# Near Earth Objects

**Stefan Wolff**

**Ph.D. Thesis**

**September 2005**

# Near Earth Objects

STEFAN WOLFF

Department of Mathematics
Technical University of Denmark

**Title of Thesis:**

Near Earth Objects

**Graduate Student:**

Stefan Wolff[*]

**Supervisors:**

Poul G. Hjorth[*] and Uffe Gråe Jørgensen[†]

**Addresses:**

[*]Department of Mathematics
Technical University of Denmark
Building 303
DK-2800 Kgs. Lyngby
Denmark

[†]Astronomical Observatory
Niels Bohr Institute
Juliane Maries Vej 30
DK-2100 Copenhagen E
Denmark

# Preface

This thesis has been prepared at the Department of Mathematics (MAT) at the Technical University of Denmark (DTU) in partial fulfilment of the requirements for the Ph.D.-degree in the DTU study programme of Mathematics, Physics and Informatics. The work was funded by a Ph.D. scholarship from the Technical University of Denmark.

A significant part of the work was carried out while visiting F. Mignard at the Observatoire de la Côte d'Azur, France, and I would like to thank everyone there for the friendly atmosphere. In particular, I want to thank my host for taking time out of a busy schedule and for numerous rewarding discussions.

I also wish to thank everyone at MAT/DTU; staff and students, always willing to lend an ear and a hand.

Finally, I wish to express particular gratitude towards Poul Hjorth and Uffe G. Jørgensen, my thesis advisors, and towards François Mignard and Christian Henriksen – *sine qua non*.

Stefan Wolff
August 31, 2005

# Summary

> The word *planet* comes from Greek *planētēs*, wanderers, because the planets
> appear to wander across the celestial sphere, contrary to the fixed stars.

This thesis presents several methods for using this motion to distinguish between stars and solar system objects in order to detect and track NEOs, Near Earth Objects: Asteroids and comets following paths that bring them near the Earth. NEOs have collided with the Earth since its formation, some causing local devastation, some causing global climate changes, yet the threat from a collision with a near Earth object has only recently been recognised and accepted.

The European Space Agency mission Gaia is a proposed space observatory, designed to perform a highly accurate census of our galaxy, the Milky Way, and beyond. Through accurate measurement of star positions, Gaia is expected to discover thousands of extra-solar planets and follow the bending of starlight by the Sun, and therefore directly observe the structure of space-time.

This thesis explores several aspects of the observation of NEOs with Gaia, emphasising detection of NEOs and the quality of orbits computed from Gaia observations. The main contribution is the work on motion detection, comprising a comparative survey of five different motion detection tests, one of which is proved to be optimal among all translation invariant and symmetric tests.

# Dansk resumé

## Jordnære objekter

> Ordet *planet* kommer af det græske *planētēs*, vandringsmænd, idet planeterne synes at vandre henover himmelhvælvet, i modsætning til fiksstjernerne.

Denne afhandling præsenterer adskillige metoder der bruger denne bevægelse til at skelne mellem stjerner og objekter fra vort solsystem, med henblik på at opdage og observere NEOer, *Near Earth Objects*: asteroider og kometer, hvis baner fører dem tæt på Jorden. NEOer har kollideret med Jorden siden dens tilblivelse. Nogle har blot forårsaget lokal ødelæggelse, andre har forårsaget globale klimaforandringer, men først for nyligt er NEO-truslen blevet anerkendt og accepteret.

Gaia er et foreslået rumobservatorium (drevet af det europæiske rumagentur ESA) der har til formål at skabe et tredimensionalt stjernekort af hidtil uset nøjagtighed. Baseret på disse nøjagtige positionsmålinger forventes Gaia at opdage tusinder af planeter uden for vort solsystem, samt at følge lysets bøjning forårsaget af Solens tyngdekraft, og herigennem foretage en direkte observation af rum-tidens struktur.

Denne afhandling undersøger adskillige aspekter af observation af NEOer med Gaia, med særlig vægt på detektion af NEOer og kvaliteten af baneparametre beregnet ud fra Gaia-observationer. Det primære bidrag er inden for detektion af bevægelse, og består af en sammenlignende oversigt over fem forskellige metoder til bevægelsesdetektion, hvorunder en vises at være optimal blandt alle symmetriske og translationsinvariante tests.
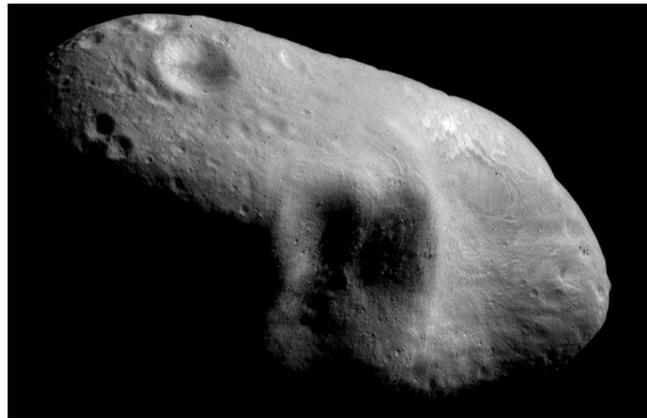
# Contents

# Chapter 1

# Introduction



**Figure 1.1:** Near Earth Asteroid Eros at a distance of 200 km, imaged by the NEAR-Shoemaker spacecraft less than a year before it landed on the asteroid. Eros is a large NEA, measuring about $33 \times 13 \times 13$ kilometers (image courtesy of APL/NASA).

Each year during the recent history of the Earth, an average of approximately $10^8$ kg of meteoritic material has been falling onto it, ranging in size from microscopic dust particles to asteroids several kilometers across [Ceplecha 1992]. Most of the influx comes from bodies more massive than $10^3$ kg. An object whose orbit brings it sufficiently close to that of the Earth, thus having a non-zero long-term probability of impacting it, is called a *Near Earth Object*, abbreviated NEO.

Although the annual probability of the Earth being struck by a large asteroid or comet is extremely small, the consequences of such a collision are so catastrophic that it is prudent to assess the nature of the hazard.

This thesis illuminates some aspects of observing NEOs and proposes elements of a method to facilitate the computation of orbits using data from the Gaia satellite.

## 1.1   Near Earth Objects

NEOs are objects that have been "nudged" out of their stable origin, typically due to gravitational perturbation by one or more of the major objects of the solar system. Depending on that origin, they may be divided into two main categories: Near Earth Comets (NEC) and Near Earth Asteroids (NEA). It is customary to impose a lower size limit, typically a diameter of 50 meters, to distinguish between "space rubble" and NEOs capable of penetrating the Earth's atmosphere.

### Near Earth Asteroids

Near Earth Asteroids originate in the Main Asteroid Belt (or the Main Belt), a region in space between the orbits of Mars and Jupiter. Several hundred thousand asteroids are known and catalogued. The Main Belt extends from about 2.1 Astronomical Units (AU) from the Sun to about 3.3 AU.

Near Earth Asteroids are divided into the following three families of asteroid: Atens, Apollos and Amors. Each family is named after the first asteroid discovered, belonging to that family. Figure 1.2 shows typical orbits for each family.

**Atens**  have semi-major axes smaller than Earth's. Their aphelion distance is larger than that of the Earth. They were named for asteroid 2062 Aten.

**Apollos**  have semi-major axes larger than Earth's. Their perihelion distance is smaller than that of the Earth. These asteroids were named for 1862 Apollo.

**Amors**  have semi-major axes larger than Earth's. Their perihelion distance is between 1.017 AU and 1.3 AU, placing these objects outside the orbit of Earth. The family is named after 1221 Amor. The asteroid 433 Eros (see figure 1.1) belongs to the Amor family of asteroids.

The orbits of Atens and Apollos cross that of the Earth, whereas the orbits of Amors do not. Amors can be considered to be "Earth-approachers", rather than "Earth-crossers". Apollos spend most of their orbital period outside the orbit of the Earth, whereas Atens spend most of theirs inside the orbit of the Earth. Asteroids with orbits always inside that of the Earth (IEO, Inner Earth Object) also exist, but only a few are known, since such objects are difficult to observe from the Earth, always being near the direction of the Sun.

The near Earth asteroids are by far the most frequently observed NEOs.

### Near Earth Comets

Comets come from the outer reaches of our solar system. Hence, Earth orbit crossing comets must have highly elliptical orbits. Like the asteroids, comets are also subdivided
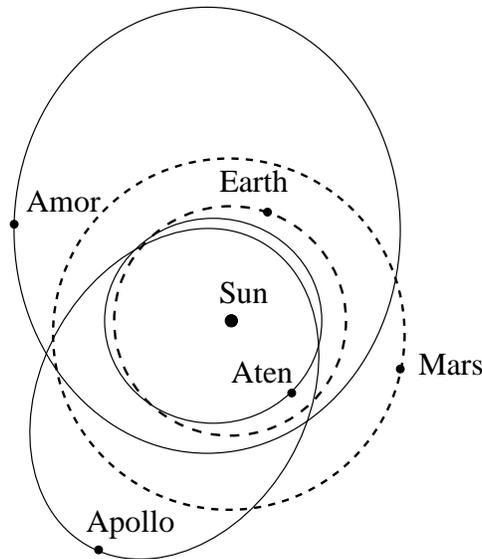
**Figure 1.2:** The orbits of the Earth, Mars and typical Aten, Amor and Apollo asteroids.

into groups: Short-period comets and long-period comets. The former group is believed to originate in the Edgeworth-Kuiper belt, a region of space just beyond the orbit of Neptune, now known to be occupied by thousands of icy bodies. The Edgeworth-Kuiper belt is thought to be a thick band around the ecliptic at a distance between 30 AU and 50 AU [Allen 2001] from the Sun. The long-period comets are believed to originate in the Oort Cloud, a region of space much farther away (50,000 AU) from the Sun. As opposed to the Edgeworth-Kuiper Belt, observations of long-period comets show no preferential direction of origin, suggesting that their region of origin is of spherical, rather than toroidal shape. Comets having an orbital period shorter than 200 Earth years are considered short-period comets, comets having an orbital period longer than 200 Earth years are considered long-period comets. The comets coming close to the orbit of the Earth are presumably perturbed out of the stable Edgeworth-Kuiper Belt or Oort Cloud orbits by the resonant gravitational influence of the giant outer planets.

## 1.2   Impact Risk and Consequence

The Minor Planet Center of the International Astronomical Union considers an object "potentially hazardous" when its minimum orbit intersection distance (MOID, see glossary in appendix A) is less than 0.05 AU (about 20 lunar distances) *and* the absolute magnitude (see section 2.2) of this object is $H < 22$, roughly corresponding to a diameter of 150 m or larger. As of August 2005, there are more than 700 known *Potentially Hazardous Objects* (PHO) according to NASA/JPL[1]. There are more than 150 PHOs of

---

[1]http://neo.jpl.nasa.gov/

absolute magnitude $H < 18$, roughly corresponding to a diameter of 1 km or larger. The minimum orbit intersection distance threshold corresponds roughly to the maximum orbit perturbation that could be caused by the gravitational influence of other solar system objects within the next century [Virtanen 2005].

To estimate the order of magnitude of the amounts of energy involved in an impact, assume a small spherical NEO of 50 m diameter and a density of $\rho = 2\text{g/cm}^3$, impacting the Earth. We will assume that all of the kinetic energy is transformed immediately when the NEO strikes the Earth with an impact velocity of 20 km/s (a typical impact velocity according to [Chyba 1991] and [Ceplecha 1992]). The mass of the object is thus approximately 131,000 tons, which yields a kinetic energy of about $3 \times 10^{16}$ J, equivalent to the explosive energy of about 6 megatonnes of TNT, or about 300 Hiroshima bombs.

An object of this size is assumed to have exploded several kilometers above Tunguska, Siberia in 1908, flattening more than 2000 square kilometers of forest. It is estimated that one such object impacts the Earth every few hundred years [NEO Taskforce 2000]. Even small impactors may cause significant damage, albeit only locally. An impact in an extended urban area will cause an enormous death toll. Impacts of global consequences to climate, corresponding to impactors greater than 600 m, are estimated to occur on the average every 70,000 years. A more recent estimate [Morbidelli et al. 2002] points to impact frequencies about one fourth of this, proposing a mean time between impactors greater than 600 m of 240,000 years. Despite their relative rarity, an actuarial assessment estimates that the 2 km objects pose the greatest risk [Ceplecha 1992].

The threat of impacts has only recently been recognised through advances in telescope technology and the collision of fragments of the comet Shoemaker-Levy 9 with Jupiter in 1994. In May 1998, NASA committed to discovering 90% of all kilometer-sized NEOs within ten years, the so-called *Spaceguard Goal*. According to [Jedicke et al. 2003], this goal is not feasible given the current effort. The same paper proposes a space-based survey as a means to achieve the goal, or, alternatively, an immediate significant increase of the limiting magnitude of existing Earth-based survey programmes. Continuing at the level of performance of the period 1999-2000, the authors estimate it would take another $33 \pm 5$ years to reach 90% completeness.

The main task of the European Space Agency mission Gaia is to measure the positions, distances and other physical characteristics of about one billion stars in the Milky Way and beyond. The Gaia satellite is scheduled to launch in year 2011-2012, and will not help achieving the Spaceguard goal within ten years of the 1998 commitment, but will add significantly to our knowledge of NEOs. This thesis explores several aspects of the observation of NEOs with Gaia.

## 1.3 Thesis Organisation and Contributions

Following this brief introduction to near Earth objects, the next chapter will provide an introduction to astrometry and celestial mechanics, with emphasis on three-body orbits and

Lagrange points. Chapter 3 compares the capabilities of the main Earth-based NEO search programmes to those of ESAs space-based survey mission Gaia, described in further detail in chapter 4, exploring the potential for observing NEOs with Gaia. This chapter also contains new results regarding the probability of losing observations of fast-moving objects. The main contribution of this thesis is presented in chapter 5. Four methods of motion detection are introduced and their relative performance analysed.

A fifth, novel, method is also presented and shown to be optimal among all translation invariant methods assuming a symmetric velocity distribution. The relative performance of all five tests is compared, and their individual advantages and disadvantages are discussed. The optimal method is then applied to simulated Gaia observations. Finally, the properties of the velocity estimate emerging from two of the methods are examined with reference to its use in orbit computation. Chapter 5 also describes how motion detection and motion estimation may be used to reduce the workload when linking observations to determine a preliminary orbit. Classical and modern methods for preliminary orbit computation are presented in the penultimate chapter, covering the so-called Gauss-Encke-Morton method and introducing orbit computation by statistical inversion methods. Possible avenues of future work are presented in chapter 7, along with a summary of the work presented in the thesis. Appendix A contains a glossary of relevant terms.

# Chapter 2

# Astrometry and Orbital Dynamics

## 2.1   Coordinate Systems in Astronomy

The most commonly used coordinate systems in astronomy are spherical coordinate systems originating in a *heliocentric* (Sun-centered), *geocentric* (Earth-centered) or *topocentric* (observer-centered) view. The celestial sphere is an imaginary spherical surface on which all the celestial bodies have apparently been placed. In the case of the topocentric coordinate system, the boundary between the visible and invisible parts of the celestial sphere is called the horizon. The poles of the horizon, i.e., the points on the celestial sphere directly overhead and straight down, are called the zenith and nadir, respectively. The celestial sphere appears to rotate about a point in the sky. This point is called the North Celestial Pole for an observer on the Earth's northern hemisphere. For an observer on the southern hemisphere, the corresponding point would be the South Celestial Pole. The great circle intersecting the celestial poles as well as the observer's zenith and nadir is called the celestial meridian.

The planets appear move nearly on the same plane on the celestial sphere. This plane is that of the *ecliptic*: the plane of the Earth's orbit around the Sun. The ecliptic is tilted about $\epsilon = 23.5°$ with respect to the celestial equator. The two points where celestial equator intersects the ecliptic plane are called the equinoxes. The equinox that the Sun appears to pass as it appears to move northward is called the vernal equinox ♈, since this happens near the 21st of March. It is also called the spring equinox. Six months later, the Sun appears to pass the opposite intersection point, called the autumnal equinox. This connecting of the equinoxes to a particular season may be seen as an unfortunate association, as the seasons on the Earth's southern hemisphere are the opposite of those on the Northern hemisphere, e.g., the spring equinox happens during the autumn on the southern hemisphere.

The following sections briefly describe the most commonly used coordinate systems in astronomy.

## The Horizon System

The astronomical horizon is defined as the intersection of the celestial sphere with the plane whose normal is given by the direction of the observer's local gravity field. The direction of this gravity field is called the astronomical vertical and its point of intersection with the celestial sphere is called the astronomical zenith. The definition of the origin of longitudes varies. The altitude $a$ of a point $P$ on the celestial sphere is the angular distance measured positive towards the astronomical zenith from the astronomical horizon along the great circle passing through $P$ and the astronomical zenith. If $P$ is below the astronomical horizon, the altitude $a$ is measured negative from the astronomical horizon towards the astronomical nadir.

The altitude of the North Celestial Pole is the observer's astronomical latitude.

The azimuth $A$ is the angular distance from the origin of longitudes in a clockwise manner (north-east-south-west) along the astronomical horizon to the intersection of the great circle passing through the point $P$ and the astronomical zenith with the astronomical horizon.

In the horizon system, the altitude $a$ is a representation of latitude and the azimuth $A$ is a representation of longitude. The azimuth is ambiguous at the poles.

## The Equatorial System
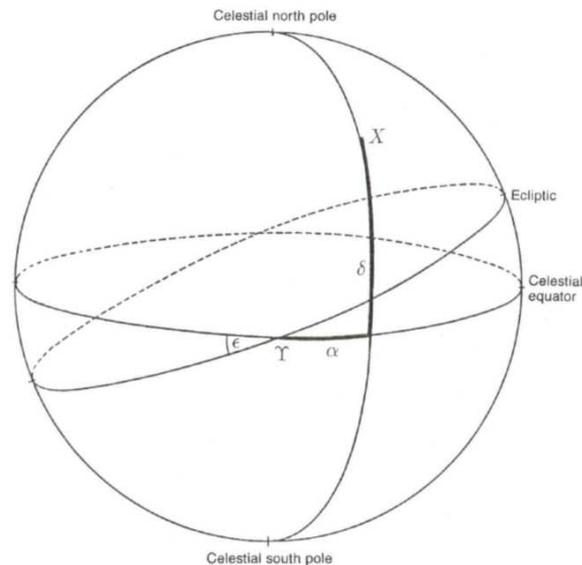


**Figure 2.1:** The equatorial reference system. Positions are designated by their *right ascension $\alpha$* and *declination $\delta$*. From [Danby 1988].

Instead of using the astronomical horizon as the fundamental circle, the equatorial system (figure 2.1) uses the celestial equator, i.e., the great circle, the poles of which are the North and South Celestial Poles, found by extending the Earth's axis of rotation to the celestial sphere. Corresponding to the altitude, there is the declination $\delta$, defined as the angular distance measured positive toward the North Celestial Pole from the celestial equator along the great circle passing through the point in question and the North Celestial Pole. For a point on the south celestial hemisphere, the declination is measured negative toward the South Celestial Pole along the great circle passing through the point in question and the South Celestial Pole.

The right ascension $\alpha$ of the point $P$ is the angular distance from the vernal equinox $\Upsilon$, measured toward the east along the celestial equator to the intersection of the celestial equator, and the great circle passing through the point $P$ and the North Celestial Pole.

## The Ecliptic System
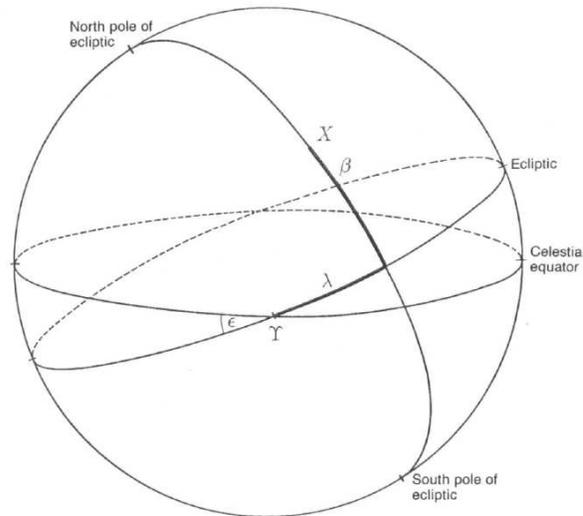


**Figure 2.2:** The ecliptic reference system. Positions are designated by their *ecliptic longitude* $\lambda$ and *ecliptic latitude* $\beta$. From [Danby 1988].

The ecliptic system (figure 2.2) uses the ecliptic as the reference plane. The ecliptic (or celestial) latitude $\beta$ of the point $P$ is the angular distance measured positive toward the north pole of the ecliptic from the ecliptic along the great circle passing through $P$ and the north pole of the ecliptic. The ecliptic latitude $\beta$ of a point $P$ on the southern ecliptic hemisphere is measured negative from the ecliptic toward the south pole of the ecliptic along the great circle passing through $P$ and the north (and south) pole of the ecliptic.

The ecliptic (or celestial) longitude $\lambda$ of the point $P$ is the angular distance measured toward the east, from the vernal equinox $\Upsilon$, along the ecliptic to the intersection of the great circle passing through the points $P$ and the north pole of the ecliptic with the celestial equator.

## 2.2 On Magnitudes

Hipparchus was among the first to classify stars according to their brightness. He divided the visible stars into six classes, the brightest in class 1 and the faintest in class 6. As technological progress has enabled astronomers to observe ever fainter objects, a need to extend and formalise this classification emerged. By introducing a logarithmic scale, such that five steps in magnitude corresponded to a factor of 100 in intensity, a classification embodying and extending the original ancient Greek system was introduced. A magnitude difference of one corresponds to an intensity ratio of $\sqrt[5]{100} \approx 2.51$. In this way, the original classification could be retained while enabling fainter stars to be classified. Since fainter objects have higher magnitudes, very bright objects may have negative magnitudes. In this system, Polaris, the North Star, has a mean magnitude of 2.1, whereas Sirius, one of the brightest stars, has a magnitude of $-1.46$, corresponding to an intensity ratio of $\sqrt[5]{100}^{2.1-(-1.46)} \approx 27$. The intensity of Sirius thus is about 27 times greater than that of Polaris. The magnitude of the full moon is about $-13.6$, and that of the Sun is about $-26.7$. In favourable observing conditions, the faintest objects visible to the naked eye are of magnitude about 6, corresponding to the faintest class recorded by Plato and Hipparchus. This implies that the intensity of the Sun is $\approx 10^{13}$ times greater than the intensity of the faintest stars visible to the naked eye, attesting to the impressive dynamic range of the human visual system. Using the Hubble Space Telescope, stars as faint as magnitude 30 have been observed.

Since all these observations are done on or near the surface of the Earth, this classification is called the apparent (or visual) magnitude, denoted $V$. The absolute magnitude is determined by scaling the magnitude corresponding to positioning the star 10 parsecs (1 parsec equals 3.26 light years) away, thus eliminating the effect of distance. The Sun has an absolute magnitude of 4.8. If the Sun was 10 parsecs away, it would be scarcely visible to the naked eye.

Solar system objects, however, would be practically invisible when placed 10 parsecs away from the observer, so they are normalised at 1 AU. Since these objects do not emit light by themselves, but only reflects light received from a light source (the Sun), they add the complexity of distance to the light source as well as the *phase angle*, the angle between the observer and the light source, as seen from the observed object. The absolute magnitude $H$ of solar system objects is determined by normalising the distance from the observer to the object as well as the distance from the light source to the object to 1 AU, while having a zero phase angle. This corresponds to putting the light source and observer at the same place while observing an object 1 AU away. An important figure relating the

diameter of an object to its absolute magnitude is the *albedo*, the ratio of the reflected light to the received light. Near Earth objects reflect between 3% and 50% of the incident light, depending on taxonomic class. A typical value for a near Earth asteroid is about 15% [Morbidelli et al. 2002].

It has been deemed useful to introduce a special magnitude scale $G$ for use with the Gaia satellite. The relation between $G$ and $V$ depend on the spectrum of the received radiation. For asteroids, having spectral parameter $V - I = 1$ according to [Høg & Knude 2001], $V - G \approx 0.25$ according to the latest design[1]. This means that the Gaia's limiting magnitude (the brightness of the faintest objects fully treated by Gaia) of $G = 20$ corresponds to a visual magnitude of $V \approx 20.25$ for asteroids and NEOs. However, because Gaia's limiting magnitude has not yet been fixed, we will assume $G_{\text{lim}} \approx V_{\text{lim}} \approx 20$ for the remainder of this thesis.

## 2.3   Keplerian Orbits

This section presents and derives Kepler's three famous empirical laws [Danby 1988] and provides an essential basis for chapter 6 on the computation of orbits [Murray & Dermott 1999]. In this and the following chapters, overdot (e.g., $\dot{x}$) refers to differentiation with respect to time $t$. Circumflex ("hat") refers to a normalised vector, e.g., $\hat{\mathbf{x}}$ is a unit vector parallel to $\mathbf{x}$. We will assume masses $m > 0$ and distances $r > 0$. Recall also that the scalar (or dot) product of a vector and itself equals the magnitude squared. The vector (or cross) product of two perpendicular vectors (such as the position and velocity vectors of an object undergoing circular motion) is the product of the magnitudes of these vectors.

### Kepler's Empirical Laws

By meticulously studying Tycho Brahe's observations of the planets, Johannes Kepler discovered the following three laws of planetary motion at the beginning of the 17th century:

1. The orbits of the planets are ellipses, with the Sun at one focus of the ellipse.

2. The line joining the planet to the Sun sweeps out equal areas in equal times as the planet travels around the ellipse.

3. The square of the period of a planet's orbit is proportional to the cube of the semi-major axis of that planet's orbit; the constant of proportionality is the same for all planets.

---

[1]Gaia Parameter Database (restricted access), Astro:AF:Magnitude_VMinG, contains an approximate expression, dated February 2005, for $V - G$ as a power series in $V - I$, derived by C. Jordi.

In the following sections, these three empirical laws will be shown to hold true in a Newtonian universe.

## Two-body dynamics

Assume two particles of mass $m_1$ and $m_2$ are affected only by their mutual gravitational force, inversely proportional to the square of the distance between the particles. The force acting on particle 1 is directed towards particle 2 and vice versa. Satisfying Newton's third law, the forces are of equal magnitude and opposite directions. Letting $\mathbf{r}_{o1}$ and $\mathbf{r}_{o2}$ denote the positions of the particles with respect to some fixed origin in inertial space, and denoting the displacement with $\mathbf{r} = \mathbf{r}_{o2} - \mathbf{r}_{o1}$, the forces acting on the particles may be written

$$
\begin{aligned}
\mathbf{F}_1 = m_1 \ddot{\mathbf{r}}_{o1} &= \mathcal{G} m_1 m_2 \frac{\mathbf{r}}{|\mathbf{r}|^3} \\
\mathbf{F}_2 = m_2 \ddot{\mathbf{r}}_{o2} &= -\mathcal{G} m_1 m_2 \frac{\mathbf{r}}{|\mathbf{r}|^3}.
\end{aligned}
\tag{2.1}
$$

Denoting the sum of the masses $M = m_1 + m_2$, the center of gravity is defined as:

$$
\mathbf{c}_g = \frac{1}{M} (m_1 \mathbf{r}_{o1} + m_2 \mathbf{r}_{o2}) = \frac{m_1}{M} \mathbf{r}_{o1} + \frac{m_2}{M} \mathbf{r}_{o2}
$$

The position vectors $\mathbf{r}_1 = \mathbf{r}_{o1} - \mathbf{c}_g$ and $\mathbf{r}_2 = \mathbf{r}_{o2} - \mathbf{c}_g$ are vectors from the center of gravity to object 1 and 2, respectively.

$$
\begin{aligned}
\mathbf{r}_1 = \mathbf{r}_{o1} - \mathbf{c}_g = \mathbf{r}_{o1} - \left( \frac{m_1}{M} \mathbf{r}_1 + \frac{m_2}{M} \mathbf{r}_2 \right) = \frac{m_2}{M} (\mathbf{r}_{o1} - \mathbf{r}_{o2}) \\
\mathbf{r}_2 = \mathbf{r}_{o2} - \mathbf{c}_g = \mathbf{r}_{o2} - \left( \frac{m_1}{M} \mathbf{r}_1 + \frac{m_2}{M} \mathbf{r}_2 \right) = \frac{m_1}{M} (\mathbf{r}_{o2} - \mathbf{r}_{o1})
\end{aligned}
$$

By differentiating $\mathbf{r}_1$ twice:

$$
\ddot{\mathbf{r}}_1 = \ddot{\mathbf{r}}_{o1} - \ddot{\mathbf{c}}_g = \frac{m_2}{M} (\ddot{\mathbf{r}}_{o1} - \ddot{\mathbf{r}}_{o2}) = \frac{m_2}{M} (\mathcal{G} m_2 + \mathcal{G} m_1) \frac{\mathbf{r}}{|\mathbf{r}|^3} = \ddot{\mathbf{r}}_{o1},
$$

we see that $\ddot{\mathbf{c}}_g = \mathbf{0}$ meaning that the center of gravity does not accelerate. Since both of the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are "attached" to the center of mass, it follows that

$$
\frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2} = \mathbf{0} \Leftrightarrow m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 = \mathbf{0}
\tag{2.2}
$$

This implies

$$
\mathbf{r}_1 = -\frac{m_2}{m_1} \mathbf{r}_2 = -\frac{m_2}{m_1} (\mathbf{r} + \mathbf{r}_1) \Leftrightarrow \mathbf{r}_1 \left( 1 + \frac{m_2}{m_1} \right) = \mathbf{r}_1 \frac{M}{m_1} = -\frac{m_2}{m_1} \mathbf{r}
$$

which gives an expression of $\mathbf{r_1}$ as a function of $\mathbf{r}$:

$$\mathbf{r_1} = -\frac{m_2 m_1}{m_1 M}\mathbf{r} = -\frac{m_2}{M}\mathbf{r}$$

A similar expression may be derived for $\mathbf{r_2}$:

$$\mathbf{r_2} = \frac{m_1}{M}\mathbf{r}$$

Differentiating these equations twice:

$$\ddot{\mathbf{r}}_1 = -\frac{m_2}{M}\ddot{\mathbf{r}}$$

$$\ddot{\mathbf{r}}_2 = \frac{m_1}{M}\ddot{\mathbf{r}}$$

and inserting these expressions in the above differential equations yields:

For object 1:

$$m_1\ddot{\mathbf{r}}_1 = -m_1\frac{m_2}{M}\ddot{\mathbf{r}} = -\frac{\mathcal{G}m_1 m_2}{r^2}\hat{\mathbf{r}}_1 = \frac{\mathcal{G}m_1 m_2}{r^2}\hat{\mathbf{r}}$$

$$\Updownarrow$$

$$\ddot{\mathbf{r}} = -\frac{\mathcal{G}M}{r^2}\hat{\mathbf{r}}$$

And object 2:

$$m_2\ddot{\mathbf{r}}_2 = m_2\frac{m_1}{M}\ddot{\mathbf{r}} = -\frac{\mathcal{G}m_1 m_2}{r^2}\hat{\mathbf{r}}$$

$$\Updownarrow$$

$$\ddot{\mathbf{r}} = -\frac{\mathcal{G}M}{r^2}\hat{\mathbf{r}}$$

Giving the exact same equation, showing, that this problem is identical to the one-body problem of a particle of negligible mass orbiting an object of mass $M$.

**Kepler's First Law**

We assume a particle acted on by a central force:

$$\ddot{\mathbf{r}} = -\mathcal{G}M\frac{\hat{\mathbf{r}}}{r^2} = -\mu\frac{\hat{\mathbf{r}}}{r^2} \tag{2.3}$$

The product $\mu = \mathcal{G}M$ is the standard gravitational parameter, also called the heliocentric (or geocentric, depending on the central object) gravitational constant. Apart from being a convenient abbreviation, the product $\mu$ is known to a much greater accuracy than the individual factors $\mathcal{G}$ and $M$ for the cases where the central object is the Earth or the Sun.

As shown in the previous section, (2.3) also describes two-body motion.

The angular momentum is usually given as the cross product of the position $\mathbf{r}$ and linear momentum $\mathbf{p}$ vectors. Letting $\mathbf{h}$ denote the angular momentum per unit mass:

$$\mathbf{h} = \frac{1}{m}\mathbf{r} \times \mathbf{p} = \mathbf{r} \times \dot{\mathbf{r}},$$

the conservation of angular momentum may be shown as:

$$\dot{\mathbf{h}} = \dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{0} + \mathbf{r} \times \left(-\mu\frac{\hat{\mathbf{r}}}{r^2}\right) = \mathbf{0}$$

This shows that the position vector and the velocity vector are always in the same plane, perpendicular to $\mathbf{h}$, which means that the orbit is in that plane.

We will now show that orbits described by (2.1) are conic sections, thus verifying and extending Kepler's empirical first law.

Taking the cross product of both sides of (2.3) with $\mathbf{h}$ and using $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}\,(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}\,(\mathbf{a} \cdot \mathbf{b})$ yields:

$$\ddot{\mathbf{r}} \times \mathbf{h} = -\frac{\mu}{r^2}\hat{\mathbf{r}} \times (\mathbf{r} \times \dot{\mathbf{r}}) = -\frac{\mu}{r^2}\left(\mathbf{r}\,(\hat{\mathbf{r}} \cdot \dot{\mathbf{r}}) - \dot{\mathbf{r}}\,(\hat{\mathbf{r}} \cdot \mathbf{r})\right) \tag{2.4}$$

The dot products are:

$$\begin{aligned} \hat{\mathbf{r}} \cdot \dot{\mathbf{r}} &= \dot{r}\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = \dot{r} \\ \hat{\mathbf{r}} \cdot \mathbf{r} &= r\hat{\mathbf{r}}\hat{\mathbf{r}} = r \end{aligned}$$

Using these, and the fact that $\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\hat{\mathbf{r}}}$ in (2.4) yields:

$$\ddot{\mathbf{r}} \times \mathbf{h} = -\frac{\mu}{r^2}\left(\mathbf{r}\dot{r} - \dot{\mathbf{r}}r\right) = -\frac{\mu}{r^2}\left(\mathbf{r}\dot{r} - r\left(\dot{r}\hat{\mathbf{r}} + r\dot{\hat{\mathbf{r}}}\right)\right) = \mu\dot{\hat{\mathbf{r}}}$$

Since $\dot{\mathbf{h}} = \mathbf{0}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{r}} \times \mathbf{h}) = \ddot{\mathbf{r}} \times \mathbf{h} + \dot{\mathbf{r}} \times \dot{\mathbf{h}} = \ddot{\mathbf{r}} \times \mathbf{h} = \mu\dot{\hat{\mathbf{r}}} = \mu\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathbf{r}} \qquad (2.5)$$

Integrating (2.5) with respect to time, we get

$$\dot{\mathbf{r}} \times \mathbf{h} = \mu\hat{\mathbf{r}} + \mathbf{c}$$

for some constant of integration $\mathbf{c} \in \mathbb{R}^3$ which is independent of time. Dividing by $\mu$ for convenience, we introduce another conserved quantity called the *Laplace-Runge-Lenz vector* (or just the *Runge-Lenz vector*) $\mathbf{e}$:

$$\mathbf{e} = \frac{\mathbf{c}}{\mu} = \frac{\dot{\mathbf{r}} \times \mathbf{h}}{\mu} - \hat{\mathbf{r}} \qquad (2.6)$$

Whereas conservation of angular momentum holds because gravity is a central force, the conservation of the Runge-Lenz vector $\mathbf{e}$ is a direct consequence of the inverse-square law of gravitation.

Taking the dot product of $\mathbf{r}$ and (2.6) and using the relation $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$, we get:

$$\mathbf{r}\cdot\mathbf{e} = \mathbf{r}\cdot\left(\frac{\dot{\mathbf{r}} \times \mathbf{h}}{\mu} - \hat{\mathbf{r}}\right) = \frac{1}{\mu}\mathbf{r}\cdot(\dot{\mathbf{r}} \times \mathbf{h}) - r = \frac{1}{\mu}\mathbf{h}\cdot(\mathbf{r} \times \dot{\mathbf{r}}) - r = \frac{\mathbf{h} \cdot \mathbf{h}}{\mu} - r = \frac{h^2}{\mu} - r \quad (2.7)$$

The dot product can also be written as

$$\mathbf{r} \cdot \mathbf{e} = re\cos v \, ,$$

where $v$ denotes the angle between the vectors $\mathbf{r}$ and $\mathbf{e}$. This angle, $v$, is also called the *true anomaly*. Using this and (2.7), we get the orbit in polar form:

$$re\cos v = \frac{h^2}{\mu} - r \Leftrightarrow r = \frac{h^2}{\mu\,(1 + e\cos v)} \qquad (2.8)$$

This polar equation describes a *conic section*, the intersection of a cone and a plane (see figure 2.3). By changing the angle and location of intersection, the conic section changes type. Omitting the degenerate cases, the conic section may be a circle, an ellipse, a parabola or a hyperbola. If the *eccentricity e* (the magnitude of the Runge-Lenz vector) is equal to zero, the radius is constant, resulting in a circular orbit. For $0 < e < 1$, the orbit is an ellipse, for $e = 1$ a parabola and $e > 1$ indicates an hyperbolic orbit.

This shows, that the solution to (2.3) and to the two-body problem described above are circular, elliptic, parabolic or hyperbolic orbits, thus verifying and extending Kepler's first law.
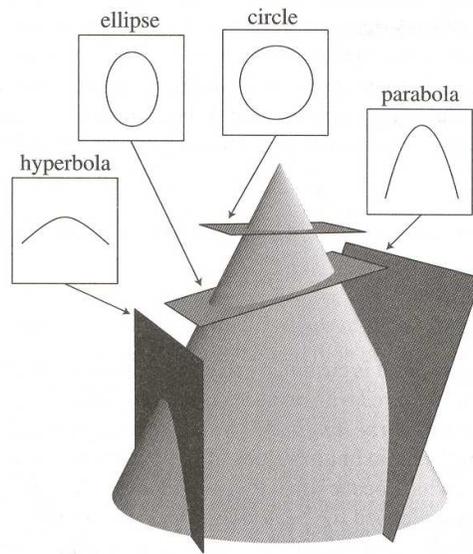
**Figure 2.3:** *Conic sections*, the intersection of a cone and a plane. Depending on the angle and location of intersection, the conic section changes type. The inverse-square law of gravitational force implies orbits shaped like conic sections. From [Murray & Dermott 1999].

**On Elliptic Orbits**
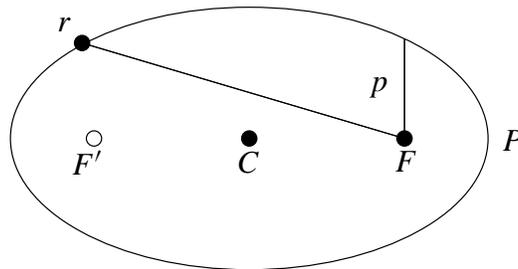


**Figure 2.4:** A particle in an elliptic orbit about a parent body in the focus $F$. The periapsis is denoted by $P$, the geometric center by $C$ and the empty focus by $F'$. The *semi-latus rectum*, $p$, is also shown.

In the following, we will examine the elliptic orbits ($0 < e < 1$).

