Universität der Bundeswehr München
Fakultät für Luft- und Raumfahrttechnik
Institut für Raumfahrttechnik und Weltraumnutzung

Object correlation and orbit determination for geostationary satellites using optical measurements

Jan Siminski

Vollständiger Abdruck der bei der Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr.-Ing. Ferdinand Svaricek
1. Gutachter:	Prof. Dr.-Ing. Felix Huber
2. Gutachter:	Priv.-Doz. Dr. rer. nat. Oliver Montenbruck
3. Gutachter:	Prof. Dr. phil. nat. Thomas Schildknecht

Die Dissertation wurde am 21. April 2016 bei der Universiät der Bundeswehr München eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am 20. Oktober 2016 angenommen. Die mündliche Prüfung fand am 20. Oktober 2016 statt.

Abstract

Due to the increasing threat of space debris to active satellites, space object catalogs must be build up and maintained. Accurate and up-to-date catalog data allow the prediction of close conjunctions and consecutively the planning of avoidance maneuvers.

Objects in lower altitudes are typically observed with radars. Objects in higher orbits, such as the geostationary one, are observed with optical telescopes due to the limiting range capabilities of radars. The geostationary orbit is extensively used in various applications, e.g. communication, navigation, and weather monitoring. Hence, it needs special protection and should be regularly scanned in order to assure accuracy and currentness of data. The German Space Operations Center builds up a ground-based telescope network for that purpose and to protect its own assets.

Due to the limited number of telescopes which regularly scan the complete region around the geostationary orbit in a limited observation time, each object can only be observed for a short duration. The possible observation time is limited by the length of the night and visibility constraints such as clouds covering the sky.

The short observation arcs, so-called tracklets, must be either associated to already cataloged objects or used for an initial orbit determination to create new catalog entries. However, the tracklets contain incomplete state information and cannot be used alone to calculate an orbital solution. Hence, they are correlated against other tracklets, i.e. tested if they belong to a common object. After a successful association, the tracklet pairs provide enough information to determine the full orbital state.

This thesis develops a methodology to perform this association and initial orbit determination using the available information of two tracklets, namely the line-of-sight and its time derivative. The method is formulated as an initial-value and boundary-value problem corresponding to the two orbital representations. The association is performed by minimizing a loss function, which describes the statistical distance between the measured arc and the combined orbital solution. If the minimum distance is below a threshold, the two tracklets are associated.

The sensitivity to errors is compared for both formulations, where the symmetric boundary-value formulation is shown to use the available information better and is generally more robust than the initial-value formulation. It offers an easier calibration, i.e. a threshold can be defined, using a set of observations, which assures that a certain percentage of the measurement distribution is successfully associated. Additionally, the expected accuracy of the found solutions is assessed and shows promising results as well. Lastly, an observability analysis is performed to find out which re-observation time is advantageous for the successful association.

Keywords: space situational awareness, space debris, data association, initial orbit determination

Zusammenfassung

Die Gefahr durch die steigende Anzahl an Trümmerteilen in den Erdumlaufbahnen begründet den Aufbau und Erhalt eines Kataloges. Die genaue Erfassung von Objekten in der Erdumgebung ermöglicht dann die Vorhersage von gefährlichen Annäherungen zwischen Satelliten und Trümmern und die Planung der nötigen Ausweichmanöver.

Objekte in niedrigen Bahnen können von bodengebundenen Radarstationen detektiert werden. Wegen der begrenzten Reichweite von Radarstationen werden weiter entfernte Objekte, wie z.B. auf der geostationären Umlaufbahn, stattdessen mit optischen Teleskopen beobachtet. Aufgrund seiner besonderen Lage stellt der geostationäre Orbit eine besonders schützenswerte Ressource da, die unverzichtbar für den Betrieb von Kommunikations-, Wetter und Navigationssatelliten ist. Das Deutsche Raumfahrt-Kontrollzentrum baut deswegen ein Teleskopnetzwerk auf, um die Umgebung des geostationären Orbits präzise und zeitaktuell zu erfassen und die eigenen Satelliten zu schützen.

