrease of uncertainty in the estimation of target parameters, there are also cases where partial measurements provide no additional information (posterior PDFs may not change after the incorporation of partial information into the analysis), or can, potentially, worsen the quality of estimation, since the process of imputing values also introduces some degree of uncertainty. Although partial astrometric measurements intrinsically have a degree of spatial dispersion, the tests performed suggest that an interesting improvement in estimation might arise from restricting their geometric boundaries during the image resolution process, thus feeding estimation routines as the one presented in this work with more specific information than that typically provided.

Future work points in the direction of addressing problems that have been already hinted in this work, but have not been given a categorical solution yet:

- As suggested in Chapter 5 in the context of incomplete observations, the use of information-theoretic measures might be a useful tool to evaluate how much the estimation improves (or worsen) after the incorporation of partial measurements. However, the suitability of that approach might extend to a wider set of information-related challenges, for example, cracking the problem of observation planning: how to place future measurements in order to achieve a target level of accuracy of precision about the orbital parameters.

- Constraining the sets of feasible solutions by incorporating additional sources of knowledge of the systems under study, such as spectral type, in the form of prior probabilities. This approach might be crucial in cases of objects with very limited observational coverage (such as **COU499**), where astronomers are prevented from estimating orbits with adequate precision, but one might desire to have more reliable tentative ephemeris for observational planning.
- Adapting the Bayesian approach presented in this work to related but slightly different orbital estimation problems, such as single-line spectroscopic binaries or hierarchical multiple systems. This comprehends extending the basic, complete information-oriented routine presented in Chapter 4, as well as the partial information-oriented algorithm developed in Chapter 5.

Bibliography

[Andrieu et al., 2003] Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to mcmc for machine learning. *Machine learning*, 50(1):5–43.

[Braak, 2006] Braak, C. J. T. (2006). A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249.

[Burgasser et al., 2012] Burgasser, A. J., Luk, C., Dhital, S., Gagliuffi, D. B., Nicholls, C. P., Prato, L., West, A. A., and Lépine, S. (2012). Discovery of a very low mass triple with late-M and T dwarf components: LP 704-48/SDSS J0006–0852AB. *The Astrophysical Journal*, 757(2):110.

[Burgasser et al., 2015] Burgasser, A. J., Melis, C., Todd, J., Gelino, C. R., Hallinan, G., and Gagliuffi, D. B. (2015). Radio emission and orbital motion from the close-encounter star–brown dwarf binary WISE J072003. 20–084651.2. *The Astronomical Journal*, 150(6):180.

[Cid Palacios, 1958] Cid Palacios, R. (1958). On the necessary and sufficient observations for determination of elliptic orbits in double stars. *The Astronomical Journal*, 63:395.

[Damien et al., 1999] Damien, P., Wakefield, J., and Walker, S. (1999). Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[Docobo, 1985] Docobo, J. (1985). On the analytic calculation of visual double star orbits. *Celestial mechanics*, 36(2):143–153.

[Duane et al., 1987] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.

[Eddington, 1924] Eddington, A. S. (1924). On the relation between the masses and luminosities of the stars. *Monthly Notices of the Royal Astronomical Society*, 84:308–332.

[Eggen, 1967] Eggen, O. (1967). Masses of visual binary stars. *Annual Review of Astronomy and Astrophysics*, 5(1):105–138.

[Ford, 2005] Ford, E. B. (2005). Quantifying the uncertainty in the orbits of extrasolar planets. *The Astronomical Journal*, 129(3):1706.

[Gautier III et al., 2012] Gautier III, T. N., Charbonneau, D., Rowe, J. F., Marcy, G. W., Isaacson, H., Torres, G., Fressin, F., Rogers, L. A., Désert, J.-M., Buchhave, L. A., et al. (2012). Kepler-20: A sun-like star with three sub-neptune exoplanets and two earth-size candidates. *The Astrophysical Journal*, 749(1):15.

[Gelfand and Smith, 1990] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

[Gelman et al., 2003] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC.

[Gelman and Rubin, 1992] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6(6):721–741.

[Geweke, 1989] Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339.

[Gillon et al., 2017] Gillon, M., Triaud, A. H., Demory, B.-O., Jehin, E., Agol, E., Deck, K. M., Lederer, S. M., De Wit, J., Burdanov, A., Ingalls, J. G., et al. (2017). Seven temperate terrestrial planets around the nearby ultracool dwarf star trappist-1. *Nature*, 542(7642):456–460.

[Green, 1995] Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, pages 711–732.

[Gregory, 2005] Gregory, P. (2005). A bayesian analysis of extrasolar planet data for hd 73526. *The Astrophysical Journal*, 631(2):1198.

[Gregory, 2007a] Gregory, P. (2007a). A bayesian kepler periodogram detects a second planet in hd 208487. *Monthly Notices of the Royal Astronomical Society*, 374(4):1321–1333.

[Gregory, 2007b] Gregory, P. (2007b). A bayesian periodogram finds evidence for three planets in hd 11964. *Monthly Notices of the Royal Astronomical Society*, 381(4):1607–1616.

[Gregory, 2011] Gregory, P. C. (2011). Bayesian exoplanet tests of a new method for mcmc sampling in highly correlated model parameter spaces. *Monthly Notices of the Royal Astronomical Society*, 410(1):94–110.