The distance from the focus ($r = 0$) to the elliptic orbit, in a direction perpendicular to the Runge-Lenz vector is called the *semi-latus rectum*, $p$:

$$p = r\left(\frac{\pi}{2}\right) = \frac{h^2}{\mu} \tag{2.9}$$

To minimise $r$ in (2.8), it is necessary to maximise $\cos v$. This means that the point closest to the focus has $v = 0$. Remembering that $v$ is the angle between $\mathbf{r}$ and $\mathbf{e}$, this shows that the Runge-Lenz vector $\mathbf{e}$ points in the direction of the point of closest approach, the *periapsis*. For an object orbiting the Sun or the Earth, this is called the *perihelion* or *perigee*, respectively.

Conversely, the point of farthest distance is achieved when $\mathbf{r}$ is antiparallel to $\mathbf{e}$, i.e., when $v = \pi$. This point is called the *apoapsis*, *apohelion* or *apogee*, depending on the object which is orbited.

Half the distance between these extrema is called the (magnitude of the) *semi-major axis*, denoted by $a$:

$$a = \frac{1}{2}(r(0) + r(\pi)) = \frac{1}{2}\left(\frac{p}{1+e} + \frac{p}{1-e}\right) = \frac{p}{1-e^2} \tag{2.10}$$

Using the semi-major axis $a$, we can write the distance of periapsis as $d_p = a(1-e)$ and the distance of apoapsis as $d_a = a(1+e)$.

The point between these extremities is the *geometric center $C$*. The distance between the focus and the geometric center is:

$$d_C = \frac{1}{2}(r(\pi) - r(0)) = \frac{1}{2}\left(\frac{p}{1-e} - \frac{p}{1+e}\right) = \frac{p}{1-e^2}e = ae$$

Reflecting one focus with respect to an axis through $C$ and perpendicular to $\mathbf{e}$ yields the other focus $\mathbf{f}_2$:

$$\mathbf{f}_2 = -2a\mathbf{e}$$

Combining (2.7) and (2.10), we see that $\mathbf{r} \cdot \mathbf{e} = p - r = a(1 - e^2) - r$. This can be employed to show that the distance between a point on the ellipse $\mathbf{r}$ and $\mathbf{f}_2$ is:

$$f_2 = |\mathbf{r} + 2a\mathbf{e}| = \sqrt{(\mathbf{r} + 2a\mathbf{e}) \cdot (\mathbf{r} + 2a\mathbf{e})} = 2a - r,$$

implying $f_1 + f_2 = r + f_2 = r + 2a - r = 2a$, introducing a way to define the ellipse: The locus of points $\mathbf{r}$, satisfying $|\mathbf{r} - \mathbf{p}_1| + |\mathbf{r} - \mathbf{p}_2| =$ constant. This demonstrates the symmetry of the ellipse with respect to an axis through $C$ and perpendicular to $\mathbf{e}$. Because of this symmetry, the point $\mathbf{r}_b$ on the ellipse having the greatest distance to a line through $C$ and parallel to $\mathbf{e}$ will be on the aforementioned axis of symmetry. This distance, called the *semi-minor* axis, can be found by regarding a right triangle $\mathbf{Cr}_b\mathbf{f}_1$. Since $\mathbf{r}_b$ is on the axis of symmetry, $|\mathbf{r}_b| = f_1 = f_2 = a$. The distance between $C$ and a focus has been shown above to be $d_C = ae$. Using the Pythagorean Theorem to find $b$:

$$b^2 = a^2 - (ae)^2 = a^2(1 - e^2) \Leftrightarrow b = a\sqrt{1 - e^2}$$

Given an ellipse whose geometric center is at $(x, y) = (0, 0)$ and whose major axis is aligned with the $x$-axis satisfies:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

This can be used to find the area by direct integration:

$$A = \int_{-a}^{a} \int_{-b\sqrt{1-\frac{x^2}{a^2}}}^{-b\sqrt{1-\frac{x^2}{a^2}}} \mathrm{d}y \mathrm{d}x = \frac{2b}{a} \int_{-a}^{a} \sqrt{a^2 - x^2} \mathrm{d}x = \frac{2b}{a} \frac{a^2 \pi}{2} = \pi ab$$

**Kepler's Second Law**

Using $\hat{\mathbf{r}}$ and $\hat{\theta}$ to denote unit vectors along and perpendicular to the radius vector, the velocity $\dot{\mathbf{r}}$ is:

$$\dot{\mathbf{r}} = \frac{\mathrm{d}}{\mathrm{d}t} r \hat{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\theta} \tag{2.11}$$

Using (2.11) to write the polar form of the angular momentum per unit mass yields:

$$\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}} = r^2\dot{\theta}\left(\hat{\mathbf{r}} \times \hat{\theta}\right),$$

The magnitude of the cross product of two perpendicular unit vectors is unity, so the magnitude of $\mathbf{h}$ is $h = r^2\dot{\theta}$. Since $\mathbf{h}$ is constant, so is $h$.

The area swept out by an infinitesimal increase in $\theta$ is (see figure 2.5):

$$\mathrm{d}A = \frac{1}{2}r(r\mathrm{d}\theta) \Leftrightarrow \dot{A} = \frac{1}{2}r^2\dot{\theta} = \frac{h}{2} \tag{2.12}$$

Since $r^2\dot{\theta}$ is constant, $\dot{A}$ is constant, showing Kepler's Second law: The radius vector from the Sun to the planet sweeps over equal areas in equal amounts of time.

**Kepler's Third Law**

As shown in (2.12), the swept area per time is:

$$\dot{A} = \frac{1}{2}r^2\dot{\theta} = \frac{1}{2}h$$
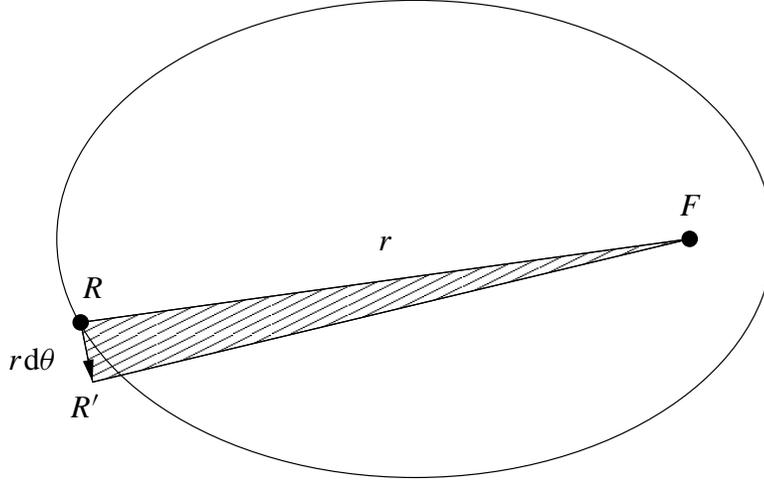
**Figure 2.5:** Area swept out by an infinitesimal increase in $\theta$ (the angle $R'FR$). The area of the shaded isosceles triangle equals $dA = r\,(r\,d\theta)\,/2$.

Knowing the area of an ellipse of semi-major axis and semi-minor axis $a$ and $b$, respectively, to be $A = \pi ab$, it is possible to determine the siderial period $P$: the time needed to complete one revolution around the focus on the elliptic orbit.

$$A = \pi ab = \int_0^P \dot{A}\,dt = \int_0^P \frac{1}{2}h\,dt = \frac{Ph}{2} \Leftrightarrow P = \frac{2\pi ab}{h} \tag{2.13}$$

According to Kepler's Third Law, the semi-major axis cubed should be proportional to the siderial period squared. From (2.13), the latter may be expressed as:

$$P^2 = \frac{(2\pi ab)^2}{h^2} \tag{2.14}$$

Isolating $h^2$ from (2.9):

$$h^2 = p\mu = a\left(1 - e^2\right)\mu \tag{2.15}$$

and inserting this in (2.14):

$$P^2 = \frac{(2\pi ab)^2}{a\left(1 - e^2\right)\mu} = \frac{\left(2\pi a^2\sqrt{1 - e^2}\right)^2}{a\left(1 - e^2\right)\mu} = 4\pi^2\frac{a^3}{\mu}$$

This shows that, in accordance with Kepler's Third Law, the siderial period squared is proportional to the semi-major axis cubed.

**The Orbital Reference System**
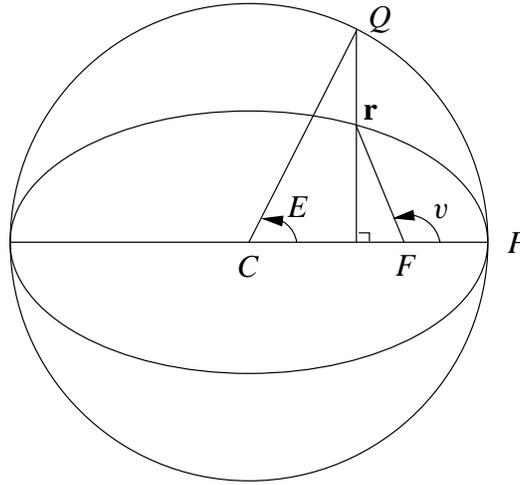


**Figure 2.6:** The relation between the true anomaly $v$ and the eccentric anomaly $E$.

Given a point **r** on an elliptic orbit, a line may be constructed, perpendicular to the Runge-Lenz vector **e**, which, pointing in the direction of the periapsis $P$, is parallel to a the line connecting the focus $F$ and the periapsis $P$. Constructing this perpendicular line from **r** and to its intersection $Q$ with a circle circumscribing the orbit (see figure 2.6). The angle between the radius vector from the geometric center $C$ to this point of intersection $Q$ and the direction of periapsis is called the *eccentric anomaly*, denoted $E$. The relation between the true anomaly $v$ and the eccentric anomaly $E$ is:

$$r \cos v = a \cos E - ae = a (\cos E - e) \tag{2.16}$$

Isolating $r$ in (2.16) and equating the result and (2.8), using $p = a(1 - e^2)$ from (2.10), yields:

$$\frac{a (\cos E - e)}{\cos v} = \frac{a \left(1 - e^2\right)}{1 + e \cos v} \Leftrightarrow \cos v = \frac{\cos E - e}{1 - e \cos E} \tag{2.17}$$

Inserting (2.17) in (2.8) leads to a simple relation:

$$r = p \frac{1 - e \cos E}{1 - e^2} = a (1 - e \cos E) \tag{2.18}$$

Using the Pythagorean Theorem to find $\sin v$ expressed as a function of $E$ leads to:

$$\sin^2 v = 1 - \cos^2 v = 1 - \left(\frac{\cos E - e}{1 - e \cos E}\right)^2 = \frac{\left(1 - e^2\right)\left(1 - \cos^2 E\right)}{(1 - e \cos E)^2} = \frac{1 - e^2}{(1 - e \cos E)^2} \sin^2 E$$

Thus, we find $\sin v$, enabling the expression of $r \sin v$ as a simple function of $E$:

$$\sin v = \frac{\sqrt{1 - e^2}}{1 - e \cos E} \sin E$$

$$r \sin v = a \left(1 - e \cos E\right) \frac{\sqrt{1 - e^2}}{1 - e \cos E} \sin E = a\sqrt{1 - e^2} \sin E = b \sin E$$

The *orbital reference system* denotes a frame of reference in the orbital plane with the $X$-axis pointing toward periapsis and the $Z$-axis parallel to $\mathbf{h}$. Completing a right-handed triad, the $Y$-axis then points in the direction corresponding to a true anomaly of 90 degrees.

$$
\begin{aligned}
X &= r \cos v = a \left(\cos E - e\right) \\
Y &= r \sin v = b \sin E = a\sqrt{1 - e^2} \sin E
\end{aligned}
\tag{2.19}
$$

The time derivatives are:

$$
\begin{aligned}
\dot{X} &= -a\dot{E} \sin E \\
\dot{Y} &= a\dot{E}\sqrt{1 - e^2} \cos E
\end{aligned}
$$

In this system, the angular momentum per unit mass $\mathbf{h}$ may be expressed as:

$$\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}} = \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} \times \begin{bmatrix} \dot{X} \\ \dot{Y} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ X\dot{Y} - \dot{X}Y \end{bmatrix}$$

The magnitude of $\mathbf{h}$ is thus

$$h = |X\dot{Y} - \dot{X}Y| = a^2\sqrt{1 - e^2} \left(1 - e \cos E\right) |\dot{E}| \tag{2.20}$$

This may be compared to the square root of (2.15):

$$h = \sqrt{a \left(1 - e^2\right) \mu} = na^2\sqrt{1 - e^2}, \tag{2.21}$$

introducing the *mean motion n*:

$$n = \frac{2\pi}{P} = \sqrt{\frac{\mu}{a^3}} \tag{2.22}$$

The time derivative of the eccentric anomaly (henceforth assumed to be positive) is derived by equating (2.20) and (2.21):

$$a^2\sqrt{1-e^2}\,(1-e\cos E)\,\dot{E} = na^2\sqrt{1-e^2} \Leftrightarrow \dot{E} = \frac{n}{1-e\cos E} \qquad (2.23)$$

According to (2.18), $1 - e\cos E = r/a$, providing an alternative way of expressing $\dot{E}$, namely

$$\dot{E} = \frac{n}{1-e\cos E} = \frac{an}{r},$$

which leads to alternative ways of writing the time derivatives of $X$ and $Y$:

$$\begin{aligned}
\dot{X} &= -\frac{a^2 n}{r}\sin E \\
\dot{Y} &= \frac{a^2 n\sqrt{1-e^2}}{r}\cos E \qquad (2.24)
\end{aligned}$$

By integrating equation (2.23), we obtain *Kepler's Equation*:

$$n\,(t-T) = E - e\sin E \qquad (2.25)$$

where $T$ is the constant of integration satisfying the boundary condition $E = 0$ when $t = T$. In other words $T$ is the time of pericenter passage. The left side of (2.25) is called the *mean anomaly*:

$$M = n\,(t-T)$$

The mean anomaly expresses the position of an object in its orbit as a fraction of one revolution. Although $M$ has the dimensions of an angle, and it increases linearly with time at a constant rate equal to the mean motion, it has no simple geometrical interpretation. However, it is clear that at periapsis, when $t = T + kP$ (for integer $k$), $M = v = 0$, and at apoapsis, when $t = T + P/2 + kP$, $M = v = \pi$.

## The Three-Body Problem

After Kepler, Newton, Hooke and their contemporaries solved the problem of the orbit of a single planet around the Sun, the natural next challenge was to find the solution for two planets orbiting the Sun. Many of the best minds in mathematics and physics worked on this problem in the following 200 years.

The first work went into finding an exact solution in analogy with the two-body problem. It was quickly recognised that the key was to find a sufficient number of conserved quantities. Energy, momentum, angular momentum, and the Laplace-Runge-Lenz vector (2.6) provide enough information to solve the two-body problem. For problems where there are enough integrals, the motion is quasiperiodic: roughly speaking, there are several interdependent periodic motions, leading to a motion in phase space which lies on a multi-dimensional torus. It has since been shown that, for the three-body problem, there is not a sufficient number of conserved quantities: the three-body problem is not "integrable".

**The Restricted Three-Body Problem**

Gradually, the problem was simplified in order to explore the kernel of the difficulty. The original eighteen-dimensional problem (three bodies, each having six degrees of freedom) becomes twelve-dimensional when transformed to center-of-mass coordinates. The planar three-body problem, simplified by restricting the planets to a plane, is in eight dimensions. The restricted three-body problem sets one mass to zero, and restricts the two major objects to being in circular orbits about their center of mass.

The approach taken in the following is similar to that of [Murray & Dermott 1999].

Consider three objects of mass $m_1$, $m_2$ and $m_3$. Let $m_1$ and $m_2$ be in circular orbits about their center of mass. Furthermore, let $m_3$ be a massless particle and let the mass of $m_1$ be greater than that of $m_2$. We assume that $m_1$ and $m_2$ exert a force on the particle $m_3$ although the particle cannot affect the two masses. Let the unit of mass be chosen such that $\mu = \mathcal{G}(m_1 + m_2) = 1$.

Following the above definitions, it holds that $\mathcal{G}m_1 = 1 - \bar{\mu}$ and $\mathcal{G}m_2 = \bar{\mu}$, where

$$\bar{\mu} = \frac{m_2}{m_1 + m_2} \tag{2.26}$$

If the coordinates of the particle in an inertial system are $(\xi, \eta, \zeta)$ and the positions of objects $m_1$ and $m_2$ are $(\xi_1, \eta_1, \zeta_1)$ and $(\xi_2, \eta_2, \zeta_2)$, respectively, the particle's equations of motion are:

$$\begin{aligned}
\ddot{\xi} &= (1 - \bar{\mu}) \frac{\xi_1 - \xi}{r_1^3} + \bar{\mu} \frac{\xi_2 - \xi}{r_2^3} \\
\ddot{\eta} &= (1 - \bar{\mu}) \frac{\eta_1 - \eta}{r_1^3} + \bar{\mu} \frac{\eta_2 - \eta}{r_2^3} \\
\ddot{\zeta} &= (1 - \bar{\mu}) \frac{\zeta_1 - \zeta}{r_1^3} + \bar{\mu} \frac{\zeta_2 - \zeta}{r_2^3}
\end{aligned} \tag{2.27}$$

Where $r_1 = \sqrt{(\xi_1 - \xi)^2 + (\eta_1 - \eta)^2 + (\zeta_1 - \zeta)^2}$ indicates the distance from the particle to object $m_1$ and $r_2 = \sqrt{(\xi_2 - \xi)^2 + (\eta_2 - \eta)^2 + (\zeta_2 - \zeta)^2}$ indicates the distance from the particle to object $m_2$.

If the two large objects are moving in circular orbits about their mutual center of mass, the distance between them remains fixed and their rotation occur at a fixed, common angular velocity.

In a coordinate system $(x, y, z)$, which rotates with the two main objects, centered on the center of mass, the coordinates of the two main objects remain fixed. The transformation from one coordinate system to another may be written as:

$$
\begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
$$

Differentiating with respect to time $t$ yields:

$$
\begin{bmatrix} \dot{\xi} \\ \dot{\eta} \\ \dot{\zeta} \end{bmatrix} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x} - y \\ \dot{y} + x \\ \dot{z} \end{bmatrix}
$$

Differentiating yet again yields:

$$
\begin{bmatrix} \ddot{\xi} \\ \ddot{\eta} \\ \ddot{\zeta} \end{bmatrix} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \ddot{x} - 2\dot{y} - x \\ \ddot{y} + 2\dot{x} - y \\ \ddot{z} \end{bmatrix}
$$

Switching to a rotating reference frame introduces extra terms, in $\dot{x}$ and $\dot{y}$, corresponding to Coriolis acceleration, and in $x$ and $y$, corresponding to the centrifugal acceleration.

Let $\mathbf{R}$ denote the rotation matrix:

$$
\mathbf{R} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

Inserting the expressions for $\xi$, $\eta$ and $\zeta$, and the corresponding time derivatives, in the equations of motion (2.27) yields:

$$
\mathbf{R} \begin{bmatrix} \ddot{x} - 2\dot{y} - x \\ \ddot{y} + 2\dot{x} - y \\ \ddot{z} \end{bmatrix} = \mathbf{R} \begin{bmatrix} -\frac{1-\bar{\mu}}{r_1^3}(x - x_1) - \frac{\bar{\mu}}{r_2^3}(x - x_2) \\ -\frac{1-\bar{\mu}}{r_1^3}(y - y_1) - \frac{\bar{\mu}}{r_2^3}(y - y_2) \\ -\frac{1-\bar{\mu}}{r_1^3}(z - z_1) - \frac{\bar{\mu}}{r_2^3}(z - z_2) \end{bmatrix}
$$

By left multiplying the rotation matrix with its inverse (which is equal to its transpose, since any rotation matrix is orthogonal), rearranging, and assuming $m_1$ is located at $(x_1, y_1, z_1) = (-\bar{\mu}, 0, 0)$ and $m_2$ at $(x_2, y_2, z_2) = (1 - \bar{\mu}, 0, 0)$, thus satisfying (2.2), completes the transformation of the particle's equations of motion into this rotating reference frame:

$$
\begin{aligned}
\ddot{x} &= 2\dot{y} + x - (1 - \bar{\mu})\frac{x + \bar{\mu}}{r_1^3} - \bar{\mu}\frac{x - 1 + \bar{\mu}}{r_2^3} \\
\ddot{y} &= -2\dot{x} + y - (1 - \bar{\mu})\frac{y}{r_1^3} - \bar{\mu}\frac{y}{r_2^3} \\
\ddot{z} &= -\left(\frac{1 - \bar{\mu}}{r_1^3} + \frac{\bar{\mu}}{r_2^3}\right)z
\end{aligned}
\tag{2.28}
$$

where $r_1 = \sqrt{(x + \bar{\mu})^2 + y^2 + z^2}$ and $r_2 = \sqrt{(x - 1 + \bar{\mu})^2 + y^2 + z^2}$. Since the performed coordinate transformation is a pure rotation, $r_1$ and $r_2$ are equal in the two reference frames.

Introducing the scalar function $U$:

$$
U = \frac{x^2 + y^2}{2} + \frac{1 - \bar{\mu}}{r_1} + \frac{\bar{\mu}}{r_2}
\tag{2.29}
$$

The gradient of this scalar function yields another way of writing the above equations of motion in the rotating reference frame:

$$
\begin{aligned}
\ddot{x} - 2\dot{y} &= \frac{\partial U}{\partial x} \\
\ddot{y} - 2\dot{x} &= \frac{\partial U}{\partial y} \\
\ddot{z} &= \frac{\partial U}{\partial z}
\end{aligned}
\tag{2.30}
$$

$$
\tag{2.31}
$$

Adding these three equation after multiplying with $\dot{x}$, $\dot{y}$ and $\dot{z}$, respectively, yields:

$$
\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = \frac{\partial U}{\partial x}\dot{x} + \frac{\partial U}{\partial y}\dot{y} + \frac{\partial U}{\partial z}\dot{z} = \frac{dU}{dt}
$$

This expression may be integrated with respect to time:

$$
\dot{x}^2 + \dot{y}^2 + \dot{z}^2 = 2U - C_J
$$

where $C_J$ is a constant of integration.

The quantity $C_J$, called the Jacobi Integral, is a constant of the motion, and may be expressed explicitly:

$$C_J = 2U - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 = x^2 + y^2 + 2\left(\frac{1-\bar{\mu}}{r_1} + \frac{\bar{\mu}}{r_2}\right) - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$$

The Jacobi Integral is the only integral of motion known to exist in the restricted three-body system, so a general solution of this problem cannot be expressed in closed form. However, although the Jacobi Integral cannot provide an orbit by itself, it may provide information on regions of space into which the object in question will never venture, i.e., the Jacobi Integral may place bounds on the motion of a given particle. The boundaries between the domain in which the particle may appear and the domain in which it cannot, are the zero-velocity surfaces given by

$$C_J = 2U = x^2 + y^2 + 2\left(\frac{1-\bar{\mu}}{r_1} + \frac{\bar{\mu}}{r_2}\right)$$

## 2.4   Equilibrium Points
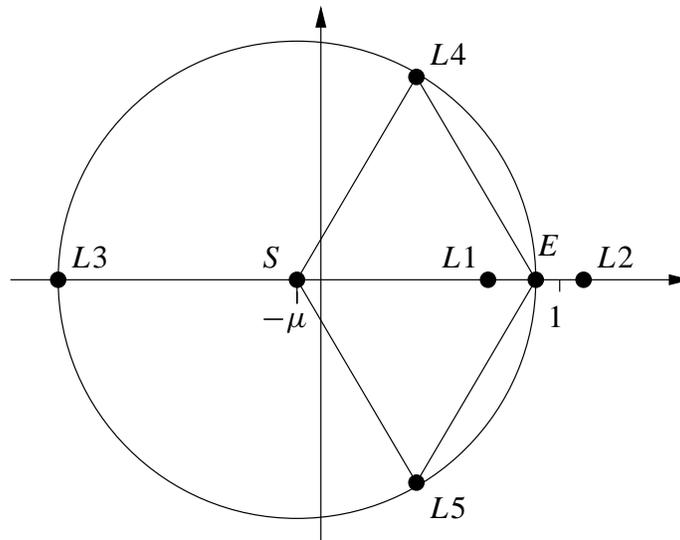


**Figure 2.7:** The five equilibrium points, known as Lagrange points $L1$ to $L5$, in the restricted three-body problem. This diagram shows the Lagrange points in the case of the Earth ($E$) orbiting the Sun ($S$). Not to scale.

By further restricting the previous chapter's zero-velocity surfaces to having zero acceleration, equilibrium points may be found, i.e., points where a particle could be placed,

with the appropriate velocity in the inertial reference frame, where it remains stationary in the rotating frame. These equilibrium points are often called *Lagrange points*, after the discoverer, the French-Italian mathematician Joseph-Louis Lagrange. This section shows the location of each of the five Lagrange points, denoted $L1$, $L2$, $L3$, $L4$ and $L5$ (see figure 2.7). The stability of each of these fixed points is also examined.

Assuming $\ddot{x} = \ddot{y} = \ddot{z} = \dot{x} = \dot{y} = \dot{z} = 0$ in (2.28) yields:

$$0 = x - (1 - \bar{\mu})\frac{x + \bar{\mu}}{r_1^3} - \bar{\mu}\frac{x - 1 + \bar{\mu}}{r_2^3} \tag{2.32}$$

$$0 = y - (1 - \bar{\mu})\frac{y}{r_1^3} - \bar{\mu}\frac{y}{r_2^3} = \left(1 - \frac{1 - \bar{\mu}}{r_1^3} - \frac{\bar{\mu}}{r_2^3}\right) y \tag{2.33}$$

$$0 = -\left(\frac{1 - \bar{\mu}}{r_1^3} + \frac{\bar{\mu}}{r_2^3}\right) z \tag{2.34}$$

Any equilibrium point must be in the $x$-$y$-plane in order to satisfy equation (2.34). Letting $y = 0$ obviously satisfies equation (2.33). This leads to the three *collinear* equilibrium points. The case of $y \neq 0$, the *off-axis* equilibrium points, will be treated below.

## The Collinear Equilibrium Points

Restricting the problem to the $x$ axis by imposing $y = 0$, leaves equation (2.32) to be solved:

$$0 = x - (1 - \bar{\mu})\frac{x + \bar{\mu}}{r_1^3} - \bar{\mu}\frac{x - 1 + \bar{\mu}}{r_2^3} = x - \frac{1 - \bar{\mu}}{(x + \bar{\mu})|x + \bar{\mu}|} - \frac{\bar{\mu}}{(x - 1 + \bar{\mu})|x - 1 + \bar{\mu}|}$$

Assume that the location of the equilibrium point is between the two masses, i.e., $-\bar{\mu} < x < 1 - \bar{\mu}$. It follows, that $|x + \bar{\mu}| = x + \bar{\mu}$ and $|x - 1 + \bar{\mu}| = -(x - 1 + \bar{\mu})$. Using the distance to $m_2$, denoted by $r_2$, as a variable instead of $x$, by using $x = 1 - \bar{\mu} - r_2$, yields:

$$0 = (1 - \bar{\mu})\left(1 - r_2 - \frac{1}{(1 - r_2)^2}\right) - \bar{\mu}\left(r_2 - \frac{1}{r_2^2}\right) \Leftrightarrow$$

$$\frac{\bar{\mu}}{1 - \bar{\mu}} = 3r_2^3 \left[\frac{1 - r_2 + \frac{r_2^2}{3}}{\left(1 - r_2^3\right)(1 - r_2)^2}\right] \tag{2.35}$$

For small $r_2$, the expression in the square brackets is approximately equal to 1, and a solution is therefore expected near

$$\frac{\bar{\mu}}{1 - \bar{\mu}} = 3r_2^3 \Leftrightarrow r_2 = \sqrt[3]{\frac{\bar{\mu}}{3\,(1 - \bar{\mu})}}$$

To facilitate reading, the parameter $\alpha$ is defined as

$$\alpha = \sqrt[3]{\frac{\bar{\mu}}{3\,(1 - \bar{\mu})}} \qquad (2.36)$$

Inserting (2.36) in equation (2.35) yields

$$\alpha = \sqrt[3]{r_2^3 \left[ \frac{1 - r_2 + \frac{r_2^2}{3}}{\left(1 - r_2^3\right)\left(1 - r_2\right)^2} \right]}$$

To get an approximate solution to the above equation, the above expression for $\alpha$ is Taylor expanded, centered on $r_2 = 0$:

$$\alpha = r_2 + \frac{r_2^2}{3} + \frac{r_2^3}{3} + \frac{53}{81}r_2^4 + O\left(r_2^5\right)$$

A series of the form $\kappa = k + c\phi\,(\kappa)$, where $c < 1$ is a constant, may be inverted by:

$$\kappa = k + \sum_{j=1}^{\infty} \frac{c^j}{j!} \frac{\mathrm{d}^{j-1}}{\mathrm{d}k^{j-1}} [\phi\,(k)]^j$$

This inversion method is due to Lagrange (see e.g., [Whittaker & Watson 1927]).

In this case, the series may be rearranged:

$$r_2 = \alpha + \left(-\frac{1}{3}\right)\phi\,(r_2)$$

Here, $c = -\frac{1}{3}$ and

$$\phi\,(r_2) = r_2^2 + r_2^3 + \frac{53}{27}r_2^4 + O\left(r_2^5\right)$$

and

$$[\phi(\alpha)]^2 = \left(\alpha^2 + \alpha^3 + \frac{53}{27}\alpha^4 + O\left(\alpha^5\right)\right)^2 = \alpha^4 + 2\alpha^5 + O\left(\alpha^6\right)$$

$$\frac{d}{d\alpha}[\phi(\alpha)]^2 = 4\alpha^3 + 10\alpha^4 + O\left(\alpha^5\right)$$

$$[\phi(\alpha)]^3 = \alpha^6 + O\left(\alpha^7\right)$$

$$\frac{d^2}{d\alpha^2}[\phi(\alpha)]^3 = 30\alpha^4 + O\left(\alpha^5\right)$$

Performing the inversion yields,

$$r_2 = \alpha - \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{23}{81}\alpha^4 + O\left(\alpha^5\right) , \tag{2.37}$$

The position of Lagrange point L1 in $(x, y, z)$ coordinates is thus:

$$L1 \approx (1 - \bar{\mu} - (\alpha - \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{23}{81}\alpha^4), 0, 0)$$

where $\alpha$ is defined by (2.36).

*Example* 1. Assume $\bar{\mu} = \frac{1}{10}$. Hence, $\alpha = \sqrt[3]{\frac{\bar{\mu}}{3(1-\bar{\mu})}} = \frac{1}{3}$. The above series (2.37) yields $r_2 \approx 0.2886755068$, corresponding to $x = 1 - \bar{\mu} - r_2 \approx 0.6113244932$, whereas numerical solution of equation (2.32) yields $x_{num} = 0.6090351100$.

$\triangle$

Looking beyond object $m_2$ (i.e., for $x > 1-\bar{\mu}$), $|x+\bar{\mu}| = x+\bar{\mu}$ and $|x-1+\bar{\mu}| = x-1+\bar{\mu}$. Using the distance $r_2$, defined in this interval as $r_2 = x-(1-\bar{\mu})$, equation (2.32) becomes:

$$0 = (1 - \bar{\mu})\left(1 + r_2 - \frac{1}{(1 - r_2)^2}\right) - \bar{\mu}\left(\frac{1}{r_2^2} - r_2\right) \Leftrightarrow$$

$$\frac{\bar{\mu}}{1 - \bar{\mu}} = 3r_2^3\left[\frac{1 + r_2 + \frac{r_2^2}{3}}{\left(1 - r_2^3\right)(1 + r_2)^2}\right]$$

In analogy with the above derivation, the auxiliary variable $\alpha$, defined in equation (2.36), is used:

$$\alpha = r_2^3\left[\frac{1 + r_2 + \frac{r_2^2}{3}}{\left(1 - r_2^3\right)(1 + r_2)^2}\right]$$

and a Taylor series expansion performed:

$$\alpha = r_2 - \frac{r_2^2}{3} + \frac{r_2^3}{3} + \frac{r_2^4}{81} + O\left(r_2^5\right),$$

which is inverted:

$$r_2 = \alpha + \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{31}{81}\alpha^4 + O\left(\alpha^5\right) \qquad (2.38)$$

The position of Lagrange point L2 in $(x, y, z)$ coordinates is thus:

$$L2 \approx (1 - \bar{\mu} + \alpha + \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{31}{81}\alpha^4, 0, 0)$$

where $\alpha$ is defined by (2.36).

*Example* 2. Assume $\bar{\mu} = \frac{1}{333000}$, approximately equal to the Sun-Earth mass ratio. Hence, $\alpha = \sqrt[3]{\frac{\bar{\mu}}{3(1-\bar{\mu})}} \approx 0.01$. The above series (2.38) yields $r_2 \approx 0.01$, corresponding to $x = 1 - \bar{\mu} + r_2 \approx 1.01$. This tells us that Lagrange point L2 in the Sun-Earth system is 1.01 AU from the Sun, or approximately 1.5 million kilometers from the Earth.

$$\triangle$$

The last of the collinear equilibrium points may be found to the "left" of object $m_1$, i.e., for $x < -\bar{\mu}$. Here, $|x + \bar{\mu}| = -(x + \bar{\mu})$ and $|x - 1 + \bar{\mu}| = -(x - 1 + \bar{\mu})$, hence, introducing $r_1 = -\bar{\mu} - x$ as variable:

$$0 = (1 - \bar{\mu})\left(\frac{1}{r_1^2} - r_1\right) - \bar{\mu}\left(1 + r_1 - \frac{1}{(1 + r_1)^2}\right) \Leftrightarrow$$

$$\frac{\bar{\mu}}{1 - \bar{\mu}} = \frac{1}{3r_1^3}\left[\frac{\left(1 - r_1^3\right)(1 + r_1)^2}{1 + r_1 + \frac{r_1^2}{3}}\right] \qquad (2.39)$$

Introducing the variable $\beta = r_1 - 1$ and Taylor expanding equation (2.39) about $\beta = 0$ yields:

$$\frac{\bar{\mu}}{1 - \bar{\mu}} = -\frac{12}{7}\beta + \frac{144}{49}\beta^2 - \frac{1567}{343}\beta^3 + O\left(\beta^4\right)$$

Inverting as previously, this time using $\bar{\mu}/(1 - \bar{\mu})$ as a variable, yields:

$$\beta = -\frac{7}{12}\left(\frac{\bar{\mu}}{1-\bar{\mu}}\right) + \frac{7}{12}\left(\frac{\bar{\mu}}{1-\bar{\mu}}\right)^2 - \frac{13223}{20736}\left(\frac{\bar{\mu}}{1-\bar{\mu}}\right)^3 + O\left(\left(\frac{\bar{\mu}}{1-\bar{\mu}}\right)^4\right) \quad (2.40)$$

The position of Lagrange point L3 in $(x, y, z)$ coordinates is thus:

$$L3 \approx (-\bar{\mu} - (1+\beta), 0, 0)$$

where $\beta$ is defined in (2.40).

*Example* 3. Using $\bar{\mu} = \frac{1}{10}$, i.e., $\frac{\bar{\mu}}{1-\bar{\mu}} = \frac{1}{9}$ yields $\beta \approx -0.05848790570$ corresponding to $r_1 = 1 + \beta \approx 0.9415120943$. Solving equation (2.39) numerically yields $r_{num} = 0.9416089086$.

$\triangle$

**The Off-Axis Equilibrium Points**

We now turn to the case of $y \neq 0$. To satisfy equation (2.34), $z = 0$ still holds. In the following, the problem will be regarded in the $x$-$y$-plane only.

Consider a massless particle, stationary in the rotating reference frame. In the inertial frame, the particle will describe a circular orbit around the origin. The resulting force $F$ acting on the particle is at all times directed towards the center of mass. If the particle is located at $(x, y)$, the gravitational pull of mass $m_1$ will be acting in a direction parallel to $\mathbf{r}_1 = (-\bar{\mu} - x, 0 - y)$, whereas the gravitational pull of mass $m_2$ will be acting in a direction parallel to $\mathbf{r}_2 = (1 - \bar{\mu} - x, 0 - y)$. If $F_1$ and $F_2$ denote the magnitudes of the gravitational forces exerted by $m_1$ and $m_2$, respectively, the resulting force will be equal to:

$$\frac{F_1}{|\mathbf{r}_1|}\begin{pmatrix} -\bar{\mu} - x \\ 0 - y \end{pmatrix} + \frac{F_2}{|\mathbf{r}_2|}\begin{pmatrix} 1 - \bar{\mu} - x \\ 0 - y \end{pmatrix}$$

Because the resulting force is directed towards the center of mass, it is parallel to $(-x, -y)$, which means the result of taking the scalar product of the resulting force and a vector perpendicular to $(-x, -y)$, such as $(y, -x)$, should be zero:

$$\left(\frac{F_1}{r_1}\begin{pmatrix} -\bar{\mu} - x \\ 0 - y \end{pmatrix} + \frac{F_2}{r_2}\begin{pmatrix} 1 - \bar{\mu} - x \\ 0 - y \end{pmatrix}\right) \cdot \begin{pmatrix} y \\ -x \end{pmatrix} = 0 \Leftrightarrow \frac{F_2}{F_1} = \frac{\bar{\mu}}{1-\bar{\mu}}\frac{r_2}{r_1} \quad (2.41)$$

Recalling from (2.26) the definition of $\bar{\mu}$, we write the object masses as $m_1 = (m_1 + m_2)(1 - \bar{\mu})$ and $m_2 = (m_1 + m_2)\bar{\mu}$. Because the gravitational force $F_1$ is proportional

to $m_1$ and $r_1^{-2}$, and $F_2$ is proportional to $m_2$ and $r_2^{-2}$, each with the same proportionality coefficient $k$, we can write:

$$
\begin{aligned}
F_1 &= k\frac{m_1}{r_1^2} = k(m_1 + m_2)\frac{1-\bar{\mu}}{r_1^2} \\
F_2 &= k\frac{m_2}{r_2^2} = k(m_1 + m_2)\frac{\bar{\mu}}{r_2^2}
\end{aligned}
$$

Dividing $F_2$ by $F_1$ yields:

$$
\frac{F_2}{F_1} = \frac{\bar{\mu}}{1-\bar{\mu}}\frac{r_1^2}{r_2^2} \tag{2.42}
$$

To satisfy both equations (2.41) og (2.42), the distances from the particle to each of the main bodies must be equal: $r_1 = r_2$. Inserting this in (2.33), and recalling that $y \neq 0$, yields:

$$
0 = \left(1 - \frac{1-\bar{\mu}}{r_1^3} - \frac{\bar{\mu}}{r_2^3}\right)y = \left(1 - \frac{1-\bar{\mu}}{r_1^3} - \frac{\bar{\mu}}{r_1^3}\right)y \Leftrightarrow r_1 = r_2 = 1
$$

The distance between $m_1$ and $m_2$ is always equal to 1. The remaining two equilibrium points therefore form two equilateral triangles with $m_1$ and $m_2$. These equilibrium points can be said to be $+60$ degrees and $-60$ degrees out of phase. By convention, the leading equilibrium point is taken to be $L_4$ and the trailing point $L_5$.