Da der gesamte Bereich um die geostationäre Bahn von wenigen Teleskopen abgesucht wird und gleichzeitig die Beobachtungszeit begrenzt ist, können nur kurze Messbögen von jedem Objekt gesammelt werden. Die nächtliche Beobachtungszeit variiert je nach Jahreszeit und wird zusätzlich durch schlechte Beobachtungsverhältnisse (z.B. Wolken) reduziert.

Die Messungen werden dann entweder einzelnen Objekten im Katalog zugeordnet oder sie erfordern die Erstellung von neuen Katalogobjekten nach vorangehender Erstbahnbestimmung. Da die kurzen Bögen meist nicht für die Bestimmung aller Bahnelemente ausreichen, müssen mehrere Bögen miteinander kombiniert werden. Dazu müssen sie jedoch vorher mit anderen Messbögen korreliert werden, d.h. es wird getestet, ob zwei Messungen zu einem Objekt gehören. Wenn dieser Test erfolgreich ist, kann eine gemeinsame Bahn berechnet werden.

Die Zuordnung der einzelnen Messungen und die dazugehörige Erstbahnbestimmung wird in dieser Arbeit als Optimimierungsproblem formuliert und numerisch gelöst. Dabei werden zwei mögliche Ansätze jeweils mit einer anderen Parametrisierung des Problems verfolgt. Einmal wird die Bahn über die Randwerte und einmal über einen Startwert parametrisiert. Für die Zuordnung wird eine statistische Distanz zwischen beobachteten Messungen und gemeinsamer Bahnlösung minimiert. Wenn das Ergebnis der Minimierung unter einem Schwellwert liegt, gehören die Beobachtungen zueinander.

Anschließend werden die beiden entwickelten Methoden mit einer Sensitivitätsanalyse verglichen. Dabei stellt sich heraus, dass die Randwertformulierung robuster ist und die Informationen aus den Messungen besser nutzt. Die Methode kann mit Messdatensätzen kalibriert werden, d.h. es wird sichergestellt, dass ein definierter Prozentsatz der Messungen erfolgreich zugeordnet werden kann. Zusätzlich wird die Genauigkeit der Bahnlösungen untersucht. Die Resultate sind vielversprechend und ermöglichen eine erfolgreiche Katalogisierung. Zuletzt wird noch eine Beobachtbarkeitsanalyse durchgeführt, welche optimale Zeitpunkte zur Wiederbeobachtung berechnet und damit eine erfolgreiche Zuordnung erleichtert.

Schlagwörter: Weltraumüberwachung, Weltraumschrott, Korrelation, Erstbahnbestimmung

Contents

1. Introduction

Satellites are a valuable and important infrastructure with impacts on various applications and services. Uncontrolled defunct satellites, upper stages, and other debris populating the Earth-bound orbits pose a threat to active satellites. The debris is caused by explosions and collisions but also due to mission related practice. That is, leaving rocket bodies of upper stages behind and dropping covers and other material after placing the satellite in its destined orbit.

Comprehensive and frequently updated databases of all possibly hazardous objects orbiting the Earth are required to guarantee safe operation in the future. This thesis deals with one of the major tasks during the build-up and maintenance of such a catalog, namely the correlation of observations. Observation correlation describes the process of finding a positive or negative relationship between different observations or between new measurements and cataloged objects (Everitt and Skrondal, 2010, pp. 23, 107). If the relationship is positive, the observations and catalog entries originate from the same observed object. The process is also denoted as data association (Bar-Shalom et al., 2004) or linkage (Milani et al., 2011). The successful association of observations increases the accuracy of the cataloged states and thus allows their operational use.

The orbits around Earth are commonly classified into different types, where the most common ones are the low, medium height, and the geostationary orbit. Each type is used for different applications due to the special characteristics of each domain. Likewise, each domain requires different sensors and surveillance methods. While lower orbits can readily be observed with radars, this is not the case for the much higher geostationary orbit due to the limited range capability of most radars. Hence, higher orbits are observed using optical telescopes. Optical observations come with certain challenging properties, such as missing information on the distance and a high sensitivity to bad viewing conditions. These particular challenges are the motivation of this work, i.e. the special case of correlating optical observations with each other and obtain an orbital solution.