[Gregory and Fischer, 2010] Gregory, P. C. and Fischer, D. A. (2010). A bayesian periodogram finds evidence for three planets in 47 ursae majoris. *Monthly Notices of the Royal Astronomical Society*, 403(2):731–747.

[Hartkopf et al., 1989] Hartkopf, W. I., McAlister, H. A., and Franz, O. G. (1989). Binary star orbits from speckle interferometry. ii-combined visual-speckle orbits of 28 close systems. *The Astronomical Journal*, 98:1014–1039.

[Hastings, 1970] Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

[Henry and McCarthy Jr, 1993] Henry, T. J. and McCarthy Jr, D. W. (1993). The mass-luminosity relation for stars of mass 1.0 to 0.08 solar mass. *The Astronomical Journal*, 106:773–789.

[Horch et al., 2017] Horch, E. P., Casetti-Dinescu, D. I., Camarata, M. A., Bidarian, A., van Altena, W. F., Sherry, W. H., Everett, M. E., Howell, S. B., Ciardi, D. R., Henry, T. J., et al. (2017). Observations of binary stars with the differential speckle survey instrument. vii. measures from 2010 september to 2012 february at the wiyn telescope. *The Astronomical Journal*, 153(5):212.

[Horch et al., 2011] Horch, E. P., Gomez, S. C., Sherry, W. H., Howell, S. B., Ciardi, D. R., Anderson, L. M., and Van Altena, W. F. (2011). Observations of binary stars with the differential speckle survey instrument. ii. hipparcos stars observed in 2010 january and june. *The Astronomical Journal*, 141(2):45.

[Horch et al., 2015] Horch, E. P., Van Altena, W. F., Demarque, P., Howell, S. B., Everett, M. E., Ciardi, D. R., Teske, J. K., Henry, T. J., and Winters, J. G. (2015). Observations of binary stars with the differential speckle survey instrument. v. toward an empirical metal-poor mass–luminosity relation. *The Astronomical Journal*, 149(5):151.

[IPAC Caltech, 2017] IPAC Caltech (2017). Exoplanet Archive. [Retrieved on May 05, 2017 from https://exoplanetarchive.ipac.caltech.edu].

[Kabanikhin, 2008] Kabanikhin, S. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, 16(4):317–357.

[Kippenhahn et al., 1990] Kippenhahn, R., Weigert, A., and Weiss, A. (1990). *Stellar structure and evolution*, volume 282. Springer.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., et al. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

[Lailly, 1983] Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In *Conference on inverse scattering: theory and application*, pages 206–220. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[Lavori et al., 1995] Lavori, P. W., Dawson, R., and Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in medicine*, 14(17):1913–1925.

[Levenberg, 1944] Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.

[Little and Rubin, 1987] Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data*. Wiley.

[Loredo et al., 2010] Loredo, T., Liu, B., Clyde, M., and Berger, J. (2010). Bayesian methods for exoplanet science: Planet detection, orbit estimation, and adaptive observing. In *ADA 6-Sixth Conference on Astronomical Data Analysis*, volume 1, page 13.

[Lucy, 2014] Lucy, L. (2014). Mass estimates for visual binaries with incomplete orbits. *Astronomy & Astrophysics*, 563:A126.

[MacKnight and Horch, 2004] MacKnight, M. and Horch, E. (2004). Calculating visual binary star orbits with the downhill simplex algorithm (amoeba). In *Bulletin of the American Astronomical Society*, volume 36, page 788.

[Marquardt, 1963] Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441.

[Mayor and Queloz, 1995] Mayor, M. and Queloz, D. (1995). A jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359.

[Mendez et al., 2017] Mendez, R. A., Claveria, R. M., Orchard, M. E., and Silva, J. F. (2017). Orbits for eighteen visual binaries and two double-line spectroscopic binaries observed with HRCAM on the CTIO SOAR 4m telescope, using a new Bayesian orbit code based on Markov-Chain Monte-Carlo. *The Astronomical Journal*.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

[Metropolis and Ulam, 1949] Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247):335–341.

[Nelson et al., 2014] Nelson, B. E., Ford, E. B., Wright, J. T., Fischer, D. A., von Braun, K., Howard, A. W., Payne, M. J., and Dindar, S. (2014). The 55 cancri planetary system: fully self-consistent n-body constraints and a dynamical analysis. *Monthly Notices of the Royal Astronomical Society*, 441(1):442–451.

[Pascual-Marqui, 1999] Pascual-Marqui, R. D. (1999). Review of methods for solving the eeg inverse problem. *International journal of bioelectromagnetism*, 1(1):75–86.

[Pourbaix, 1994] Pourbaix, D. (1994). A trial-and-error approach to the determination of the orbital parameters of visual binaries. *Astronomy and Astrophysics*, 290:682–691.

[Pourbaix, 1998] Pourbaix, D. (1998). Simultaneous least-squares adjustment of visual and spectroscopic observations of binary stars. *Astronomy and Astrophysics Supplement Series*, 131(2):377–382.

[Pourbaix, 2000] Pourbaix, D. (2000). Resolved double-lined spectroscopic binaries: A neglected source of hypothesis-free parallaxes and stellar masses. *Astronomy and Astrophysics Supplement Series*, 145(2):215–222.

[Robert and Casella, 2004] Robert, C. and Casella, G. (2004). Monte carlo statistical methods. *Springer, New York*.