The coordinates of these off-axis equilibrium positions are

$$
\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} - \bar{\mu} \\ \pm\frac{\sqrt{3}}{2} \end{pmatrix}
$$

## Stability of Equilibrium Points

Linearising the equations of motion (2.30) at an equilibrium point and converting them to a system of first order differential equations yields:

$$
\begin{pmatrix} \dot{x} \\ \dot{y} \\ \ddot{x} \\ \ddot{y} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ U_{xx} & U_{xy} & 0 & 2 \\ U_{xy} & U_{yy} & -2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{pmatrix} \tag{2.43}
$$

where $U_{xx}$, $U_{xy}$ and $U_{yy}$ denote the second derivatives of (2.29) evaluated at the equilibrium point in question:

$$U_{xx} = \left(\frac{\partial^2 U}{\partial x^2}\right)_0$$

$$U_{xy} = \left(\frac{\partial^2 U}{\partial x \partial y}\right)_0$$

$$U_{yy} = \left(\frac{\partial^2 U}{\partial y^2}\right)_0$$

The characteristic polynomial for the $4 \times 4$ matrix in equation (2.43) is:

$$\lambda^4 + (4 - U_{xx} - U_{yy})\lambda^2 + U_{xx}U_{yy} - U_{xy}^2 = 0$$

By substituting $s = \lambda^2$, this equation transforms into a quadratic equation in $s$. The four roots, i.e., the eigenvalues, are:

$$\lambda_{1,2} = \pm\sqrt{\frac{U_{xx} + U_{yy} - 4}{2} - \frac{\sqrt{\left(4 - U_{xx} - U_{yy}\right)^2 - 4\left(U_{xx}U_{yy} - U_{xy}^2\right)}}{2}}$$

$$\lambda_{3,4} = \pm\sqrt{\frac{U_{xx} + U_{yy} - 4}{2} + \frac{\sqrt{\left(4 - U_{xx} - U_{yy}\right)^2 - 4\left(U_{xx}U_{yy} - U_{xy}^2\right)}}{2}}$$

To a complex eigenvalue $a + ib$, where $i$ denotes the imaginary unit, there is a corresponding solution of the form:

$$F(t) = e^{(a+ib)t} = e^{at}\left(\cos bt + i \sin bt\right)$$

If $a > 0$, the $e^{at}$ factor will make this solution tend to infinity as $t \to \infty$. We are, however, looking for periodic solutions, so the real part of every eigenvalue must be non-positive. Since all four eigenvalues are of the form $\lambda = \pm(a + ib)$, this implies $a = 0$, purely imaginary eigenvalues.

Let the quantities $A$, $B$, $C$ and $D$ be defined by:

$$A = \frac{1-\bar{\mu}}{(r_1^3)_0} + \frac{\bar{\mu}}{(r_2^3)_0} \tag{2.44}$$

$$B = 3\left(\frac{1-\bar{\mu}}{(r_1^5)_0} + \frac{\bar{\mu}}{(r_2^5)_0}\right) y_0^2 \tag{2.45}$$

$$C = 3\left((1-\bar{\mu})\frac{x_0+\bar{\mu}}{(r_1^5)_0}\bar{\mu} + \frac{x_0-1+\bar{\mu}}{(r_2^5)_0}\right) y_0 \tag{2.46}$$

$$D = 3\left((1-\bar{\mu})\frac{x_0+\bar{\mu}}{(r_1^5)_0}\bar{\mu} + \frac{x_0-1+\bar{\mu}}{(r_2^5)_0}\right) \tag{2.47}$$

Using these, the second derivatives may be expressed as:

$$U_{xx} = 1 - A + D$$
$$U_{yy} = 1 - A + B$$
$$U_{xy} = C$$

The following sections will use these tools to describe the stability of the collinear and off-axis equilibrium points.

**Stability of Collinear Equilibrium Points**

The collinear equilibrium points are all positioned on the $x$-axis, which implies $y = z = 0$. This means that $B = C = 0$ and $r_1^2 = (x_0 + \bar{\mu})^2$ and $r_2^2 = (x_0 - 1 + \bar{\mu})^2$, yielding:

$$U_{xx} = 1 + 2A$$
$$U_{yy} = 1 - A$$
$$U_{xy} = 0$$

The characteristic equation may thus be written

$$\lambda^4 + (2 - A)\lambda^2 + (1 + 2A)(1 - A) - 0 = 0$$

The product of the four roots of the characteristic equation (the eigenvalues) is equal to the constant term of that equation, i.e.:

$$\lambda_1 \lambda_2 \lambda_3 \lambda_4 = (1 + 2A)(1 - A)$$

Since the eigenvalues must be purely imaginary, and since $\lambda_1 = -\lambda_2$ and $\lambda_3 = -\lambda_4$, the product of all four eigenvalues must be positive. To satisfy this, $-\frac{1}{2} < A < 1$ must hold.

However, substituting the values of $r_1$ and $r_2$ for the collinear equilibrium points into equation (2.44) shows that $A > 1$. This shows that the collinear equilibrium points are unstable.

It is, however, possible to find quasiperiodic orbits near these unstable equilibrium points. The Gaia satellite (see chapter 4) is to be placed in such an orbit near equilibrium point L2 in the Sun-Earth system.

**Stability of Off-Axis Equilibrium Points**

As derived above, the location of the off-axis equilibrium points is given by $r_1 = r_2 = 1$, $x = \frac{1}{2} - \bar{\mu}$, $y = \pm \frac{\sqrt{3}}{2}$. This yields,

$$
\begin{aligned}
U_{xx} &= \frac{3}{4} \\
U_{yy} &= \frac{9}{4} \\
U_{xy} &= \pm 3\sqrt{3}\frac{1 - 2\bar{\mu}}{4}
\end{aligned}
$$

The characteristic equation

$$\lambda^4 + \lambda^2 + \frac{27}{4}\bar{\mu}(1 - \bar{\mu}) = 0$$

the roots of which are

$$
\begin{aligned}
\lambda_{1,2} &= \pm\frac{\sqrt{-1 - \sqrt{1 - 27(1 - \bar{\mu})\bar{\mu}}}}{\sqrt{2}} \\
\lambda_{3,4} &= \pm\frac{\sqrt{-1 + \sqrt{1 - 27(1 - \bar{\mu})\bar{\mu}}}}{\sqrt{2}}
\end{aligned}
$$

To ensure that the eigenvalues will be purely imaginary, the following must hold

$$1 - 27(1 - \bar{\mu})\bar{\mu} \geq 0 \Leftrightarrow \bar{\mu} \leq \frac{27 - \sqrt{621}}{54} \approx 0.0385 \tag{2.48}$$

If the mass ratio $\bar{\mu}$ is less than 0.0385 Lagrange points L4 and L5 *should* be stable. How-ever, due to the effects of resonance, instabilities may occur at a finite number of mass ratios that satisfy equation (2.48). See also [Murray & Dermott 1999].

## 2.5   Radiation Forces

So far, only the effects of the gravitational forces have been treated in this thesis. Solar system objects are, however, affected by other forces, such as the radial force exerted by the Sun's radiation, and collisions with other solar system objects. The effects of the latter are impulsive in nature and thus difficult to quantify. This section presents the direct and the more subtle indirect effects of the Sun's radiation on the orbit of an object.

### Radiation Pressure

Every electromagnetic wave carries momentum. A plane wave traveling in the direction given by the unit vector **d**, striking a body having a frontal area $A$ facing the wave, and being absorbed by this body, will transfer momentum to this body. Since a change in momentum over time equals a force, the electromagnetic wave will exert a force on the body:

$$\mathbf{F} = A\frac{S}{c}\mathbf{d} \tag{2.49}$$

Here, $c$ is the speed of light, and $S$ designates the energy flux of the wave[2]. If the wave is totally reflected, rather than totally absorbed, the magnitude of the force is twice that given in equation (2.49).

Since the energy flux of a wave oscillates in time, it may be more practical to introduce the time-averaged energy flux $\overline{S}$. Rearranging the above expression to express force per area, yields the (time-averaged) *radiation pressure*:

$$\frac{F}{A} = \frac{\overline{S}}{c} \tag{2.50}$$

The energy emitted by the Sun is globally in the form of a spherical wave. However, when comparing the radius of that sphere (say, 1 AU) to the radius of the body hit by the wave, these waves may be regarded as planar, so we can use equation (2.50).

The time-averaged energy flux in the sunlight, as a function of distance $r$ from the Sun is:

$$\overline{S}_\odot(r) = \frac{L_\odot}{4\pi r^2}$$

---

[2] $S$ is the magnitude of the so-called *Poynting vector.*

where $L_\odot = 3.9 \times 10^{26}$ W is the luminosity of the Sun.

Much of the energy radiated by the sun is contributed by waves with frequencies outside the visible spectrum. Therefore, the use in this thesis of the term "sunlight" also implies non-visible frequencies.

The force exerted by sunlight is proportional to the energy flux, and thus it is proportional to $r^{-2}$, similar to the gravitational force.

Since the force exerted by the radiation pressure on a body is proportional to the frontal area of the body facing the Sun, and thus roughly proportional to the square of the radius of the body, whereas the gravitational force is proportional to the mass, and thus to the radius cubed, it would be interesting to find the radius at which the magnitudes of these forces were equal.

Letting $F_{\text{rp}}$ and $F_G$ denote the force contribution of the radiation pressure and gravitation, respectively, and denoting the distance to the Sun by $r_d$ and the radius of the body (assumed to be spherical) by $r_o$, the desired quantity may be obtained by solving the following inequality:

$$F_G = \frac{\mathcal{G}M_\odot m_o}{r_d^2} = \frac{\mathcal{G}M_\odot}{r_d^2}\frac{4}{3}\pi r_o^3 \rho > F_{\text{rp}} = A\frac{\overline{S}}{c} = \pi r_o^2 \frac{\overline{S}}{c}$$

$$\Updownarrow$$

$$r_o > \frac{3}{16}\frac{L_\odot}{\pi c\mathcal{G}M_\odot \rho}$$

where $A = \pi r_o^2$ denotes the disk-shaped silhouette area, $\rho$ is the density of the body, $m_o = 4\pi r_0^3 \rho/3$ is the mass of the body and $M_\odot$ is the mass of the Sun.

Assuming a density equal to the mean bulk density of ordinary chondrite meteorites [Consolmagno et al. 1998], $\rho = 3.3 \cdot 10^3 \frac{\text{kg}}{\text{m}^3}$, the critical radius is:

$$r_o > \frac{3}{16}\frac{3.9 \cdot 10^{26}\text{W}}{3.14 \cdot 3.0 \cdot 10^8 \frac{\text{m}}{\text{s}} \cdot 6.67 \cdot 10^{-11}\text{N}\frac{\text{m}^2}{\text{kg}^2} \cdot 1.99 \cdot 10^{30}\text{kg} \cdot 3.3 \cdot 10^3 \frac{\text{kg}}{\text{m}^3}} = 1.8 \cdot 10^{-7}\text{m}$$

Depending on the density, grains of dust smaller than about $10^{-6}$ m across are affected more by the radiation pressure than by the gravitational pull towards the Sun, and are subsequently "blown" out of the solar system. Particles smaller than $10^{-7}$ m across tend to scatter light, rather than absorb it, and hence these particles are not affected by the radiation pressure to the same extent as larger objects.

On NEOs, having a lower size limit of 50 m, many orders of magnitude greater than the critical radius, the effect of the radiation pressure is very modest, and is normally only measurable when observing across several siderial periods.

## Other Forces

Apart from collision forces, NEOs are also affected by the so-called Poynting-Robertson and Yarkovsky effects. The former acts as a dragging force caused by the uneven reemission of absorbed solar radiation. From the perspective of the NEO, the Sun's radiation appears to have a non-zero composant in the direction opposed to the motion of the NEO, thus decelerating it in its orbit. This effect was first described in [Poynting 1904].

The Yarkovsky effect is a consequence of the Sun's warming of the NEO's surface as it rotates: the face exposed to the Sun warms up, and then rotates to the night side where it cools off. The "sunset" point will be warmer than the "sunrise" point and therefore will radiate a little more. This anisotropic thermal re-radiation will subject the NEO to a thrust, accelerating or decelerating it in its orbit, depending on the orientation of the axis of rotation. The Yarkovsky effect is described in [Hartmann et al. 1999], and has been directly measured using radar ranging [Chesley et al. 2003].

While the radiation pressure, Poynting-Robertson drag and the Yarkovsky effect do not have a great impact on the short-term evolution of asteroid orbits, and as such are only peripherally connected to the topic of this thesis, it has been proposed that they may be responsible for the "generation" of near Earth objects out of Main Belt asteroids by perturbing orbits [Morbidelli & Vokrouhlicky 2003].

The long-term effect of radiation forces have been estimated in [Giorgini et al. 2002], in the case of asteroid (29075) 1950 DA, reported to have a non-negligible impact probability in March 2880.

For more information on the effect of solar radiation, refer to [Burns et al. 1979] and [Mignard 1982].

# Chapter 3

# NEO Search Programmes

Several NEO detection programmes are currently in operation or in a preparatory phase. To facilitate a comparison between detection programmes, the following sections emphasise three parameters

**Sky coverage.** The larger the area covered, the higher the probability of detecting NEOs.

**Limiting magnitude.** The fainter the limiting magnitude, the higher probability of detection NEOs.

**Accuracy** in determining epherimides.

The following sections present a selection of the most prolific NEO detection programmes currently in operation, responsible for more than 90% of new NEO discoveries at the time of this writing (2003). All the major search programmes are based in the USA. The Catalina Sky Survey has been included, being the only one to survey the sky of the southern hemisphere. The Pan-STARRS project is the most ambitious ground-based NEO search programme currently in development.

## LINEAR

The *Lincoln Near-Earth Asteroid Research* (LINEAR) is a cooperation between Massachusetts Institute of Technology (MIT) and the US Air Force, using two one-meter class telescopes and a 0.5 meter telescope for follow-up observations, all located in New Mexico, USA. Currently, each main telescope employs two CCDs[1], one $1024 \times 1024$ pixel CCD covering one fifth of the telescope's field of view, and one $1960 \times 2560$ pixel CCD covering the full two square degree field of view. In fair observing conditions, the LINEAR programme telescopes has a limiting magnitude of about V=19.5.

---

[1] *Charge-Coupled Device*, see glossary.

The sensitivity of the CCDs, and particularly the relatively rapid readout rates, allows LINEAR to cover large areas of sky each night. Each field of about 2 square degrees is scanned five times. Every night, about 600 fields are covered, totalling about 1200 square degrees. The programme searches as close as 60 degrees from the Sun. Currently, the LINEAR program is responsible for the majority of NEO discoveries. Information from J. B. Evans and [LINEAR 2005].

# NEAT

NASA's Jet Propulsion Laboratory and the US Air Force cooperate in the *Near Earth Asteroid Tracking* (NEAT) programme, currently using a 1.2 metre Schmidt telescope (designated *NEAT/P* in table 3.1) located at Palomar Mountain, Southern California, USA. The limiting magnitude for this telescope is about $V = 20.5$ and each image covers 3.75 square degrees. The telescope at Palomar Mountain is dedicated to NEO search for about 130 hours each month.

The NEAT programme also uses a 1.2 metre class telescope (designated *NEAT/M* in table 3.1), located at the Maui Space Surveillance Site (MSSS), Maui, Hawaii. This telescope has a limiting magnitude of about $V = 19.5$, but is dedicated to the search for NEOs twice as many hours per month as the Palomar telescope. Both of these telescopes perform NEO searches at solar elongations as low as 75 degrees. Information from S. H. Pravdo and [NEAT 2005].

# Catalina Sky Survey

The *Catalina Sky Survey* is a search programme based in the USA, which has telescopes at several sites, including a collaboration between the Research School of Astronomy and Astrophysics (RSAA) of the Australian National University and the University of Arizona Lunar and Planetary Laboratory (LPL) to search for NEOs from Siding Springs Observatory in Australia. A 0.5 m Schmidt telescope is currently (2003) undergoing modification to provide added sky coverage in regions of the southern sky unreachable from the currently active NEO search stations. The field of view is about 8 square degrees, projected onto a 4k×4k pixel CCD.

A 0.68 m Schmidt telescope (designated *Catalina/C* in table 3.1) is already operating full-time under the University of Arizona on Mt. Bigelow, Arizona, USA. The Siding Springs telescope (designated *Catalina/S* in table 3.1) is to be dedicated to NEO search full-time. A 1.5 metre telescope (designated *Catalina/L* in table 3.1) at Mt. Lemmon, Arizona, USA, is available for follow-up observations. This telescope is anticipated to be performing NEO searches approximately half of the available nights.

All three telescopes have a position accuracy of approximately 0.2 arcseconds, and are able to search at solar elongations as low as 60 degrees. Information from S. Larson and [CSS 2005].

# LONEOS

LONEOS, the Lowell Observatory NEO Search is situated near Flagstaff, Arizona, USA. It uses a 0.6 m class fully-automated Schmidt telescope to conduct a full-time search for NEOs (approximately 200 nights per year are sufficiently clear). Using two 2K × 4K pixel CCD detectors to cover a field of view of 2.85 × 2.85 degrees, the telescope is designed to make four scans per region over the entire visible sky each month down to a limiting magnitude of about $V = 19.5$, although asteroids as faint as $V = 19.8$ have been detected. The telescope has the capability to scan the entire sky accessible from the site every month. Each clear night, the telescope covers approximately 1000 square degrees. The accuracy is approximately 0.5 arcsecond. The LONEOS telescope regularly observes at a solar elongation of 70 degrees. Information from B. Koehn.

# ADAS

The Asiago DLR Asteroid Survey is a joint programme among the Department of Astronomy and Astronomical Observatory of Padova, Italy, and the DLR (Deutsches Zentrum für Luft- und Raumfahrt) Institute of Space Sensor Technology and Planetary Exploration, Berlin, Germany. The program conducts the search using a 67/92 cm Schmidt telescope at Asiago - Cima Ekar, Italy. The telescope is equipped with a 2048 × 2048 pixel CCD, and the field of view is 0.67 square degrees. The search has mainly been conducted in a strip from −5° to +15° around the celestial equator. The limiting magnitude is about $V = 21$, and the typical astrometric position accuracy is better than 0.4 arcseconds. The project is currently at a standstill due to lack of personel. Information from C. Barbieri.

# Japan Spaceguard Association

The JSGA, financed by Japan's National Space Development Agency (NASDA), the National Aeronautic Laboratory, and the Space and Technology Agency, uses an observatory near Bisei, Japan. There, the NEO search program has access to one 0.5 m class telescope and one 1.0 m Cassegrain instrument, reaching limiting magnitudes of $V = 17.5$ and $V = 19.5$, respectively, in typical seeing conditions, using an integration time of 60 seconds. Having about 60% of the observable nights available, the JSGA is able to cover the entire visible sky in about three months, taking advantage of the 1.0 m telescopes field of

view of nearly seven square degrees. For a single set of observations, the astrometric accuracy is about 0.5 arcseconds. The JSGA usually observe near opposition. Information from S. Isobe.

## CINEOS

The Campo Imperatore Near Earth Object Survey (CINEOS) is a dedicated search and follow-up program of near Earth objects, born in 1996 from a collaboration between the Observatory of Rome (OAR-INAF), and the Istituto di Astrofisica Spaziale (IASF-CNR). The CINEOS program uses a 0.9 m class Schmidt telescope situated at the Campo Imperatore Observatory about 130 km North-East of Rome, Italy, and currently emphasises observing at solar angles as low as 40 degrees, aiming at a ground-based system optimised towards the discoveries of Aten family asteroids and other inner-Earth objects. The field of view is 52 by 52 arcminutes, covering an area of about 0.75 square degrees, enabling the survey of about 150 square degrees per lunation. Using a 60 second integration time, the limiting magnitude is about $V = 20.5$. Single observations on the $2048 \times 2048$ pixel CCD have an astrometric accuracy of about 0.4 arcseconds. Information from A. Di Paola.

## Spacewatch

The Spacewatch project uses the 0.9 m and 1.8 m telescopes on Kitt Peak, 45 miles southwest of Tucson, Arizona, USA. The former telescope (designated *Spacewatch I* in table 3.1) is dedicated to NEO surveying, and has a limiting magnitude of $V = 21.7$, whereas that of the latter (designated *Spacewatch II* in table 3.1) has a limiting magnitude of $V = 22.4$, and the observable hours are split evenly between surveying and targeted follow-up observations. The 1.8 m telescope has, however, recovered objects as faint as $V = 23.3$ "at considerable effort". The Spacewatch Project is unique in emphasising faint objects (e.g., $V > 20.5$) rather than sky coverage. The sky coverage of the telescopes is 1500 square degrees and 150 square degrees, respectively. The astrometric accuracy of the 0.9 m Schmidt telescope is about 0.5 arcseconds, whereas the 1.8 m telescope seems slightly more accurate at 0.4 arcseconds. Although able to observe as close as 60 degrees in cases of urgent follow-up observations, surveying is not generally done less than 90 degrees from the Sun. Information from R. McMillan.

## Pan-STARRS

Pan-STARRS, the *Panoramic Survey Telescope And Rapid Response System*, is a wide-field search programme under development at the University of Hawaii's Institute for

Astronomy. By combining four 1.8 m telescopes, a limiting magnitude of $V = 24$ is expected. In survey mode Pan-STARRS will cover 6000 square degrees per night, surveying the whole available sky as seen from Hawaii three times each lunation. Pan-STARRS PS1, essentially one quarter of Pan-STARRS, will be completed ahead of the full observatory. It will have the same optics design and camera design as anticipated for the full version of Pan-STARRS. First light for Pan-STARRS PS1 is scheduled for January 2006, with deployment of the full array within a further two years. See also [PanSTARRS 2005].

## 3.1    Search Programme Comparison

Table 3.1 contains information for comparing the most prolific NEO search programmes currently (2003) in operation. Each observatory is designated by the observatory code assigned by the Minor Planet Center[2] (*MPC code*). The table facilitates comparison between ground-based search programmes and GAIA, in terms of limiting magnitude ($V_{lim}$), monthly sky coverage (*coverage/mon*), astrometric accuracy (*accuracy*, measured as the average residual, not bias-corrected) as well as minimum solar elongation (*min. elong.*). The monthly sky coverage are optimistic estimates, disregarding downtime due to hardware failure, *et cetera*. Several of the minimum solar elongation reported are only attained during urgent follow-up observations, not during regular surveying. The Gaia data is taken from chapter 4.

| Observatory | MPC code | $V_{lim}$ | coverage / month | accuracy | min. elong. |
|---|---|---|---|---|---|
| Catalina/C | 703 | $\approx 20.2$ | 10, 000-17, 000 deg$^2$ | $\approx 0.2$ as | 60 deg |
| Catalina/L | G96 | $\approx 22.5$ | 2500-4500 deg$^2$ | $\approx 0.2$ as | 60 deg |
| Catalina/S | E12 | $\approx 20.0$ | 600-1100 deg$^2$ | $\approx 0.2$ as | 60 deg |
| LINEAR | 704 | $\approx 19.5$ | 17, 000 deg$^2$ | $\approx 0.6$ as | 60 deg |
| NEAT/P | 644 | $\approx 20.5$ | 8000-10, 000 deg$^2$ | $\approx 0.5$ as | 75 deg |
| NEAT/M | 566 | $\approx 19.5$ | 8000-10, 000 deg$^2$ | $\approx 0.5$ as | 75 deg |
| Spacewatch I | 691 | $\approx 21.7$ | 1500 deg$^2$ | $\approx 0.5$ as | 60 deg |
| Spacewatch II | 291 | $\approx 22.4$ | 150 deg$^2$ | $\approx 0.4$ as | 60 deg |
| *Gaia* | *N/A* | $\approx 20$ | *25,000 deg$^2$* | $\approx 5$ mas | *45 deg* |

**Table 3.1:** A list of the most prolific ground-based NEO observatories currently in operation, comparing limiting visual magnitude, monthly sky coverage, position accuracy and minimum solar elongation. Gaia space observatory information (see chapter 4) added for comparison.

Jedicke et al. [Jedicke et al. 2003] examines the prospects of achieving the Spaceguard Goal (the discovery of 90% of all 1-km NEOs by the year 2008), comparing ground-based and space-based observations. Perhaps surprisingly, the paper concludes that there is little need for distributing ground-based survey telescopes in latitude and longitude

---

[2]http://cfa-www.harvard.edu/iau/mpc.html

as long as the sky coverage is sufficient. The bias of NEO survey programmes on the northern hemisphere apparently does not constitute a handicap. While admittedly not considering a cost-benefit analysis, the authors consider a space-based survey to offer an advantage over Earth-based counterparts.
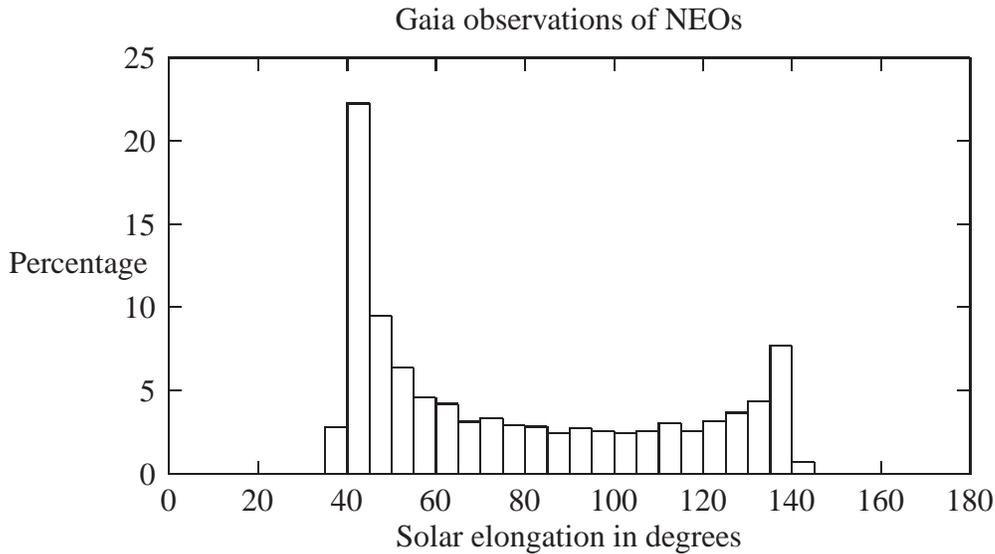


**Figure 3.1:** The simulated percentage of NEOs discovered by Gaia as a function of solar elongation. From [Mignard 2001]. See also chapter 4.

If the Pan-STARRS project proceeds according to plan, Gaia is not expected to discover many new NEOs because of its limiting magnitude. By the time Gaia is launched, Pan-STARRS will have been surveying for several years at limiting magnitude 24.

While it is possible to observe at solar elongations as small as 30 degrees from Earth [Høg & Knude 2001], it is not practical for surveying, since it is only possible to do so for a very short time per day.

Due to the regular observation at low solar elongations, Gaia is expected to discover several of the elusive IEOs, objects having an orbit entirely interior to the Earth's orbit. While theories of existence had been widely accepted, the first of these objects was not discovered until 2003 [Evans et al. 2003].

Ultimately, since more of the sky is available to the space-based survey at any time, and search can (and will, in the case of Gaia) take place 24 hours a day, the space-based survey has an advantage. Also, because the sky-plane density of NEOs increases in the direction toward the Sun (figure 3.1), Gaia's regular low solar elongation observation makes it an excellent candidate for observing not only Atens and IEOs, but NEOs in general.

# Chapter 4

# Gaia
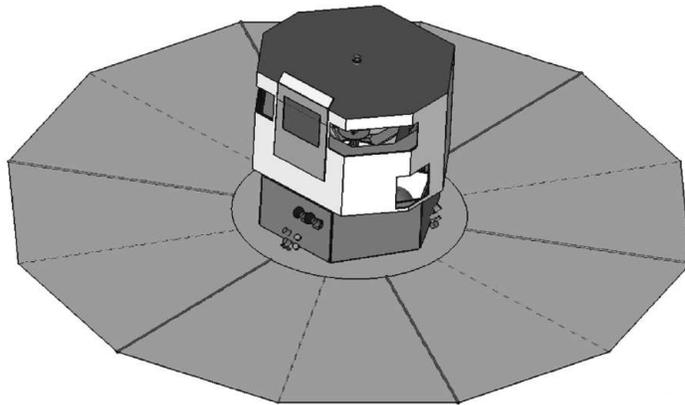


**Figure 4.1:** Schematic view of a possible design of the Gaia spacecraft. Diameter of deployed solar array ≈ 9 m.

Gaia is a European Space Agency mission aimed at making a complete census of all objects down to apparent magnitude 20. The positions of an expected 1 billion objects are to be determined with unprecedented accuracy from the Gaia space observatory. The resulting scientific harvest will provide detailed information on stellar evolution and star formation, as well as a clarification of the origin and formation history of our galaxy. Gaia is expected to discover thousands of extra-solar planets and follow the bending of starlight by the Sun, and therefore directly observe the structure of space-time. Relativistic parameters and the solar quadrupole moment will be determined with unprecedented precision. All this is achieved through the accurate measurement of star positions. Designated an ESA cornerstone mission, the Gaia spacecraft is expected to be launched in 2011-2012.

Although the main goal of Gaia is to clarify the origin and history of our Galaxy, this chapter will explore its capabilities for observing near Earth objects. Section 4.2 describes the Gaia instruments relevant for NEO observation. Following this section, the Astro

instrument is explained in further detail, as is a simulator by F. Mignard, predicting Gaia observations. In the penultimate section of this chapter, this simulator is used to predict the number of objects that "evade" observation due to their proper motion. Section 4.7 describes a proposed method for observing NEOs by dedicating four detectors in the Spectro instrument to this task.

## 4.1 Orbit and Scanning Principle

Gaia will perform its observations from a quasi-periodical orbit about Lagrange point L2 of the Sun-Earth system (see figure 2.7 and section 2.4). As shown in example 2, this point is situated 1.5 million kilometers from the Earth along the Sun-Earth line, opposite the Sun. Owing to the dynamical properties of this point, Gaia can maintain a roughly constant distance to the Earth for several years, using only modest corrective manoeuvres. To avoid steep temperature gradients and to ensure sufficient sunlight reaching the solar panels, Gaia must stay out of the Earth's shadow, a circular zone of radius only slightly larger than that of the Earth – approximately 6,500 kilometers. Hence, Gaia is placed in a so-called Lissajous orbit about L2 of an extent of about 300,000 kilometers, completing a full cycle every six months. The thermal stability of the region is very important, since a temperature variation of less than one thousandth of a degree over a few hours would disturb the alignment of the mirrors and thus significantly degrade the images [Mignard 2003].

The duration of the operational phase of the Gaia mission is five years. During its lifetime, the satellite will continuously spin with a constant speed corresponding to one full revolution every six hours. The spin axis is kept at a constant angle with respect to the Sun, precessing about the direction of the Sun approximately every 70 days. The image of a spinning top, revolving around its axis while the axis precesses around the vertical, is appropriate. The spin and precession of Gaia, coupled with its orbital motion around the Sun, enables the scanning of the entire celestial sphere.

The angle $\xi$ between the spin axis and the direction towards the Sun is determined as a compromise between thermal stability, power requirements and astrometric accuracy. It is currently 50 degrees, but is likely to be reduced to 45 degrees to enable the use of a smaller, and thus lighter, sunshield. For $\xi = 50°$, Gaia will be able to observe objects as close to the Sun as $\approx 40°$. If $\xi$ is decreased to 45°, the minimum solar elongation is correspondingly increased to 45°.

Figure 4.4 shows an example of the monthly sky coverage of Gaia, using the a solar aspect angle of $\xi = 45°$ and an across-scan field of view of 0.737° (see section 4.2). Based on a simulation of 10,000 random positions uniformly distributed over the sky, the plot shows the 6068 positions that were observed at least once. The fraction of the sky covered in one month varies between 51% and 68% depending on the starting time of the month. The average coverage is close to 60%. The scanning leaves two big "holes" (the blind spots
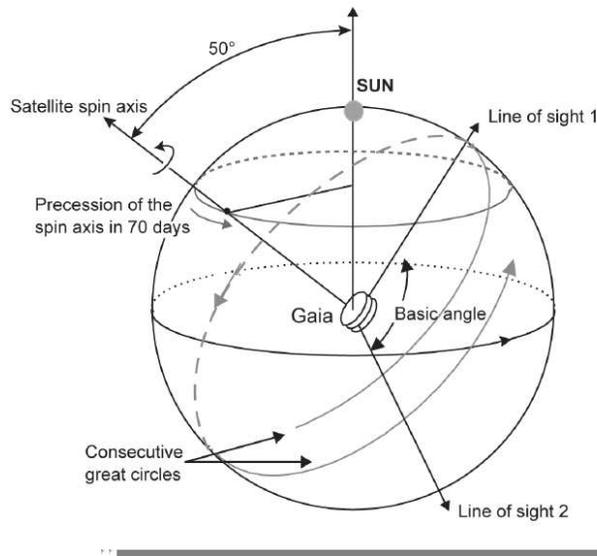
**Figure 4.2:** The scanning principle employed by Gaia. Line-of-sight 1 corresponds to the preceding field of view (Astro 1), line-of-sight 2 corresponds to the following field of view (Astro 2). Gaia makes a full revolution every 6 hours, while the spin axis precesses about the direction of the Sun once every 70 days. From [de Bruijne 2003-II].

illustrated in figure 4.3) centered on the direction of the Sun and on the direction opposite the Sun. Data from L. Lindegren, Lund Observatory (personal correspondence).

## 4.2   Gaia Instruments

Inside Gaia's payload module are three telescopes, two of which are identical. These two *Astro* instruments are dedicated to the accurate measurement of the stellar positions. Each consists of three curved, rectangular mirrors to focus the starlight. The largest mirror in each telescope system is 1.4 metres long. Each will focus its starlight onto the focal plane, an array of CCDs. They will measure the position and brightness of the celestial objects that Gaia detects. The lines of sight of the two Astro telescopes are separated by a *basic angle* of 99.4 degrees. Due to the 6-hour spin period, the second Astro telescope (Following Field Of View, *FFOV*) will observe approximately the same as the first Astro telescope (Preceding Field Of View, *PFOV*), only 99.4 minutes later.

The third telescope is of a different design. It is called the *Spectro* instrument and is designed to perform photometry (the detection of the brightness of celestial objects in a number of different colour bands) used to determine the physical parameters of celestial objects. It will also measure the radial velocity of celestial objects. This information can then be combined with that from the Astro instruments, to give a full picture of how the
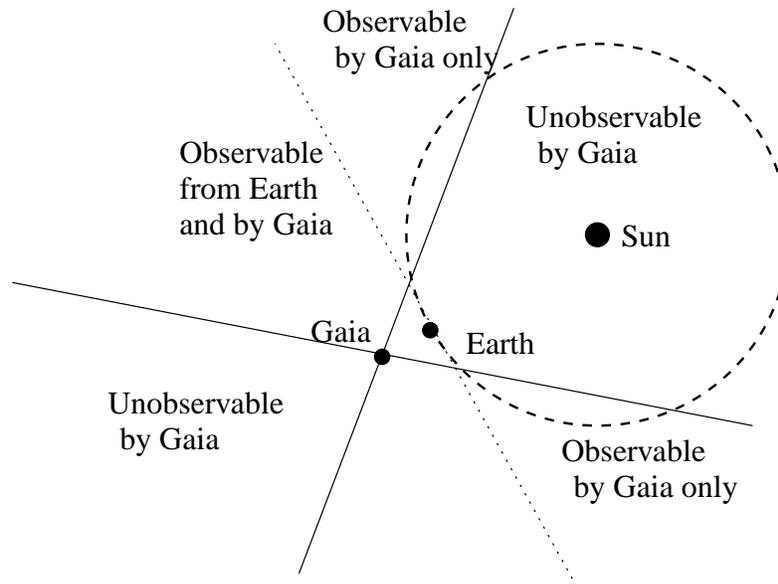
**Figure 4.3:** Diagram illustrating the "blind spots" of Gaia versus that of Earth-based telescopes. Gaia does not observe in the directions towards and away from the Sun, but does observe closer to the Sun than is routinely done from Earth, which is important for discovering NEOs inside the orbit of the Earth. While it is possible to observe closer than 90° from the Sun from Earth-based telescopes, this is typically only done in case of follow-up observations. Cf. figure 4.4. The diagram is not to scale.

celestial object is moving through space.

## Focal Planes

The light from the Astro and Spectro telescopes is projected onto the Astro and Spectro focal planes, shown in figures 4.5 and 4.7, respectively. These focal planes are arrays of CCDs, each consisting of a regular grid of several million pixels. Each box in the figures represents a vertical (i.e., in the across-scan direction) column of CCDs, sometimes also called a *CCD strip* or a *CCD row*.

### The Astro Focal Plane

The Astro focal plane, shown in figure 4.5, consists of two sky mappers (Astro Sky Mapper, ASM) and eleven astrometric fields (AF1-AF11). The light from both of the Astro telescopes is projected onto one focal plane, with the exception of ASM1 and ASM2 receiving light only from the preceding and following field of view, respectively. As the satellite spins, the light from celestial objects moves from left to right. As an object enters the preceding field of view, it is detected in ASM1. To avoid spurious observations caused

Gaia monthly sky coverage, example
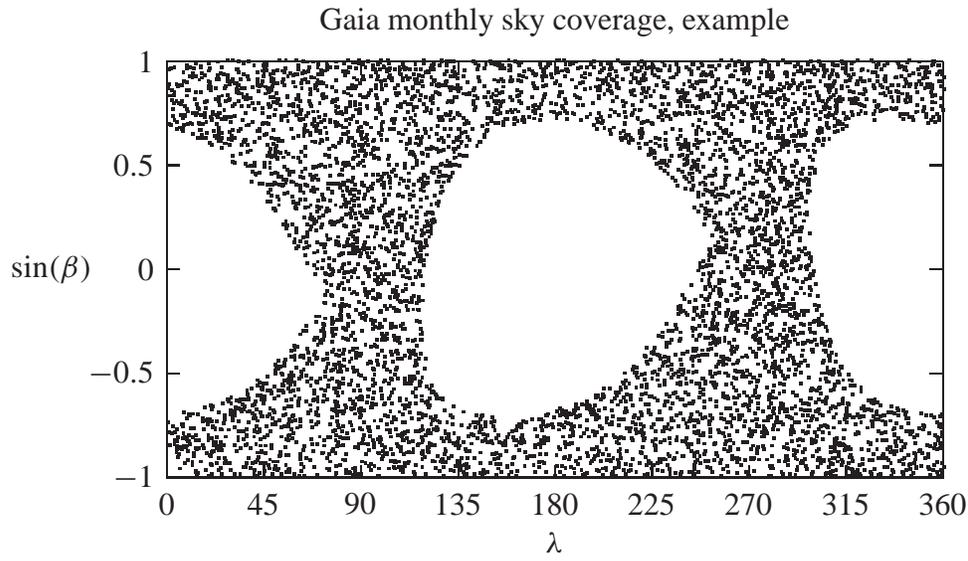


**Figure 4.4:** Gaia's monthly sky coverage in ecliptic coordinates. This example shows a typical sky coverage of $\approx 60\%$. Note the blind spots in the direction of the Sun and opposition (cf. figure 4.3). Data from L. Lindegren, Lund Observatory (personal correspondence)

by defective pixels or cosmic radiation, the observation must be confirmed as it reaches AF1 a few seconds later. This confirmation happens by examining a group of pixels (a *window*) corresponding to a position in the sky centered on the direction where the object was detected in the ASM1. The shape and size of the window, which may differ for each of the astrometric fields AF1-AF11, are determined based on the brightness of the detected object. Because windows are fixed in the sky, they should be large enough to make sure moving objects do not pass outside them. Contrarily, because windows containing multiple objects are of little value, windows should be as small as possible to avoid crowding. Once the object is confirmed, windows are recorded in each of the ten remaining astrometric fields. To reduce the amount of data transmitted to Earth, the pixel data can be summed into samples to be transmitted instead. For faint objects in AF2-AF11, the sampling includes a projection, or *binning*, onto the along-scan direction leading to a significant reduction in across-scan accuracy (see figure 4.6). The assumed windowing scheme is detailed in section 4.3. For details on the Astro sampling and windowing schemes, refer to the deep, comprehensive and dedicated studies by E. Høg in [Høg, et al. 2003-II], [Høg 2004], [Høg 2004-II], [Høg & de Bruijne 2005] and [Høg 2005].

