CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

1

THE PROBLEM OF ORBIT DETERMINATION

In this chapter we define the problem of orbit determination, by specifying its three basic mathematical elements: the dynamics, the observations and the error model. We state the minimum principle, the least squares principle as the main case, and attempt a classification of the types of orbit determination found in astronomy and astrodynamics. The last section contains suggestions on the reading sequence, to adapt this book to different needs.

1.1 Orbits and observations

The two essential elements of an orbit determination problem are orbits and observations. Orbits are solutions of an equation of motion:

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, t, \boldsymbol{\mu})$$

which is an ordinary differential equation; $\mathbf{y} \in \mathbb{R}^p$ is the **state vector**, $\boldsymbol{\mu} \in \mathbb{R}^{p'}$ are the **dynamical parameters**, such as the geopotential coefficients, $t \in \mathbb{R}$ is the time. In the asteroid case the equation of motion is the N-body problem, the asteroid orbit being perturbed by the gravitational attraction of the planets; for many comets and some exceptionally accurate orbits of asteroids the non-gravitational effects are also relevant. For an artificial satellite the equation of motion is the satellite problem, the orbit being mostly perturbed by the asymmetric part of the geopotential, but also by non-gravitational perturbations.

The **initial conditions** are the value of the state vector at an epoch t_0:

$$\mathbf{y}(t_0) = \mathbf{y_0} \in \mathbb{R}^p.$$

In the two simple cases cited above we have $p = 6$, i.e., the vector of the initial condition is just formed by the position and velocity of the small

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

body in some inertial reference system. The orbits are specific solutions, for a given value of $\mathbf{y_0}$ and $\boldsymbol{\mu}$, of the equation of motion (initial condition problem). All the orbits together form the general solution

$$\mathbf{y} = \mathbf{y}(t, \mathbf{y_0}, \boldsymbol{\mu}),$$

also known as **integral flow** when considered as a mapping from the initial conditions (and dynamical parameters) to the current state at time t:

$$\mathbf{y}(t) = \Phi_{t_0}^{t}(\mathbf{y_0}, \boldsymbol{\mu}).$$

For the second element we introduce an **observation function**

$$R(\mathbf{y}, t, \boldsymbol{\nu})$$

depending on the current state, directly upon time, and also upon a number of **kinematical parameters** $\boldsymbol{\nu} \in \mathbb{R}^{p''}$. The function R is assumed to be differentiable. The composition of the general solution with the observation function is the **prediction function**

$$r(t) = R(\mathbf{y}(t), t, \boldsymbol{\nu})$$

which is used to predict the outcome of a specific observation at some time t_i, with $i = 1, \ldots, m$. However, the observation result r_i is generically not equal to the prediction, the difference being the **residual**

$$\xi_i = r_i - R(\mathbf{y}(t_i), t_i, \boldsymbol{\nu}), \qquad i = 1, \ldots, m.$$

The observation function can depend also upon the index i, the most common case being the use of a two-dimensional observation function like (right ascension, declination) or (range, range-rate), in which case R has two different analytical expressions, one for i even, the other for i odd. All the residuals can be assembled forming a vector in $\mathbb{R}^m$

$$\boldsymbol{\xi} = (\xi_i)_{i=1,\ldots,m}$$

which is in principle a function of all the $p + p' + p''$ variables $(\mathbf{y_0}, \boldsymbol{\mu}, \boldsymbol{\nu})$.

The above equations define a fully *deterministic* model: each residual is a single valued function of the $p+p'+p''$ parameters. This function is obtained from the observation function, for which we assume an explicit analytical expression, by using the general solution, which is not known as an analytical expression but is uniquely defined by the differential equations; both functions are assumed to be differentiable, see Chapter 2. These assumptions may not be the whole truth, as we shall see in Chapters 14 and 17, but we shall work with them for now.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

The random element is introduced by the assumption that every observation contains an error. Even assuming we know with perfect accuracy all the true values $(\mathbf{y_0}^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ of the parameters, that our model is perfectly complete (both for the equation of motion and for the observations), and that our explicit computations are perfectly accurate (they are computed in "exact arithmetic", not with a realistic computer), nevertheless the residuals

$$\xi_i^* = r_i - R(\mathbf{y}(\mathbf{y_0}^*, t_i, \boldsymbol{\mu}^*), t_i, \boldsymbol{\nu}^*, i) = \epsilon_i$$

would not be zero but random variables. The joint distribution of $\boldsymbol{\epsilon} = (\epsilon_i)_{i=1,\ldots,m}$ needs to be modeled, that is we need some assumptions, either in the form of a probability density function or as a set of inequalities, describing the observation errors we rate as acceptable. The probabilistic approach in most cases uses Gaussian distributions, discussed in Chapter 3.