[Rowe et al., 2014] Rowe, J. F., Bryson, S. T., Marcy, G. W., Lissauer, J. J., Jontof-Hutter, D., Mullally, F., Gilliland, R. L., Issacson, H., Ford, E., Howell, S. B., et al. (2014). Validation of kepler's multiple planet candidates. iii. light curve analysis and announcement of hundreds of new multi-planet systems. *The Astrophysical Journal*, 784(1):45.

[Rubin, 1987] Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

[Rubin, 1976] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, pages 581–592.

[Rubinstein and Kroese, 2016] Rubinstein, R. Y. and Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley & Sons.

[Rudy, 1987] Rudy, Y & Messinger-Rapport, B. (1987). The inverse problem in electrocardiography: solutions in terms of epicardial potentials. *Critical reviews in biomedical engineering*, 16(3):215–268.

[Sahlmann et al., 2013] Sahlmann, J., Lazorenko, P., Ségransan, D., Martín, E., Queloz, D., Mayor, M., and Udry, S. (2013). Astrometric orbit of a low-mass companion to an ultracool dwarf. *Astronomy & Astrophysics*, 556:A133.

[Smith et al., 1997] Smith, P. J., Shafi, M., and Gao, H. (1997). Quick simulation: A review of importance sampling techniques in communications systems. *IEEE Journal on Selected Areas in Communications*, 15(4):597–613.

[Storn and Price, 1995] Storn, R. and Price, K. (1995). Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces, berkeley. *CA: International Computer Science Institute*.

[Storn and Price, 1997] Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

[Swendsen and Wang, 1986] Swendsen, R. H. and Wang, J.-S. (1986). Replica monte carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607.

[Tanner and Wong, 1987] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.

[Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.

[Thiele, 1883] Thiele, T. N. (1883). Neue methode zur berechung von doppelsternbahnen. *Astronomische Nachrichten*, 104:245.

[Tierney, 1994] Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, pages 1701–1728.

[Tinney et al., 2003] Tinney, C. G., Butler, R. P., Marcy, G. W., Jones, H. R., Penny, A. J., McCarthy, C., Carter, B. D., and Bond, J. (2003). Four new planets orbiting metal-enriched stars. *The Astrophysical Journal*, 587(1):423.

[Tokovinin, 1992] Tokovinin, A. (1992). Complementary approaches to double and multiple star research. *IAU Colloquium 135, ASP Conference Series, ed, HA McAlister & WI Hartkopf,(ASP, San Francisco)*, 32:573.

[Tokovinin et al., 2010] Tokovinin, A., Mason, B. D., and Hartkopf, W. I. (2010). Speckle interferometry at the Blanco and SOAR telescopes in 2008 and 2009. *The Astronomical Journal*, 139(2):743.

[Tokovinin et al., 2014] Tokovinin, A., Mason, B. D., and Hartkopf, W. I. (2014). Speckle interferometry at SOAR in 2012 and 2013. *The Astronomical Journal*, 147(5):123.

[Tokovinin et al., 2015] Tokovinin, A., Mason, B. D., Hartkopf, W. I., Mendez, R. A., and Horch, E. P. (2015). Speckle interferometry at SOAR in 2014. *The Astronomical Journal*, 150(2):50.

[Tokovinin et al., 2016] Tokovinin, A., Mason, B. D., Hartkopf, W. I., Mendez, R. A., and Horch, E. P. (2016). Speckle interferometry at SOAR in 2015 based on observations obtained at the southern astrophysical research (SOAR) telescope. *The Astronomical Journal*, 151(6):153.

[Von Neumann, 1951] Von Neumann, J. (1951). Various techniques used in connection with random digits. *National Bureau of Standards Applied Math Series*, 12:36–38.

[Vrugt et al., 2009] Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., and Higdon, D. (2009). Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290.

[Wright and Howard, 2009] Wright, J. and Howard, A. (2009). Efficient fitting of multiplanet keplerian models to radial velocity and astrometry data. *The Astrophysical Journal Supplement Series*, 182(1):205.

[Wunsch, 1996] Wunsch, C. (1996). *The ocean circulation inverse problem*. Cambridge University Press.

[Yuan, 2010] Yuan, Y. C. (2010). Multiple imputation for missing data: Concepts and new development (version 9.0). *SAS Institute Inc, Rockville, MD*, 49:1–11.

[Zhang, 2003] Zhang, P. (2003). Multiple imputation: theory and method. *International Statistical Review*, 71(3):581–592.

Appendix A

Derivation of radial velocity equation

Consider the expression for position's radial component z as a starting point:

$$z = r\sin(\omega + \nu)\ \sin i. \tag{A.1}$$

Since inclination i does not vary over time, it can be left out from the derivative:

$$\dot{z} = \frac{dz(t)}{dt} = \sin i \cdot \frac{d}{dt}\left(r\sin(\omega+\nu)\right). \tag{A.2}$$

Using the product rule for derivatives, we obtain:

$$\dot{z} = \sin i \cdot (\dot{r}\sin(\omega+\nu) + r\cos(\omega+\nu)\dot{\nu}) \tag{A.3}$$