Since the windows sampled in the astrometric fields are fixed on the sky, centered on the position where the object was detected in the sky mapper, a moving object may "escape" and thus evade observation. The probability of such an escape happening is examined in section 4.6.

For each transit of a sufficiently slow object, we can obtain the following data from each
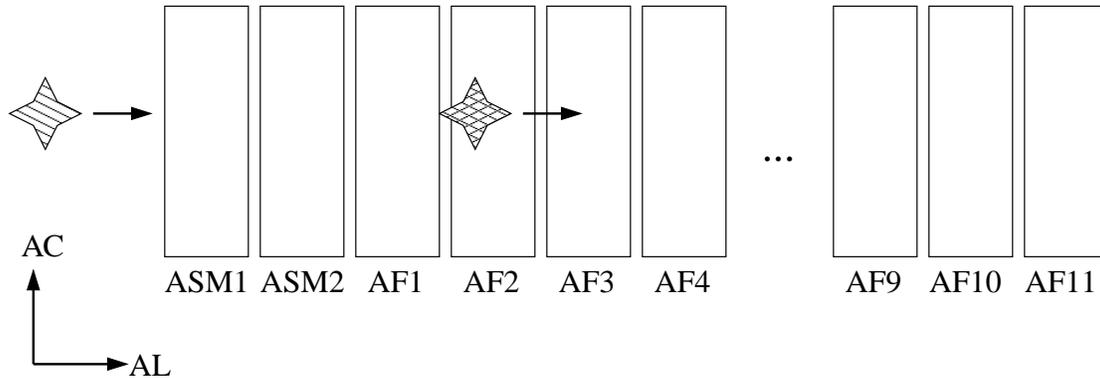
**Figure 4.5:** Simplified diagram of the Astro focal plane. As the satellite sweeps the sky, celestial objects appear to move from left to right, in the along-scan (AL) direction. AC: Across-scan direction. ASM: Astro Sky Mapper. AF: Astro Field.

CCD strip: object position, position standard deviation and a time tag. The position standard deviation is an estimate of the accuracy of the position observation. It depends on the brightness of the object, but also on the sampling scheme employed and on whether the result of the on-board centroiding (the process of estimating the true position based on samples) is transmitted to ground. Such triples of information will be extensively used for motion detection in chapter 5.

The Astro focal plane also contains a group of CCDs dedicated to broad-band photometry.

**The Spectro Focal Plane**

The Spectro focal plane, shown in figure 4.7, is where much of much of Gaia's photometric data originate. Since not every pixel can be transmitted to ground, due to the limited telemetry[1] budget, in analogy with the astrometric focal plane, a set of sky mappers is used to detect when an object is about to transit the Spectro focal plane. These Spectro sky mappers, SSM1, SSM2, SSM3 and SSM4 in figure 4.7, can be used to observe NEOs. An object entering the field of view of the Spectro telescope is detected in SSM1 and confirmed in SSM2 to avoid false detections, e.g., those caused by cosmic rays. This detection/confirmation is repeated in the SSM3/SSM4 pair of CCD columns. The motion of the object between the SSM1/SSM2 and SSM3/SSM4 is estimated. If the object has moved significantly, it is likely to be a near Earth object. A similar method for performing NEO observations in the Spectro instrument is described in section 4.7.

---

[1]Telemetry: the data transmitted to ground, and the process of transmitting it.
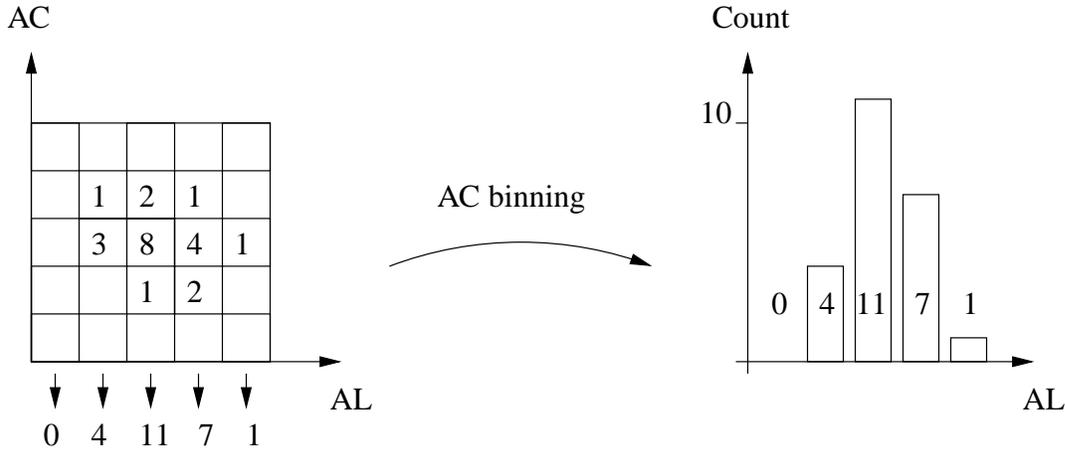
**Figure 4.6:** To reduce the amount of data transmitted to ground, pixel data is sometimes "binned" in the across-scan (AC) direction before transmitting. This data reduction comes at the price of reduced on-ground AC accuracy.

## 4.3   Astro Telescope Technical Data

In the remainder of this thesis, we assume the use of the Astro telescopes, following the design in [Høg, et al. 2003-II] for faint objects, i.e., Gaia visual magnitude 16 to 20. This implies a Sun aspect angle of 50 degrees and a basic angle between the Astro telescopes of 106 degrees, rather than the current basic angle of 99.4 degrees.

Regarding the focal plane, the width[2] of an ASM and the width of an AF (both including CCD interspace) is 0.03684 degrees and 0.06017 degrees, respectively. Assuming an ASM of 2600 pixels and an AF of 4500 pixels and an along-scan pixel size of $10\mu$m, this corresponds to an ASM width of 26 mm + 4 mm interspace and an AF width of 45 mm + 4 mm interspace, in accordance with [Pouny et al. 2003]. This data, coupled with the 6 hour spin period, leads to the transit times given in table 4.1.

The size of a pixel, projected onto the celestial sphere, is 44.2 mas along-scan and 133 mas across-scan, corresponding to a focal length of about 46.6 m.

The assumed windowing scheme is described in [Høg, et al. 2003-II] and summarised in table 4.2. Note, that these windows may be truncated and/or binned before being transmitted to ground. The philosophy behind this windowing scheme is to have a relatively large AF1 window to ensure confirmation of even the fastest objects. Many moving objects will also be observed in the large AF11 window, facilitating an accurate velocity estimation because of the large timebase. The latest scheme has a large AF6 window instead of the AF11, in order to observe faster-moving objects at the cost of a smaller timebase.

---

[2]*width* refers to the on-sky extent in the along-scan direction.
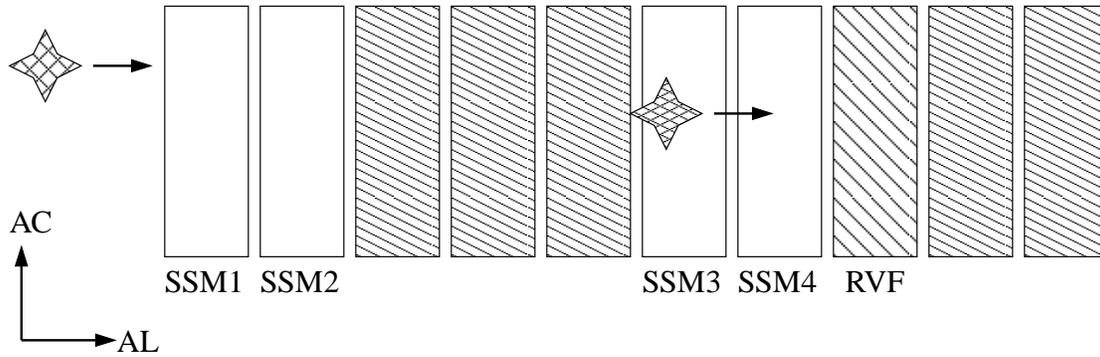
**Figure 4.7:** Simplified diagram of part of the Spectro focal plane. As the satellite sweeps the sky, celestial objects appear to move from left to right, in the along-scan (AL) direction. AC: across-scan direction. SSM: Spectro Sky Mapper. The remaining six detectors are used for photometric observations.

## 4.4  Gaia Simulator

To investigate various properties of the observations of Gaia, the author has made use of a Gaia software simulation written by F. Mignard of the Observatoire de la Côte d'Azur [Mignard 2001], [Mignard 2001-II]. This simulation takes as input the orbital elements of a population of solar system objects and computes all Gaia observations of each object in the population during the predetermined mission duration. The output includes the time of transit, the position and position standard deviation, the apparent magnitude and instantaneous inertial velocity[3] for each observation of each object.

Originally, the simulator output consisted of one observation per telescope. In February 2004, F. Mignard and the author modified the simulator to provide output for each of the twelve CCD strips in each Astro telescope. The windowing scheme is not simulated, so if the object is within the field of view, the observation is recorded. This enables subsequent analysis of the recorded observations, imposing the restrictions of the windowing scheme, see section 4.6. Figure 4.8 shows an example of the output of the simulator.

The simulator is also able to calculate for each observation a position standard deviation, i.e., the position standard deviation a similar observation by Gaia is expected to have. The error model determines the standard deviation based on the object's apparent magnitude, angular size, velocity and the phase angle. When modifying the simulator, the measurements from each of the CCDs were implemented to output the same position standard deviation. Owing to the difference in windowing and sampling, this is not the case. The error model (see [Mignard 2003-II] and [Hestroffer et al. 2003]) is essentially only correct for the astrometric sky mappers, but lacking a final fixed version of the windowing and sampling scheme, it was decided to keep this incomplete implementation as

---

[3]The inertial velocity is the motion of the object on the sky with respect to the fixed stars, i.e., corrected for the motion and attitude change of Gaia

|           | PFOV   | FFOV   |
|-----------|--------|--------|
| ASMx→AF1  | 5.1 s  | 2.9 s  |
| ASMx→AF2  | 8.7 s  | 6.5 s  |
| ASMx→AF3  | 12.3 s | 10.1 s |
| ASMx→AF4  | 16.0 s | 13.7 s |
| ASMx→AF5  | 19.6 s | 17.3 s |
| ASMx→AF6  | 23.2 s | 21.0 s |
| ASMx→AF7  | 26.8 s | 24.6 s |
| ASMx→AF8  | 30.4 s | 28.2 s |
| ASMx→AF9  | 34.0 s | 31.8 s |
| ASMx→AF10 | 37.6 s | 35.4 s |
| ASMx→AF11 | 41.2 s | 39.0 s |

**Table 4.1:** Assumed transit times: The time it takes for a fixed object to move from the sky mapper (ASM1 in the case of the Preceding FOV, ASM2 for the Following FOV) to each astrometric field, CCD center to CCD center.

| CCD      | Size in pixels        | Angular size           |
|----------|-----------------------|------------------------|
| AF1      | 12 pixels × 12 pixels | 530.4 mas × 1596 mas   |
| AF2-AF10 | 6 pixels × 12 pixels  | 265.2 mas × 1596 mas   |
| AF11     | 68 pixels × 14 pixels | 3005.6 mas × 1862 mas  |

**Table 4.2:** Assumed read window sizes (AL × AC). From [Høg, et al. 2003-II].

an approximation.

## 4.5 Simulator Input Data

The input data for the simulation is a file containing the orbital elements and absolute magnitude of each object in the population. Two populations have been used:

The first population consists of the first 20,000 numbered asteroids. Being the first 20,000 numbered, one would expect them to be the 20,000 brightest, which constitutes a bias. Although this population is designated MBO (Main Belt Objects) in the following, it also contains a number of NEOs, such as 433 Eros. A truncated population of the first 2000 asteroids has also been used.

The second population consists of a simulated population of NEOs, based on the work presented in [Bottke et al. 2000] and kindly provided by F. Mignard [Mignard 2001]. It constitutes a roughly complete NEO population for absolute magnitudes $H < 22$. Figure 4.9 shows the number of NEOs as a function of absolute magnitude. As shown in the figure, the number of bright objects is fairly low. To remedy this, in order to obtain a large

```
                  1              1 Ceres                 720
                200                 201                 202  ...
      53.8354758013      53.8355094717      53.8355512579  ...
             8.8462              8.8462              8.8462  ...
            -8.6787             -8.6787             -8.6787  ...
             8.9889              8.9889              8.9889  ...
           221.3800            221.3800            221.3800  ...
     263.64223961488    263.64224969653    263.64226220830  ...
       2.96659863466      2.96659787407      2.96659693015  ...
             4.8917              4.8917              4.8917  ...
             4.9270              4.9270              4.9270  ...


                  2              2 Pallas               1260
                100                 101                 102  ...
        (continued)
```

**Figure 4.8:** Simulator output example. The first line is a header, indicating object number, name "1 Ceres" and the number of observations of this object during the mission (720). The second line is the CCD id (200: ASM2, 201: FFOV AF1, 202: FFOV AF2, etc). The third line is the time of transit (in days). The fourth line contains the apparent magnitude (V). The fifth and sixth lines contain the object's instantaneous inertial velocity in the along-scan and across-scan directions, respectively (in mas/s). Line seven contains the current inclination (in degrees) of the scan circle with respect to the ecliptic plane. Lines eight and nine are ecliptic longitude and latitude (in degrees), and lines ten and eleven contain along-scan and across-scan position standard deviations in mas. Each of the lines 2-11 have 720 entries, as indicated in the header line.

number of simulated observations for statistical stability, each object is made brighter by the simulator by reducing the value of $H$ upon loading the orbital elements. Because of this increased brightness, many NEOs will be observed at a greater distance, thus introducing a bias reducing the average instantaneous inertial velocity observed.

## 4.6   Escape Statistics

Because the observation windows sampled in the astrometric fields AF1-AF11 are fixed on the sky, centered on the position where the object was detected in the sky mapper ASM1/ASM2, a moving object may "escape", i.e., move outside the window, and thus evade observation. This section describes the computation of these "escape probabilities". Based on the window sizes and transit times given in section 4.3, it is possible to compute the critical "escape velocities" for each CCD, $v_e(\text{CCD})$, as:
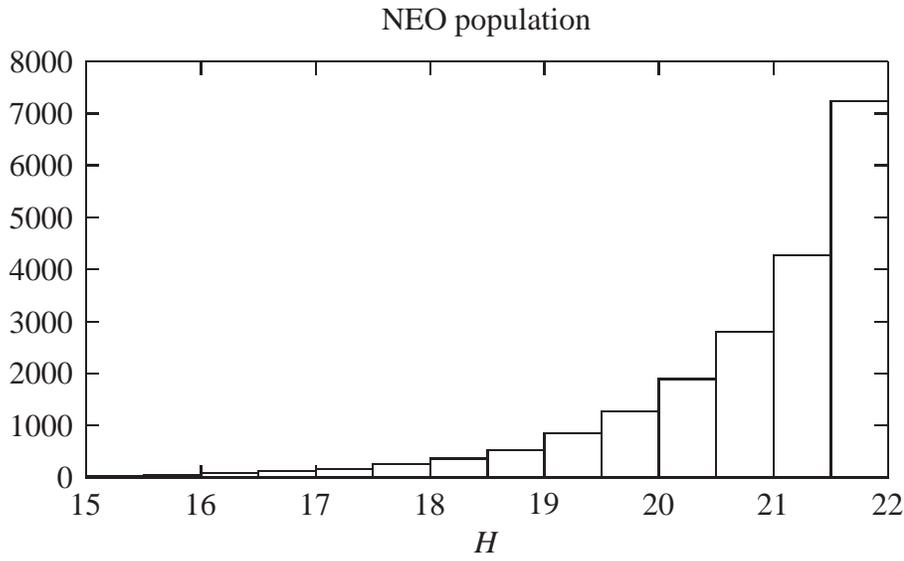
NEO population



**Figure 4.9:** The number of NEOs as a function of absolute magnitude $H$ for the simulated NEO population. The brightest object is of absolute magnitude $H = 13.1$, the faintest objects have $H = 22.0$. From [Mignard 2001].

$$v_e(\text{CCD}) \; = \; \frac{W(\text{CCD})}{2T(\text{CCD})}$$

where $W(\text{CCD})$ is the window size (in the along-scan or across-scan direction, as appropriate), given in table 4.2, and $T(\text{CCD})$ is the transit time for the appropriate CCD column, given in table 4.1. If the object displacement (velocity multiplied by transit time) exceeds half the window width, the object will escape. Table 4.3 lists the critical velocities.

Defining a *set of observations* as the observations obtained from a single object crossing a single telescope, such a set may contain up to twelve observations (one ASM and eleven AF observations). Because a detection in the ASM without confirmation in the AF1 is not recorded, the smallest number of observations in a set will be two.

By examining each set of observations, it is possible to generate escape probabilities. Since the windowing scheme is not implemented in the simulator, the probabilities are generated using a software tool, written by the author, implementing the windowing scheme in section 4.3.

Figure 4.10 shows the probability that an object's inertial velocity is sufficiently low for it to be observed in a CCD strip, based on simulations of a population of the 2000 first numbered asteroids. Because of the larger timebases in the preceding field of view (table 4.1), the probability of "survival" is slightly lower than in the following field. Note that

|       | PFOV AL    | PFOV AC    | FFOV AL    | FFOV AC    |
|-------|------------|------------|------------|------------|
| AF1   | 51.8 mas/s | 156 mas/s  | 91.1 mas/s | 274 mas/s  |
| AF2   | 15.2 mas/s | 91.4 mas/s | 20.3 mas/s | 122 mas/s  |
| AF3   | 10.7 mas/s | 64.7 mas/s | 13.1 mas/s | 78.8 mas/s |
| AF4   | 8.31 mas/s | 50.0 mas/s | 9.65 mas/s | 58.1 mas/s |
| AF5   | 6.78 mas/s | 40.8 mas/s | 7.64 mas/s | 46.0 mas/s |
| AF6   | 5.72 mas/s | 34.4 mas/s | 6.33 mas/s | 38.1 mas/s |
| AF7   | 4.95 mas/s | 29.8 mas/s | 5.40 mas/s | 32.5 mas/s |
| AF8   | 4.36 mas/s | 26.3 mas/s | 4.71 mas/s | 28.3 mas/s |
| AF9   | 3.90 mas/s | 23.5 mas/s | 4.17 mas/s | 25.1 mas/s |
| AF10  | 3.53 mas/s | 21.2 mas/s | 3.75 mas/s | 22.5 mas/s |
| AF11  | 36.5 mas/s | 22.6 mas/s | 38.5 mas/s | 23.9 mas/s |

**Table 4.3:** "Escape velocities". Objects moving faster than the indicated velocities in either the along-scan or across-scan directions will not be observed in that particular CCD. Example: An object moving at 5 mas/s AL and 22 mas/s AC, observed with the PFOV, will be observed in ASM1 and AF1 to AF6, but not in AF7, AF8 and AF9 because of the AL velocity exceeding the critical velocities listed. The critical velocities in both the AC and AL directions rule out observation in the AF10, whereas the object will be observed in AF11. Thus, the object's transit of the PFOV will lead to a set of eight observations. Extension of the table in [Wolff 2004].

because of the larger AF11 window, observations not observed in one or more of the astrometric fields may "reincarnate" in AF11.

The same population of asteroids is compared to a simulation run with the NEO population in figure 4.11, where the survival probabilities of preceding and following fields of view have been averaged. Because of their greater average inertial velocity, the survival probability of the NEOs is significantly lower. Less than 10% of the NEOs provide observations from all astrometric fields, but more than 40% are nevertheless observed in the big AF11 window, facilitating accurate velocity estimation because of the large timebase.

To get a larger number of accurate NEO velocity estimations, it has been suggested to change the windowing scheme described in section 4.3 to a scheme exchanging the window sizes of CCDs AF5 and AF11, such that AF5 would have a large window, whereas AF11 would have one of normal size. This change has been implemented in the software tool mentioned above, leading to the results shown in figure 4.12. The "big-AF5" windowing scheme led to a 10% increase in the total number of NEO observations whereas the number of asteroid observations decreased by 4%, when compared to the original scheme. At first glance, this decrease might seem odd. However, due to the relatively low inertial speed of the asteroids, the gain from having a big window in AF5 is almost negligible, whereas the loss caused by a small window in AF11 is not.

Recalling the bias introduced by the artificial brightening of the NEO population (see section 4.5), reducing the average velocity, the figures involving this population are believed to be slightly optimistic.
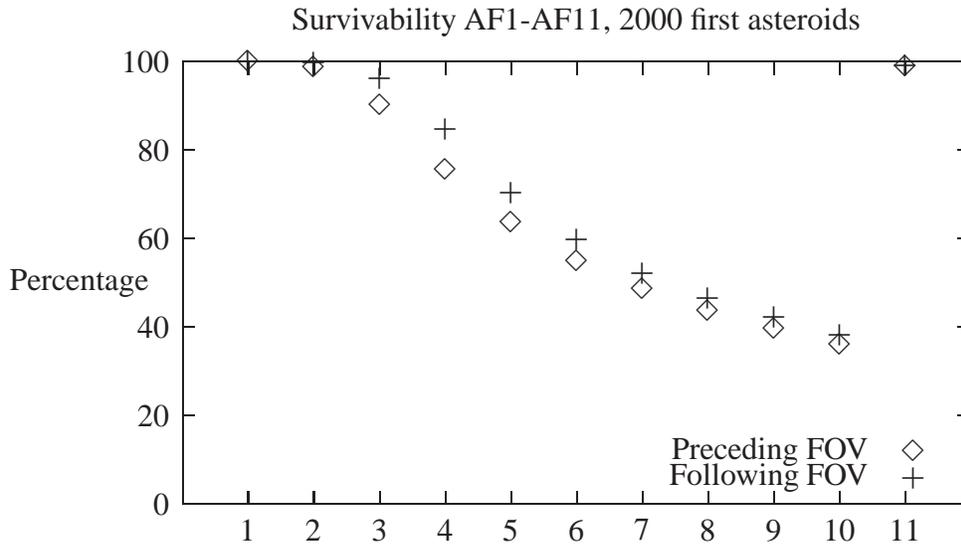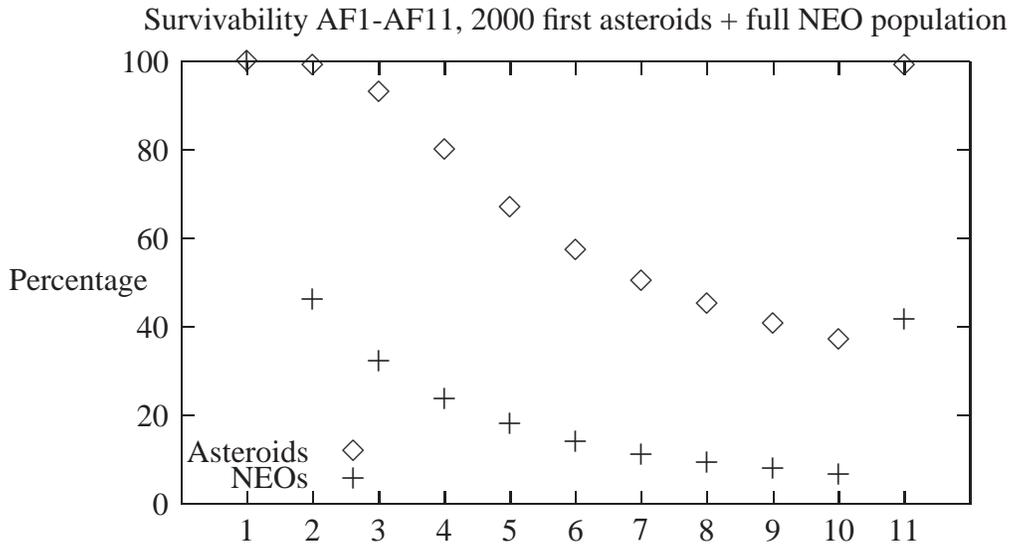
**Figure 4.10:** "Survivability" of observations of the 2000 first asteroids in each of the astrometric fields AF1 to AF11. The probability that an object's inertial velocity is sufficiently low for it to be observed in a CCD strip. The preceding FOV loses more observations because of the greater distance to its sky mapper, ASM1. Almost all asteroids are observed in the larger AF11 window. From [Wolff 2004].

Note, that this method disregards any possible problems caused by the centroid being near the edge of the CCD. This and other problems are touched upon in [de Bruijne 2005-III], which contains a simpler method for doing a similar calculation assuming normal velocity distributions and independent AL and AC velocities. When using the same focal plane design, the results obtained with the simple statistical method are comparable (within 10 percentage points) to the results presented here. It is believed that the difference can be attributed to the difference in assumptions of the velocity distributions.

## 4.7 NEO Observation in the Spectro Instrument

The method described in this section was developed jointly by E. Høg, F. Arenou, P. Hjorth, U. G. Jørgensen, F. Mignard and the author of this thesis [Høg, et al. 2003]. It was originally envisioned for a slightly different design of the Spectro focal plane, described in [Høg, et al. 2003-II], but is straight-forward to transfer to the current design, which was approved by the Gaia Science Team in March 2004 [Pace 2004]. It makes use of four CCD columns, namely RVSM (Radial Velocity Sky Mapper) #1, #2, #7 and #8, corresponding roughly in position to SSM1, SSM2, SSM4 and RVF in figure 4.5.

The previous design of the Spectro instrument consists of several clusters of CCDs, including an eight-CCD "auxiliary" cluster. We propose (in [Høg, et al. 2003]) to use some

Survivability AF1-AF11, 2000 first asteroids + full NEO population



**Figure 4.11:** "Survivability" of observations of the 2000 first asteroids and the simulated NEO population in each of the astrometric fields AF1 to AF11. Less than 10% of the faster-moving NEOs provide observations from all astrometric fields, but more than 40% are nevertheless observed in the big AF11 window. From [Wolff 2004].

of the CCDs in this cluster for detecting faint moving objects, including NEOs. The integration time for each CCD is 5.5 seconds and the interspacing is 2.1 seconds, yielding 7.6 seconds between each CCD. CCDs #1, #2, #7 and #8 are used to maximise time base. The CCDs are used in pairs to reduce false detections from cosmic rays.

1. Detect object in CCDs #1 and #2. This redundancy is necessary to reject false detections from spurious cosmic rays.

2. Rediscover the same object using CCDs #7 and #8.

3. Reject object if it is not moving.

4. Otherwise, transmit data to ground.

Ad 1: The basic assumptions on RVSM are listed in table 4.4, where $G$ is the apparent visual magnitude from the point of view of Gaia. The detection probability for a single field transit is denoted by $P$. The standard deviation in the determination of position is given by $\sigma_{AL}$ and $\sigma_{AC}$ in the along-scan direction and across-scan direction, respectively. The probability of successful detection in all four CCDs is denoted by $P^4$, and $\sigma_{v_{AL}}$ and $\sigma_{v_{AC}}$ give the standard deviations on the velocity approximation in the along-scan and across-scan directions.

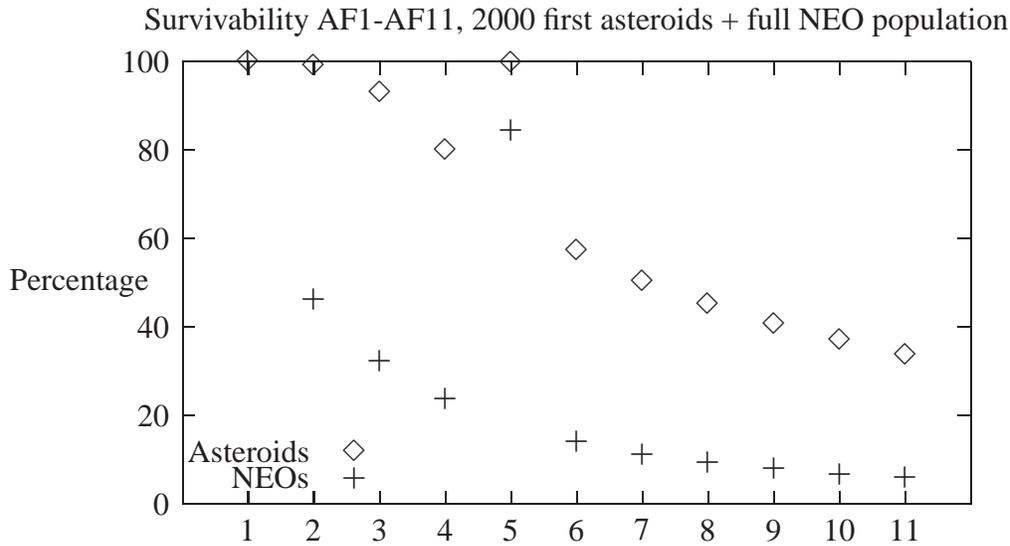Survivability AF1-AF11, 2000 first asteroids + full NEO population



**Figure 4.12:** "Survivability" of observations of the 2000 first asteroids and the simulated NEO population in each of the astrometric fields AF1 to AF11, "big AF5" windowing scheme. This windowing scheme gives rise to a 10% increase in NEO observations at the cost of a 4% decrease in asteroid observations, when compared to the original windowing scheme (figure 4.11). From [Wolff 2004].

Ad 2: This rediscovery should be performed by examining an area centered on the expected position in the sky, i.e., the position at which it was observed using CCDs #1 and #2. The radial extent of this area equals the expected maximum velocity (according to figure 5.9, 60 mas/s for NEOs appears to be a reasonable cutoff, minimising the window size while still retaining more than 75% of the observations[4]) multiplied by the time base. What to do when several objects are detected inside this area has not yet been decided upon. This typically happens in areas of great star densities (say within 10 degrees of the galactic plane, constituting 20% of the sky), where telemetry peaks. However, in these high density areas, the priority of NEO search can be limited. It can be seen in table 4.4 that the probability for detecting an object of magnitude G=21 in all four CCDs is 41%, indicating the faintest feasible magnitude for detection using this method.

Ad 3: To reduce telemetry, objects that are deemed not to be moving (with respect to the fixed stars) are rejected, since they are expected to be observed in the Astro instruments. This rejection is performed by calculating the standard deviation of the velocity, which in turn is approximated by the displacement divided by the time base. Across-scan and along-scan standard deviations on velocity for a time base of 46 seconds may be found in table 4.4 for various G magnitudes. The proposed method suggests rejecting objects not satisfying $v > 3\sigma_v$. Using the appropriate estimates, this is shown in section 5.6 to be an

---

[4]mas: millisecond of arc.

| G | P | $\sigma_{AL}$ | $\sigma_{AC}$ | $P^4$ | $\sigma_{v_{AL}}$ | $\sigma_{v_{AL}}$ |
|---|---|---|---|---|---|---|
| 20.0 | 91% | 100 mas | 240 mas | 68% | 2 mas/s | 5 mas/s |
| 20.5 | 88% | 140 mas | 280 mas | 60% | 3 mas/s | 6 mas/s |
| 21.0 | 80% | 210 mas | 360 mas | 41% | 5 mas/s | 8 mas/s |
| 21.5 | 46% | 260 mas | 405 mas | 4% | 6 mas/s | 9 mas/s |
| 22.0 | 10% | 310 mas | 495 mas | 0% | 7 mas/s | 11 mas/s |

**Table 4.4:** RVSM assumptions. Time base (#1 to #7): $6 \times (5.5 + 2.1)$s $= 45.6$ s. Does not include blurring due to motion.

optimal rejection strategy.

# Chapter 5

# Motion Detection and Estimation

This chapter, constituting the main contribution of this thesis, presents five different methods of motion detection, based on the fields of data-fitting and theoretical statistics. The last of the five is proved to be optimal among all translation invariant method, assuming a symmetric velocity distribution. The relative performance of all five tests is compared, and their individual advantages and disadvantages are discussed. The optimal test is applied to simulated Gaia observations in section 5.10. In the final section of the chapter, the properties of the velocity estimate emerging from two of the methods are examined with reference to its use in orbit computation.

## 5.1   Overview

In order to accurately compute the orbit of an object, it is necessary to obtain many observations, spanning a significant portion of the orbit. This typically involves distinguishing observations belonging to the object in question from observations belonging to other objects. In other words, *linking* a series of observations of that particular object. The Gaia mission is expected to observe on the order of one billion objects on the average of about 80 times each [de Bruijne 2005], leading to a significant amount of work when linking observations. Although temporal and spatial limitations may be imposed to reduce the search space, the total task of linking all observations for each of the expected half a million solar system objects [ESA 2000] is obviously a non-negligible undertaking. Introducing a filtering step to discard observations of non-solar system objects would reduce the needed work by many orders of magnitude. Such a filter could be approximated by thresholding according to the apparent instantaneous velocity, since nearby objects would, in general, appear to be moving faster than remote objects.

Whenever an object passes through the field of view of one of Gaia's astrometric telescopes, up to twelve observations are recorded within about forty seconds. These observations may be used to estimate the object's velocity at the time of the crossing of the

field of view. Thus, every field of view crossing may be interpreted as leading to a position measurement as well as a velocity measurement, contrary to conventional astrometry, yielding only a position measurement.

The star having the highest known *proper motion* is Barnard's Star, the fifth closest known star[1] to Earth, moving 10.3 seconds of arc per year. Pluto, one of the "slowest" members of the solar system, having a siderial period of 248 years, moves more than 5000 seconds of arc per year. Distinguishing solar system objects from non-solar system objects based on instantaneous proper motion thus seems feasible[2]. However, the method is not flawless: A solar system object having an instantaneous velocity vector (almost) parallel to its topocentric position vector, will have an instantaneous proper motion that is very small, and may thus be regarded as a star and erroneously excluded from the set of observations of solar system objects. However, this is expected to occur only rarely. Figure 5.9 displays the inertial speed of simulated Gaia observations of the synthesized NEO population described in the previous chapter.

The proper motion of Barnard's star, 10.3 seconds of arc per year, corresponds to about 0.3 microsecond of arc per second. The along-scan centroiding error of a single CCD transit for a bright object, such as Barnard's star, is approximately 0.04 milliseconds of arc [de Bruijne 2005-II]. In section 5.9 it is shown, that even this, the fastest-moving star, will be regarded as a fixed object by Gaia, whereas nearly all observations of NEOs and Main Belt Asteroids reveal the underlying motion.

This chapter will show that, owing to Gaia's impressive (yet finite) astrometric accuracy, an approximate distinction between observations of solar system objects and observations of objects outside the solar system may be obtained by distinguishing between moving and fixed objects. Because of the limited astrometric accuracy, a slow-moving object will be regarded as a fixed object.

The across-scan pixel binning (see section 4.2) leads to reduced accuracy in the across-scan direction. Because of this, as well as to initially simplify the problem, we will only consider position data in the along-scan direction. Thus, we assume we have obtained a *set of observations*, consisting of $N$ along-scan position observations (between two and twelve, from ASM and AF1-AF11) $x_1, x_2, \ldots, x_N$, at observation times $t_1, t_2, \ldots, t_N$, along with an approximate residual variance for each observation: $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. The following sections describe several methods to use such a *set of observations* to obtain a Boolean variable (a flag) indicating whether or not the observed object is (currently) moving, and hence, indirectly indicating whether or not the observed object is a solar system object.

The assumed model is:

$$x_i = P_0 + v t_i + \epsilon_i , \tag{5.1}$$

---

[1]The Sun being one of the five

[2]*Transneptunian Objects*, objects with an orbit beyond that of Neptune may move slower than Pluto. However, at Gaia's limiting magnitude, not many TNO observations are expected

where $x_i$ is the $i$th observed position, $t_i$ is the time of the $i$th observation, $P_0$ is the true position the time of the first observation, which implies that the time of the first observation equals zero. The velocity, defined as the instantaneous proper motion, assumed to be constant during a field crossing, is denoted by $v$. We will assume that the measurement errors $\epsilon_i$ are independent normal variables, of zero mean and variance $\sigma_i^2$. According to the discussion on the error statistics of asteroid observations in [Carpino et al. 2003], this is a valid assumption. We assume constant $\sigma_i^2 = \sigma^2$ for all but the generalised method presented in section 5.8. All spatial quantities are projections on to the along-scan direction. Detecting motion is tantamount to analysing the observed positions to test whether or not $v$ equals zero.

We assume as a null hypothesis that the observed object is fixed in the sky, i.e., that the fluctuations observed are due to measurement error. The following sections present four methods, co-developed by F. Mignard[3] and the author, for testing this hypothesis; four statistical tests for *trend*, a (linear) change in position over time. The philosophy behind these four tests is to assume as little as possible about the distribution of the position residuals when devising the test statistic. The last part of this chapter describes an optimal method for motion detection, arising from a fundamentally different approach to the one followed when developing the first four tests: To calculate a test statistic based directly on the comparison of the computed probability of the measurements given a zero velocity and a known (non-zero) velocity, assuming full knowledge of the distribution of the position residuals.

Applying a test we can commit two type of errors. We can label a fixed object as moving, in which case we say we commit a type I error, and we can label a moving object as fixed, and in that case we commit a type II error. The situation is summed up in the following diagram:

|                    | Object fixed | Object moving |
|--------------------|--------------|---------------|
| Flagged as fixed   | Correct      | type II error |
| Flagged as moving  | type I error | Correct       |

In the following, the probability of erroneously labelling a fixed object as moving is denoted by $p_\mathrm{I}$, and, conversely, the probability of erroneously labelling a moving object as fixed is denoted by $p_\mathrm{II}$.

## 5.2   Linearity Assumption

In this chapter we assume that the underlying motion can be assumed to be linear during the short periods of observation. This section investigates the plausibility of this assumption.

---

[3]Observatoire de la Côte d'Azur, France.

A way of examining whether or not it is plausible that the short-term motion of an object can be described by a straight line, is by using simulation. From one field crossing, the orbit is propagated assuming (I) linear motion and (II) Keplerian motion. A statistical analysis of the difference in position may then be performed and interpreted as a measure of non-linearity. The work in [Mignard 2005], based on the simulated observations of Main Belt Asteroids, indicates that the mean distance between the two predictions after 60 days is about 4.9 degrees.The mean distance is described as approximately proportional to the square of the time. Scaling these values according to this proportionality yields a mean distance of 1.33 $\mu$as after the 45 seconds needed for a transit of the astrometric focal plane, thus providing a rough order of magnitude of effect of non-linearity.

An alternative way of testing the linearity assumption is to fit observations to a straight line in a least-squares sense, and subsequently examining the magnitude of the residuals. Observations from the crossing of one field of view were generated using the simulator described in chapter 4, and subsequently filtered using the tool that implements the windowing scheme. These filtered observations were then fitted to a straight line in a least-squares sense, and the maximum residual extracted. The root-mean-square of these maximal residuals proved to be less than half a microarcsecond for the NEO population, which was expected to have the most non-linear behaviour. Even the largest of the maximal residuals was less than 1/60th millisecond of arc.