1.2 The minimum principle

The basic tool of the classical theory of orbit determination (Gauss 1809) is the definition of a **target function** $\mathcal{Q}(\boldsymbol{\xi})$ depending on the vector of residuals $\boldsymbol{\xi}$. The target function cannot be chosen arbitrarily, but needs to satisfy suitable conditions of regularity and convexity. We shall focus on the simplest case, in which $\mathcal{Q}$ is proportional to the sum of squares of all the residuals:

$$\mathcal{Q}(\boldsymbol{\xi}) = \frac{1}{m}\,\boldsymbol{\xi}^T\,\boldsymbol{\xi} = \frac{1}{m}\,\sum_{i=1}^{m}\xi_i^2.$$

A quadratic form of general type, provided it is non-negative, can be handled with exactly the same formalism (see Chapter 5) and often needs to be used in practical applications. Since each residual is a function of all the parameters,

$$\xi_i = \xi_i(\mathbf{y_0}, \boldsymbol{\mu}, \boldsymbol{\nu}),$$

the target function is also a function of $(\mathbf{y_0}, \boldsymbol{\mu}, \boldsymbol{\nu})$. The next step is to select the parameters to be fit to the data: let $\mathbf{x} \in \mathbb{R}^N$ be a subvector of $(\mathbf{y_0}, \boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathbb{R}^{p+p'+p''}$, that is $\mathbf{x} = (x_i), i = 1, N$, with each x_i either a component of the initial conditions, or a dynamical parameter, or a kinematical parameter. Then we consider the target function

$$Q(\mathbf{x}) = \mathcal{Q}(\boldsymbol{\xi}(\mathbf{x}))$$

as a function of $\mathbf{x}$ only, leaving the vector of the **consider parameters** $\mathbf{k} \in \mathbb{R}^{p+p'+p''-N}$ (all the parameters not included in $\mathbf{x}$) fixed at the assumed value.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

The **minimum principle** selects as **nominal solution** the point $\mathbf{x}^* \in \mathbb{R}^N$ where the target function $Q(\mathbf{x})$ has its minimum value Q^*. The principle of **least squares** is the minimum principle with as target function the sum of squares $\mathcal{Q}(\boldsymbol{\xi}) = \boldsymbol{\xi}^T\, \boldsymbol{\xi}/m$, or some other quadratic form.

1.3 Two interpretations

The minimum principle should not be understood as if the "real" solution needs to be the point of minimum $\mathbf{x}^*$. Two interpretations can be used.

According to the **optimization interpretation**, $\mathbf{x}^*$ is the optimum point but values of the target function immediately above the minimum are also acceptable. The set of acceptable solutions can be described as the **confidence region**

$$Z(\sigma) = \left\{ \mathbf{x} \in \mathbb{R}^N \,\middle|\, Q(\mathbf{x}) \leq Q^* + \frac{\sigma^2}{m} \right\}$$

depending upon the **confidence parameter** $\sigma > 0$. For least squares

$$Z(\sigma) = \left\{ \mathbf{x} \in \mathbb{R}^N \,\middle|\, \sum_{i=1}^{N} \xi_i^2 \leq m\, Q^* + \sigma^2 \right\}.$$

The intuitive meaning of the confidence region is clear: the solutions $\mathbf{x}$ in $Z(\sigma)$ correspond to observation errors larger than those for $\mathbf{x}^*$, but still compatible with the available information on the observation procedure. The choice of the value of σ bounding the acceptable errors is not easy.

The alternative **probabilistic interpretation** describes the observation errors ϵ_i as random variables with an assumed probability density, which should be the result of an error model, justified by a priori knowledge of the observation process and/or a posteriori statistical tests. The vector $\boldsymbol{\epsilon} = (\epsilon_i), i = 1, m$, is then a set of jointly distributed random variables (see Section 3.1), and also the joint probability density function needs to be known; in particular, independence of the errors for observations at different times cannot be assumed, but needs to be justified by statistical tests.

Then the probabilistic model of the observation errors can be mapped in a probabilistic model of the result of orbit determination, with a probability density for the random variables $\mathbf{x}$ which in principle exists and can be, at least under some hypotheses, explicitly computed. The probability that the true orbit coincides exactly with the nominal solution $\mathbf{x}^*$ is zero, although under reasonable hypotheses $\mathbf{x}$ could be both the mode (point of maximum of the probability density) and the expected value.