It is left to find closed-form expressions for $\dot{r}$ and $\dot{\nu}$. Since term $a(1-e^2)$ does not time-dependent, the expression for $\dot{r}$ can be obtained as:

$$\begin{aligned}\dot{r} = \frac{d\ r(t)}{dt} = &= \frac{d}{dt}\left(\frac{a(1-e^2)}{1+e\cos\nu}\right) \\ &= a(1-e^2)\cdot\frac{d}{dt}\left(1+e\cos\nu\right)^{-1} \\ &= a(1-e^2)\cdot(1+e\cos\nu)^{-2}\cdot -1\cdot(-e\sin\nu)\ \dot{\nu}.\end{aligned} \tag{A.4}$$

Rearranging the equation above and identifying term r, the expression for $\dot{r}$ can be reduced to:

$$\begin{aligned}\dot{r} &= \frac{a(1-e^2)}{1+e\cos\nu}\cdot\frac{e\sin\nu\ \dot{\nu}}{1+e\cos\nu} \\ &= r\ \frac{e\sin\nu\ \dot{\nu}}{1+e\cos\nu}.\end{aligned} \tag{A.5}$$

To find $\dot{\nu}$, we make use of Kepler's Second Law ("the line that connect both bodies sweeps equal areas in equal times"). An interpretation of that law is that the derivative of the swept area is constant. If A denotes the swept area, integration over one period P yields:

$$\frac{dA}{dt} = \frac{1}{2}r^2\dot{\nu} \rightarrow \frac{A}{P} = \frac{1}{2}r^2\dot{\nu}. \tag{A.6}$$

Since the area of an ellipse is calculated as $A = \pi ab$ (b being the semi-minor axis) and $b = a\sqrt{1-e^2}$, the equation above can be expressed as:

$$\frac{A}{P} = \frac{\pi a^2\sqrt{1-e^2}}{P} = \frac{1}{2}r^2\dot{\nu}, \tag{A.7}$$

from where we isolate quantity $r\dot{\nu}$. The aforementioned quantity is preferred over $\dot{\nu}$ for convenience, since it can

be directly substituted in equations A.5 and A.3 and lead to the final result.

$$\begin{aligned} r\dot{\nu} &= r^{-1} \cdot \frac{2\pi a^2\sqrt{1-e^2}}{P} \\ &= \frac{1+e\cos\nu}{a(1-e^2)} \cdot \frac{2\pi a^2\sqrt{1-e^2}}{P} \\ &= \frac{2\pi a(1+e\cos\nu)}{P\sqrt{1-e^2}}. \end{aligned} \tag{A.8}$$

Substituting $r\ \dot{\nu}$ in Equation A.5 yields:

$$\begin{aligned} \dot{r} &= r\ \frac{e\sin\nu\ \dot{\nu}}{1+e\cos\nu} \\ &= r\dot{\nu} \cdot \frac{e\sin\nu}{1+e\cos\nu} \\ &= \frac{2\pi a\ e\sin\nu}{P\sqrt{1-e^2}} \end{aligned} \tag{A.9}$$

Using equations A.8, A.9 in A.3 gives:

$$\begin{aligned} \dot{z} &= \sin i \cdot (\dot{r} \cdot \sin(\omega+\nu) + (r\dot{\nu}) \cdot \cos(\omega+\nu)) \\ &= \sin i \cdot \left(\frac{2\pi a\ e\sin\nu}{P\sqrt{1-e^2}} \cdot \sin(\omega+\nu) + \frac{2\pi a(1+e\cos\nu)}{P\sqrt{1-e^2}} \cdot \cos(\omega+\nu) \right) \\ &= \frac{2\pi a\ \sin i}{P\sqrt{1-e^2}} \cdot (e\sin\nu\ \sin(\omega+\nu) + (1+e\cos\nu)\cos(\omega+\nu)), \end{aligned} \tag{A.10}$$

which, with the aid of the trigonometric identities shown in Equation A.12, leads to the final expression for radial velocity:

$$V_r = \dot{z} = \frac{2\pi a\sin i}{P\sqrt{1-e^2}} \left[\cos(\omega+\nu) + e\cos\omega\right]. \tag{A.11}$$

$$\begin{aligned} \cos(\nu+\omega) &= \cos\nu\ \cos\omega - \sin\nu\ \sin\omega \\ \sin(\nu+\omega) &= \sin\nu\ \cos\omega + \cos\nu\ \sin\omega \end{aligned} \tag{A.12}$$

Appendix B

On Thiele-Innes and Campbell elements

B.1 Least-squares estimate

The starting point is the sum of individual errors:

$$\sum_{k=1}^{N_x} \frac{1}{\sigma_x^2(k)}[X(k) - X^{model}(k)]^2 + \sum_{k=1}^{N_y} \frac{1}{\sigma_y^2(k)}[Y(k) - Y^{model}(k)]^2 \tag{B.1}$$

Equation 2.8 enables us to replace X_{model}, Y_{model} with their analytic expression for any epoch (indexed by k):

$$\begin{aligned} X_{obs}(k) - X_{model}(k) &= X_{obs}(k) - [B \cdot x(k) + G \cdot y(k)] \\ Y_{obs}(k) - Y_{model}(k) &= Y_{obs}(k) - [A \cdot x(k) + F \cdot y(k)]. \end{aligned} \tag{B.2}$$