This introduction will extend this motivation by first listing benefits and uses of the geostationary orbit. After highlighting its importance and the threat by debris, current space surveillance efforts are outlined. The already mentioned research task is explained in more detail with a comprehensive literature review. The review compares the different published approaches along with their advantages and shortcomings. This helps to identify and also justify the contribution of this work. Lastly, the general structure of the thesis is described with a brief summary of all chapters.

1.1 Geostationary satellites

The geostationary orbit is an equatorial, circular, and geosynchronous orbit at approximately 35,786 km altitude. The special characteristics of this orbital height has been already identified by H. Potočnik (Noordung, 1929). An object in this orbit revolves around Earth in one sidereal day, i.e. the same period as the Earth is rotating around itself in space-fixed coordinates (around 23 hours, 56 minutes, 4.09 seconds). This configuration causes an unique geometry: the satellites remain at a fixed location on the sky for an observer on Earth.

Three important applications of this orbit type are:

1. *Telecommunication*: The constant position w.r.t. to an on-ground observer allows a simple fixed antenna pointing. The benefits of this geometry were already described by Clarke, (1945). It enables uninterrupted data transfer with no need of antenna steering. Consequently, transmitting and receiving hardware is cheaper. Multiple satellites distributed around the Earth provide a global data coverage. Geostationary communication satellites are extensively used by commercial but also military operators. Companies like Inmarsat and Thuraya provide voice and data communication to special satellite phones or transceivers. They offer access to the internet and other services in remote places (e.g. in maritime environment) or in regions lacking of infrastructure (e.g. rural areas). Commercial companies such as SES ASTRA, Eutelsat, Intelsat, and others broadcast television and radio stations.

2. *Meteorology*: The geostationary orbit also offers a fixed view from space on the Earth surface (illustrated in Figure 1.1). Almost a complete hemisphere is visible from an orbit with this altitude. This allows frequent observations of the atmosphere without the need to wait for the next satellite pass over the region of interest. The atmospheric data collected by meteorological satellites, such as Meteosat (see Figure 1.2) or GOES, are the source of information for weather forecasting methods. They predict the daily weather but also catastrophic disasters (e.g. hurricanes). Based on the predictions the local population can be warned and evacuated from critical areas.

3. *Navigation*: Global navigation satellite systems (typically located on medium Earth orbits around 20,000 km altitude) supply geospatial positioning for public and military services. They are used for navigation in, amongst other domains, maritime, aviation, and automotive environment. So called augmentation systems (e.g. EGNOS, WAAS, MSAS) provide external information to assure accuracy, reliability, and availability of such systems (Kee et al., 1991). The satellites are positioned in the geostationary orbit in order to guarantee constant access to this data. Additionally, navigation satellites orbiting on inclined geosynchronous orbits increase the visibility locally (e.g. the Japanese QZSS system) while also providing augmentation services. Geosynchronous orbits share the same revolution period with geostationary orbits but do not stay stationary at the sky. They stay in a fixed longitude range but also reach higher latitudes, which improves the observation conditions. Circular geosynchronous orbits intersect with the geostationary ring and are consequently also monitored when protecting the geostationary domain.

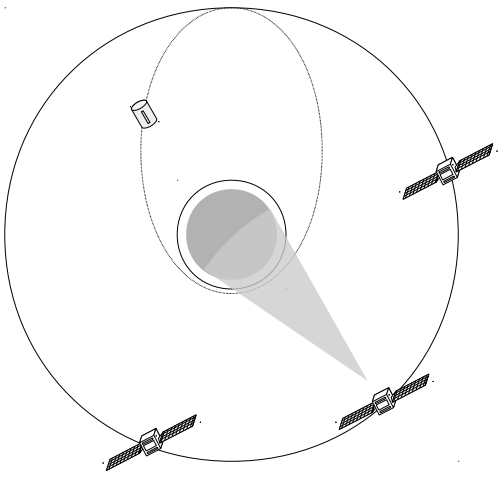


Figure 1.1: *Illustration of three geostationary satellites and an upper rocket stage on an eccentric geostationary transfer orbit. Due to the same rotation period, the broadcasting beam of the satellite (illustrated in light gray) constantly covers the same regions/hemisphere.*