© in this web service Cambridge University Press & Assessment www.cambridge.org

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

In other words, the optimization interpretation describes the possible solutions as a subset of the $\mathbf{x}$ space where the target function has an acceptable value, surrounding the nominal solution which is the minimum point. The probabilistic interpretation regards the solutions as a probability density cloud, surrounding the point of highest probability density. Both interpretations can be useful, having different advantages and limitations.

1.4 Classification of the problem

Orbit determination appears as a number of different problems, with different dynamical systems and observation techniques. One way to classify the dynamical systems is to decompose the right-hand side of the equation of motion into three parts:

$$\frac{d\mathbf{y}}{dt} = \mathbf{f_0}(\mathbf{y}, t, \boldsymbol{\mu}) + \mathbf{f_1}(\mathbf{y}, t, \boldsymbol{\mu}) + \mathbf{f_2}(\mathbf{y}, t, \boldsymbol{\mu});$$

the *unperturbed* equation of motion has only the main term $\mathbf{f_0}$, with $|\mathbf{f_0}| \gg |\mathbf{f_1}|$. The main term may not contain unknown parameters, or very few. The perturbations are subdivided into the most relevant ones $\mathbf{f_1}$ and the negligible ones $\mathbf{f_2}$. Negligible means not only that $|\mathbf{f_1}| \gg |\mathbf{f_2}|$ but also that the effects of the $\mathbf{f_2}$ terms on the general solution are small (with respect to the observational accuracy), thus the equation of motion actually solved to compute the predictions contains only $\mathbf{f_0} + \mathbf{f_1}$. The choice of the terms to be neglected in each specific case is therefore a delicate issue, discussed in Sections 4.6, 15.3, and 17.3.

Let us focus on the main term $\mathbf{f_0}$. For a satellite of the Earth it is the monopole gravitational attraction of the Earth; for an object in heliocentric orbit it is the monopole attraction from the Sun, and so on. In most cases the unperturbed equation of motion is a two-body problem. Only in a few exceptional examples is there no dominant two-body term.

Thus we can classify orbit determination problems by the central body:

- Earth satellite orbits, for the Moon, artificial satellites, and space debris;
- heliocentric orbits, including the planets, the smaller asteroids, comets, meteoroids, trans-neptunian objects, and artificial interplanetary probes;
- satellite orbits of other planets, for the natural satellites, planetary orbiters, binary asteroids, and asteroid/comet orbiter missions;
- the orbits around another star, for binary stars and extrasolar planets;
- the cases without a dominant central body, such as orbits near the Lagrangian equilibrium points, temporary satellite captures, very small interplanetary dust with motion dominated by radiation pressure.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

The orbit determination problems may differ also in the observation method, in the number and timing of the data, and in their accuracy. The main difference is between the *collaborative* and the *population* orbit determination problems.

Tracking

In **collaborative orbit determination** the object whose orbit has to be determined has a man-built device specifically intended to assist the observer. In this case the observation procedure is usually called **tracking**.

The most common case is tracking by radio waves: artificial satellites are normally equipped with a device called a **transponder**, which receives, amplifies, and retransmits the radio signal received from a ground station in a given frequency band.[1] Then the **range-rate**, the time derivative of the distance between the spacecraft and the ground station, can be measured by the Doppler shift between the signal emitted from the ground station and the one received back. If the signal also contains, beside the carrier, an encoded signal and the transponder is *regenerative*, that is it can send back this encoded signal on top of the return carrier, then also the **range**, or distance from the ground station, can be measured. This is possible also at interplanetary distance, thus the spacecraft could be in heliocentric orbit but also orbiting around another planet, or around an asteroid/comet.

In the above example the spacecraft needs to consume energy in the transponder, thus it has to be active, with a power system and possibly with attitude control to suitably point some antenna. There are examples in which the spacecraft is totally passive, such as the Earth satellites specifically launched for satellite laser ranging: they are only equipped with a special class of mirror, the *corner cubes*, to return a light ray in the same direction it came from with minimal dispersion. The ground stations are equipped with lasers capable of powerful but short-duration pulses of monochromatic light: the time interval between the emission of each pulse and the return signal detection measures the distance to the satellite.