Given the linear dependency of X_{model}, Y_{model} with respect to normalized coordinates x, y, it is possible to calculate the least-squares estimate for unknown variables B, G, A, F in a non-iterative way. Moreover, first term of Equation B.1 depends only on (B, G), whereas second term depends on (A, F). Therefore, the estimate for (B, G) is obtained by minimizing the first term and the estimate for (A, F) by minimizing the second one. The problem is thus reduced to a pair of uncoupled linear equations. By calculating the derivatives of the expression of the error with respect to each of the Thiele-Innes constants and making the results equal to zero, one can obtain the following formulae (for the sake of briefness, a set of auxiliary terms is introduced first):

$$\begin{aligned} \alpha &= \sum_i w_i\ x(i)^2 & \beta &= \sum_i w_i\ y(i)^2 & \gamma &= \sum_i w_i\ x(i)\ y(i) \\ & & r_{11} &= \sum_i w_i\ X_{obs}(i)\ x(i) & r_{12} &= \sum_i w_i\ X_{obs}(i)\ y(i) \\ & & r_{21} &= \sum_i w_i\ Y_{obs}(i)\ x(i) & r_{22} &= \sum_i w_i\ Y_{obs}(i)\ y(i) \end{aligned} \tag{B.3}$$

Then, the least-squares estimate for the Thiele-Innes constant is calculated as follows:

$$\begin{aligned} \hat{B} &= \frac{\beta \cdot r_{11} - \gamma \cdot r_{12}}{\Delta}, & \hat{G} &= \frac{\alpha \cdot r_{12} - \gamma \cdot r_{11}}{\Delta}, \\ \hat{A} &= \frac{\beta \cdot r_{21} - \gamma \cdot r_{22}}{\Delta}, & \hat{F} &= \frac{\alpha \cdot r_{22} - \gamma \cdot r_{21}}{\Delta}, \end{aligned} \tag{B.4}$$

where $\Delta = \alpha \cdot \beta - \gamma^2$. An alternative matrix representation is given as follows:

$$\begin{bmatrix} \hat{B} \\ \hat{G} \\ \hat{A} \\ \hat{F} \end{bmatrix}^T = [\vec{X}_{obs}\vec{Y}_{obs}]\ \mathbf{WF}^T(\mathbf{FWF}^T)^{-1}, \text{ where } \mathbf{F} = \begin{bmatrix} x(1) & \dots & x(N) & 0 & \dots & 0 \\ y(1) & \dots & y(N) & 0 & \dots & 0 \\ 0 & \dots & 0 & x(1) & \dots & x(N) \\ 0 & \dots & 0 & y(1) & \dots & y(N) \end{bmatrix}, \tag{B.5}$$

where $\{x(k), y(k)\}_{k=1,\ldots,N}$ are the normalized coordinates given P, T, e and epochs $\{\tau_k\}_{k=1,\ldots,N}$. Matrix $\mathbf{W}$ is a diagonal matrix containing the weight of each observation (usually the reciprocal of the observational error variance).

B.2 Conversion from Thiele-Innes to Campbell

Once that the estimates for Thiele-Innes constants are obtained ($\hat{B}$, $\hat{G}$, $\hat{A}$, $\hat{F}$), it is necessary to recover the equivalent Campbell elements representation (a, ω, Ω, i). For ω and Ω, one must solve the following equations:

$$\begin{aligned} \omega + \Omega &= \arctan\left(\frac{B-F}{A+G}\right), \\ \omega - \Omega &= \arctan\left(\frac{-B-F}{A-G}\right), \end{aligned} \tag{B.6}$$

choosing the solution that satisfies that $\sin \omega + \Omega$ has the same sign as $B - F$ and that $\sin \omega - \Omega$ has the same sign as $-B - F$. If that procedure outputs a value of Ω that does not satisfy the convention that $\Omega \in (0, \pi)$, it must be corrected in the way explained next: if $\Omega < 0$, values of ω and Ω are modified as $\omega = \pi + \omega$, $\Omega = \pi + \Omega$; whereas if $\Omega > \pi$, values of ω and Ω are modified as $\omega = \omega - \pi$, $\Omega = \Omega - \pi$.

For semi-major axis a and inclination i, the following auxiliary variables must be calculated first:

$$\begin{aligned} k &= \frac{A^2 + B^2 + F^2 + G^2}{2}, \\ m &= A \cdot G - B \cdot F, \\ j &= \sqrt{k^2 - m^2}. \end{aligned} \tag{B.7}$$

Then, a and i are determined with the following formulae:

$$\begin{aligned} a &= \sqrt{j + k} \\ i &= \arccos\left(\frac{m}{a^2}\right) \end{aligned} \tag{B.8}$$

Appendix C

Dependency between $(\sigma_\rho, \sigma_\theta)$ and (σ_x, σ_y)

In this work, the basic assumption about observational error is that the difference between the real position (in the apparent orbit) and the observed value is Gaussian-distributed for both axis X and axis Y. This is the reason why the algorithms proposed use Cartesian coordinates to represent the relative position of binary stars. Furthermore, the (X, Y) representation avoids the drawbacks derived from the "cyclic" nature[1] of the angular values –for example, angles near 0 have a similar geometrical meaning to those close to 2π, but the numerical difference between them does not give an account of that.