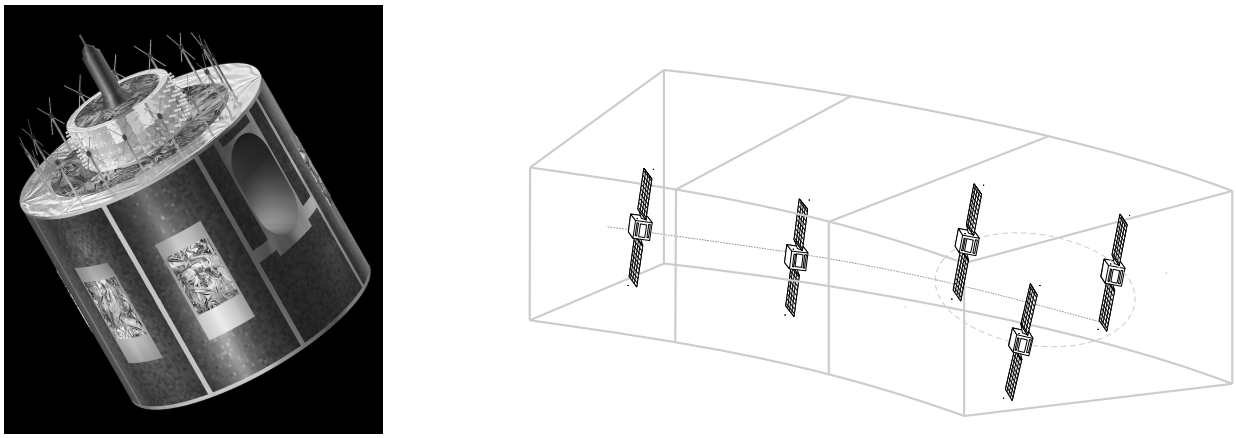


Figure 1.2: *The left side shows an artistic view of a second generation Meteosat satellite, developed by ESA and EUMETSAT (copyright by ESA and Ducros). The right side illustrates the close proximity operations in geostationary slots.*

To summarize the list: geostationary satellites are an essential infrastructure required by various applications and services. That is why the orbit is considered a limited natural resource, comparable to frequency bands, and is similarly organized and restricted by the International Telecommunications Union (ITU). The ITU organizes the World Administrative Radio Conference and since 1977 allocates and assigns longitude slots (cf. Soop, (1994)). The size of the assigned slots is typically 0.1° in longitude and latitude, which corresponds to around 100 km in along-track and cross-track direction. Satellite operators fly multiple satellites co-located within one slot for redundancy and better coverage (Dorsey et al., 1986). The situation is illustrated in Figure 1.2. The extensive exploitation of the geostationary orbit and the resulting dense population causes challenging difficulties for the correlation tasks performed in this work. The problem will be analyzed in Chapter 9.

The object density and growing interest in the orbit is illustrated in Figure 1.3. The left side shows the annual increase in launches, while the right side shows the geostationary domain as seen from Earth. The latter illustrates the density along European longitudes. The density of objects is considerably smaller than for the lower orbits (even considering the limited size of the region) and hence also the collision risk is smaller. Refer to the book by Klinkrad, (2006) and the text by Kessler, (1994) for density estimates, risk analyses, and growth models. The collision probability specifically for the geostationary domain is provided by Flury, (1991). Even if two objects collide the resulting cloud of fragments will not be as large comparably to events on lower orbits considering the same mass due to smaller relative velocities. In contrast to lower orbits, however, objects and defunct satellites will not automatically decay. The decelerating effect of the atmosphere acts as a natural sink in lower altitudes and causes the satellite to re-enter and eventually burn up in the atmosphere or hit the Earth's surface. Additionally, objects in lower orbits perform end-of-life maneuvers to accelerate the decay. Such maneuvers are economically unfeasible for higher orbits and thus retired satellites are placed in disposal orbits instead. In the disposal strategy described by Flury, (1991), satellites are typically placed a few hundred km above the geostationary ring. Re-entry maneuvers are unfeasible as too much propellent w.r.t. to object mass is required to reach the Earth atmosphere. Various ideas are proposed to remove hazardous objects from orbits. In the geostationary orbit defunct satellites are pushed away to the disposal orbits or repaired if possible (Hirzinger et al., 2004). The robotics involved in the servicing must be mature enough to guarantee a success. Otherwise even more debris will be created by the mission.