The above examples are about artificial celestial bodies, that is man-made spacecraft. However, a tracking device can be planted on a natural body: e.g., corner cubes have been placed on the Moon by American and Soviet missions in the 1970s, thus **lunar laser ranging** has been regularly performed for more than 30 years, and the orbit of our natural satellite is known with centimeter level accuracy, actually more accurately than the

[1] The return signal can be shifted in frequency with respect to the received one, but this is done with *phase locking*, preserving very accurately the timing information.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

orbit of any artificial satellite, affected by non-gravitational perturbations. The Viking landers have been on the surface of Mars for more than five years with operational transponders, and this has allowed the computation of the orbit of Mars with an accuracy of a few tens of meters. The interplanetary space probes like Voyager can be used to constrain the orbit of the planets they encounter. Planetary orbiters like Cassini (now around Saturn) and the future BepiColombo (around Mercury) will provide very accurate orbit determination for these planets and for the natural satellites of Saturn, thanks to the very accurate transponders on these spacecraft. Thus the main difference is not between *natural* and *artificial* orbits.

The specific properties of the collaborative cases are three.

First, the body has some built in capability to respond to tracking; thus the number of observations, their distribution in time and their accuracy are planned in the design phase of the mission. A simulation of orbit determination is a compulsory phase of **mission analysis**, the study showing that some proposed space mission is feasible from the astrodynamics point of view. If the simulated orbit determination gives poor results, the required frequency and accuracy of the observations has to be improved. Thus the most difficult cases of divergent orbit determination should not occur in the collaborative case; even strong nonlinearity and chaos should not happen. However, if there is some failure, either hardware like an antenna failing to deploy, or software like a faulty on-board computer program, or planning like an orbit determination simulation providing illusory results, then a tracking case may show some problems of the non-collaborative case, including divergence, excessive nonlinearity, and chaos.

Second, the observation data contain information on which object is being tracked. In the simplest case, there is only one spacecraft answering in a given frequency band in a given direction (within a given solid angle). Frequency bands and orbit slots (e.g., in the geosynchronous belt) are allocated by international authorities to avoid confusion and interference between signals to and from the satellites. In other cases (e.g., satellite constellations, such as navigation satellites) the satellite encodes its identity in the signal sent back to the ground. Thus we can assume we always know to which spacecraft each batch of tracking data belongs.[2] In most cases it is possible to treat each spacecraft as a separate problem of orbit determination; the exceptions are the cases of satellite-to-satellite tracking, where the radio/laser beam travels between two (or more) satellites, in which case the orbits of the two (or more) have to be solved simultaneously.

[2] Of course also this can occasionally fail, making orbit determination quite messy.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

Third, if the amount of observational data and their accuracy exceeds what is required for the determination of the orbit in the strictest sense, that is for fitting the initial conditions $\mathbf{y_0}$, the additional information can be used to fit other parameters, either dynamical or kinematical, and in fact this is often the case. This is the key idea of **satellite geodesy**, where the gravity field of the central planet (the Earth, the Moon, another planet, an asteroid) is determined from the tracking data, rather than from the inhomogeneous ground-based gravimetry. In satellite geodesy around the Earth also the position of the ground stations can be determined with an accuracy far superior to that possible with ground-based measurements.

Catalogs

In the case of **population orbit determination** the observations are a scarce resource because the objects do not assist the observer. The total number of observations may not be small; actually it can be comparable to that of the tracking data points for a scientific space mission, e.g., tens of millions. The problem is that they refer to objects of a large population, and the average number of observations per object is small: e.g., 10^7 observations of a population of 10^6 objects (down to the minimum size observable).

The example most extensively discussed in this book is the orbit determination of the small bodies of the Solar System, including asteroids, comets, meteoroids and trans-neptunian objects. The number of objects needs to be qualified by a class of orbits and a minimum size: e.g., there are of the order of 10^6 main belt asteroids of size ≥ 1 km in diameter (this is just an estimate, extrapolated from the orbits already determined). A **survey** consists of a number of telescopes scanning the sky and looking for objects with stellar appearance which move with respect to the approximately fixed stars; this is the origin of the name asteroid, as proposed by Herschel. When such a **moving object** is detected the amount of information is minimal, typically only *astrometry*, that is angular positions, and *photometry*, that is apparent magnitude. There is a signature neither to identify the object with the ones already discovered, nor to decide it is new.