In data sets of real binary stars, observations may be provided as (X, Y) as well as (ρ, θ) pairs. Data from modern, interferometric telescopes is usually presented in polar coordinates. Although transformation from one representation to the other is straightforward when it comes to particular points (Equation C.1), translating the error characterization of one representation (say, σ_ρ, σ_θ) into an error characterization of the other (σ_x, σ_y) is somewhat more complicated.

$$\begin{aligned} x &= \rho\cos\theta & \rho &= \sqrt{x^2+y^2} \\ y &= \rho\sin\theta & \theta &= \arctan\frac{y}{x} \end{aligned} \tag{C.1}$$

Let X_0, Y_0 be a real value of the apparent orbit (i.e., unaffected by noise). If X_{obs}, Y_{obs} are the coordinates of an observation of X_0, Y_0, then:

$$\begin{aligned} X_{obs} &= X_0 + v_x, \\ Y_{obs} &= Y_0 + v_y, \end{aligned} \tag{C.2}$$

where v_x, v_y are random variables that follow a normal distribution with mean value $\mu = 0$ and standard deviations σ_x, σ_y (it is assumed that $\sigma_{xy} = 0$). By applying the transformations of Equation C.1 to X_{obs}, Y_{obs}, we obtain:

$$\begin{aligned} \rho_{obs} &= \sqrt{X_0^2 + v_x^2 + Y_0^2 + v_y^2 + 2X_0v_x + 2Y_0v_y}, \\ \theta_{obs} &= \arctan\frac{Y_0 + v_y}{X_0 + v_x}. \end{aligned} \tag{C.3}$$

One may want to express the terms above as $\rho_{obs} = \rho_0 + v_\rho$ and $\theta_{obs} = \theta_0 + v_\theta$; however, the non-linearity of the formulae obtained does not allow to isolate the "real" and the noise component analytically. Similarly, if the starting point is an error characterization of ρ and θ:

$$\begin{aligned} \rho_{obs} &= \rho_0 + v_\rho, \text{ with } v_\rho \sim \mathcal{N}(0, \sigma_\rho), \\ \theta_{obs} &= \theta_0 + v_\theta, \text{ with } v_\theta \sim \mathcal{N}(0, \sigma_\theta), \end{aligned} \tag{C.4}$$

[1] Angle θ is often confined to $[0, 2\pi)$, with addition and subtraction operations being performed according to modular arithmetic.

then the transformations of Equation C.1 lead to:

$$X_{obs} = (\rho_0 + v_\rho) \cdot \cos(\theta_0 + v_\theta), \qquad (C.5)$$
$$Y_{obs} = (\rho_0 + v_\rho) \cdot \sin(\theta_0 + v_\theta).$$

As in Equation C.3, terms X_{obs} and Y_{obs} obtained from ρ and θ cannot be easily expressed as the addition of a "real" and a noise component. As a way to address that problem, a numerical approach is proposed. The results contained in the following paragraphs are presented as an insight.

In order to explore the dependency of σ_ρ, σ_θ relative to σ_x, σ_y, magnitude of separation ρ and position angle θ, a set of representative points was selected in the Cartesian plane, with varying values of ρ (three fixed levels of separation) and θ (chosen to cover most part of the circumference). The aim of this selection is to cover a representative array of observational cases. Under the assumption that the error is Gaussian-distributed along axes X and Y, σ_x and σ_y were fixed to 0.01 $[arcsec]$. For each combination of (ρ, θ) (i.e., each point in Figure C.1), a synthetic data set of 4000 points is generated in order to simulate feasible observations around the "real" value. After transforming each of the resulting points to the (ρ, θ) representation, the covariance matrix $\Sigma_{\rho\theta}$ is calculated. Matrix $\Sigma_{\rho\theta}$ shows the following general traits:

1. Cross-covariance term $\sigma_{\rho\theta}$ is in most cases negligible or, at least, one order of magnitude below $\min(\sigma_\rho^2, \sigma_\theta^2)$.
2. As separation ρ increases, θ has less variability (σ_θ decreases as shown in Figure C.2). Thus, being ρ-dependent, error characterization for θ could not be reduced to a single value of σ_θ in a certain data set.
3. Term σ_ρ has values close to σ_x (σ_y) in all cases. The highest percentage error values obtained are around 4% ($err = 100 \times (\sigma_x - \sigma_\rho)/\sigma_x$ [%]).
4. Unlike separation ρ, changes in position angle θ do not have any major consequence on the components of matrix $\Sigma_{\rho\theta}$.

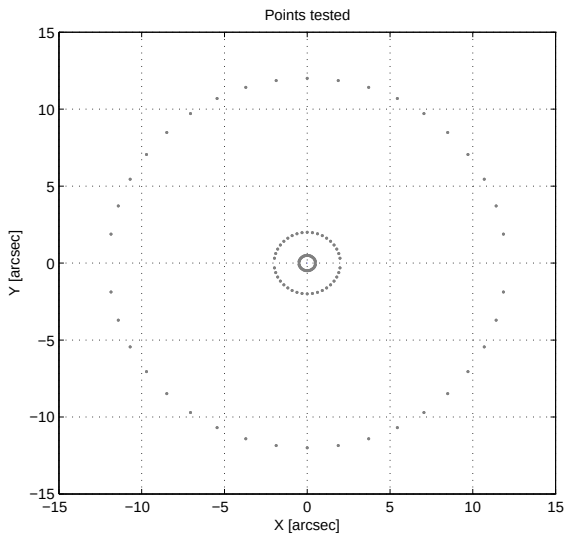

Figure C.1 Points selected for testing.