Because most NEO observations will be very faint, we expect NEO observations to have position errors of several milliseconds of arc. Hence, the non-linearity is expected to be far smaller than the measurement error, thus justifying the linearity assumption.

## 5.3 Evaluation

In order to compare different methods of motion detection, a Monte Carlo software tool was developed by the author. Given the number of observations, the velocity $v$ of the observed object and the residual variance, a simulated set of observations is generated as randomly perturbed observations of linear motion, according to (5.1). A given method can then be applied on this data set, yielding a flag that indicates if the method has detected motion. This sequence of generation and application is then repeated a large number of times to obtain the probability $p_{\text{fix}}$ that the method cannot reject the null hypothesis, given $v$. In other words, $p_{\text{fix}}$ is the probability that the set of observations is labelled as belonging to a fixed object. This probability may subsequently be plotted as a function of the velocity, facilitating visual comparison of the methods. The ideal plot would be similar to a $\delta$-function, such that the null hypothesis accepted if and only if $v = 0$:

$$p_{\text{fix,ideal}}(v) = \begin{cases} 1 & \text{for } v = 0 \\ 0 & \text{for } v \neq 0 \end{cases}$$

Because of random measurement errors, such an ideal plot cannot be achieved.

The tool scales the velocity $v$ according to the observational error $\sigma$, which, as mentioned above, is assumed to be the same for all observation in a set. In figures 5.1 to 5.8 the simplified timing $t_i = i$ (in seconds) is used. Figures 5.10 and 5.11 are based on the transit times for Gaia's preceding field of view (see table 4.1).

## 5.4 Non-parametric Tests

This section describes two tests for motion. Both are independent of the distribution of the residuals, and are thus said to be non-parametric. We only assume that, under the null hypothesis, each positional error is independent and identically distributed. The tests make no use of the relative measurement times, only the order in which the position measurements are made.

### Successive Squared Differences

Under the null hypothesis (i.e., assuming the object is fixed in the sky), let $x_1, x_2, \ldots, x_N$ denote a series of $N$ independent and identically distributed elements, with mean $\mu$ and variance $\sigma^2$. The expected value of the square of the difference between any two elements $x_i$ and $x_j$, $i \neq j$, is:

$$\mathrm{E}\left((x_i - x_j)^2\right) = \mathrm{E}\left(x_i^2 + x_j^2 - 2x_i x_j\right) = 2\left(\mathrm{E}\left(x_i^2\right) - (\mathrm{E}(x_i))^2\right) = 2\sigma^2,$$

where the definition of variance is used: $\mathrm{Var}(x) = \mathrm{E}\left((x - \mathrm{E}(x))^2\right) = \mathrm{E}(x^2) - (\mathrm{E}(x))^2$. In short, half the squared distance is expected to equal the variance. We now pair element $i$ with its neighbour, element $i + 1$, for $i = 1, 2, \ldots, N - 1$. Half the average of the squared distance between the neighbours in each of the $N - 1$ pairs, is also expected to equal the variance. We write this as

$$\mathrm{E}\left(q^2(N)\right) = \sigma^2, \tag{5.2}$$

where

$$q^2(N) = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2.$$

The test statistic is the ratio of $q^2(N)$ to the variance:

$$\gamma_{\mathrm{SD}}(N) = \frac{q^2(N)}{S^2(N)},$$

where the variance is expressed as an unbiased sample variance estimate:

$$S^2(N) = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \,,$$

and where $\bar{x}$ is the estimated mean:

$$\bar{x} = \bar{x}(N) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Because the expected value of the variance estimate $S^2$ equals the variance $\sigma^2$, the expected value of the test statistic $\gamma_{SD}(N)$, under the null hypothesis, would be 1, as seen from (5.2). Conversely, if there was a trend, one would expect an element to be, on the average, closer to its neighbouring element than to the mean, and thus, the test statistic $\gamma_{SD}(N)$ would be less than 1.

By calculating the test statistic $\gamma_{SD}(N)$ and comparing it to a threshold value, one may estimate whether or not the object in question was moving at the time of observation. If the test statistic exceeds the threshold value, the null hypothesis cannot be rejected and the object is assumed to be fixed. The threshold value may be determined from statistical tables or computed using, e.g., a Monte Carlo method.

This method has some interesting asymptotic properties for large sample sizes, but since the sample size never exceeds twelve elements in this application, the reader is referred to [Aïvazian 1978], where the method is described under *le critére des carrés des différences successives*.

Figure 5.1 shows $p_{\text{fix}}$ as a function of the velocity $v$ for the method of successive squared differences. The method has been calibrated such that $p_{\text{I}} = 4.5\%$. This seemingly arbitrary value is caused by a limitation inherent in the Mann-Kendall method, described in the next section.

## Mann-Kendall

The Mann-Kendall method is based on Kendall's $\tau$ statistic, described in [Kendall 1938], used as a measure of correlation in a bivariate population. If we treat the element number (1,2,3,4,...) as one of the variables, and the time-ordered position observations as the other variable, then the correlation between the two populations can be considered as an indication of a trend.

Under the null hypothesis, i.e., when the observations are independent and identically distributed, a positive or negative sign of the difference between any unique pair of observations is expected to be equally likely. This observation forms the basis of the Mann-Kendall test [Mann 1945]. The test statistic of this test is:

Probability for a 'fixed' label, based on 12 observations
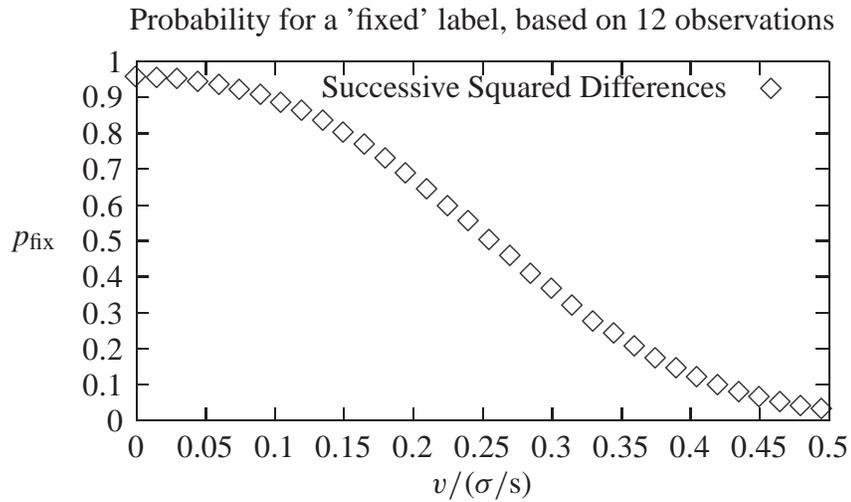


**Figure 5.1:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations and calibrated such that $p_I = 4.5\%$. For velocities $v \geq 0.47\sigma/s$, the probability of erroneously labelling an object as fixed, $p_{II}$, is less than 5%. From [Wolff 2005-III].

$$\gamma_{MK}(N) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} \text{sign}(X_i - X_j)$$

where

$$\text{sign}(z) = \begin{cases} -1, & z < 0 \\ 0, & z = 0 \\ 1, & z > 0 \end{cases}$$

Ignoring the possibility of identical observations ($X_i = X_j, i \neq j$), the mean and variance of the test statistic may be derived as follows:

Introducing the variable $Y_{ij}$, defined for $i > j$, as

$$Y_{ij} = \begin{cases} 1, & X_i > X_j \\ -1, & X_i < X_j \end{cases}$$

Under the null hypothesis, $Y_{ij}$ is equally likely to be 1 or $-1$ and therefore $E(Y_{ij}) = 0$ and $E(Y_{ij}^2) = 1$. The mean of the test statistic $\gamma_{MK}(N)$ is

$$E(\gamma_{MK}(N)) = E\left(\sum_{i=2}^{N} \sum_{j=1}^{i-1} Y_{ij}\right) = 0$$

The variance may be found by evaluating

$$\text{Var}\left(\gamma_{\text{MK}}\right) = E\left(\gamma_{\text{MK}}^2\right) - (E\left(\gamma_{\text{MK}}\right))^2 = E\left(\gamma_{\text{MK}}^2\right) = E\left(\sum_{i>j,i'>j'} Y_{ij}Y_{i'j'}\right) \qquad (5.3)$$

This is may be done by splitting up the sum into six cases:

| Case no. | Property | No. of terms |
|---|---|---|
| 1 | $i, j, i', j'$ all distinct | $\binom{N}{2}\binom{N-2}{2}$ |
| 2 | $i = i', j = j'$ | $\binom{N}{2}$ |
| 3 | $i = i', j \neq j'$ | $2\binom{N}{3}$ |
| 4 | $i \neq i', j = j'$ | $2\binom{N}{3}$ |
| 5 | $i > j = i' > j'$ | $\binom{N}{3}$ |
| 6 | $i' > j' = i > j$ | $\binom{N}{3}$ |

Summing the number of terms yields $\frac{1}{4}N^2(N-1)^2$ as expected, the total number of terms in the sum (5.3).

*Case 1. $i, j, i', j'$ all distinct.* By independence, $E\left(Y_{ij}Y_{i'j'}\right) = 0$. Total contribution is zero.

*Case 2. $i = i', j = j'$.* In this case, $E\left(Y_{ij}Y_{i'j'}\right) = 1$. Since this happens in $\binom{N}{2} = \frac{N(N-1)}{2}$ terms, the total contribution is $\frac{N(N-1)}{2}$.

*Case 3. $i = i', j \neq j'$.* If and only if $X_i$ is either greater than or less than both $X_j$ and $X'_j$, then $Y_{ij}Y_{i'j'} = 1$. The probability of this, under the null hypothesis, is $\frac{2}{3}$. Therefore, $E\left(Y_{ij}Y_{i'j'}\right) = (+1)\frac{2}{3} + (-1)\frac{1}{3} = \frac{1}{3}$. Since the number of terms of this case equals $2\binom{N}{3}$, the contribution is $\frac{2}{3}\binom{N}{3}$.

*Case 4. $i \neq i', j = j'$.* This case is similar to case 3. The contribution is $\frac{2}{3}\binom{N}{3}$.

*Case 5. $i > j = i' > j'$.* If and only if $X_j$ is between $X_i$ and $X'_j$, then $Y_{ij}Y_{i'j'} = 1$. The probability of this, under the null hypothesis, is $\frac{1}{3}$. Therefore, $E\left(Y_{ij}Y_{i'j'}\right) = (+1)\frac{1}{3} + (-1)\frac{2}{3} = -\frac{1}{3}$. Since the number of terms of this case equals $\binom{N}{3}$, the contribution is $-\frac{1}{3}\binom{N}{3}$.

*Case 6. $i' > j' = i > j$.* This case is similar to case 5. The contribution is $-\frac{1}{3}\binom{N}{3}$.

Summing the total contributions of each case gives the variance of $\gamma_{MK}(N)$ according to (5.3):

$$\text{Var}\left(\gamma_{MK}(N)\right) = \text{E}\left(\sum_{i>j,i'>j'} Y_{ij}Y_{i'j'}\right) = 0 + \frac{N(N-1)}{2} + \frac{2}{3}\binom{N}{3} = \frac{N(N-1)(2N+5)}{18}$$

For large sample sizes, the test statistic converges to a normal random variable under the null hypothesis [Mann 1945]. The test statistic divided by the square root of its variance is $\mathcal{N}(0, 1)$-distributed, which leads to the easy determination of a suitable threshold.

For smaller values of $N$, as is the case in the present problem, the thresholds can be found in statistical tables or computed using, e.g., Monte Carlo methods. These small values of $N$ also lead to an additional difficulty using this method: Since the test statistic can only take on values $-N(N-1)/2, -N(N-1)/2+2, \ldots, N(N-1)/2-2, N(N-1)/2$, it is not possible to test against arbitrary levels of confidence. For example, for $N = 8$, the double sided test $|\gamma_{MK}(N)| < \alpha$ has confidence levels 98.6%, 96.9%, 93.9% and 89.2% for $\alpha$ equal to 18, 16, 14 and 12, respectively. Testing against a 95% confidence level is thus not possible for $N = 8$. The use of the Mann-Kendall test for an observation set of, say, $N = 2$ observations, is also of little value, since, in this case, the test statistic will only assume the values $\gamma_{MK}(2) = \pm 1$. This limited threshold resolution is the reason for the apparently arbitrary levels of confidence of $p_{\text{I}} = 4.5\%$ and $p_{\text{I}} = 8.3\%$ in figures 5.1 to 5.8.

Figure 5.2 shows $p_{\text{fix}}$ as a function of the velocity $v$ for the method of successive squared differences and the Mann-Kendall method. The latter performs significantly better at intermediate velocities.

## 5.5 Parametric Tests

In the following sections, the residuals are assumed to be distributed according to normal law with mean $\mu = 0$ and constant variance $\sigma_i^2 = \sigma^2$. The first test compares the bias-corrected sample variance to the expected variance, and, similar to the previous two, ignores any details about the specific measurement times, using only the ordering of the measurements. The second test fits the measurement data to a straight line $x_i = a + bt_i$ and subsequently tests whether the slope $b$ (being a velocity estimate) is significantly different from zero.

### Variance Ratio

This test is based on the ratio of the empirical variance, $S^2$, to the expected variance, i.e., the expected position standard deviation squared. If the observed variance (the empirical variance) can be explained by the expected variance alone, no trend is observed.

Probability for a 'fixed' label, based on 12 observations
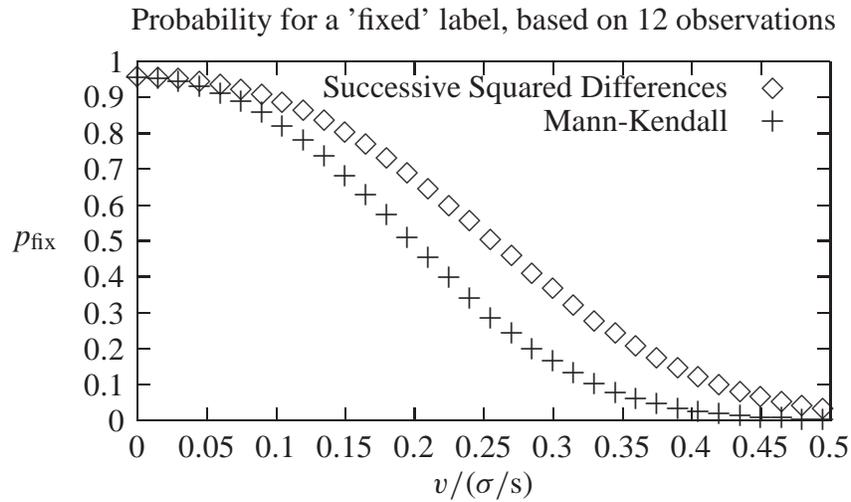


**Figure 5.2:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations and calibrated such that $p_I = 4.5\%$. The Mann-Kendall method is the better method in this case – it is closer to the ideal plot, as described in section 5.3. The probability of erroneously labelling an object as fixed, $p_{II}$, is less than 5% for $v > 0.36\sigma/\text{s}$ for the Mann-Kendall method, whereas in the case of the method of successive squared differences, this is only true for $v > 0.47\sigma/\text{s}$. From [Wolff 2005-III].

Conversely, the case of the observed variance exceeding the expected variance can be explained by a trend in the series of observations.

Under the null hypothesis, i.e., when the observed variance is fully explained by the expected variance, $\sigma^2$, it holds that

$$\gamma_{\chi^2}(N) = (N-1)\frac{S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

is distributed according to a $\chi^2(N-1)$ law (see e.g., [Kendall & Stuart 1961]). Thus, testing for a trend may be done by performing a one-sided test of the above statistic against a $\chi^2(N-1)$ distribution.

Figure 5.3 shows $p_{\text{fix}}$ as a function of the velocity $v$ in a comparison of the method of successive squared differences, the Mann-Kendall method and the variance ratio method. The difference between the first and last methods is slight. Compare figure 5.3, based on twelve observations (e.g., Gaia's ASM and AF1-AF11), to figure 5.4, based on only four (e.g., a Gaia set of observations reduced to four observations because of a large across-scan velocity). Apart from the fact that the velocity needs to be greater to properly distinguish moving from fixed objects, based on four observations rather than twelve, the relative performance of the three methods differ significantly between the two figures. The variance ratio method seems better adapted to coping with modest datasets.

Probability for a 'fixed' label, based on 12 observations
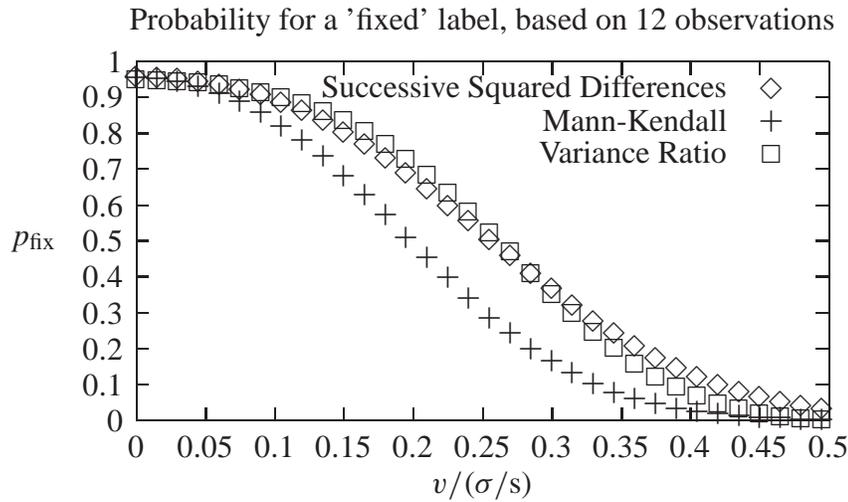


**Figure 5.3:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations and calibrated such that $p_I = 4.5\%$. The variance ratio appears slightly inferior to the other methods for $v < 0.3\sigma/s$. Exceeding this velocity threshold, it performs slightly better than the method of successive squared differences. For relatively large velocities, the performance of the three methods is very similar. From [Wolff 2005-III].

## Regression-based Test

A different approach to trend testing may be taken by performing a best fit, in a least-squares sense, of the observed data onto a straight line $x_i = a + bt_i$, and subsequently testing whether the linear coefficient (the slope) can be assumed to be zero. This assumes that the alternative to the null hypothesis is a linear trend, and not, e.g., a quadratic trend. In section 5.2 it is shown that this can be safely assumed when dealing with observations from one field transit.

The linear coefficient (and thus, the velocity) is estimated as follows:

$$\tilde{b} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(t_i - \bar{t})}{\sum_{i=1}^{N}(t_i - \bar{t})^2} \tag{5.4}$$

Using this, the test statistic may be written as (see [Kendall & Stuart 1961]):

$$\gamma_R(N) = \frac{\tilde{b}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2 - \tilde{b}^2\sum_{i=1}^{N}(t_i-\bar{t})^2}{(N-2)\sum_{i=1}^{N}(t_i-\bar{t})^2}}}$$

Under the null hypothesis, this can be shown to be distributed according to a Student's $t$-distribution with $N - 2$ degrees of freedom.

Probability for a 'fixed' label, based on 4 observations
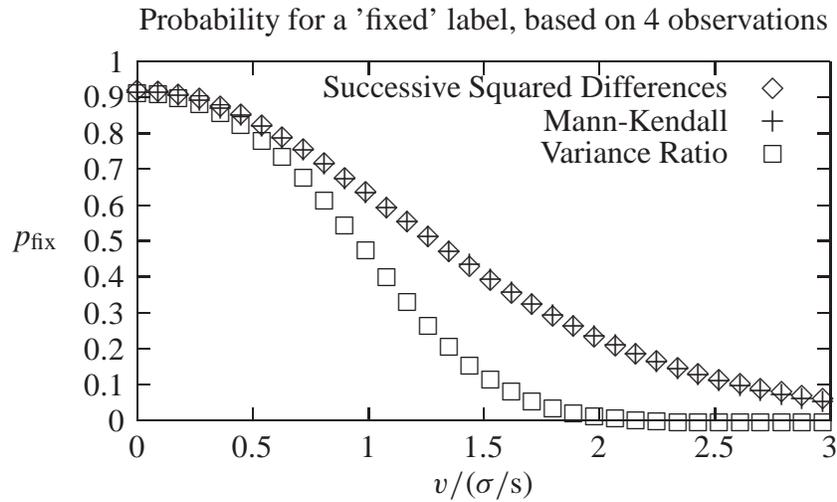


**Figure 5.4:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on four observations and calibrated such that $p_{\mathrm{I}} = 8.3\%$. As opposed to figure 5.3, the variance ratio method is clearly superior (closer to the ideal plot, as described in section 5.3), whereas the method of successive squared differences and the Mann-Kendall method are almost indistinguishable. In the case of the variance ratio method, $p_{\mathrm{II}}$ is less than 5% for $v > 1.75\sigma/\mathrm{s}$, whereas for the two other methods, the velocity must exceed $3\sigma/\mathrm{s}$. From [Wolff 2005-III].

Assuming that the observations may be described by a straight line (see section 5.2), the null hypothesis, i.e., the case of a zero slope, may be tested for by performing a two-sided test in a $t(N-2)$ distribution.

*Example* 4. Given the set of $N = 7$ observations $(t_i, x_i)$:

| i | $t_i$ | $x_i$ |
|---|-------|-------|
| 1 | $-3$  | $-2.6$ |
| 2 | $-2$  | $-2.2$ |
| 3 | $-1$  | $0.9$  |
| 4 | $0$   | $1.5$  |
| 5 | $1$   | $0.2$  |
| 6 | $2$   | $0.4$  |
| 7 | $3$   | $3.3$  |

we wish to determine whether or not this series of observations can be assumed to come from a noisy linear process with a non-zero slope. First, we determine $\bar{t}$, $\bar{x}$ and, for convenience, we find the value of the sums involved:

$$\bar{t} = 0$$

$$\bar{x} = \frac{1.5}{7} \approx 0.21$$

$$\sum_{i=1}^{N}(t_i - \bar{t})^2 = 28$$

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 \approx 25.43$$

$$\sum_{i=1}^{N}(t_i - \bar{t})(x_i - \bar{x}) = 22.2$$

The slope estimate is:

$$\tilde{b} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \approx \frac{22.2}{28} \approx 0.79$$

The test statistic is:

$$\gamma_R \approx \frac{0.79}{\sqrt{\frac{25.43 - 0.79^2 \cdot 28}{(7-5) \cdot 28}}} = 3.31$$

This should be compared with a $t(7-2)$ distribution at the required threshold $\alpha$. For $7 - 2 = 5$ degrees of freedom, a selection of confidence intervals for the Student's t-distribution is shown in the following table

| Level | Threshold |
|-------|-----------|
| 95.0% | 2.02 |
| 97.5% | 2.57 |
| 99.0% | 4.03 |

In this example, the null hypothesis will be rejected in the case of 95% and 97.5% confidence intervals, but accepted for a 99% confidence interval.

$\triangle$

The four methods thus far presented are compared in figure 5.5, showing $p_{\text{fix}}$ as a function of the velocity $v$. The regression-based method appears slightly better than the Mann-Kendall method, $p_{\text{II}}$ being less than 5% for $v > 0.34\sigma$, as opposed to $v > 0.36\sigma$ for the Mann-Kendall method. Comparing figure 5.5, based on twelve observations, to figure 5.6, based on only four, it is perhaps surprising that the regression-based method, despite being the only method that includes temporal information, fares as badly as the method of successive squared differences and the Mann-Kendall method for four observations.

Probability for a 'fixed' label, based on 12 observations
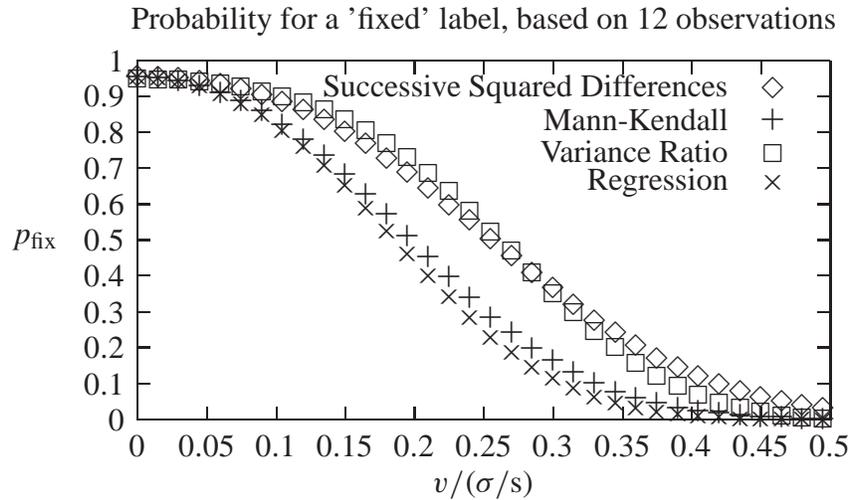


**Figure 5.5:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations, using transit times defined as $t_i = i$ (in seconds), and calibrated such that $p_{\text{I}} = 4.5\%$. The regression-based method appears slightly better than the Mann-Kendall method, $p_{\text{II}}$ being less than 5% for $v > 0.34\sigma$. For relatively large velocities, the performance of the four methods is very similar. From [Wolff 2005-III].

Based on figures 5.5 and 5.6, none of the four methods presented thus far is unequivocally the "best" method. It would seem that the best results would have to come from applying different methods depending on the number of observations. As a consequence of an attempt to avoid this relatively complex composite method, and to answer the question of whether there is a theoretical limit to the quality of a motion detection method, an optimal motion detection method was developed.

## 5.6   A New, Optimal, Motion Detection Method

This section presents a new and optimal motion detection method, developed following a methodology essentially different from the one previously employed, namely attempting to create an optimal method from the outset. This method, co-developed by C. Henrik-
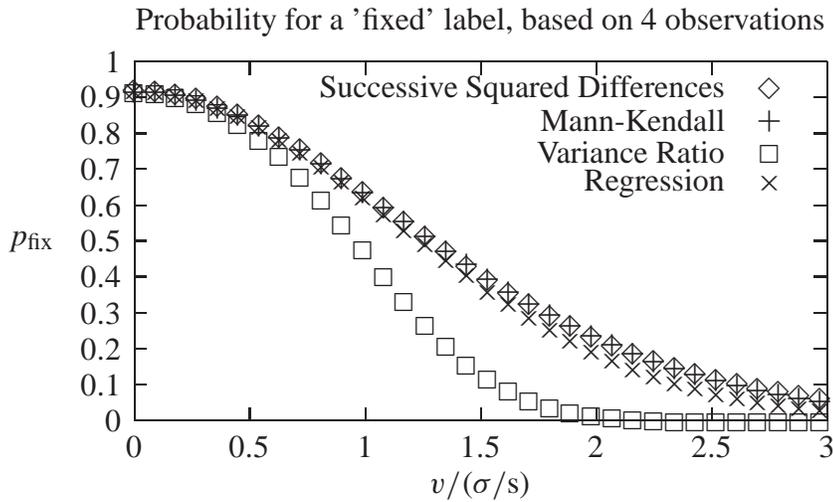
Probability for a 'fixed' label, based on 4 observations

**Figure 5.6:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on four observations, using transit times defined as $t_i = i$ (in seconds), and calibrated such that $p_I = 8.3\%$. As opposed to figure 5.5, the variance ratio method is clearly superior (closer to the ideal function, as described in section 5.3), whereas, perhaps surprisingly, the regression-based method fares almost as badly as the remaining two methods. From [Wolff 2005-III].

sen[4] and the author, is also described in [Wolff 2005-II] and, for the simplified case, in [Henriksen & Wolff 2005].

The following sections describe an optimal technique for categorising groups of observations in a way that is computationally inexpensive and easy to implement. Initially, we regard a simplified case having unit time steps and unit variance. This is subsequently generalised to encompass arbitrary observation times and individual residual variances in section 5.8.

## Introduction and results

This section presents an optimal method of distinguishing fixed and moving objects based on short-timebase astrometric observations. The method belongs to the class of *likelihood ratio* (LR) tests, using the ratio of two probabilities of the same event under different hypotheses as a test statistic. The proofs of the results can be found in the next section. A generalisation of the method can be found in section 5.8.

When we in the following say that some function is measurable, we understand with respect to the sigma algebra of Borel sets. When we talk about a probability distribution it is defined on this set. The symbol $\lambda$ denotes the Lebesgue measure (see, e.g., [Rudin 1988]).

---

[4]Department of Mathematics, Technical University of Denmark.

As in the previous sections, we restrict ourselves to the one-dimensional movement, i.e., movement on the real line. Suppose we have exactly two kinds of objects: fixed objects and uniformly moving objects, moving at constant speed $v$. It was shown in section 5.2, that the motion of celestial bodies, observed over short periods of time, may be regarded as linear.

Changing the indexing, we imagine that we have measured the position of an object at times $0, 1, \ldots, N$, thus obtaining a vector of $N+1$ measurements. Note, that this equidistant sampling means $t_i = i$. The $i$th measurement, denoted $X^i$, is the sum of the true position of the object and an error term $\epsilon_i$, according to (5.1). We assume that the error terms $\epsilon_i$ are independent stochastic variables, each normally distributed with mean 0 and variance 1, i.e., $\epsilon_i \in \mathcal{N}(0, 1)$. Since the variance of the error terms are typically well known (as a function of visual magnitude, see chapter 4), unit variance may be obtained by appriopriate scaling.

Denote by $\mathcal{X}_0 \in \mathbb{R}^{N+1}$ the stochastic variable that corresponds to the measured data of a fixed object. For $v \geq 0$ denote by $\mathcal{X}_v$ the stochastic variable that corresponds to data coming from an object moving with velocity $-v$ with probability $1/2$ and $v$ with probability $1/2$. Since we do not want to make any assumption on the initial position of an object, we introduce the stochastic variable $\mathcal{Y}_v \in \mathbb{R}^N$, describing relative positions: We write $\mathcal{X}_v = (X_v^0, X_v^1, \ldots, X_v^N)$ and set $\mathcal{Y}_v = (Y_v^1, \ldots, Y_v^N) = (X_v^1 - X_v^0, X_v^2 - X_v^0, \ldots, X_v^N - X_v^0)$. Then we can compute the density function associated with $\mathcal{Y}_v$.

**Proposition 1.** *Let* $n(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$ *denote the density function of the normal distribution with mean* $\mu = 0$ *and variance* $\sigma^2 = 1$. *The distribution of the stochastic variable* $\mathcal{Y}_0$ *is given by the density function*

$$f_0(\mathbf{y}) = \int_{\mathbb{R}} n(s)\, n(y_1 + s)\, n(y_2 + s) \cdots n(y_N + s)\, \mathrm{d}s,$$

*and the distribution of* $\mathcal{Y}_v$ *given by the density function*

$$f_v(\mathbf{y}) = \frac{1}{2} f_0(y_1 - v, y_2 - 2v, \ldots, y_N - Nv)$$
$$+ \frac{1}{2} f_0(y_1 + v, y_2 + 2v, \ldots, y_N + Nv)$$

*The likelihood ratio* $f_v/f_0$ *is given by*

$$\frac{f_v(\mathbf{y})}{f_0(\mathbf{y})} = \exp\left(-\frac{v^2}{24} N(N+1)(N+2)\right) \cosh\left(\frac{v}{2} \sum (2i - N)\, y_i\right).$$

The proof of this proposition is in the next section.

We now define the concept of a *test*. A test is a measurable subset $M$ of $\mathbb{R}^{N+1}$, where we label an object as moving if $\mathcal{X} \in M$ and as fixed if $\mathcal{X} \notin M$. We say that a test $M$ is *translation invariant* if $\mathbf{x} \in M \Rightarrow \mathbf{x} + (\tau, \tau, \ldots, \tau) \in M$, for all $\tau \in \mathbb{R}$.

In the following, we assume that the velocity $v$ is known *a priori*. Later we show that the derived test is independent of the distribution of $v$.

To a translation invariant test $M$ we associate a number $p_I$ and an error probability function $Q_v(M) = p_{II}(v)$. As before, the number $p_I$ is the probability that the test commits a type I error, and $Q_v(M) = p_{II}(v)$ is the probability that the test commits a type II error, for an object moving with speed $v$. More formally, we define

$$\tilde{M} = \{\mathbf{y} \in \mathbb{R}^N \mid (0, y_1, \ldots, y_N) \in M\},$$

$$p_I = \int_{\tilde{M}} f_0(\mathbf{y}) \, d\mathbf{y},$$

$$Q_v(M) = p_{II}(v) = 1 - \int_{\tilde{M}} f_v(\mathbf{y}) \, d\mathbf{y}.$$

Clearly the lower the number $p_I$ and the function values $p_{II}(v)$ the better the test. The main result is that there is an optimal test, which we now describe.

The following expression defines an estimator of the (linear) velocity, derived in the next section:

$$\tilde{v}(\mathbf{x}) = \frac{6}{N(N+1)(N+2)} \sum_{i=0}^{N} (2i - N)x_i. \tag{5.5}$$

This estimator is translation invariant in the sense that $\tilde{v}(\mathbf{x} + (\tau, \tau, \ldots, \tau)) = \tilde{v}(\mathbf{x})$, for all $\tau \in \mathbb{R}$.

Using the velocity estimate (5.5), we define the test $K_\alpha$ by

$$K_\alpha = \{\mathbf{x} \in \mathbb{R}^{N+1} \mid |\tilde{v}(\mathbf{x})| \geq \alpha\}. \tag{5.6}$$

This means that according to this test an object is labelled as moving if the absolute value of the estimated velocity exceeds some threshold value $\alpha$.

Recall that $p_I(K_\alpha)$ is the probability that the test $K_\alpha$ commits the mistake of labelling a fixed object as moving, which is equivalent to the probability of $|\tilde{v}|$ being greater than or equal to $\alpha$, given $v = 0$:

$$p_I(K_\alpha) = P(|\tilde{v}(\mathbf{x})| \geq \alpha \mid v = 0) \tag{5.7}$$

According to (5.5), $\tilde{v}(\mathbf{x})$ is a linear combination of stochastic variables, $\tilde{v}(\mathbf{x}) = \sum_{i=0}^{N} \gamma_i \, x_i$, where $x_i \in \mathcal{N}(P_0 + iv, 1)$. Therefore, the mean and variance of $\tilde{v}(\mathbf{x})$ are:

$$\mu_{\tilde{v}} = \sum_{i=0}^{N} \gamma_i \, \mu_{x_i} = v$$

$$\sigma_{\tilde{v}}^2 = \sum_{i=0}^{N} \gamma_i^2 \sigma_{x_i}^2 = \frac{12}{N(N+1)(N+2)}$$

Owing to the symmetry about $v = 0$, we may write that $P(|\tilde{v}(\mathbf{x})| \geq \alpha) = 2\,P(\tilde{v}(\mathbf{x}) \geq \alpha)$, and $p_{\mathrm{I}}(K_\alpha)$ may therefore also be written as:

$$p_{\mathrm{I}}(K_\alpha) = 2P\left(\frac{\tilde{v}(\mathbf{x})}{\sigma_{\tilde{v}}} \geq \frac{\alpha}{\sigma_{\tilde{v}}}\right) = 2P\left(z \geq \frac{\alpha}{\sigma_{\tilde{v}}}\right) = 2P(z \geq u(\alpha)),$$

where $z$ is a stochastic variable, distributed according to the standardised normal distribution: $z \in \mathcal{N}(0, 1)$.

Suppose we have a maximum acceptable value $\delta$ of $p_{\mathrm{I}}(K_\alpha)$. A corresponding value for $\alpha$ can now be found by solving for $u$

$$\delta = p_{\mathrm{I}}(K_\alpha) = 2P(z \geq u)$$

and finding $\alpha$ as follows:

$$\alpha = u\,\sigma_{\tilde{v}} = u\sqrt{\frac{12}{N(N+1)(N+2)}}.$$

*Example* 5. As an example, suppose we have measured 4 positions, implying $N = 3$: $x_0 = 13.9$, $x_1 = 17.2$, $x_2 = 15.3$, $x_3 = 15.4$. The velocity estimate is:

$$\tilde{v}(\mathbf{x}) = \frac{1}{10}(-3x_0 - 1x_1 + 1x_2 + 3x_3) = 0.26$$

Supposing we accept 10% type I errors ($\delta = 0.1$), a statistical table yields the value $u \approx 1.96$, leading to threshold of $\alpha \approx 0.88$. Since $|\tilde{v}(\mathbf{x})| \leq \alpha$, this set of observations is flagged as belonging to a fixed object.

$\triangle$

As the example shows, the test (5.6) is easy to implement and computationally inexpensive. It is also optimal in the following strong sense.

**Theorem 1.** *Suppose $0 < \delta < 1$. There exists an $\alpha$ such that $p_{\mathrm{I}}(K_\alpha) = \delta$ and for any symmetric and translation invariant test $M$ with $p_{\mathrm{I}}(M) \leq \delta$ we have $Q_v(K_\alpha) \leq Q_v(M)$, for any velocity $v$.*

Notice that $\alpha$ does not depend on $v$. We say that a measure $\eta$ on $\mathbb{R}$ is *symmetric* if for any interval $I \subset \mathbb{R}$ we have $\eta(I) = \eta(-I)$. We have the following corollary.

**Corollary 1.** *Fix $\alpha$. For any symmetric distribution of object velocities, and any translation invariant test $M$ committing type I errors with probability $p_{\mathrm{I}}(M) \leq p_{\mathrm{I}}(K_\alpha)$, the test $M$ commits type II errors with a probability equal to or greater than that of the test $K_\alpha$.*

Loosely speaking, (5.7) and the theorem says that for any given probability of type I errors we choose to accept, there is a test $K_\alpha$ that commits this amount of type I errors, and any other test that performs as well in this regard will commit at least as many type II errors.

Notice that we can say that the test is optimal without any *a priori* knowledge of the velocity distribution except that it be symmetric. This is significant, since in many applications one typically does not know much about the velocity distribution, except that it is symmetric.

## 5.7 Proofs of the results

First, we prove Proposition 1, then we show how the main theorem implies the corollary. The proofs of the main theorem and the lemma then follow in that order.