As we will see in Chapter 8, orbit determination is typically not possible with the discovery data alone. Thus the orbit determination problem cannot be disentangled from the **identification** problem, that is to find the independent discoveries referring to the same physical object: only by joining the information, contained in such separate discoveries, we can gather enough data for a solution. The output of the identification/orbit determination procedure is a *catalog* containing the list of distinct objects discovered, their

© in this web service Cambridge University Press & Assessment www.cambridge.org

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

best fit orbits, an estimate of their uncertainty and the little physical information available, in most cases just the absolute magnitude, a measure of the intrinsic capability of the object to reflect sunlight.[3]

The above example refers to *passive* observations detecting photons of reflected sunlight. Active observations are used in planetary radar observations, where a powerful beam of microwaves is directed towards a celestial body such as a major planet, a natural planetary satellite, an asteroid, a comet. At the present state-of-the-art, given that the signal-to-noise ratio at distance r is proportional to $1/r^4$, only the major inner planets, some very large satellites (e.g., Titan), and large asteroids can be observed by radar at interplanetary distances. Most of the targets therefore are **near-Earth asteroids**, which have the possibility of comparatively close approaches to the Earth.[4] Radar observations are a complex subject, because the radar return signal contains photons reflected from different parts of the asteroid surface, each with a different range and range-rate with respect to the radar antenna. In fact, the radar astrometry data are normal points obtained from a large fit providing also information on the size, shape, radar reflectivity, and rotation state of the object. The information constraining the orbit can be synthesized into an equivalent observation of range and range-rate. The accuracy of radar astrometry is between two and three orders of magnitude better than conventional astrometry.

The above examples are about natural bodies, but a very similar problem is obtained by considering spacecraft whose operational life is over. They can be observed in a non-collaborative way, with exactly the same techniques as asteroids, that is by astrometry and by radar. In most cases, however, these observations do not allow us to discriminate one dead spacecraft from another (actually, some care needs to be used to identify among the observations the ones belonging to operational spacecraft). As the search for this **space debris** progresses towards smaller and smaller Earth-orbiting objects, the list of bodies increases by adding spent rocket stages, pieces of exploded satellites and rocket motors, screws, bolts, and small pieces released during stage separation and antenna deployment, as well as particles of fuel, of frozen cooling liquid, all kinds of trash. A current estimate places at about 350 000 the number of orbiting debris above 1 cm of diameter. Thus the space debris problem is a population orbit determination problem, and surveys have to be set up to compile catalogs of all the particles above

[3] The absolute magnitude gives an indication of the diameter and mass, but the correspondence between these quantities contains unknown parameters such as the albedo and the density.

[4] With the current technology, radar astrometry for small asteroids (diameter < 1 km) is possible up to a distance of 0.2–0.3 AU.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-0-521-87389-5 — Theory of Orbit Determination
Andrea Milani, Giovanni Gronchi
Excerpt
More Information

a given size. The analogy is striking because there is an impact monitoring problem: the objects larger than a few mm could seriously damage the International Space Station by colliding at a relative speed of several km/s.

Thus the specific properties of the population cases are three, and they are opposite to those of the collaborative case.

First, the number of observations is not under our control. A survey can be designed to obtain a very large number of observations, but unavoidably the larger the data set, the larger the set of distinct objects for which the orbit has to be determined. Thus the average number of observations per object is small, typically of the order of 10.

Second, the batches of observations which can be immediately assigned to a single object are not enough to compute an orbit, thus the identification problem needs to be solved before orbit determination is possible. On the other hand, an identification can be considered reliable only if an orbit can be consistently fit to all the data believed to be of the same physical object. Thus orbit determination and identification are just a single algorithm, necessarily complex.

Third, the dynamical and kinematical parameters are normally not determined. After the reliable identifications have been established, each orbit can be solved individually, fitting just $N = p = 6$ parameters. Additionally, a separate fit of the photometric data can provide the absolute magnitude. However, this has to be performed for millions of bodies.

Planetary systems

There are a few examples of orbit determination which do not fit well into the binary classification collaborative/population. Interesting examples are the **planetary systems**. There are two main cases.

Our Solar System contains a small number N_P of planets.[5] The equation of motion for the planets needs to take into account the perturbations from the other planets, relativistic corrections, the perturbations from the larger satellites (especially the Moon), and the larger asteroids. The masses of the major planets appear as dynamical parameters $\boldsymbol{\mu}$, together with the post-Newtonian parameters describing general relativity effects.

Thus the orbits of the planets have to be determined all at once, including

[5] The exact definition of planet has been controversial, e.g., Pluto has size and mass comparable to those of other trans-neptunian bodies previously classified as minor planets, and it is significantly smaller than some satellites such as the Moon, Ganymede, and Titan. What matters in our discussion is the number of bodies whose masses are large enough to produce observable perturbations in the orbits of other planets, as discussed in Section 4.6; for the current accuracy in astrometric observations Pluto does not need to be included.

© in this web service Cambridge University Press & Assessment www.cambridge.org