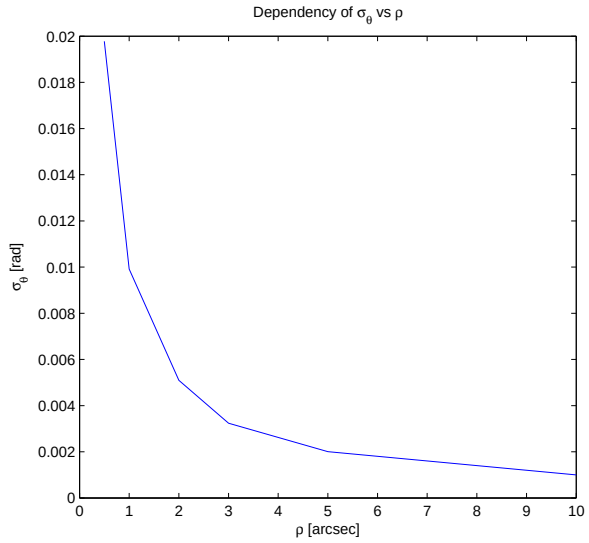

Figure C.2 Dependency between ρ and σ_θ.

The analogous experiment can be performed if the error characterization is provided in terms of σ_ρ and σ_θ. Since position angle is not a major factor in the results, it is fixed to an arbitrary value ($\theta = \pi/12$). Instead, a set of different values of σ_θ is explored: $\sigma_\theta \in \{0.02, 0.010, 0.0050, 0.0035, 0.0020, 0.0010\}$ (note that this set was selected, approximately, according to the values observed in Figure C.2). In this case, a set of 4000 points is generated for each combination of ρ and σ_θ (σ_ρ fixed at 0.01 $[arcsec]$), and then the covariance matrix Σ_{xy} is calculated. Table C.1 shows the numerical estimates of Σ_{xy} for each combination of (ρ, σ_θ). Matrices approximate equal to $\Sigma_{xy} = [\sigma_\rho,\ 0;\ 0,\ \sigma_\rho] = [0.01^2,\ 0; 0,\ 0.01^2]$ are highlighted in gray.

Table C.1 Covariance matrices for different values of ρ, σ_θ

ρ [arcsec] \ σ_θ [rad]	0.02		0.010		0.0050		0.0035		0.0020		0.0010	
0.5	1.001e-04	5.420e-07	9.420e-05	1.818e-05	9.259e-05	2.326e-05	9.384e-05	2.451e-05	9.521e-05	2.532e-05	9.468e-05	2.526e-05
	5.420e-07	9.835e-05	1.818e-05	2.999e-05	2.326e-05	1.257e-05	2.451e-05	9.747e-06	2.532e-05	7.802e-06	2.526e-05	7.015e-06
1	1.219e-04	-7.941e-05	9.822e-05	-3.212e-08	9.523e-05	1.804e-05	9.424e-05	2.143e-05	9.328e-05	2.374e-05	9.237e-05	2.451e-05
	-7.941e-05	3.824e-04	-3.212e-08	9.949e-05	1.804e-05	3.045e-05	2.143e-05	1.777e-05	2.374e-05	1.038e-05	2.451e-05	7.559e-06
2	1.994e-04	-3.702e-04	1.168e-04	-7.428e-05	9.702e-05	1.755e-07	9.683e-05	1.320e-05	9.604e-05	2.123e-05	9.524e-05	2.446e-05
	-3.702e-04	1.481e-03	-7.428e-05	3.835e-04	1.755e-07	1.004e-04	1.320e-05	5.225e-05	2.123e-05	2.133e-05	2.446e-05	1.065e-05
3	3.336e-04	-8.728e-04	1.553e-04	-2.061e-04	1.074e-04	-3.183e-05	1.010e-04	-4.544e-06	9.353e-05	1.548e-05	9.552e-05	2.338e-05
	-8.728e-04	3.386e-03	-2.061e-04	8.334e-04	-3.183e-05	2.196e-04	-4.544e-06	1.123e-04	1.548e-05	3.901e-05	2.338e-05	1.520e-05
5	7.547e-04	-2.460e-03	2.576e-04	-5.876e-04	1.380e-04	-1.376e-04	1.140e-04	-5.008e-05	9.896e-05	-8.168e-07	9.638e-05	1.859e-05
	-2.460e-03	9.365e-03	-5.876e-04	2.296e-03	-1.376e-04	5.908e-04	-5.008e-05	2.900e-04	-8.168e-07	1.003e-04	1.859e-05	3.033e-05
10	2.776e-03	-1.001e-02	7.699e-04	-2.501e-03	2.608e-04	-6.088e-04	1.743e-04	-2.769e-04	1.185e-04	-7.374e-05	9.737e-05	-5.853e-07
	-1.001e-02	3.767e-02	-2.501e-03	9.454e-03	-6.088e-04	2.355e-03	-2.769e-04	1.111e-03	-7.374e-05	3.864e-04	-5.853e-07	1.004e-04

Table C.1 indicates that not all combinations of ρ and σ_θ generate distributions that can be considered symmetrical or with uncorrelated components. In fact, the only matrices that exhibit such properties are those whose values of σ_θ and ρ are close to the curve shown in Figure C.2. Intuitively, the reason of this is that if ρ is too large, small variations in the position angle would produce significant changes in position, so the variability of θ must be confined to a narrow range in order to keep the empirical distribution of X, Y similar to that used in the first experiment.