We start by proving Proposition 1, beginning with the derivation of $f_0(\mathbf{y})$:

When $v = 0$, the residuals may be expressed as $\epsilon_i = x_i - p$, according to (5.1). Since the residuals are independent and distributed according to $\mathcal{N}(0, 1)$, we may express $f_0(x_0, x_1, \ldots, x_N)$ as (see [Kendall & Stuart 1961]):

$$f(x_0, x_1, \ldots, x_N) = n(x_0 - p)n(x_1 - p) \cdots n(x_N - p)$$

We now perform the following mapping:

$$
\begin{bmatrix} x_0 \\ y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}
=
\begin{bmatrix} x_0 \\ x_1 - x_0 \\ x_2 - x_0 \\ \vdots \\ x_N - x_0 \end{bmatrix}
= \mathbf{A}
\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}
$$

The transformation matrix $\mathbf{A}$ is easily seen to have a unit determinant, and hence:

$$f_0(x_0, y_1, y_2, \ldots, y_N) = n(x_0 - p)n(y_1 + x_0)n(y_2 + x_0) \cdots n(y_N + x_0)$$

The marginal probability density function $f_0(\mathbf{y}) = f_0(y_1, y_2, \ldots, y_N)$ can be found by "integrating out" $x_0$ (see [Kendall & Stuart 1961]):

$$f_0(\mathbf{y}) = \int_{\mathbb{R}} n(s-p)n(y_1+s)n(y_2+s) \cdots n(y_N+s)\mathrm{d}s = \int_{\mathbb{R}} n(s)n(y_1+s) \cdots n(y_N+s)\mathrm{d}s \tag{5.8}$$

We continue with deriving the likelihood ratio $f_v/f_0$. Define

$$
\begin{aligned}
h(\mathbf{y}, v) &= f_0(y_1 + v, y_2 + 2v, \ldots, y_N + Nv) \\
&= \int_{\mathbb{R}} n(s)n(y_1 + v + s)\ldots n(y_N + Nv + s) \, ds \\
&= \int_{\mathbb{R}} k \exp(-as^2 - bs - c) \, ds
\end{aligned}
$$

where

$$
\begin{aligned}
k &= \left(\frac{1}{\sqrt{2\pi}}\right)^{N+1} \\
a &= \frac{N+1}{2} \\
b(v) &= \sum_{i=1}^{N}(y_i + iv) \\
c(v) &= \frac{1}{2}\sum_{i=1}^{N}(y_i + iv)^2
\end{aligned}
$$

Performing the integration yields

$$
h(\mathbf{y}, v) = k\sqrt{\frac{\pi}{a}} \exp\left(d(v)\right),
$$

where

$$
d(v) = \frac{b(v)^2}{4a} - c(v).
$$

Since $f_0(\mathbf{y}) = h(\mathbf{y}, 0)$ and $f_v(\mathbf{y}) = \frac{1}{2}\left(h(\mathbf{y}, v) + h(\mathbf{y}, -v)\right)$ because of the symmetric velocity distribution, the ratio $f_v/f_0$ is equal to:

$$
\begin{aligned}
\frac{f_v(\mathbf{y})}{f_0(\mathbf{y})} &= \frac{1}{2}\frac{h(\mathbf{y}, v) + h(\mathbf{y}, -v)}{h(\mathbf{y}, 0)} \\
&= \frac{1}{2}\frac{\exp\left(d(v)\right) + \exp\left(d(-v)\right)}{\exp\left(d(0)\right)}
\end{aligned}
$$

Observe, that $d(v)$ may be written as a quadratic polynomial in $v$: $d(v) = lv^2 + mv + n$, where

$$l = \frac{\left(\sum_{i=1}^{N} i\right)^2}{4a} - \frac{\sum_{i=1}^{N} i^2}{2} = -\frac{1}{24} N(N+1)(N+2),$$

$$m = \frac{\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} i\right)}{2a} - \sum_{i=1}^{N} i y_i = -\frac{1}{2} \sum_{i=1}^{N} (2i - N) y_i,$$

$$n = \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{4a} - \frac{\sum_{i=1}^{N} y_i^2}{2}.$$

This means the ratio $f_v/f_0$ may be written as

$$
\begin{aligned}
\frac{f_v(\mathbf{y})}{f_0(\mathbf{y})} &= \frac{1}{2} \frac{\exp(lv^2 + mv + n) + \exp(lv^2 - mv + n)}{\exp(n)} \\
&= \exp(lv^2) \cosh(-mv) \\
&= \exp\left(-\frac{v^2}{24} N(N+1)(N+2)\right) \cosh\left(\frac{v}{2} \sum_{i=1}^{N} (2i - N) y_i\right).
\end{aligned}
$$

$\square$

A large numerical value of the ratio $f_v/f_0$ indicates that the object is likely to be moving. Conversely, a small absolute value indicates that the object is likely to be fixed. Since $v$ and $N$ are known constants and $\cosh(x)$ increases as $|x|$ increases, the magnitude of the ratio $f_v/f_0$ depends monotonically on the magnitude of $\sum_{i=1}^{N} (2i - N) y_i$. Expressing this using $x_i$ rather than $y_i$:

$$\sum_{i=1}^{N}(2i-N)y_i = \sum_{i=1}^{N}(2i-N)(x_i-x_0) = \sum_{i=0}^{N}(2i-N)x_i - x_0 \sum_{i=0}^{N}(2i-N) = \sum_{i=0}^{N}(2i-N)x_i ,$$

since $\sum_{i=0}^{N}(2i-N) = 2\sum_{i=0}^{N} i - N(N+1) = 0$. Using $t_i = i$ implied by the equidistant sampling, the expected value of this sum is:

$$\mathrm{E}\left(\sum_{i=0}^{N}(2i-N)\,x_i\right) = \mathrm{E}\left(\sum_{i=0}^{N}(2i-N)\,(P_0+vt_i+\epsilon_i)\right) \qquad (5.9)$$

$$= \mathrm{E}\left(v\sum_{i=0}^{N}(2i-N)\,i\right)+\mathrm{E}\left(P_0\sum_{i=0}^{N}(2i-N)\right) \quad (5.10)$$

$$+\mathrm{E}\left(\sum_{i=0}^{N}(2i-N)\,\epsilon_i\right) \qquad (5.11)$$

$$= v\frac{N(N+1)(N+2)}{6}+0+0 \qquad (5.12)$$

This shows, that if the sum $\sum_{i=0}^{N}(2i-N)x_i$ is scaled by $6/\left(N\,(N+1)\,(N+2)\right)$, we have an unbiased estimator of $v$. Its translation invariance is seen from $\sum_{i=0}^{N}(2i-N)=0$, as shown above. We have now derived the velocity estimate (5.5):

$$\tilde{v}=\frac{6}{N(N+1)(N+2)}\sum_{i=0}^{N}(2i-N)x_i$$

This scaling does not affect the test in any way. It does, however, furnish a physical interpretation of the test: For large numerical values of the velocity estimator, it is likely that the object is moving, and vice versa. This could hardly be more intuitive. However, it is important to stress, that this intuition is not the origin of the test. The origin is the ratio of the probability density functions $f_v$ to $f_0$, the likelihood radio.

Using Lagrange multipliers, it is straight-forward to show that (5.5) has minimum variance among all translation invariant, linear, unbiased velocity estimators. The general formulation for a linear velocity estimator is:

$$\tilde{v}=\sum_{i=0}^{N}\alpha_i x_i \qquad (5.13)$$

Because $\sigma^2=1$, the variance is

$$\mathrm{Var}\,(\tilde{v})=\sum_{i=0}^{N}\alpha_i^2\,. \qquad (5.14)$$

To ensure that (5.13) is translation invariant, we impose:

$$\sum_{i=0}^{N}\alpha_i=0 \qquad (5.15)$$

Furthermore, we impose the following to ensure an unbiased estimator:

$$\sum_{i=0}^{N} i\alpha_i = 1 \tag{5.16}$$

If a function $f : A \to \mathbb{R}$ has an extremum at $\mathbf{u} \in A^\circ$ and $f$ is differentiable in $\mathbf{u}$, then $\nabla f(\mathbf{u}) = \mathbf{0}$. Introducing the two constraints (5.15) and (5.16) using the Lagrangian multipliers $\lambda_1$ and $\lambda_2$, respectively, we can minimise (5.14) by solving:

$$0 = \frac{\partial}{\partial \alpha_i} \left( \text{Var}(\tilde{v}) + \lambda_1 \sum_{i=0}^{N} \alpha_i + \lambda_2 \sum_{i=0}^{N} i\alpha_i \right)$$

$$\Updownarrow$$

$$0 = 2\alpha_i + \lambda_1 + i\lambda_2 \tag{5.17}$$

for $i = 0, 1, \ldots, N$. Adding (5.17) for $i = 0, 1, \ldots, N$, and using (5.15) yields:

$$0 = 2\sum_{i=0}^{N} \alpha_i + \sum_{i=0}^{N} \lambda_1 + \sum_{i=0}^{N} i\lambda_2$$

$$\Updownarrow$$

$$0 = 0 + (N+1)\lambda_1 + N\frac{N+1}{2}\lambda_2 \tag{5.18}$$

By multiplying (5.17) by $i$ before adding for $i = 0, 1, \ldots, N$ yields an expression we can reduce using (5.16):

$$0 = 2\sum_{i=0}^{N} i\alpha_i + \sum_{i=0}^{N} i\lambda_1 + \sum_{i=0}^{N} i^2\lambda_2$$

$$\Updownarrow$$

$$0 = 2 + N\frac{N+1}{2}\lambda_1 + N\frac{(N+1)(2N+1)}{6}\lambda_2 \tag{5.19}$$

By simultaneously solving (5.18) and (5.19) for $(\lambda_1, \lambda_2)$, we get:

$$\begin{cases} \lambda_1 = \frac{12}{(N+1)(N+2)} \\ \lambda_2 = -\frac{24}{N(N+1)(N+2)} \end{cases}$$

Inserting these values for $\lambda_1$ and $\lambda_2$ in (5.17) yields:

$$\alpha_i = -\frac{1}{2}(\lambda_1 + i\lambda_2) = \frac{6}{N(N+1)(N+2)}(2i-N)$$

Subsequently inserting these values for $\alpha_i$ in (5.13) yields the velocity estimator in (5.5), proving that, among all linear, translation invariant, unbiased velocity estimators, (5.5) has minimum variance.

Before proving the main theorem, we show how it implies the corollary. Let $M$ be an arbitrary translation invariant test committing type I errors with probability $p_I(M) \leq p_I(K_\alpha)$. We must show that the test $M$ commits type II errors with a probability equal to or greater than that of the test $K_\alpha$. Define $\tilde{M} = \{\mathbf{y} \in \mathbb{R}^N \mid (0, y_1, \ldots, y_N) \in M\}$. Since $M$ is translation invariant we have $\mathbf{x} \in M \Leftrightarrow \mathbf{y} \in \tilde{M}$ when $\mathbf{y} = (x_1 - x_0, \ldots, x_N - x_0)$. Define in a similar way the set $\tilde{K}_\alpha$ from $K_\alpha$. Let $\eta$ be the symmetric probability measure on $\mathbb{R}$ given the distribution of velocities among moving objects. Define a probability measure $\mu$ on the non-negative reals by letting $\mu(U) = \eta(U \cup -U)$, for each measurable subset $U \subset [0, \infty)$. If $\mathcal{X}$ denotes the measured positions of a randomly chosen moving object then $\mathcal{Y} = (X^1 - X^0, \ldots, X^N - X_0)$ has the density function $g(\mathbf{y}) = \int_0^\infty f_v(\mathbf{y})\, d\mu(v)$.

We now introduce $Q(M)$ and $Q(K_\alpha)$, the probabilities that the tests $M$ and $K_\alpha$, respectively, commit type II errors. These probabilities depend on the velocity distribution rather than on the *a priori* knowledge of the object velocity, which was the case with $Q_v(M)$.

To prove the corollary, we need to prove that $Q(M) \geq Q(K_\alpha)$:

$$
\begin{aligned}
Q(M) &= 1 - \int_{\tilde{M}} \int_0^\infty f_v(\mathbf{y})\, d\mu(v)\, d\lambda(\mathbf{y}) \\
&= 1 - \int_0^\infty \int_{\tilde{M}} f_v(\mathbf{y})\, d\lambda(\mathbf{y})\, d\mu(v) \\
&= 1 - \int_0^\infty 1 - Q_v(M)\, d\mu(v).
\end{aligned}
$$

According to the theorem, $Q_v(K_\alpha) \leq Q_v(M)$, so

$$
\begin{aligned}
Q(M) &= 1 - \int_0^\infty 1 - Q_v(M)\, d\mu(v) \\
&\geq 1 - \int_0^\infty 1 - Q_v(K_\alpha)\, d\mu(v) \\
&= 1 - \int_0^\infty \int_{\tilde{K}_\alpha} f_v(\mathbf{y})\, d\lambda(\mathbf{y})\, d\mu(v) \\
&= 1 - \int_{\tilde{K}_\alpha} \int_0^\infty f_v(\mathbf{y})\, d\mu(v)\, d\lambda(\mathbf{y}) \\
&= Q(K_\alpha).
\end{aligned}
$$

That we can exchange the order of integration follows from Tonelli's Theorem. We have finished the proof of the corollary.

□

The proof of the theorem hinges on the following lemma:

**Lemma 1.** *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be non-negative representatives of $L^1$ functions such that $\int f \, d\lambda = \int g \, d\lambda = 1$. Define $A_\beta = \{\mathbf{y} \mid f(\mathbf{y}) \geq \beta g(\mathbf{y})\}$, for $\beta \geq 0$. Then the set $A_\beta$ has the following extremal property: For any measurable subset $B \subset \mathbb{R}^N$ with $\int_B g \, d\lambda \leq \int_{A_\beta} g \, d\lambda$ we have $\int_B f \, d\lambda \leq \int_{A_\beta} f \, d\lambda$.*

Equipped with this lemma, the proof of the theorem is straight-forward. Let $\delta$ be given and let $\alpha$ be chosen such that $p_I(K_\alpha) = \delta$. Such an $\alpha$ exists due to continuity. Let $M$ be an arbitrary translation invariant test with a probability of type I errors $p_I(M)$ less than or equal to $p_I(K_\alpha)$, the probability of type I errors of $K_\alpha$ for some fixed $\alpha$. We must show that for $M$ the probability of type II errors is greater than or equal to the corresponding probability for $K_\alpha$. We have $p_I(M) = \int_{\tilde{M}} f_0 \, d\lambda$, and $p_I(K_\alpha) = \int_{\tilde{K}_\alpha} f_0 \, d\lambda$, so by assumption

$$\int_{\tilde{M}} f_0 \, d\lambda \leq \int_{\tilde{K}_\alpha} f_0 \, d\lambda. \tag{5.20}$$

By Proposition 1

$$\frac{f_v(\mathbf{y})}{f_0(\mathbf{y})} = \exp\left(-\frac{v^2}{24} N(N+1)(N+2)\right) \cosh\left(\frac{v}{2} \sum (2i - N) \, y_i\right)$$

Now, $\sum_{i=1}^N (2i - N) y_i = \sum_{i=0}^N (2i - N) x_i$, and it follows that the set $\tilde{K}_\alpha$ is equal to $A_\beta = \{\mathbf{y} \mid f_v(\mathbf{y}) \geq \beta f_0(\mathbf{y})\}$, for some $\beta = \beta(\alpha)$. By the lemma and (5.20) we have that

$$\int_{\tilde{M}} f_v \, d\lambda \leq \int_{A_\beta} f_v \, d\lambda = \int_{\tilde{K}_\alpha} f_v \, d\lambda$$

which concludes the proof of the theorem.

□

All that remains is to prove the lemma. First note that

$$\int_{A_\beta \cap B} g \, d\lambda + \int_{B \setminus A_\beta} g \, d\lambda = \int_B g \, d\lambda \leq$$

$$\int_{A_\beta} g \, d\lambda = \int_{A_\beta \cap B} g \, d\lambda + \int_{A_\beta \setminus B} g \, d\lambda.$$

It follows that

$$\int_{B \setminus A_\beta} g \, d\lambda \leq \int_{A_\beta \setminus B} g \, d\lambda.$$

We compute

$$\int_{A_\beta \setminus B} f \, d\lambda \geq \int_{A_\beta \setminus B} \beta g \, d\lambda \geq \int_{B \setminus A_\beta} \beta g \, d\lambda \geq \int_{B \setminus A_\beta} f \, d\lambda$$

Adding $\int_{A_\beta \cap B} f \, d\lambda$ on both side of the inequality we see that $\int_{A_\beta} f \, d\lambda \geq \int_B f \, d\lambda$, which ends the proof of the lemma.

$\square$

Figures 5.7 and 5.8 show the relative performance of each of the five tests presented in this chapter. The optimal method performs better than the other tests: It is closer to the ideal curve, as described in section 5.3. This was to be expected, because it is optimal among all symmetric and translation invariant methods according to corollary 1, proved above.

Thus far, we have assumed transit times (in seconds) described by $t_i = i$. For the fifth, optimal, method, we have assumed a constant position standard deviation of unity, $\sigma_i = \sigma = 1$. The following section will generalise the optimal method to enable arbitrary transit times and position standard deviation.
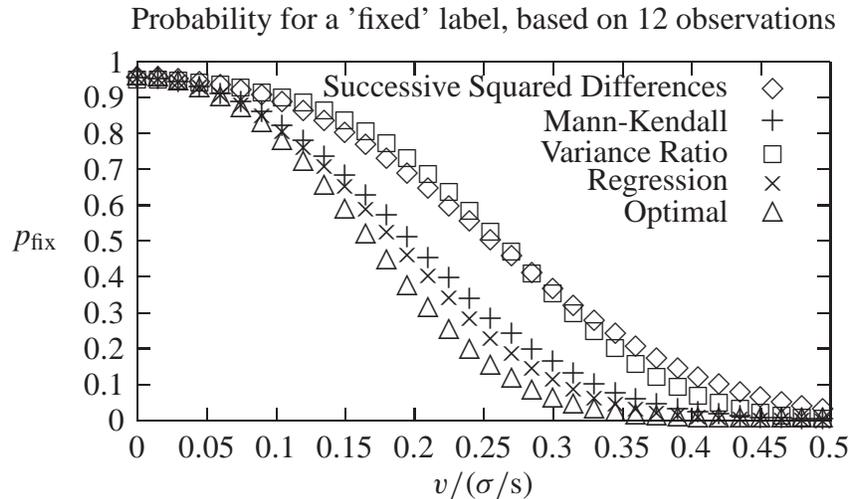


**Figure 5.7:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations, using transit times defined as $t_i = i$ (in seconds), and calibrated such that $p_I = 4.5\%$. The optimal method appears slightly better (closer to the ideal function, as described in section 5.3) than the regression-based method, $p_{II}$ being less than 5% for $v > 0.30\sigma/s$ in the case of the former, as opposed the latter, requiring $v > 0.34\sigma/s$ for this to be true. See figure 5.10 for a similar example using Gaia transit times. From [Wolff 2005-II].

Probability for a 'fixed' label, based on 4 observations
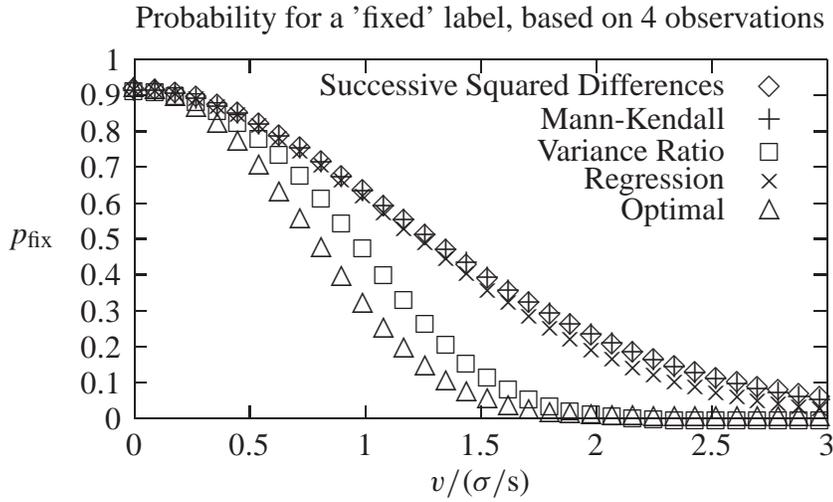


**Figure 5.8:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on four observations, using transit times defined as $t_i = i$ (in seconds), and calibrated such that $p_I = 8.3\%$. As expected, the optimal method is superior (closer to the ideal function, as described in section 5.3) to even the variance ratio method. In the case of the former, $p_{II}$ is less than 5% for $v > 1.5\sigma/s$, whereas the latter requires $v > 1.75\sigma/s$ for this to be true. See figure 5.11 for a similar plot using Gaia transit times. From [Wolff 2005-II].

## 5.8 Generalising the results

In this section, we will be generalising the one-dimensional simplified case of $t_i = i$ and $\sigma_i^2 = 1$ to arbitrary sampling times and variances. The derivation will be analogous to the derivation of the simplified case, leading to the recycling of several variables to facilitate the transition. As an implication of this analogy, the proofs of optimality and independence of velocity distribution shown in the previous section also hold for the generalised case.

As in the previous section, the speed is initially assumed to be known, and any direction of motion, $+|v|$ or $-|v|$ is assumed to be equally likely. The probability density function of a random variable $X \in \mathcal{N}(\mu, \sigma^2)$ is

$$f(x) = \frac{1}{\sigma} n \left( \frac{x - \mu}{\sigma} \right)$$

where $n(x)$ is the probability density function of a random variable distributed according to $\mathcal{N}(0, 1)$. Rewriting (5.1) as $\epsilon_i = x_i - vt_i - P_0$, the marginal probability distribution, given the velocity $v$, is (cf. (5.8)):

$$h(\mathbf{y} - v\mathbf{t}) = \prod_{i=0}^{N}\left(\frac{1}{\sigma_i}\right)\int_{\mathbb{R}} n\left(\frac{s}{\sigma_0}\right) n\left(\frac{y_1 - vt_1}{\sigma_1}\right) n\left(\frac{y_2 - vt_2}{\sigma_2}\right)\cdots n\left(\frac{y_N - vt_N}{\sigma_N}\right) ds$$

$$= \prod_{i=0}^{N}\left(\frac{1}{\sigma_i}\right)\int_{\mathbb{R}}\left(\frac{1}{\sqrt{2\pi}}\right)^{N+1}\exp\left(-as^2 - bs - c\right)ds \qquad (5.21)$$

where

$$a = \frac{1}{2}\sum_{i=0}^{N}\frac{1}{\sigma_i^2} \qquad (5.22)$$

$$b = \sum_{i=1}^{N}\frac{y_i - vt_i}{\sigma_i^2} \qquad (5.23)$$

$$c = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - vt_i}{\sigma_i}\right)^2 \qquad (5.24)$$

Performing the integration in (5.21) yields:

$$h(\mathbf{y} - v\mathbf{t}) = \prod_{i=0}^{N}\left(\frac{1}{\sigma_i}\right)\left(\frac{1}{\sqrt{2\pi}}\right)^{N+1}\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a} - c\right) \qquad (5.25)$$

where

$$\frac{b^2}{4a} - c = \frac{\left(\sum_{i=1}^{N}\frac{y_i - vt_i}{\sigma_i^2}\right)^2}{2\sum_{i=0}^{N}\sigma_i^{-2}} - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - vt_i}{\sigma_i}\right)^2$$

$$= \frac{\left(\sum_{i=1}^{N}\frac{y_i}{\sigma_i^2}\right)^2 + v^2\left(\sum_{i=1}^{N}\frac{t_i}{\sigma_i^2}\right)^2 - 2v\left(\sum_{i=1}^{N}\frac{y_i}{\sigma_i^2}\right)\left(\sum_{i=1}^{N}\frac{t_i}{\sigma_i^2}\right)}{2\sum_{i=0}^{N}\sigma_i^{-2}}$$

$$-\frac{1}{2}\sum_{i=1}^{N}\frac{y_i^2}{\sigma_i^2} - \frac{v^2}{2}\sum_{i=1}^{N}\frac{t_i^2}{\sigma_i^2} + v\sum_{i=1}^{N}\frac{y_i t_i}{\sigma_i^2}$$

$$= lv^2 + mv + n \qquad (5.26)$$

We introduce the following short-hand notation:

$$\Psi_1 = \sum_{i=0}^{N} \frac{1}{\sigma_i^2} \quad \Psi_t = \sum_{i=1}^{N} \frac{t_i}{\sigma_i^2} \quad \Psi_y = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}$$

$$\Psi_{yt} = \sum_{i=1}^{N} \frac{y_i t_i}{\sigma_i^2} \quad \Psi_{t^2} = \sum_{i=1}^{N} \frac{t_i^2}{\sigma_i^2} \quad \Psi_{y^2} = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2}$$

Note, that $\Psi_{y^2}$ in general is different from $\Psi_y^2 = \Psi_y \Psi_y$. We may now write $l$, $m$ and $n$ from (5.26) as:

$$l = \frac{\Psi_t^2}{2\Psi_1} - \frac{1}{2}\Psi_{t^2}$$

$$m = \Psi_{yt} - \frac{\Psi_y \Psi_t}{\Psi_1}$$

$$n = \frac{\Psi_y^2}{2\Psi_1} - \frac{1}{2}\Psi_{y^2}$$

We are now ready to express the ratio of the probability density functions $f_v/f_0$:

$$\begin{aligned}
\frac{f_v}{f_0} &= \frac{\frac{1}{2}h(\mathbf{y} - v\mathbf{t}) + \frac{1}{2}h(\mathbf{y} + v\mathbf{t})}{h(\mathbf{y})} \\
&= \frac{1}{2}\frac{\exp(lv^2 + mv + n) + \exp(lv^2 - mv + n)}{\exp(n)} \\
&= \exp(lv^2)\cosh(mv)
\end{aligned} \quad (5.27)$$

As seen before, the magnitude of the probability density ratio depends on the magnitude of $m$. We write $m$ as a linear combination of the elements of $\mathbf{y}$:

$$\begin{aligned}
m &= \Psi_{yt} - \frac{\Psi_y \Psi_t}{\Psi_1} \\
&= \sum_{i=1}^{N} \frac{y_i t_i}{\sigma_i^2} - \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}\frac{\Psi_t}{\Psi_1} \\
&= \sum_{i=1}^{N} y_i \left( \frac{t_i}{\sigma_i^2} - \frac{\Psi_t}{\sigma_i^2 \Psi_1} \right)
\end{aligned}$$

Since $y_i = x_i - x_0 = vt_i + \epsilon_1 - \epsilon_0$, the expected value of $m$ is:

$$
\begin{aligned}
\mathrm{E}(m) &= \mathrm{E}\left(\sum_{i=1}^{N} y_i \left(\frac{t_i}{\sigma_i^2} - \frac{\Psi_t}{\sigma_i^2 \Psi_1}\right)\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{N} (vt_i + \epsilon_1 - \epsilon_0)\left(\frac{t_i}{\sigma_i^2} - \frac{\Psi_t}{\sigma_i^2 \Psi_1}\right)\right) \\
&= v\mathrm{E}\left(\sum_{i=1}^{N} t_i \left(\frac{t_i}{\sigma_i^2} - \frac{\Psi_t}{\sigma_i^2 \Psi_1}\right)\right) \\
&= v\left(\Psi_{t^2} - \frac{\Psi_t^2}{\Psi_1}\right)
\end{aligned}
$$

Thus, by scaling appropriately, we have a velocity estimate:

$$
\tilde{v} = \frac{m}{\Psi_{t^2} - \frac{\Psi_t^2}{\Psi_1}} = \frac{\Psi_1 \Psi_{yt} - \Psi_y \Psi_t}{\Psi_1 \Psi_{t^2} - \Psi_t^2} = \frac{1}{\Psi_1 \Psi_{t^2} - \Psi_t^2} \sum_{i=0}^{N} x_i \frac{\Psi_1 t_i - \Psi_t}{\sigma_i^2} \tag{5.28}
$$

As before, this scaling is merely done to assist in the understanding of the test.

To show that this velocity estimate is the most efficient (has the least variance) among all translation invariant, unbiased estimators, we will compute the most efficient velocity estimate and show that it equals (5.28), in analogy with the simplified case.

The general form of a linear velocity estimate is:

$$
\tilde{v} = \sum_{i=0}^{N} \alpha_i x_i = \sum_{i=0}^{N} \alpha_i (vt_i + p + \epsilon_i) \tag{5.29}
$$

The variance of which is:

$$
\mathrm{Var}(\tilde{v}) = \sum_{i=0}^{N} \alpha_i^2 \sigma_i^2
$$

The expected value is:

$$
\mathrm{E}(\tilde{v}) = \mathrm{E}\left(\sum_{i=0}^{N} \alpha_i (vt_i + p + \epsilon_i)\right) = v\sum_{i=0}^{N} \alpha_i t_i + p\sum_{i=0}^{N} \alpha_i
$$

To make sure the estimator is translation invariant, the second term on the right hand side must be zero for any $p$, so:

$$\sum_{i=0}^{N} \alpha_i = 0$$

An unbiased estimator is defined as having an expected value equal to the value it estimates, so $E(\tilde{v}) = v$. We ensure this property by imposing the following constraint:

$$\sum_{i=0}^{N} \alpha_i t_i = 1$$

As before, we introduce two Lagrange multipliers $\lambda_1$ and $\lambda_2$. The following must now hold for $i = 0, 1, 2, \ldots, N$:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \alpha_i} \left( \text{Var}(\tilde{v}) + \lambda_1 \sum_{i=0}^{N} \alpha_i + \lambda_2 \sum_{i=0}^{N} \alpha_i t_i \right) \\
&\Updownarrow \\
0 &= 2\alpha_i \sigma_i^2 + \lambda_1 + \lambda_2 t_i \\
&\Updownarrow \\
0 &= 2\alpha_i + \lambda_1 \frac{1}{\sigma_i^2} + \lambda_2 \frac{t_i}{\sigma_i^2}
\end{aligned}
\tag{5.30}
$$

Summing (5.30) over $i = 0, 1, 2, \ldots, N$ yields:

$$0 = 2 \sum_{i=0}^{N} \alpha_i + \lambda_1 \sum_{i=0}^{N} \frac{1}{\sigma_i^2} + \lambda_2 \sum_{i=0}^{N} \frac{t_i}{\sigma_i^2} = \lambda_1 \sum_{i=0}^{N} \frac{1}{\sigma_i^2} + \lambda_2 \sum_{i=0}^{N} \frac{t_i}{\sigma_i^2} \tag{5.31}$$

Multiplying (5.30) by $t_i$ before summing yields:

$$0 = 2 \sum_{i=0}^{N} \alpha_i t_i + \lambda_1 \sum_{i=0}^{N} \frac{t_i}{\sigma_i^2} + \lambda_2 \sum_{i=0}^{N} \frac{t_i^2}{\sigma_i^2} = 2 + \lambda_1 \sum_{i=0}^{N} \frac{t_i}{\sigma_i^2} + \lambda_2 \sum_{i=0}^{N} \frac{t_i^2}{\sigma_i^2} \tag{5.32}$$

By simultaneously solving (5.31) and (5.32) for $(\lambda_1, \lambda_2)$, we get:

$$
\begin{cases}
\lambda_1 &= \frac{2\Psi_t}{\Psi_1 \Psi_{t^2} - \Psi_t^2} \\
\lambda_2 &= -\frac{2\Psi_1}{\Psi_1 \Psi_{t^2} - \Psi_t^2}
\end{cases}
$$

Inserting these in (5.30) and solving for $\alpha_i$ yields:

$$\alpha_i = \frac{\Psi_1 t_i - \Psi_t}{\sigma_i^2 \left( \Psi_1 \Psi_{t^2} - \Psi_t^2 \right)}$$

When inserting this expression for $\alpha_i$ in (5.29), we get the same velocity estimator as in (5.28). This shows that (5.28) is the velocity estimate having minimum variance among all linear, translation invariant, unbiased velocity estimates.

A least-squares estimator, such as the velocity estimator in the regression-based approach (5.4), also has minimum variance [Kendall & Stuart 1961]. It is straight-forward to show that (5.29) reduces to (5.4) for constant variance $\sigma_i^2 = \sigma^2$.

Inserting $t_i = i$ and $\sigma_i = 1$ in the general formulation, one obtains the simplified test based on the velocity estimate in (5.5).

## 5.9   Discussion

Figures 5.10 and 5.11 show the performance of each of the five tests for transit times corresponding to the Gaia preceding field of view, see table 4.1. Because the new method is optimal, not just among linear methods, but among all symmetric and translation invariant methods, it is expected to perform better than the four other methods. In both the case of four and the case of twelve observations, this is the case. The improvement, compared with the other methods, is not dramatic, thus attesting to the quality of the other methods. However, knowing that the method is optimal provides the advantage of needing only one method, instead of using one method or another, depending on the number of observations.

The optimal method is general in the sense that it is not limited to treating data from the Astro telescopes, or even to Gaia data. It can be used for the task of reducing the number of observations before attempting to link them when computing orbits. It can also be used as a NEO detection strategy in the method for NEO observation using the Spectro instrument, described in section 4.7.

Unlike the Mann-Kendall method, the optimal method enables the user to choose an arbitrary confidence level. Unlike both the Mann-Kendall method (because of its integer statistic) and the regression-based method (because of the $t(n-2)$ distribution required), the optimal method can provide meaningful results for sets of observations containing as few as two observations.

We now return to beginning of this chapter, to the original purpose of investigating motion detection, namely distinguishing between stars and NEOs. We wish to show whether a fast-moving star, such as Barnard's Star, is detected as a moving object. The inertial speed was $v = 3 \times 10^{-4}$ mas/s and the single-CCD position standard deviation was $\sigma = 0.04$ mas. Looking at figure 5.10 for $v = 7.5 \times 10^{-3}\sigma/$s, it is clear that this velocity is so low, that Barnard's Star will be treated like any fixed star.

With NEOs, moving much faster (see figure 5.9), things are different. Assume a near Earth object, moving in the along-scan direction at $v = 3.5$ mas/s. Because of this low speed, it is observed in all of the astrometric fields. Assuming $\sigma = 7$ mas, the velocity is $v = 0.5\sigma$/s. Referring to figure 5.10 we see, that this NEO is almost certain to be labelled a moving object: $v = 0.5\sigma$/s is off the chart!

Let us consider a faster NEO. Assume a motion in the along-scan direction at $v = 10$ mas/s and, because of this greater speed, we assume the position standard deviation is larger than in the previous example: $\sigma = 10$ mas. Because of the speed, we only obtain four observations according to table 4.3, in the ASM, AF1, AF2 and AF3. Looking up $v = 1\sigma$/s in figure 5.11, we see that this object also almost certainly will be correctly labelled a moving object.
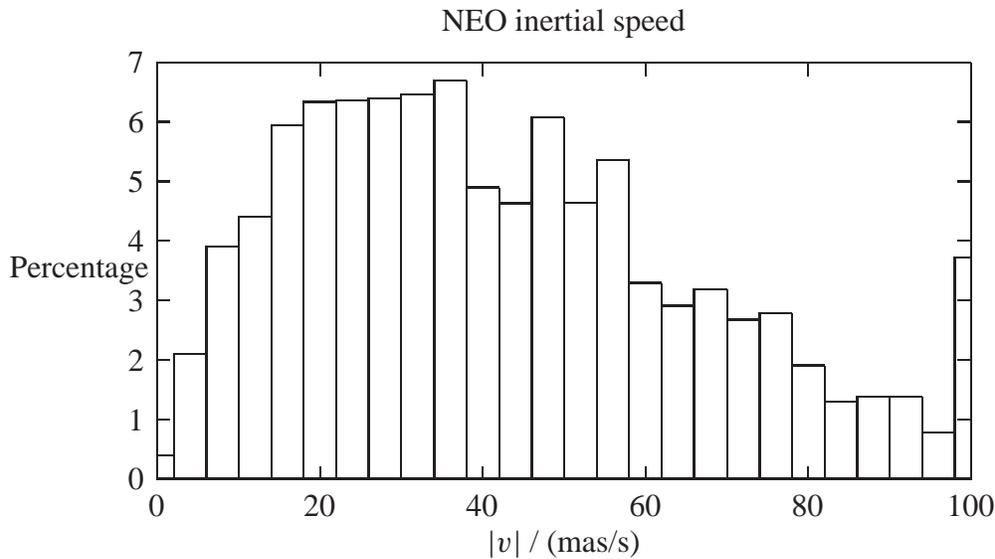


**Figure 5.9:** The inertial speed of observations of the synthesised NEO population, generated using the simulator described in section 4.4. Less than 0.5% of the observations come from objects moving at $|v| = 2$ mas/s or slower. Only the fastest of these objects will escape confirmation, and thus observation, in Gaia's AF1. See table 4.3.

## 5.10   Detecting Motion in Gaia Observations

This section deals with the application of the optimal motion detection method to simulated Gaia observations. Because Gaia position data is two-dimensional, and all the methods described above, including the optimal method, are one-dimensional, we need a solution that enables the use of a one-dimensional method to detect motion in two-dimensional data. One way of overcoming this problem is to use the motion detection

Probability for a 'fixed' label, based on 12 observations
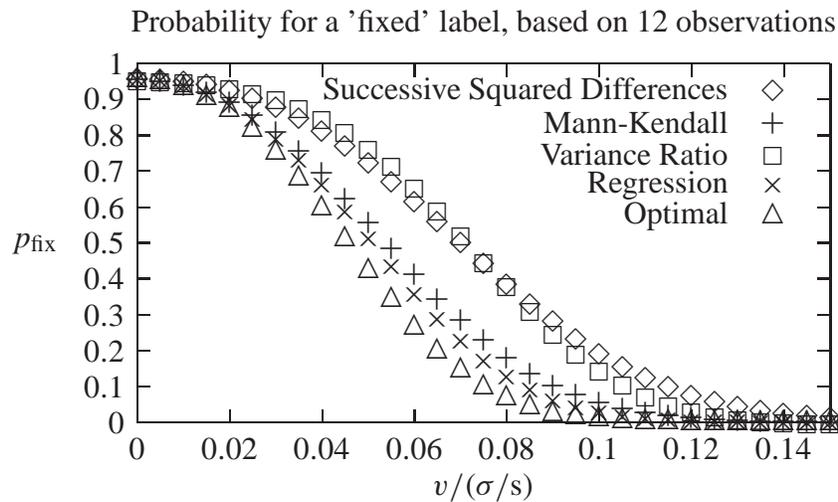


**Figure 5.10:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on twelve observations, using transit times $t_i$ corresponding to the Gaia PFOV (table 4.1), and calibrated such that $p_I = 4.5\%$.

Probability for a 'fixed' label, based on 4 observations
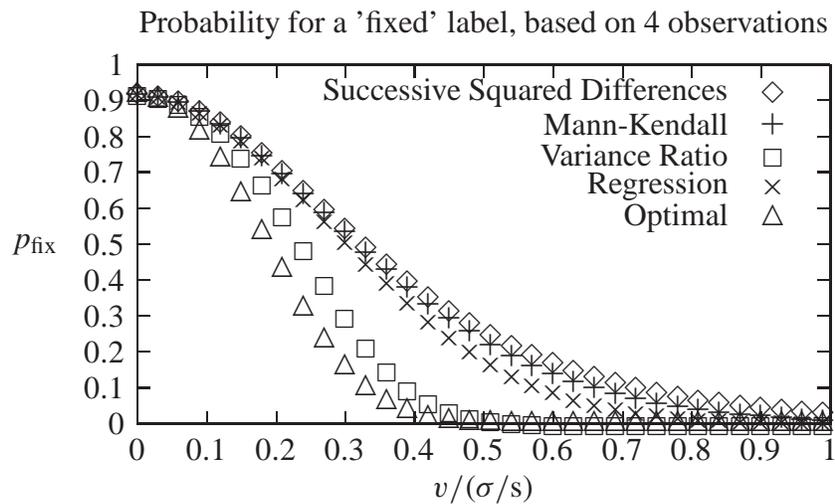


**Figure 5.11:** The probability that an object, moving at velocity $v$, is labelled as fixed. Based on four observations in Gaia's preceding field of view (ASM1, AF1, AF2, AF3), using the transit times from table 4.1, and calibrated such that $p_I = 8.3\%$.

method on the positions along each of the dimensions in turn, and flag an object as fixed if and only if none of the tests indicate motion.