Results suggest that if, for a given data set, there exist a characterization of the observational error of ρ, then σ_ρ may be a good estimate of σ_x, σ_y. A more complex option is to calculate Σ_{xy} from a set of the synthetic data (points drawn from the statistical information available, i.e., σ_ρ, σ_θ; as done throughout this section).

Appendix D

Auxiliary tables

Table D.1 Additional information (visual binaries)

WDS label	Discoverer designation	HIP label	Gr[a]	Orbit reference[b]
16115+0943	FIN 354	79337	3→ 3	Doc2013d
17305−1446	HU 177	85679	5→3	USN2002
17313+1901	COU 499	85740	5→4	Cou1999b
17533−3444	BU 1123	87567	5→4	Doc1991b
18003+2154	A 1374AB	1566-1708-1	5→5	USN2002
18108−3529	B 1352	89076	5→4	USN2002
18191−3509	OL 18	89766	X→5	NEW
18359+1659	STT 358AB	91159	4→4	Hei1995
18537−0533	A 93	92726	5→4	Hei1998
18558+0327	A 2192	92909	3→3	Doc1988c
19027−0043	STF 2434BC	93519	5→5	Alz1998a
19350+2328	A 162	96317	3→3	Ole1994
20073−5127	RST 1059	99114	X→4	NEW
20514−0538	STF 2729AB	102945	2→2	RAO2015
20597−5211	I 669AB	103620	X→3	NEW
21504−5818	HDS 3109	107806	X→4	NEW
22156−4121	CHR 187	109908	3→3	Tok2015c
23171−1349	BU 182AB	114962	4→3	Hei1991

a. Grades are defined in http://ad.usno.navy.mil/wds/orb6/wdsref.html, with 1 being binaries with final orbits and 5 binaries with great uncertainty of orbital parameters.

b. References taken from the Sixth Catalog of Orbits of Visual Binary Stars, available at http://ad.usno.navy.mil/wds/orb6/wdsref.html

Table D.2 Compiled photometry

WDS	**HIP**	V_{Sim}[a]	V_{Hip}[b]	**Source**$_{V_{Hip}}$[c]	$(V-I)_{Hip}$[b]	**Source**$_{(V-I)_{Hip}}$[d]	V_P[e]	V_S[e]	V_{Sys}	ΔI[f]
16115+0943	79337	6.519	6.52	H	0.30 ± 0.01	L	7.19	7.52	6.59	0.7
17305−1446	85679	8.76 ± 0.01	8.72	H	0.51 ± 0.02	L	8.5	9.6	8.16	1.0
17313+1901	85740	8.96 ± 0.01	8.98	G	0.42 ± 0.01	H	8.7	8.7	7.95	0.5
17533−3444	87567	6.17	6.14	G	0.00 ± 0.01	L	6.86	6.92	6.14	0.3
18003+2154	1566-1708-1[g]	8.53 ± 0.01	8.611 ± 0.012[h]	—	—	—	8.9	10.9	8.74	1.2
18099+0307	89000	5.69	5.67	H	0.56 ± 0.01	L	6.1	7.1	5.74	0.7
18108−3529	89076	9.02 ± 0.02	8.98	H	0.74 ± 0.03	L	9.87	9.88	9.12	0.3
18191−3509	89766	8.55 ± 0.02	8.51	H	0.99 ± 0.02	L	9.17	9.73	8.66	0.6
18359+1659	91159	6.21	6.21	H	0.63 ± 0.07	F	6.94	7.08	6.26	0.1
18537−0533	92726	8.78	8.78	G	0.77 ± 0.01	H	9.16	10.15	8.79	1.2
18558+0327	92909	7.610 ± 0.009	7.07	G	0.17 ± 0.01	H	7.73	8.00	7.10	0.4
19027−0043	93519	8.81	8.80	G	0.77 ± 0.00	R	8.44	8.93	7.91	1.2
19350+2328	96317	7.94 ± 0.01	7.93	H	0.23 ± 0.02	L	8.73	8.77	8.00	0.1
20073−5127	99114	8.11 ± 0.01	8.13	H	0.34 ± 0.02	L	8.89	9.03	8.21	0.5
20514−0538	102945	6.07	5.99	G	0.53 ± 0.00	H	6.40	7.43	6.04	1.1
20597−5211	103620	8.32 ± 0.01	8.33	G	0.87 ± 0.00	H	9.01	9.51	8.48	0.4
21504−5818	107806	7.89 ± 0.01	7.89	H	0.78 ± 0.01	L	8.56	9.07	8.03	0.3
22156−4121	109908	4.810	4.79	G	0.83 ± 0.02	A	5.20	6.68	4.95	2.2
22313−0633	111170	6.615	6.15	G	0.64 ± 0.02	A	6.3	8.6	6.18	1.9
23171−1349	114962	8.14	8.16	G	0.62 ± 0.01	A	8.77	9.08	8.16	0.6

a. From SIMBAD.
b. From Hipparcos catalogue.
c. G = ground-based, H=HIP, T=Tycho.
d. 'A' for an observation of $V-I$ in Cousins' system; 'F', 'H' and 'K' when $V-I$ derived from measurements in other bands/photoelectric systems; 'L' when $V-I$ derived from Hipparcos and Star Mapper photometry; 'R' when colors are unknown or uncertain.
e. From WDS.
f. From our own measurements in the I-band. When one more than one, it is the average, excluding uncertain (:) values.
g. Tycho number
h. This is the V_T mag from the Tycho catalogue. No colors provided.