The results of applying this method to simulated observations of NEOs, at different $p_I$ levels ($p_I$ being the probability of committing a type I error, i.e., that a fixed object is

labelled as moving), is shown in table 5.1. This shows that even at a $p_I$ level of 0.01%, only 0.084% of the NEO observations were not detected as coming from a moving source. Assuming all stars are fixed stars (see section 5.9 for stellar motion), this $p_I$ level means that 99.99% of all observations of stars are rejected, but only 0.084% of the simulated NEO observations.

| AL | AC | $p_I = 10\%$ | $p_I = 2\%$ | $p_I = 0.1\%$ | $p_I = 0.01\%$ |
|----|----|----|----|----|----|
| Yes | Yes | 90.063% | 86.387% | 80.118% | 75.967% |
| Yes | No | 8.057% | 11.295% | 16.805% | 20.404% |
| No | Yes | 1.857% | 2.279% | 3.008% | 3.545% |
| No | No | 0.023% | 0.038% | 0.069% | 0.084% |

**Table 5.1:** Motion detection on simulated Gaia observations of NEOs. The probability of a fixed object being labelled as moving, is denoted by $p_I$. The first two columns indicate whether or not motion was detected in the AL and AC directions, respectively. This shows that even at a $p_I$ level of 0.01%, only 0.084% of the NEO observations were not detected as coming from a moving source.

Although this method of combining the result of two one-dimensional tests provides a very efficient way of distinguishing between observations of stars and observations of NEOs, it is not necessarily optimal. An optimal method for two-dimensional motion detection is one of the entries in the list of obvious directions of future work (see section 7.2).

## 5.11  Velocity Estimation

Unlike traditional observation of the motion of celestial objects, Gaia observations will, for each field of view transit, provide both a very accurate position and velocity. Making the best possible use of this added information is a challenge to the field of orbit computation in the time to come.

This section, deviating slightly from motion detection as such, deals with the velocity estimate appearing as a side effect of both the optimal method described above (equation (5.29)) and the least-squares approach of the regression-based method (equation (5.4)).

As shown earlier, these two velocity estimates are identical for constant variance $\sigma_i^2 = \sigma^2$. Receiving such an estimate "for free", it is only natural to explore the quality of $\tilde{v}$ as a velocity estimator.

Simulating Gaia's observations of the 2000 first Main Belt asteroids and the simulated population of NEOs for the full five-year mission duration, and estimating the along-scan velocity for each field crossing, we compare the velocity estimate with the actual (simulated) object velocity and plot the corresponding velocity estimate errors as a function of actual velocity. This can be seen in figures 5.12 and 5.13.
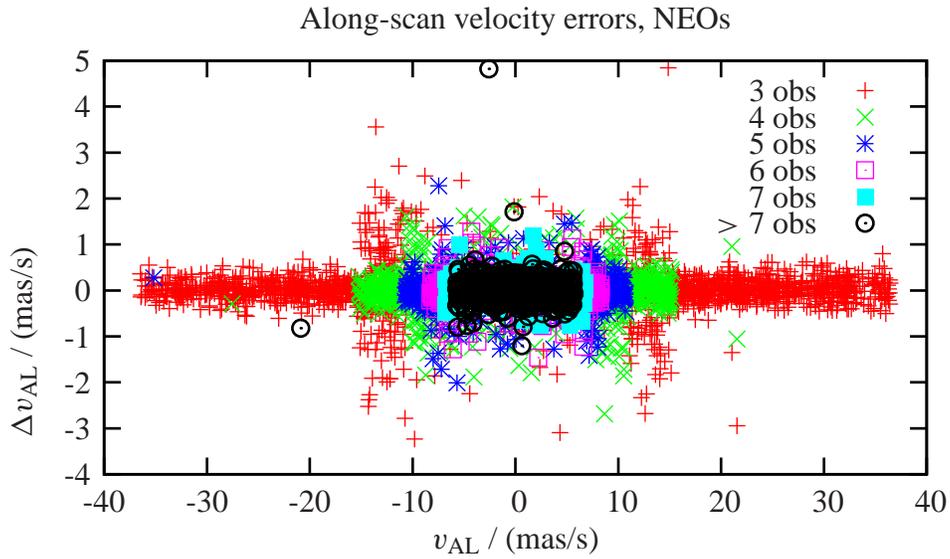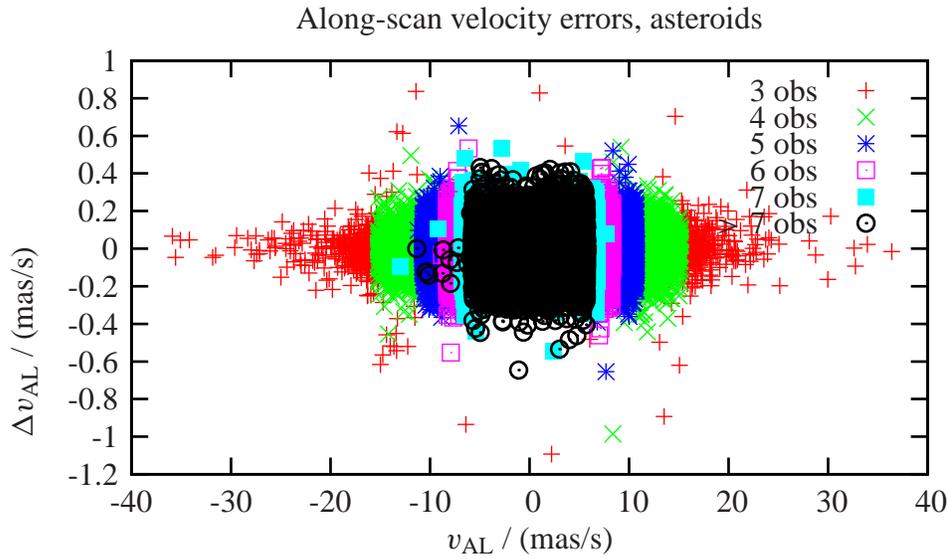
**Figure 5.12:** NEO velocity fit errors. Only observations from the preceding field of view are shown. $RMS(v_{AL}) = 15.6$ mas/s. $RMS(\Delta v_{AL}) = 0.49$ mas/s. From [Wolff 2004].

The figures show data from the preceding fields of view only, and only display sets of observations consisting of more than two observations. The critical velocities of table 4.3 are evident: No sets of observations containing more than three observations occur for $|v_{AL}| > 15.2$ mas/s, because an object moving at that velocity will not be observed in AF2-AF10 in the preceding field of view. Similarly, no sets of observations containing more than four observations occur for $|v_{AL}| > 10.7$ mas/s.

Assume a NEO is observed, moving at $|v_{AL}| = 13$ mas/s. This along-scan velocity rules out observations in AF3-AF10 (see table 4.3). If we assume the across-scan velocity is $|v_{AC}| = 25$ mas/s, this will cause the object to not be observed in AF11, either. This particular object will yield a set of three observations, from the ASM, the AF1 and the AF2, implying a very short timebase, thus resulting in a poor velocity estimate. This explains the apparent drastic deterioration in velocity estimates for three-observation sets as the speed falls below 15.2 mas/s (figure 5.12).

Keen eyes may detect that the plots exhibit a few apparently wrong measurements, e.g., having a set of four observations of an object having an along-scan velocity $|v_{AL}| > 15.2$ mas/s, which should never occur. These blemishes are caused by a limitation in the version of the Gaia simulator used (see section 4.4). The angle describing the orientation of Gaia's scan circle is approximated by a constant for each set of observations. However, since the simulator works in time steps, if a set of observations happens to straddle the transition from one time step to the next, a step in the orientation angle may be recorded. When using this recorded orientation angle along with the recorded ecliptic coordinates to obtain the observed position with respect to the scan circle, the step in the orientation

**Figure 5.13:** Asteroid velocity fit errors. Only observations from the preceding field of view are shown. $\mathrm{RMS}(v_{\mathrm{AL}}) = 6.77$ mas/s. $\mathrm{RMS}(\Delta v_{\mathrm{AL}}) = 0.08$ mas/s. From [Wolff 2004].

angle leads to inconsistent results. Fortunately, this only happens rarely, approximately one in each thousand sets of observations, and is not regarded as a problem affecting the results of the examination.

Not easily decipherable from the plots, the accuracy of the velocity estimate from one field crossing (the RMS of the errors) of the simulated NEO population and that of the 2000 first Main Belt asteroids are:

$$\sigma_{v_{\mathrm{AL}},\mathrm{NEO}} \approx 0.6 \text{ mas/s}$$
$$\sigma_{v_{\mathrm{AL}},\mathrm{MBO}} \approx 0.08 \text{ mas/s}$$

This predicts that the velocity of about 95% of the NEOs and Main Belt asteroids can be determined with an absolute error of less than 1.2 mas/s and 0.16 mas/s, respectively, based on measurements from a single astrometric field crossing.

Furthermore, these simulations indicate that the relative error on the velocity estimation based on a single crossing is less than 10% for about 80% of the simulated NEO population. For 95% of this population, the relative error on the velocity is less than 30%. As to the velocity error of the 2000 Main Belt asteroids, 95% of these have a relative error better than 10%, and 80% better than 2.5%.

Remember, though, that because of the bias in the NEO observations (see section 4.5), the figures for the NEO population are probably slightly optimistic.

# Chapter 6

# Orbit Computation

This chapter will provide a brief introduction to orbit computation, covering both classical approaches as well as modern statistical techniques. The term *orbit computation* is used to cover all three aspects into which astrometry has traditionally been divided: Finding a *preliminary orbit* to be used for the *prediction* of future positions to help obtain additional observations used to improve the orbital elements to better match all observations, a process called *orbit improvement*. Orbit computation is usually treated as an inverse problem – the corresponding direct problem is the prediction of future sky positions of an object. This chapter mainly deals with finding a preliminary orbit from a few observations. A preliminary orbit is a set of *osculating elements* describing the orbit at a certain *epoch*. If the universe was a Newtonian two-body system, the osculating elements would remain valid. However, effects such as perturbations (described in section 6.7) affect orbits, invalidating the osculating elements over time. Even for an ideal Newtonian two-body "universe", the osculating elements may be erroneous owing to observational errors. Therefore, a set of computed orbital elements is often accompanied by an assessment of the uncertainty of the orbital parameters. Orbit improvement, the process of reducing the uncertainty by fitting model parameters to better suit observations, is briefly touched upon in section 6.6.

## 6.1 Orbital Elements

The orbital elements (figure 6.1) comprise a set of parameters used to refer an orbit to a standard reference frame. When considering objects orbiting the Sun, it is customary to use a Sun-centered, or *heliocentric* reference frame, where the reference plane is that of the ecliptic. Within the reference plane, the reference line is the direction of the vernal equinox ♈. In general, orbital planes do not coincide with the reference plane. The angle between the reference plane and the orbital plane is called the *inclination $i$*. If $0° \leq i \leq 90°$, the orbit is called *prograde*. Conversely, if $i \geq 90°$, the object appears to move "backwards", and the orbit is called *retrograde*. The line of intersection between the orbital plane and the reference plane is called the *line of nodes*. The point where
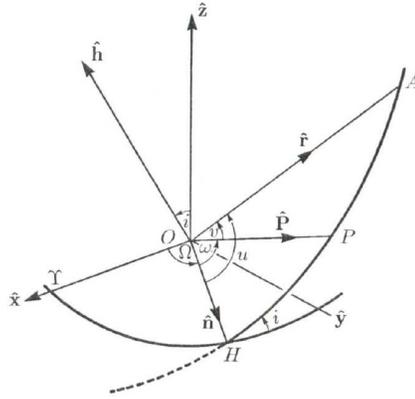
**Figure 6.1:** Orbital elements in the heliocentric reference frame. The Sun is at $O$, $\hat{\mathbf{x}}$ points toward the vernal equinox ♈, and $\hat{\mathbf{z}}$ toward the north pole of the ecliptic. The perihelion is denoted by $P$ and $H$ is the ascending node. The vector $\hat{\mathbf{h}}$ is perpendicular to the orbital plane. The position of the object on the celestial sphere is given by $A$. From [Danby 1988].

the orbit passes the reference plane moving north is called the *ascending node*. The corresponding point where the orbit moves south through the reference plane is called the *descending node*. The angle between the reference direction and the radius vector to the ascending node is called the *longitude of ascending node*, denoted $\Omega$. The angle between the ascending node and the perihelion (the point of closest approach to the Sun) is called the *argument of perihelion*, denoted $\omega$. The sum of the longitude of ascending node and the argument of perihelion is called the *longitude of perihelion*, denoted $\varpi = \Omega + \omega$. Note, that the two angles are generally in two different planes. In general, therefore, $\varpi$ is a "dogleg" angle.

The set of orbital parameters used in this chapter consists of:

$a$: The semimajor axis

$e$: The eccentricity of the orbit

$i$: The inclination of the orbital plane

$\Omega$: The longitude of ascending node

$\omega$: The argument of perihelion

$T$: The time of perihelion passage

## 6.2   Early Orbit Computation

When an abundance of observations is available, it is possible to determine the orbit of a celestial body without imposing many restrictions, e.g., without restricting the orbit to a conic section. The following method was known and used by Kepler to postulate his three famous laws of orbital motion, introduced in section 2.3 [Collins 2004].

First, the *siderial period* (the time it takes the object to make one full orbit around the Sun, relative to the stars) of the object must be found. We assume the orbit of the Earth is well known. The siderial period of the foreign body is found by observing it at some particular configuration, e.g., observing the object at opposition (when the object and the Sun are 180° apart, as seen from the Earth). The next time the object can be observed at opposition, is exactly one *synodic period* later, i.e., the synodic period is the time it takes for the object to reappear at the same spot in the sky, relative to the Sun, as observed from Earth. We denote the synodic period by $P_{\text{syn}}$. The angular distance travelled by the Earth is:

$$\Delta v_{\oplus} = 2\pi \frac{P_{\text{syn}}}{P_{\oplus}} \ ,$$

where $P_{\oplus}$ denote the siderial period of the Earth. If we denote the siderial period of the foreign body by $P_{\text{sid}}$, the angular distance travelled can be determined by

$$\Delta v = 2\pi \frac{P_{\text{syn}}}{P_{\text{sid}}} \ .$$

Since the configuration is the same, the difference in angular distance travelled must be $2\pi$, and hence:

$$\left| \Delta v_{\oplus} - \Delta v \right| = \left| \frac{2\pi \, P_{\text{syn}}}{P_{\oplus}} - \frac{2\pi \, P_{\text{syn}}}{P_{\text{sid}}} \right| = 2\pi$$

This can be rewritten to:

$$\frac{1}{P_{\text{sid}}} = \left| \frac{1}{P_{\oplus}} - \frac{1}{P_{\text{syn}}} \right| \ ,$$

from which the siderial period of the object is derived. This formula was devised by Copernicus [Collins 2004].

Two observations from Earth, $P_{\text{sid}}$ apart in time, will, in general, furnish two observations from two different vantage points, as the siderial period of the object will not in general be commensurate with that of the Earth, see figure 6.2. Knowing the position of the Earth with respect to the Sun at the two times of observation means that two sides ($r_1$ and $r_2$) and

three angles ($\gamma_1$, $\gamma_2$ and $\gamma_3$) are known of the quadrilateral defined by the position of the Sun $S$, the positions of the Earth $E_1$ and $E_2$, and the positions of the object, coinciding at $P$. This enables the computation of the remaining lengths, thus determining the position of the object. Repeating this procedure throughout the orbit of the object, the entire orbit can be determined. Obviously, this procedure requires a lot more observations than the minimum three pairs required to determine an orbit, however it enabled Kepler to determine the orbits of several objects without any assumption about the orbital shape.
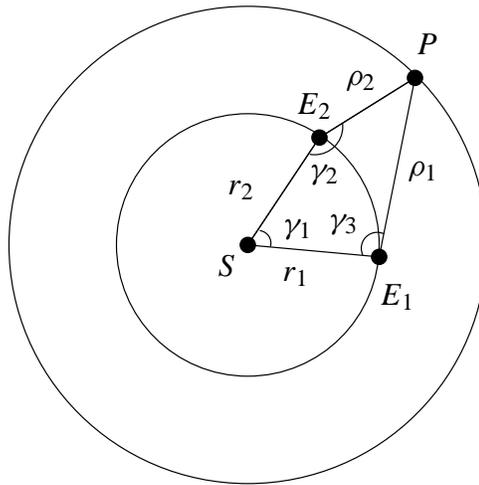


**Figure 6.2:** Early orbit computation

Whereas this method was sufficient for determining the exact orbit of Mars, it was less suitable for the orbit computation of comets, since their eccentric orbits would rapidly bring them close to the Sun, making them unobservable. A new method for the computation of orbital elements was needed. Almost a century after Newton's proof that the orbits of celestial objects was conic sections, and only a few decades before the ground-breaking work of Gauss, Laplace developed a method to determine the orbital elements from a limited number of observations on a short observational arc. If a heliocentric position vector and a corresponding velocity vector is known, all orbital elements can be derived (see section 6.4). Laplace's method is based on Taylor series approximations to derive the velocity from a series of observations. Although three observations are enough to determine the orbital elements, in order for these Taylor series approximations, and hence the orbital elements, to be accurate, more observations are needed. Nevertheless, Laplace's method has been polished by many researchers for the past two centuries, and is still popular [Marsden 1985].

# 6.3  Classical Orbit Computation

This section describes an orbit computation method originating in Gauss' work leading to the recovery of Ceres, the first asteroid discovered. Since Gauss' first work in 1801, many people, including Gauss himself, have improved upon the method. The method described, is called the Gauss-Encke-Merton (after the main contributors) in [Danby 1988] and [Marsden 1985]. We will start by introducing a few useful tools: the $f$ and $g$ functions, and the sector-triangle ratios. The presentation will follow that of [Tatum 2005] and [Danby 1988], using a notation similar to the latter.

## The $f$ and $g$ functions

A convenient tool when solving problems in celestial mechanics are the so-called $f$ and $g$ functions. These functions describe the motion of an object in its orbit as the displacement in radial and tangential directions with respect to a reference position, the position at time $t_0$. Denote the position and velocity of an object in its orbit at time $t_0$ by $\mathbf{r}_0$ and $\mathbf{v}_0$, respectively. Assuming these vectors are not parallel, the position of the object at time $t$ may be uniquely described by:

$$\mathbf{r}(t) = f(t, t_0)\mathbf{r}_0 + g(t, t_0)\mathbf{v}_0 \ . \tag{6.1}$$

Since these functions are independent of the reference system, they also hold in the orbital reference system (2.19), and hence

$$
\begin{aligned}
X(t) &= f X(t_0) + g \dot{X}(t_0) \\
Y(t) &= f Y(t_0) + g \dot{Y}(t_0) \ .
\end{aligned}
$$

Isolating $f$ and $g$ from this system of equations, we get

$$
\begin{aligned}
f &= \frac{X(t)\dot{Y}(t_0) - Y(t)\dot{X}(t_0)}{D} \\
g &= \frac{Y(t)X(t_0) - X(t)Y(t_0)}{D} \ ,
\end{aligned}
$$

where $D = X(t_0)\dot{Y}(t_0) - \dot{X}(t_0)Y(t_0)$, which is the same as $h$ (see (2.20)). Using (2.19) and (2.24) to express $f$ and $g$ as functions of the eccentric anomaly, we get:

$$f = \frac{a\,(\cos E - e)\,\frac{a^2 n\sqrt{1-e^2}}{r_0}\cos E_0 + a\sqrt{1-e^2}\sin E\frac{a^2 n}{r}\sin E_0}{D}$$

$$= \frac{a}{r_0}\,(\cos(E - E_0) - 1) + 1$$

$$g = \frac{a\sqrt{1-e^2}\sin E\,a\,(\cos E_0 - e) - a\,(\cos E - e)\,a\sqrt{1-e^2}\sin E_0}{D}$$

$$= \frac{1}{n}\,(\sin(E - E_0) - e\,(\sin E - \sin E_0)),\tag{6.2}$$

where $r_0$ and $E_0$ denote the distance to the focus and the eccentric anomaly at time $t_0$.
A first-order Taylor series approximation of $\mathbf{r}$ centered on $t_0$ would be:

$$\mathbf{r} \approx \mathbf{r}_0 + (t - t_0)\,\dot{\mathbf{r}}_0 = \mathbf{r}_0 + (t - t_0)\,\mathbf{v}_0$$

Comparing this to (6.1) we see that lowest-order approximations of $f$ and $g$ are:

$$f \approx 1$$
$$g \approx t - t_0 \tag{6.3}$$

## Sector-Triangle Ratios
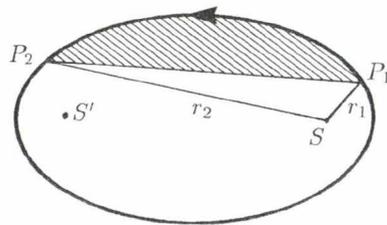


**Figure 6.3:** The sector-triangle ratio $y = A_s/A_t$, where $A_t$ is the area of the triangle $SP_1P_2$. The area of the sector $A_s$, swept by the radius vectors of the object position as it moves from $P_1$ to $P_2$, is the sum of the area of the triangle and the shaded area. From [Danby 1988].

The sector-triangle ratio $y$ is the ratio $A_s/A_t$, where $A_s$ is the area of the sector swept out by the radius vectors of two positions, and $A_t$ is the area of the triangle these same two positions make together with the focus. See figure 6.3.

The area of a sector is easily obtained, since it is proportional to the time between observations according to Kepler's Second Law. Assuming we have two heliocentric positions

$r_1$ and $r_2$, at times $t_1$ and $t_2$, respectively. The true anomalies are $v_1$ and $v_2$, and the corresponding eccentric anomalies are $E_1$ and $E_2$.

According to (2.13), the swept area is:

$$A_s = \frac{h}{2}\tau \,, \tag{6.4}$$

where $\tau = t_2 - t_1$. The area of the corresponding triangle equals half the cross product of the radius vectors, or:

$$A_t = \frac{1}{2}\mathbf{r}_1 \times \mathbf{r}_2 = \frac{1}{2}r_1 r_2 \sin(v_2 - v_1)$$

The sector-triangle ratio may be written:

$$y = \frac{A_s}{A_t} = \frac{h\tau}{r_1 r_2 \sin(v_2 - v_1)} = \frac{h\tau}{r_1 r_2 \sin 2f}, \tag{6.5}$$

where $2f = v_2 - v_1$.

According to (2.16), the true anomaly $v$ and eccentric anomaly $E$ are related as

$$r \cos v = a(\cos E - e)$$

Making use of the trigonometric identity $\cos^2 \frac{v}{2} = \frac{1+\cos v}{2}$, we find:

$$r \cos^2 \frac{v}{2} = a(1-e)\cos^2 \frac{E}{2} \tag{6.6}$$

Using the trigonometric identity $\sin^2 v + \cos^2 v = 1$, we find

$$r \sin^2 \frac{v}{2} = a(1+e)\sin^2 \frac{E}{2} \tag{6.7}$$

Substituting $r_1$, $v_1$, $E_1$ and $r_2$, $v_2$, $E_2$, respectively, into (6.6) and multiplying, yields, after taking the square root of the product:

$$\sqrt{r_1 r_2} \cos \frac{v_1}{2} \cos \frac{v_2}{2} = a(1-e)\cos \frac{E_1}{2} \cos \frac{E_2}{2} \tag{6.8}$$

A similar treatment of (6.7) yields:

$$\sqrt{r_1 r_2} \sin \frac{v_1}{2} \sin \frac{v_2}{2} = a(1+e)\sin \frac{E_1}{2} \sin \frac{E_2}{2} \tag{6.9}$$

For convenience, we introduce

$$2f = v_2 - v_1$$
$$2F = v_2 + v_1$$
$$2g = E_2 - E_1$$
$$2G = E_2 + E_1$$

Using the trigonometric identity $2 \cos x \cos y = \cos(x + y) + \cos(x - y)$ to rewrite (6.8):

$$\sqrt{r_1 r_2} \, (\cos F + \cos f) = a(1 - e) \, (\cos G + \cos g) \tag{6.10}$$

Using the trigonometric identity $2 \sin x \sin y = \cos(x - y) - \cos(x + y)$ to rewrite (6.9):

$$\sqrt{r_1 r_2} \, (\cos f - \cos F) = a(1 + e) \, (\cos g - \cos G) \tag{6.11}$$

Adding (6.10) and (6.11) and subsequently dividing by two yields:

$$\sqrt{r_1 r_2} \cos f = a \, (\cos g - e \cos G) \tag{6.12}$$

Substituting the $r_1$, $v_1$, $E_1$ into (6.6) and $r_2$, $v_2$, $E_2$ into (6.7) and multiplying, yields, after taking the square root of the product:

$$\sqrt{r_1 r_2} \cos \frac{v_1}{2} \sin \frac{v_2}{2} = a\sqrt{1 - e^2} \cos \frac{E_1}{2} \sin \frac{E_2}{2} \tag{6.13}$$

Exchanging the indices, i.e., substituting the $r_1$, $v_1$, $E_1$ into (6.7) and $r_2$, $v_2$, $E_2$ into (6.6) and multiplying, yields, after taking the square root of the product:

$$\sqrt{r_1 r_2} \cos \frac{v_2}{2} \sin \frac{v_1}{2} = a\sqrt{1 - e^2} \cos \frac{E_2}{2} \sin \frac{E_1}{2} \tag{6.14}$$

Using the trigonometric identity $2 \sin x \cos y = \sin(x + y) + \sin(x - y)$ on (6.13) and (6.14) yields, respectively:

$$\sqrt{r_1 r_2} \, (\sin F + \sin f) = a\sqrt{1 - e^2} \, (\sin G + \sin g) \tag{6.15}$$

$$\sqrt{r_1 r_2} \, (\sin F - \sin f) = a\sqrt{1 - e^2} \, (\sin G - \sin g) \tag{6.16}$$

Subtracting (6.16) from (6.15) and dividing by two:

$$\sqrt{r_1 r_2} \sin f = a\sqrt{1 - e^2} \sin g \tag{6.17}$$

Using (2.18) and the trigonometric identity $2 \cos x \cos y = \cos(x + y) + \cos(x - y)$ to express $|\mathbf{r}_1| + |\mathbf{r}_2|$:

$$
\begin{aligned}
r_1 + r_2 &= a\,(1 - e\cos E_1) + a\,(1 - e\cos E_2) \\
&= a\,(2 - e\,(\cos E_1 + \cos E_2)) \\
&= a\,(2 - e\,(\cos(G - g) + \cos(G + g))) \\
&= 2a\,(1 - e\cos G \cos g)
\end{aligned}
\tag{6.18}
$$

Now we can use (6.18) and (6.12) to eliminate $e \cos G$:

$$
r_1 + r_2 - 2\sqrt{r_1 r_2}\cos h \cos g = 2a \sin^2 g
\tag{6.19}
$$

Using (2.22) to express the mean motion in the left hand side of Kepler's Equation (2.25), we get:

$$
n(t - T) = \sqrt{\frac{\mu}{a^3}}(t - T) = E - e\sin E
\tag{6.20}
$$

Substituting $E_1$, $t_1$ and $E_2$, $t_2$, respectively, into (6.20), and subtracting the former from the latter, yields:

$$
E_2 - E_1 - e\,(\sin E_2 - \sin E_1) = \frac{t_2 - t_1}{a^{3/2}}
$$

Using the trigonometric identity $2 \cos x \sin y = \sin(x + y) - \sin(x - y)$ yields:

$$
2\,(g - e\cos G \sin g) = \frac{\tau}{a^{3/2}}
\tag{6.21}
$$

We can now eliminate $e \cos G$ from (6.21) and (6.12):

$$
2g - \sin 2g + \frac{2}{a}\sqrt{r_1 r_2}\sin g \cos f = \frac{\tau}{a^{3/2}}
\tag{6.22}
$$

From (2.21) we see that $h = na^2\sqrt{1 - e^2}$. We can use this to rewrite (6.17) to give an expression for $h$:

$$
h = n\frac{a\sqrt{r_1 r_2}\sin f}{\sin g} = \sqrt{\frac{\mu}{a^3}}\frac{a\sqrt{r_1 r_2}\sin f}{\sin g} = \frac{\sqrt{\mu r_1 r_2}\sin f}{\sqrt{a}\sin g}
\tag{6.23}
$$

Expressing the sector-triangle ratio (6.5) using this expression for $h$, reduced using the trigonometric identity $\sin 2x = 2\sin x \cos x$:

$$y = \frac{\sqrt{\mu}\tau}{2\sqrt{ar_1r_2}\cos f \sin g}$$

From this expression we may obtain an expression for $a$:

$$a = \frac{\mu\tau}{4y^2 r_1 r_2 \cos^2 f \sin^2 g}$$

Introducing, for convenience:

$$\begin{aligned} M &= \frac{\sqrt{\mu}\tau}{2\left(\sqrt{r_1 r_2}\cos f\right)^{3/2}}, \\ N &= \frac{r_1 + r_2}{2\sqrt{r_1 r_2}\cos f}, \end{aligned}$$

We may now express $a$ as:

$$a = M^2 \frac{\sqrt{r_1 r_2}\cos f}{y^2 \sin^2 g} \tag{6.24}$$

Substituting this into (6.19) and isolating $y^2$, we get:

$$y^2 = \frac{M^3}{N - \cos g} \tag{6.25}$$

Substituting (6.24) into (6.22) and isolating $y^3 - y^2$, we get:

$$y^3 - y^2 = \frac{M^2 (g - \sin g \cos g)}{\sin^3 g} \tag{6.26}$$

Given two radius vectors, we can now find the sector-triangle ratio $y$ by simultaneously solving (6.25) and (6.26) numerically, using, e.g., a Newton-Raphson method.

## The Method of Gauss

This method of orbit computation uses three two-dimensional direction vectors to determine a velocity vector corresponding to one of the observations. From this information, all orbital elements can be derived (see section 6.4). The method expects the heliocentric radius vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ (corresponding to three observations at times $t_1$, $t_2$ and $t_3$) to be coplanar. Assuming the vectors are not parallel, this means that:

$$\mathbf{r}_2 = c_1 \mathbf{r}_1 + c_3 \mathbf{r}_3 \qquad (6.27)$$

Using the $f$ and $g$ functions to describe $\mathbf{r}_1$ and $\mathbf{r}_3$ from $\mathbf{r}_2$ and $\dot{\mathbf{r}}_2 = \mathbf{v}_2$, we get:

$$\begin{aligned} \mathbf{r}_1 &= f_1 \mathbf{r}_2 + g_1 \mathbf{v}_2 \\ \mathbf{r}_3 &= f_3 \mathbf{r}_2 + g_3 \mathbf{v}_2 \end{aligned} \qquad (6.28)$$

This gives an alternative way of expressing $\mathbf{r}_2$:

$$\mathbf{r}_2 = c_1 \left( f_1 \mathbf{r}_2 + g_1 \mathbf{v}_2 \right) + c_3 \left( f_3 \mathbf{r}_2 + g_3 \mathbf{v}_2 \right) \qquad (6.29)$$

By taking the right cross product of (6.29) with $\mathbf{v}_2$ and the left cross product with $\mathbf{r}_2$, we have:

$$\begin{aligned} 1 &= c_1 f_1 + c_3 f_3 \quad \text{and} \\ 0 &= c_1 g_1 + c_3 g_3 \end{aligned}$$

from which we may express $c_1$ and $c_3$ using $f_1$, $f_3$, $g_1$ and $g_3$:

$$c_1 = \frac{g_3}{f_1 g_3 - g_1 f_3} \qquad (6.30)$$

$$c_3 = -\frac{g_1}{f_1 g_3 - g_1 f_3} \qquad (6.31)$$

Using the approximation (6.3), we get approximate values of $c_1$ and $c_3$:

$$c_1 \approx \frac{t_3 - t_2}{(t_3 - t_2) - (t_1 - t_2)} = \frac{t_3 - t_2}{t_3 - t_1}$$

$$c_3 \approx -\frac{t_1 - t_2}{(t_3 - t_2) - (t_1 - t_2)} = \frac{t_2 - t_1}{t_3 - t_1} \qquad (6.32)$$

By taking the cross product of both sides of (6.27) with $\mathbf{r}_3$ and $\mathbf{r}_1$, respectively, we get:

$$\mathbf{r}_3 \times \mathbf{r}_2 = c_1 \mathbf{r}_3 \times \mathbf{r}_1 + c_3 \mathbf{r}_3 \times \mathbf{r}_3 \Leftrightarrow c_1 = \frac{|\mathbf{r}_2 \times \mathbf{r}_3|}{|\mathbf{r}_1 \times \mathbf{r}_3|}$$

$$\mathbf{r}_1 \times \mathbf{r}_2 = c_1 \mathbf{r}_1 \times \mathbf{r}_1 + c_3 \mathbf{r}_1 \times \mathbf{r}_3 \Leftrightarrow c_3 = \frac{|\mathbf{r}_1 \times \mathbf{r}_2|}{|\mathbf{r}_1 \times \mathbf{r}_3|} \qquad (6.33)$$

Introducing the area of triangles:

$A_{t1}$ denotes the area of the triangle made by $\mathbf{r}_2$ and $\mathbf{r}_3$

$A_{t2}$ denotes the area of the triangle made by $\mathbf{r}_1$ and $\mathbf{r}_3$

$A_{t3}$ denotes the area of the triangle made by $\mathbf{r}_1$ and $\mathbf{r}_2$

We introduce the same numbering scheme to denote the areas of the sectors, e.g., $A_{s1}$ denotes the area of the sector defined by $\mathbf{r}_2$ and $\mathbf{r}_3$. Using the fact that the sector areas are proportional to the difference in time between the corresponding observations, we may write:

$$
\begin{aligned}
c_1 &= \frac{|\mathbf{r}_2 \times \mathbf{r}_3|}{|\mathbf{r}_1 \times \mathbf{r}_3|} = \frac{A_{t1}}{A_{t2}} = \frac{A_{s1} A_{t1} A_{s2}}{A_{s2} A_{s1} A_{t2}} = \frac{t_3 - t_2}{t_3 - t_1} \frac{y_2}{y_1} \\
c_3 &= \frac{|\mathbf{r}_1 \times \mathbf{r}_2|}{|\mathbf{r}_1 \times \mathbf{r}_3|} = \frac{A_{t3}}{A_{t2}} = \frac{A_{s3} A_{t3} A_{s2}}{A_{s2} A_{s3} A_{t2}} = \frac{t_2 - t_1}{t_3 - t_1} \frac{y_2}{y_3}
\end{aligned}
\tag{6.34}
$$

Equation (6.34) shows how, knowing the three heliocentric positions of the object in question, the coefficients $c_1$ and $c_3$ may be derived using the sector-triangle ratios described in the previous section. By using (6.34) to determine $c_1$ and $c_3$ we are bringing Kepler's laws into consideration.
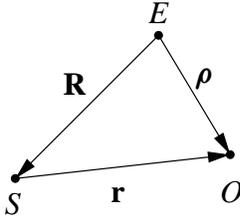


**Figure 6.4:** Diagram of the position vectors $\boldsymbol{\rho}$, $\mathbf{R}$ and $\mathbf{r}$ in relation to the Sun $S$, the Earth $E$ and the observed object $O$.

Introducing the geocentric position vectors (see figure 6.4) $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ and $\boldsymbol{\rho}_3$ and the geocentric positions of the Sun $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$, we have that $\boldsymbol{\rho}_i = \mathbf{r}_i + \mathbf{R}_i$ for $i = 1, 2, 3$. Dividing $\boldsymbol{\rho}_i$ into an unknown magnitude $\rho_i$ and a known direction unit vector $\hat{\boldsymbol{\rho}}_i$, we can rewrite (6.27):

$$
\begin{aligned}
\rho_2 \hat{\boldsymbol{\rho}}_2 - \mathbf{R}_2 &= c_1 \left( \rho_1 \hat{\boldsymbol{\rho}}_1 - \mathbf{R}_1 \right) + c_3 \left( \rho_3 \hat{\boldsymbol{\rho}}_3 - \mathbf{R}_3 \right) \\
&\Updownarrow \\
c_1 \rho_1 \hat{\boldsymbol{\rho}}_1 - \rho_2 \hat{\boldsymbol{\rho}}_2 + c_3 \rho_3 \hat{\boldsymbol{\rho}}_3 &= c_1 \mathbf{R}_1 - \mathbf{R}_2 + c_3 \mathbf{R}_3
\end{aligned}
\tag{6.35}
$$

Equation (6.35) is the fundamental equation to solve for the three unknowns $\rho_1$, $\rho_2$ and $\rho_3$ when using Gauss' method for orbit computation. Since $c_1$ and $c_3$ are not independent of

$\rho_1$, $\rho_2$ and $\rho_3$, the problem is solved by initially using approximations (6.32) for $c_1$ and $c_2$ to calculate approximations for $\rho_1$, $\rho_2$ and $\rho_3$, to be used to obtain better approximations for $c_1$ and $c_3$, *et cetera*. The convergence of this iterative procedure is by no means certain, but given sufficiently accurate data over a sufficiently long arc, the procedure usually converges to an accuracy matching that of the observations after less than 100 iterations.

Because this vector equation is ill-conditioned, numerical problems may be reduced by transforming it into a more well-suited coordinate system before attempting the iterative procedure. This new coordinate system uses the axes $\boldsymbol{\xi}$, $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, with $\boldsymbol{\xi}$ pointing toward the first observed position, and so that the direction of the third observation intersects the $\boldsymbol{\eta}$-axis. The $\boldsymbol{\zeta}$-axis completes a right-handed triad with $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$.

$$
\begin{aligned}
\boldsymbol{\xi} &= \hat{\boldsymbol{\rho}}_1 \\
\boldsymbol{\eta} &= \frac{\hat{\boldsymbol{\rho}}_1 \times (\hat{\boldsymbol{\rho}}_3 \times \hat{\boldsymbol{\rho}}_1)}{|\hat{\boldsymbol{\rho}}_1 \times (\hat{\boldsymbol{\rho}}_3 \times \hat{\boldsymbol{\rho}}_1)|} \\
\boldsymbol{\zeta} &= \boldsymbol{\xi} \times \boldsymbol{\eta}
\end{aligned}
$$

Expressing the observations in this coordinate system yields:

$$
\begin{aligned}
\hat{\boldsymbol{\rho}}_1 &= \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \mu_1 \\ \nu_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
\hat{\boldsymbol{\rho}}_2 &= \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} \lambda_2 \\ \mu_2 \\ \nu_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\rho}}_2 \cdot \boldsymbol{\xi} \\ \hat{\boldsymbol{\rho}}_2 \cdot \boldsymbol{\eta} \\ \hat{\boldsymbol{\rho}}_2 \cdot \boldsymbol{\zeta} \end{bmatrix} \\
\hat{\boldsymbol{\rho}}_3 &= \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} \lambda_3 \\ \mu_3 \\ \nu_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\rho}}_3 \cdot \boldsymbol{\xi} \\ \hat{\boldsymbol{\rho}}_3 \cdot \boldsymbol{\eta} \\ 0 \end{bmatrix}
\end{aligned}
$$

Notice the zeros on the right hand side, indicating that this system of equations is solvable using back substitution. To do this, we introduce the geocentric Sun position transformed in the same way:

$$
\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\zeta} \end{bmatrix}^T \mathbf{R}_i \ , \quad i = 1, 2, 3
$$

The transformed version of (6.35) is:

$$
c_1 \rho_1 \begin{bmatrix} \lambda_1 \\ \mu_1 \\ \nu_1 \end{bmatrix} - \rho_2 \begin{bmatrix} \lambda_2 \\ \mu_2 \\ \nu_2 \end{bmatrix} + c_3 \rho_3 \begin{bmatrix} \lambda_3 \\ \mu_3 \\ \nu_3 \end{bmatrix} = c_1 \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} - \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} + c_3 \begin{bmatrix} X_3 \\ Y_3 \\ Z_3 \end{bmatrix}
$$

$$(6.36)$$

The first unknown to determine in the back substitution is $\rho_2$. From the $\boldsymbol{\zeta}$-component of (6.36), we have:

$$
\rho_2 = \frac{-c_1 Z_1 + Z_2 - c_3 Z_3}{\nu_2}
$$

We see that for very small values of $\nu_2$, numerical problems are to be expected. Since $\nu_2$ is a measure of departure of the observed arc from a great circle, a small value indicates that the observed arc is (almost) coincident with a great circle, and thus the problem has no unique solution.

Knowing $\rho_2$, the back substitution continues by finding $\rho_3$:

$$
\rho_3 = \frac{\rho_2 \mu_2 + c_1 Y_1 - Y_2 + c_3 Y_3}{c_3 \mu_3},
$$

and, finally, $\rho_1$:

$$
\rho_1 = \frac{\rho_2 \lambda_2 - c_3 \rho_3 \lambda_3 + c_1 X_1 - X_2 + c_3 X_3}{c_1}
$$

Having obtained $\rho_1$, $\rho_2$ and $\rho_3$, the heliocentric position vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are obtained using $\mathbf{r}_i = \rho_i \hat{\boldsymbol{\rho}}_i - \mathbf{R}_i$, for $i = 1, 2, 3$.

Knowing $\mathbf{r}_1$ and $\mathbf{r}_2$, the velocity $\mathbf{v}_1$ can be computed from the following (see (6.28)):

$$
\mathbf{r}_2 = f \mathbf{r}_1 + g \mathbf{v}_1 \tag{6.37}
$$

The values of the $f$ and $g$ functions still need to be determined. Taking the cross product of $\mathbf{r}_1$ with (6.37) yields:

$$
\mathbf{r}_1 \times \mathbf{r}_2 = f \mathbf{r}_1 \times \mathbf{r}_1 + g \mathbf{r}_1 \times \mathbf{v}_1 = g \mathbf{r}_1 \times \mathbf{v}_1 = g \mathbf{h}
$$

From this we see, that the area of the triangle (half the magnitude of the cross product of the vectors) defined by $\mathbf{r}_1$ and $\mathbf{r}_2$ is $A_t = gh/2$. Using Kepler's Law of areas to express the sector area (6.4), the sector-triangle ratio $y$ (already determined in the iterative process above) may be used to determine $g$:

$$
y = \frac{A_s}{A_t} = \frac{h\tau}{hg} = \frac{\tau}{g} \Leftrightarrow g = \frac{\tau}{y}
$$

Rewriting (6.2), we get:

$$f = 1 - \frac{a}{r_1}\left(1 - \cos\left(E_2 - E_1\right)\right)$$

Using the trigonometric identity $2\sin^2\frac{x}{2} = 1 - \cos x$:

$$f = 1 - \frac{2}{r_1}a\sin^2\frac{E_2 - E_1}{2}$$

Isolating $a\sin^2\frac{E_2-E_1}{2}$ in (6.23), noting that the $f$ and $g$ used there are differences in true and eccentric anomalies:

$$a\sin^2\frac{E_2 - E_1}{2} = \frac{\mu r_1 r_2 \sin^2\frac{v_2-v_1}{2}}{h^2} = \frac{\mu r_1 r_2\left(1 - \cos\left(v_2 - v_1\right)\right)}{2h^2} = \frac{\mu\left(r_1 r_2 - \mathbf{r}_1\cdot\mathbf{r}_2\right)}{2\,\mathbf{h}\cdot\mathbf{h}}$$

Finally, we can find the value of the $f$ function:

$$f = 1 - \frac{\mu\left(r_1 r_2 - \mathbf{r}_1\cdot\mathbf{r}_2\right)}{r_1\,\mathbf{h}\cdot\mathbf{h}}$$

Using these values for $f$ and $g$, we obtain a velocity vector $\mathbf{v}_1$ corresponding to the position vector $\mathbf{r}_1$ using (6.37):

$$\mathbf{v}_1 = \frac{\mathbf{r}_2 - f\mathbf{r}_1}{g}$$

The following section shows how orbital elements may be computed from a heliocentric position vector $\mathbf{r}$ and the corresponding velocity vector $\mathbf{v}$.

## 6.4   Obtaining Orbital Elements

This section deals with elliptic orbits only. For a treatment of parabolic and elliptic orbits, refer to [Danby 1988].

Knowing an heliocentric position vector $\mathbf{r}$ and a corresponding velocity vector $\mathbf{v}$, the orbital elements (see section 6.1) may be found as follows. The angular momentum per unit mass is given by the cross product of the position vector and the velocity:

$$\mathbf{h} = \mathbf{r}\times\mathbf{v}$$

From the geometry of figure 6.1, it can be seen that the direction of $\mathbf{h}$ can be written:

$$\hat{\mathbf{h}} = \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix} = \begin{bmatrix} \sin \Omega \sin i \\ -\cos \Omega \sin i \\ \cos i \end{bmatrix}$$

The direction of $\mathbf{h}$ can thus be used to find both the longitude of the ascending node $\Omega$ and the inclination $i$.

The Runge-Lenz vector (2.6) is given by:

$$\mathbf{e} = \begin{bmatrix} e_x \\ e_y \\ e_z \end{bmatrix} = \frac{\mathbf{v} \times \mathbf{h}}{\mu} - \hat{\mathbf{r}}$$

The magnitude of $\mathbf{e}$ is the eccentricity $e$. Recall also, that the Runge-Lenz vector points toward the pericenter.

Define the unit vector pointing toward the ascending node:

$$\hat{\mathbf{n}} = \begin{bmatrix} \cos \Omega \\ \sin \Omega \\ 0 \end{bmatrix}$$

The argument of pericenter $\omega$ can be found by using the fact that $\hat{\mathbf{n}} \cdot \mathbf{e} = e \cos \omega$ and $\hat{\mathbf{n}} \times \mathbf{e} = e \sin \omega \hat{\mathbf{h}}$.

The semi-major axis $a$ is found by isolating $a$ in (2.15):

$$a = \frac{h^2}{\mu \left(1 - e^2\right)}$$

The eccentric anomaly at the time of observation can be found from (2.18) and its derivative, leading to: $e \cos E = 1 - \frac{r}{a}$ and $e \sin E = \frac{r \dot{r}}{\sqrt{a \mu}}$.

Then, finally, the time of perihelion passage $T$ may be found from Kepler's Equation (2.25) as:

$$T = t - (E - e \sin E) \sqrt{\frac{a^3}{\mu}}$$

## 6.5 Complications

For practical orbit computation, several complications are omitted in the above. Some of these will be briefly discussed in this section.

**Geocenter-Topocenter correction**

This describes the act of obtaining geocentric observations from observations obtained using a telescope placed on the surface of the Earth, and not at its center.

**Light-time correction**

The observed position of a solar system object is the result of the detection of sunlight reflected off the surface of this object. The time it takes from reflection to observation may be several hours. Thus, the observed position is not the actual position at the time of observation.

**Effects of phase**

What is usually desired when doing astrometry, is the position of the center of mass, the *barycenter*. What is observed, however, is the center of the reflected "patch" of light, the *photocenter*. For high-accuracy astrometry, the difference between the barycenter and the photocenter, the so-called *photocenter shift*, can be significant. For an extreme example of photocenter shift, imagine observing the crescent moon two days after new moon. The center of the illuminated part of the moon is clearly far away from the barycenter.

# 6.6 Orbit Improvement

A classical way of performing orbit improvement is an iterative procedure called the differential correction method. Assuming we have a set of parameters $\mathbf{X}_0 = [a, e, i, \Omega, \omega, T]$ for the preliminary orbit. The residuals corresponding to $\mathbf{X}_0$ are $\boldsymbol{\epsilon}_0$. A differential correction to $\mathbf{X}_0$ may be obtained by linearising the map between $\mathbf{X}$ and $\boldsymbol{\epsilon}$ in a neighbourhood of $\mathbf{X}_0$ and using it to find the optimal correction to $\mathbf{X}_0$ in a least squares sense. This new set of parameters is then used as input as the method is iterated.

The model used when calculating the residuals is often more detailed than the one used to obtain the preliminary orbit, typically taking relativistic effects and/or the perturbations by other celestial bodies into account.

# 6.7 Perturbations

The previous sections of this chapter deal with the dynamics of a system of two bodies. Since the mass of the Sun dominates the mass of the solar system, the two-body approach is a fairly good approximation of the dynamics of each solar system object. In general, Newton's second law says that the acceleration of an object is proportional to the sum of

the forces acting on it. These (gravitational) forces are proportional to the masses of the objects on which they act, and inversely proportional to the square of the distance between said objects:

$$m\mathbf{a} = \sum_{i=1}^{N} \mathbf{F}_i = \sum_{i=1}^{N} \frac{\mathcal{G}mm_i}{r_i^2} \hat{\mathbf{r}}_i \tag{6.38}$$

where $m$ is the mass of the observed object and $m_i$ is the mass of the $i$th other object. The distance and direction to the $i$th object is denoted by $r_i$ and $\hat{\mathbf{r}}_i$, respectively. In the solar system, the dominating object is the Sun, having a mass three orders of magnitude greater than that of the second largest solar system object, Jupiter. However, according to (6.38), once two bodies come into close proximity to eachother, their mutual gravitational forces may exceed by far that exerted by the Sun. For comets, having typically elongated orbits with transneptunian aphelia, Jupiter is an important perturber. For near Earth objects, the Earth can also constitute a significant perturbing influence.

Assume we have a primary mass, denoted by index $c$, and two secondary masses, denoted by indices $i$ and $j$, respectively. Letting $\mathbf{r}_i$ and $\mathbf{r}_j$ denote the position vectors of the secondary masses with respect to the central, primary mass:

$$\mathbf{r}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \text{ and } \mathbf{r}_j = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix},$$

we get the following laws of motion in the inertial reference frame:

$$
\begin{aligned}
m_c\ddot{\mathbf{R}}_c &= \mathcal{G}m_cm_i\frac{\mathbf{r}_i}{r_i^3} + \mathcal{G}m_cm_j\frac{\mathbf{r}_j}{r_j^3} \\
m_i\ddot{\mathbf{R}}_i &= \mathcal{G}m_im_j\frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3} - \mathcal{G}m_cm_i\frac{\mathbf{r}_i}{r_i^3} \\
m_j\ddot{\mathbf{R}}_j &= \mathcal{G}m_im_j\frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} - \mathcal{G}m_cm_j\frac{\mathbf{r}_j}{r_j^3}
\end{aligned}
$$

The accelerations of the secondary objects relative to the primary are:

$$\ddot{\mathbf{r}}_i = \ddot{\mathbf{R}}_i - \ddot{\mathbf{R}}_c = \mathcal{G}m_j \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3} - \mathcal{G}m_c \frac{\mathbf{r}_i}{r_i^3} - \left( \mathcal{G}m_i \frac{\mathbf{r}_i}{r_i^3} + \mathcal{G}m_j \frac{\mathbf{r}_j}{r_j^3} \right)$$

$$= -\mathcal{G}(m_c + m_i) \frac{\mathbf{r}_i}{r_i^3} - \mathcal{G}m_j \left( \frac{\mathbf{r}_j}{r_j^3} - \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3} \right)$$

$$\ddot{\mathbf{r}}_j = \ddot{\mathbf{R}}_j - \ddot{\mathbf{R}}_c = \mathcal{G}m_i \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} - \mathcal{G}m_c \frac{\mathbf{r}_j}{r_j^3} - \left( \mathcal{G}m_i \frac{\mathbf{r}_i}{i_j^3} + \mathcal{G}m_j \frac{\mathbf{r}_j}{r_j^3} \right)$$

$$= -\mathcal{G}(m_c + m_j) \frac{\mathbf{r}_j}{r_j^3} - \mathcal{G}m_i \left( \frac{\mathbf{r}_i}{r_i^3} - \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \right)$$

These relative accelerations can be interpreted as gradients of scalar functions:

$$\ddot{\mathbf{r}}_i = \nabla_i (U_i + \mathcal{R}_i)$$
$$\ddot{\mathbf{r}}_j = \nabla_j (U_j + \mathcal{R}_j) \; ,$$

where $\nabla_i$ and $\nabla_j$ denotes the gradient with respect to the coordinates of object $i$ and $j$, respectively.

$$U_i = \mathcal{G}\frac{m_c + m_i}{r_i} \text{ and } U_j = \mathcal{G}\frac{m_c + m_j}{r_j}$$

The parts of the total potential denoted by $U_i$ and $U_j$ correspond to the unperturbed two-body dynamics described earlier:

$$\ddot{\mathbf{r}}_i = \nabla_i U_i = -\mathcal{G}\frac{m_c + m_i}{r_i^2} \hat{\mathbf{r}}_i \; ,$$

in analogy with (2.3). The remaining part of the potential, denoted by $\mathcal{R}$, is called the *disturbing function*, representing the change arising from the gravitational influence of the other secondary object. The exact expression of $\mathcal{R}$ depends on the choice of coordinate system. In this coordinate system, with the primary object at the origin, the disturbing functions may be expressed:

$$\mathcal{R}_i = \frac{\mathcal{G}m_j}{|\mathbf{r}_j - \mathbf{r}_i|} - \mathcal{G}m_j \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_j^3} \text{ and } \mathcal{R}_j = \frac{\mathcal{G}m_i}{|\mathbf{r}_i - \mathbf{r}_j|} - \mathcal{G}m_i \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_i^3}$$

The above analysis may be extended to any number of perturbing objects. For more information, refer to [Murray & Dermott 1999].

The transition from two-body dynamics to three-or-more-body dynamics does not lead to dramatically different orbits, owing to the Sun's dominance. The orbits of the planets of the solar system are still elliptical in shape. The perturbations cause the point of perihelion to advance slowly, making the elliptical orbits seem to rotate. The orbit of Mercury, for example, is observed to precess 5600 seconds of arc per century. Le Verrier found, that approximately 5025.5 seconds of arc can be explained by the precession of the equinoxes. Taking into account the perturbations of the planets – mainly Venus and the Earth because of their proximity, and Jupiter because of its mass – and the flattening of the Sun caused by its rotation accounts for an additional 531.5 seconds of arc. The remaining 43 seconds of arc have since been explained through the theory of general relativity.

## 6.8   Modern Approaches to Orbit Computation

In the years following his remarkable contribution to the recovery of Ceres, Gauss continued to improve the method used, in practice developing the least-squares method independently of Legendre. The method of least squares was the first statistical interpretation of orbit computation, yet it was almost two centuries before the subject was given a fully statistical treatment in [Muinonen & Bowell 1993], where the orbit computation problem was treated as a problem of statistical inversion based on Bayesian inference (see, e.g., [Lehtinen 1988]). The full solution to the problem is to determine the *a posteriori* probability density function of the orbital elements. Once this has been obtained, no additional sensitivity analysis is needed.

According to Bayesian inference, the probability density function of the vector of orbital elements $\mathbf{P}$ is proportional to the *a priori* probability density function $p_{\mathrm{pr}}$ and the probability density function of the residuals $p_{\epsilon}$:

$$p_{\mathrm{p}}(\mathbf{P}) = \frac{p_{\mathrm{pr}}(\mathbf{P}) \, p_{\epsilon}(\Delta \boldsymbol{\psi}(\mathbf{P}))}{\int p_{\mathrm{pr}}(\mathbf{P}) \, p_{\epsilon}(\Delta \boldsymbol{\psi}(\mathbf{P})) \mathrm{d}\mathbf{P}}$$

where $\Delta \boldsymbol{\psi}(\mathbf{P})$ denote the observational errors projected on the sky plane. This expression may be regularised to maintain the invariance of $p_{\mathrm{p}}$ under transformations from one orbital element to another, e.g., from Keplerian to Cartesian elements [Virtanen et al. 2001].

### Prediction

Using statistical methods such as this for prediction is a matter of mapping the probability density function of the orbital elements to a probability density function of the future sky position. According to [Muinonen & Bowell 1993], the joint probability density function for the topocentric distance $R$, right ascension $\alpha$ and declination $\delta$ at the time $t$ is:

$$p(R, \alpha, \delta; t) = \frac{1}{R^2 \cos \delta} \int d\mathbf{P} \; p_\mathrm{p}(\mathbf{P}) \; \delta_\mathrm{D}(R - R(\mathbf{P}, t)) \; \delta_\mathrm{D}(\alpha - \alpha(\mathbf{P}, t)) \; \delta_\mathrm{D}(\delta - \delta(\mathbf{P}, t))$$

where $\delta_\mathrm{D}$ denotes Dirac's delta function. The integral is evaluated using a Monte Carlo method or by linearising the *a posteriori* probability density function of the orbital elements, thus obtaining a Gaussian probability density.

## Comparison of Methods

One of the main problems one faces when using the classical, deterministic methods, is that the iterative procedures involved, both in the initial orbit computation and the subsequent orbit improvement, may converge to "unphysical" or ambiguous solutions, or even not converge at all. Methods originating in statistical inversion theory, such as the one described above, always provide meaningful results. The results obtained using a modern method in the case of, e.g., two observations of an asteroid, seconds apart, and with large observational errors, may not be very useful, but nonetheless valid. In order for results from classical methods to be meaningful, an assessment of the uncertainty of orbital elements is needed. Thus, a classical solution may consist of single estimators of the orbital elements, each with its respective error estimate. A solution from a method originating in statistical inversion theory provides full error analysis, even for non-Gaussian *a priori* errors [Muinonen & Bowell 1993], a property absent from the classical methods. This full error analysis comes at a cost, however. Especially when the Monte Carlo orbits are integrated, rather than derived from two-body dynamics, the computational cost is fairly high.

The rapidly increasing computing power available, coupled with the additional knowledge provided by the *a posteriori* probability density functions themselves, make this a matter of little concern. To quote [Virtanen 2005]: "Adopting the statistical approach to inversion does not make a complex inversion simple. But it can help to discern the complexity of the problem, if not known *a priori*, and, in ambiguous cases, give more realistic estimates for the parameters, and most importantly provide meaningful estimates for their errors."

# Chapter 7

# Conclusion and Future Work

## 7.1   Conclusion

The purpose of this project has been to develop and implement algorithms for the detection of near Earth objects, with emphasis on how the Gaia space observatory might be used to that effect. Introductions to relevant topics such as near Earth objects, celestial mechanics and orbit computation have each been covered in a chapter of this thesis. A technical description of the relevant instruments of Gaia were presented, as was a comparison to the most prolific Earth-based NEO search programmes.

The main contribution of this thesis is to be found in chapter 5 on motion detection. Here, several methods for one-dimensional motion detection are presented and compared.

The method of successive squared differences is based on the fact that, in a series of position measurements of a moving object, neighbouring measurements are likely to be closer to eachother than to the arithmetic mean of the series.

The Mann-Kendall is based on the signs of every unique pair of measurements in a series. Under the null hypothesis (assuming no motion), positive and negative signs are equally likely, contrary to the case of a moving object.

The variance ratio method is based on the knowledge of an expected sample variance. In the case of astrometric observations, the expected position variance is often known, as a function of viewing conditions. If the estimated sample variance is significantly less than the expected sample variance, this indicates a trend.

The regression-based method fits the position measurements to a straight line in a least-squares sense and subsequently tests whether the slope is significantly different from zero.

A fifth, novel, method is presented and shown to be optimal. It belongs to the class of likelihood ratio (LR) tests, and we show that the test statistic, the ratio of two probabilities of the same event under different hypotheses, has a monotonic relationship with an estimate of the speed $|\tilde{v}|$. This leads to a test based on the comparison of $|\tilde{v}|$ to a threshold value. This test is shown to be optimal among all symmetric and translation invariant tests.

The application of a variant of this essentially one-dimensional method to two-dimensional simulated Gaia observations has shown that it is possible to make a clear distinction between observations of stars and observations of NEOs: The method was able to retain 99.9% of the NEO observations while rejecting 99.99% of the stellar observations.

The final sections of chapter 5 presents the results of an analysis of the velocity estimate, based on simulated Gaia observations. It is shown that, based on one transit of one of Gaia's two astrometric instruments, the relative error on the velocity is less than 30% for 95% of a synthesised NEO population. Furthermore, 95% of the observations of a test population consisting of the 2,000 first numbered Main Belt asteroids have a relative velocity error better than 10%

## 7.2   Future Work

A logical next step, in order to extract the full potential of Gaia's accurate observations, is to generalise the optimal motion detection method to two dimensions. As a temporary substitute, one can use the method described in the previous chapter, or approximate the proper speed by $\tilde{v} = \sqrt{\tilde{v}_{\mathrm{AL}}^2 + \tilde{v}_{\mathrm{AC}}^2}$. This is, however, not necessarily the best solution, particularly not in the case of anisotropic residual distributions.

Several of the motion detection methods described in chapter 5 rely on the knowledge of the position standard deviations. In practice, these standard deviations will always be estimates, based on the observing conditions. An evaluation of the effects of inaccurate position standard deviation estimates is needed.

Finally, the effect of the artificial brightening of the NEO population, performed in order to obtain a large number of simulated observations for statistical stability, also needs to be thoroughly examined, although the effect is believed to be slight.

# Appendix A

# Glossary

**AC:** Across-scan. Direction perpendicular to Gaia's instantaneous scanning plane. Parallel to Gaia's instantaneous spin axis. Perpendicular to the along-scan direction, see AL.

**AL:** Along-scan. The instantaneous direction of motion of a Gaia telescope as it sweeps over the celestial sphere. Perpendicular to Gaia's spin axis and line of sight.

**Albedo:** Surface reflectivity. The ratio of the amount of electromagnetic radiation reflected by a body to the amount incident upon it, commonly expressed as a percentage.

**Argument of perihelion:** The angle between the ascending node and the perihelion. Denoted $\omega$.

**Ascending node:** The point where an object in its orbit passes the reference plane (e.g., the ecliptic plane) moving north.

**Astro:** Gaia instrument, mainly used for astrometry. Gaia has two Astro telescopes, projected onto the same focal plane.

**AU:** Astronomical Unit, approximately equal to the average distance between the Earth and the Sun. 1 AU $\approx 1.5 \times 10^{11}$ m

**Aphelion:** The point on the orbit of an object orbiting the Sun, where the distance to the Sun is at a maximum. The opposite of perihelion.

**Arcsecond:** Second of arc. Sixty seconds of arc is one minute of arc. Sixty minutes of arc is one degree.

**as:** see Arcsecond.

**Basic Angle:** The angle between Gaia's two Astro telescopes.

**Binning:**  The process of irreversibly combining the data from a number of pixels into a smaller number of samples, to reduce telemetry.

**CCD:**  Charge-Coupled Device. Electronic detectors, used instead of photographic plates or film in modern telescopes and digital cameras.

**Chaining:**  See Linking

**Cross Matching:**  See Linking

**Descending node:**  The point where an object in its orbit passes the reference plane (e.g., the ecliptic plane) moving south.

**Ecliptic:**  The plane in which the Earth orbits the Sun.

**Ephemeris:**  (pl.  ephemerides).  A table listing specific data of a moving object, as a function of time. Ephemerides usually contain right ascension and declination, apparent angle of elongation from the Sun (in degrees), and magnitude (brightness) of the object; other quantities frequently included in ephemerides include the objects distances from the Sun and Earth (in AU), phase angle, and moon phase.

**Epoch:**  An arbitrary fixed instant of time used as a chronological reference datum for orbital motions (see osculating elements).

**FFOV:**  Following field of view. Referring to the second of Gaia's two Astro telescopes. See PFOV.

**FOV:**  Field of view. The space visible in a telescope at one view. See also window.

**Heliocentric:**  Centered on the Sun.

**Inclination:**  The angle between the reference plane (e.g., the ecliptic plane) and the orbital plane.

**Inertial motion:**  See Proper motion.

**Lagrange points:**  Five equilibrium points in the restricted three-body problem.  Gaia will be orbiting Lagrange point L2 of the Sun-Earth system.

**Line of nodes:**  The line of intersection between the orbital plane and the reference plane (e.g., the ecliptic plane). Passes through the ascending and descending nodes.

**Linking:**  The process of selecting observations from the same object for the purpose of computing an orbit. Also called chaining and cross matching.

**Longitude of descending node:**  The angle between the reference direction (e.g., the direction of the vernal equinox) and the radius vector to the ascending node. Denoted $\Omega$.

**Longitude of perihelion:** The sum of the longitude of ascending node and the argument of perihelion. Denoted $\varpi = \Omega + \omega$. Note, that the two angles are generally in two different planes. In general, therefore, $\varpi$ is a "dogleg" angle.

**Perihelion:** The point on the orbit of an object orbiting the Sun, where the distance to the Sun is at a minimum. The opposite of aphelion.

**MBO:** Main Belt Object. Asteroids from the Main Asteroid Belt between the orbits of Mars and Jupiter. See also section 4.5.

**MOID:** Minimum Orbital Intersection Distance. The minimum distance between the orbits of two objects. MOID changes not more than 0.02 AU per century, except in the case of close approaches with major objects, where large perturbations may occur.

**mas:** See Milliarcsecond.

**Milliarcsecond:** One thousandth of a second of arc. One degree equals 3,600,000 mas

**Osculating elements:** Orbital elements used to describe the unperturbed (two-body) orbit that the object would follow if perturbations were to cease instantaneously. Osculating elements are always changing with time and therefore must have a stated epoch of validity.

**PFOV:** Preceding field of view. Referring to the first of Gaia's two Astro telescopes. See FFOV.

**Phase angle:** The angle between the observer and the light source, as seen from the observed object.

**Prograde:** The "normal" way of orbiting the central object. When viewed from the ecliptic north pole, solar system object on prograde orbits move around the Sun in the counterclockwise direction. Also called *direct*. See retrograde.

**Proper motion:** Motion with respect to the fixed stars.

**Retrograde:** The "abnormal" way of orbiting the central object. When viewed from the ecliptic north pole, solar system objects on retrograde orbits move around the Sun in the clockwise direction. See prograde.

**Set:** A *set of observations* consists of the observations obtained from one object, crossing one field of view. Sets of observations from the Gaia Astro instruments consist of two to twelve observations.

**Spectro:** Gaia instrument, mainly used for photometry and radial velocity measurements.

**Telemetry:** The data transmitted from Gaia to Earth, and the process of transmitting it.

**Trend:** A change in over time. In this thesis, a trend is a linear change in position over
time.

**Window:** A group of pixels, ideally centered on an object. Once an object is detected,
a window is allocated through which the object is observed. Because windows are
fixed in the sky, they should be large enough to make sure moving objects do not
escape them. However, because windows containing multiple objects are of little
value, windows should also be as small as possible. To reduce telemetry, the pixels
of a window may be binned before being transmitted to Earth. See also FOV.

# Bibliography

[Aïvazian 1978] *Étude statistique des dépendances*, S. Aïvazian. Second edition, Edition Mir, Moscow, 1978.

[Allen 2001] *The Edge of the Solar System*, R. L. Allen, G. M. Bernstein, R. Malhotra. ApJ Letters 549 L241-244, 2001.

[Arenou et al. 2003] *Detection Performances in the ASM (1)*, F. Arenou, C. Babusiaux, S. Mignot. Gaia technical report, On-Board Detection working group, OBD-CoCo-005, 2003

[Bottke et al. 2000] *Understanding the Distribution of Near-Earth Asteroids*, W. F. Bottke Jr., R. Jedicke, A. Morbidelli, J.-M. Petit, B. Gladman. Science, Vol 288, pp. 2190-2194, 2000

[de Bruijne 2003] *GAIA astrometric accuracy - provisional GAIA-2 results*, J. H. J. de Bruijne. Gaia technical report GAIA-JdB-008, 2003

[de Bruijne 2003-II] *Gaia: Scanning Law*, J. H. J. de Bruijne, 2003. Information sheet, available from the Gaia homepage, see [ESA 2004]

[de Bruijne 2005] *Accuracy Budget and Performances*, J. H. J. de Bruijne. Proc. of symposium "The Three-Dimensional Universe with Gaia", ESA SP-576, pp. 35-42, 2005.

[de Bruijne 2005-II] *Gaia Astrometric Accuracy – March 2005 Gaia-2 ESA Baseline*, J. H. J. de Bruijne. ESA Technical Note, Gaia-JdB-020, 2005.

[de Bruijne 2005-III] *Windowing and on-board data handling of MBOs and NEOs*, J. H. J. de Bruijne. ESA Technical Note, Gaia-JdB-024, 2005

[Burns et al. 1979] *Radiation Forces on Small Particles in the Solar System*, J. A. Burns, P. Lamy, S. Soter. Icarus 40, 1-48, 1979.

[Carpino et al. 2003] *Error Statistics of Asteroid Optical Astrometric Observations*, M. Carpino, A. Milani, S. R. Chesley. Icarus 166, pp. 248-270, 2003

[CSS 2005] *The Catalina Sky Survey Home Page*, http://www.lpl.arizona.edu/css/

[Ceplecha 1992] *Influx of Interplanetary Bodies onto Earth*, Z. Ceplecha. Astron. Astrophys. 263, 361-366, 1992.

[Chesley et al. 2003] *Direct Detection of the Yarkovsky Effect by Radar Ranging to Asteroid 6469 Golevka*, S. R. Chesley, S. J. Ostro, D. Vokrouhlický, D. Čapek, J. D. Giorgini, M. C. Nolan, J.-L. Margot, A. A. Hine, L. A. M. Benner, A. B. Chamberlin. Science, Vol. 302, 1739-1742, 2003.

[Chyba 1991] *Mantle Siderophiles and the Lunar Impact Record*, C. Chyba. Icarus 92, 217-233, 1991.

[Collins 2004] *The Foundations of Celestial Mechanics*, G. W. Collins. Pachart Publishing House, 2004.

[Consolmagno et al. 1998] *The Porosities of Ordinary Chondrites: Models and Interpretation*, G. J. Consolmagno, D. T. Britt, C. P. Stoll, Meteoritics and Planetary Science 33, 1221-1229, 1998.

[Danby 1988] *Fundamentals of Celestial Mechanics*, J. M. A. Danby. Second edition 1988, Willmann-Bell, Inc. ISBN: 0-943396-20-4

[ESA 2000] *GAIA – Composition, Formation and Evolution of the Galaxy. Concept and Technology Study Report*, European Space Agency, ESA-SCI(2000)4, 2000.

[ESA 2004] *Gaia – Taking the Galactic Census* homepage. European Space Agency 2004. http://www.rssd.esa.int/Gaia

[Evans et al. 2003] *Detection and Discovery of Near-Earth Asteroids by the LINEAR Program*, J. B. Evans, F. C. Shelly, G. H. Stokes. Lincoln Laboratory Journal, Vol 14, 2, 2003.

[Giorgini et al. 2002] *Asteroid 1950 DA's Encounter with Earth in 2880: Physical Limits of Collision Probability Prediction*, J. D. Giorgini, S. J. Ostro, L. A. M. Benner, P. W. Chodas, S. R. Chesley, R. S. Hudson, M. C. Nolan, A. R. Klemola, E. M. Standish, R. F. Jurgens, R. Rose, A. B. Chamberlin, D. K. Yeomans, J.-L. Margot. Science, Vol. 296, pp. 132-136, 2002

[Hartmann et al. 1999] *Reviewing the Yarkovsky Effect: New Light on the Delivery of Stone and Iron Meteorites from the Asteroid Belt*, W. K. Hartmann, P. Farinella, D. Vokrouhlický, S. J. Weidenschilling, Al. Morbidelli, F. Marzari, D. R. Davis, E. Ryan. Meteoritics and Planetary Science 34, A161-A167, 1999

[Henriksen & Wolff 2005] *Optimal Motion Detection*, C. Henriksen, S. Wolff. Submitted to the Monthly Notices of the Royal Astronomical Society, 2005

[Hestroffer et al. 2003] *Degradation of the centroiding precision with moving objects*, D. Hestroffer, F. Arenou, C. Babusiaux. Gaia Solar System Working Group meeting in Paris, France, May 2003. Available at http://www.obs-nice.fr/tanga/SSWG/

[Høg & Knude 2001] *Detection of NEOs with GAIA and Ground-Based Follow-Up*, E. Høg, J. Knude. Copenhagen University Observatory, GAIA-CUO-84, 2001

[Høg, et al. 2003] *Faint Objects and NEOs with GAIA*, E. Høg, F. Arenou, P. Hjorth, U. G. Jørgensen, F. Mignard and S. Wolff. Copenhagen University Observatory, GAIA-CUO-118, 2003.

[Høg, et al. 2003-II] *Scientific requirements for the on-board processing*, E. Høg, F. Arenou, S. Mignot, C. Babusiaux, D. Katz, C. Jordi. Copenhagen University Observatory GAIA-CUO-117, 2003

[Høg 2004] *Sampling Scheme D for Astro and MBP*, E. Høg. Copenhagen University Observatory GAIA-CUO-143, 2004

[Høg 2004-II] *Sampling Scheme E for Astro and MBP*, E. Høg. Copenhagen University Observatory GAIA-CUO-150, 2004

[Høg 2005] *Sampling of ASM and AF11*, E. Høg. Copenhagen University Observatory GAIA-CUO-166, 2005

[Høg & de Bruijne 2005] *Summary of Sampling Schemes for ASM, AF, BBP, SSM and MBP*, E. Høg, J. H. J. de Bruijne. Copenhagen University Observatory, GAIA-CUO-151 version 3.1, 2005

[Jedicke et al. 2003] *Earth and Space-Based NEO Survey Simulations: Prospects for Achieving the Spaceguard Goal*, R. Jedicke, A. Morbidelli, T. Spahr, J.-M. Petit, W. F. Bottke. Icarus 161, 17-33, 2003

[Kendall 1938] *A New Measure of Rank Correlation*, M. G. Kendall. Biometrika 30, 81-93, 1938

[Kendall & Stuart 1961] *The advanced theory of statistics*, M. G. Kendall, A. Stuart. Griffin, London, 1961

[Lehtinen 1988] *On Statistical Inversion Theory*, M. S. Lehtinen. In *Theory and Applications of Inverse Problems* (H. Haario, Ed.), pp. 46-57. Pitman Research Notes in Mathematics Series 167, 1988.

[LINEAR 2005] *The LINEAR Program Home Page*, http://www.ll.mit.edu/LINEAR/

[Mann 1945] *Nonparametric tests against trend*, H. B. Mann. Econometrica 13, 245-259, 1945

[Marsden 1985] *Initial Orbit Determination: The Pragmatist's Point of View*, B. G. Marsden. Astron. J. 90, pp. 1541-1547 (1985)

[Mignard 1982] *Radiation pressure and dust particle dynamics*, F. Mignard. Icarus, vol. 49, pp. 347-366, 1982

[Mignard 2001] *Detection of Near Earth Objects with GAIA*, F. Mignard. Observatoire de la Côte d'Azur/CERGA, SAG-FM-007, 2001

[Mignard 2001-II] *Phase Effect in the Detection of NEOs with GAIA*, F. Mignard. Observatoire de la Côte d'Azur/CERGA, SAG-FM-008, 2001

[Mignard 2003] *Gaia: The L2 Orbit*, F. Mignard. Gaia information sheet, 2003. Available at http://www.rssd.esa.int/gaia

[Mignard 2003-II] *Update of the simulation. Current error model*, F. Mignard. Gaia Solar System Working Group meeting in Besançon, France, November 2003. Available at http://www.obs-nice.fr/tanga/SSWG/

[Mignard 2005] *Chaining Together New Planets with GAIA – II: Testing*, F. Mignard. Gaia Solar System Working Group meeting in Nice, France, May 2005. Available at http://www.obs-nice.fr/tanga/SSWG/

[Morbidelli et al. 2002] *From Magnitudes to Diameters: The Albedo Distribution of Near Earth Objects and the Earth Collision Hazard*, A. Morbidelli, R. Jedicke, W. F. Bottke, P. Michel, E. F. Tedesco. Icarus 158, 329-342, 2002

[Morbidelli & Vokrouhlicky 2003] *The Yarkovsky-Driven Origin of Near Earth Asteroids*, A. Morbidelli, D. Vokrouhlicky. Icarus, Vol. 163, pp. 120-134, 2003

[Muinonen & Bowell 1993] *Asteroid Orbit Determination Using Bayesian Probabilities*, K. Muinonen, E. Bowell. Icarus 104, pp. 255-279, 1993

[Murray & Dermott 1999] *Solar System Dynamics*, C. D. Murray, S. F. Dermott. Cambridge University Press 1999. ISBN: 0-521-57295-9

[NEAT 2005] *The NEAT Program Home Page*, http://neat.jpl.nasa.gov/

[NEO Taskforce 2000] *Report of the Task Force on Potentially Hazardous Near Earth Objects*, the UK Task Force on Potentially Hazardous NEOs, 2000. Available at http://www.nearearthobjects.co.uk/

[OBD 2003] *Detection Performances in the ASM (I)*, Draft version 1.4, On-board Detection Working Group. OBD-CoCo-05, Jan 31st, 2003.

[Pace 2004] *New Gaia Astro and MBP FPAs Layout Baseline*, O. Pace, March 2004

[PanSTARRS 2005] *The Panoramic Survey Telescope And Rapid Response System – Pan-STARRS Program Home Page*, available at http://pan-starrs.ifa.hawaii.edu/

[Pouny et al. 2003] *ASTRO FPA General Design Description*, P. Pouny and GAIA FPA team, GAIAFPA.NT.00120.T.ASTR, Dec 2003

[Poynting 1904] *Radiation in the Solar System: Its Effect on Temperature and Its Pressure on Small Bodies*, J. H. Poynting. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 202, 525-552, 1904.

[Rudin 1988] *Real and complex analysis*, W. Rudin. Third edition, McGraw-Hill, New York 1988

[Tatum 2005] *Celestial Mechanics notes*, J. B. Tatum. Frequently updated. Available at http://orca.phys.uvic.ca/~tatum/

[Virtanen et al. 2001] *Statistical Ranging of Asteroid Orbits*, J. Virtanen, K. Muinonen, E. Bowell. Icarus 154, pp. 412-431, 2001

[Virtanen 2005] *Asteroid Orbital Inversion Using Statistical Methods*, J. Virtanen. Ph.d. thesis, University of Helsinki 2005.

[Whittaker & Watson 1927] *A Course of Modern Analysis*, E. T. Whittaker, G. N. Watson. Fourth edition, Cambridge University Press, 1927

[Wolff 2004] *Detection of Moving Objects: First Results*, S. Wolff. Gaia Solar System Working Group Meeting in Helsinki, May 2004. Available at http://www.obs-nice.fr/tanga/SSWG/

[Wolff 2005] *Observing Near-Earth Objects Using the Gaia Satellite*, S. Wolff. ECMI Newsletter 37, 2005. ISSN: 1013-9338.

[Wolff 2005-II] *Optimal Motion Detection*, S. Wolff. Gaia Solar System Working Group Meeting in Nice, May 2005. Available at http://www.obs-nice.fr/tanga/SSWG/

[Wolff 2005-III] *Gaia-Assisted On-Board Detection of Moving Objects*, S. Wolff. Proc. of symposium "The Three-Dimensional Universe with Gaia", ESA SP-576, pp. 329-331, 2005.