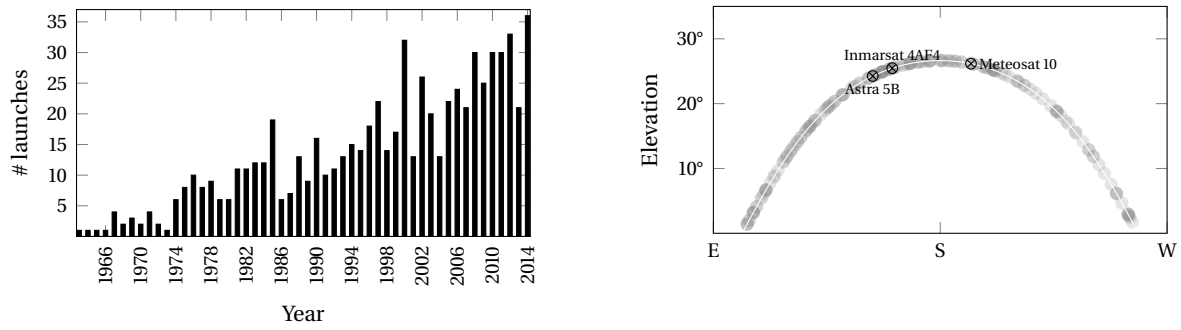


Figure 1.3: *The number of launches per year of cataloged geostationary satellites (left). Data is obtained from the publicly available two-line element catalog on space-track.org. The right side shows active geostationary objects as seen from Central Europe (South view)*

Operators regularly have to maneuver their satellites to safer orbits when facing close-proximity approaches with debris. Maneuver decisions are taken considering the collision probability. Maneuvers waste fuel and thus shorten the life-time of satellites. The German Space Operations Center (GSOC) got alerted by 21 encountering objects in the years 2011–2013 for their geostationary satellites and had to perform one avoidance maneuver in 2012 to circumvent a collision with a non-operational satellite (see summary by Aida et al., (2014) for more details). Large uncertainties attached to the cataloged debris states unnecessarily increase the false alarm rate. Precise orbits of objects must be therefore known in order to detect dangerous close approaches and evaluate the realistic collision probability.

Space surveillance

The US Joint Space Operations Center (JSpOC) maintains a database of space objects and alerts operators whenever a close conjunction with other spacecrafts or debris is predicted. The US space surveillance system, consisting of radar facilities as well as optical telescopes, delivers the necessary data to build up the catalog. JSpOC provides so-called two-line elements to the public through the *space-track.org* platform. However, the provided elements do not describe the actual orbit with sufficient accuracy as an approximate low-fidelity force model is used to represent the motion of the satellites. Furthermore, the orbital data comes with no information about the uncertainty. This hinders critical probability-based decisions such as maneuvers.

Russia operates a similar network within their space control system. The international surveillance sensors are summarized and analyzed by Weeden, (2015, pp. 990-997). The International Scientific Optical Network (ISON), headed by the Keldysh Institute for Applied Mathematics in Moscow, tracks near-Earth objects using telescopes distributed all over the world (Molotov et al., 2008). The institute recently started to publish orbital data of objects in high altitudes together with the Interstate Joint-Stock corporation “Vympel”.

GSOC collaborates with AIUB to build up a telescope network in order to protect the own assets in the geostationary region. This will allow to plan maneuvers independently of external information. The own catalog will contain precise orbits but also the uncertainty information necessary to perform decisions. The first telescope will be placed in Sutherland, South Africa, and covers the European longitude range together with the ZimSMART telescope located in Zimmerwald. The telescope setup is explained in detail in Chapter 3. For more details and first experience refer to (Fiedler et al., 2015). Both telescopes are shown in Figure 1.4.

The telescopes are placed on the northern and southern hemisphere in order to improve the visibility

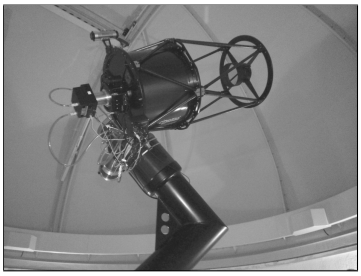



Figure 1.4: *Left photograph shows the ZimSMART telescope located in Zimmerwald, Switzerland (copyright by AIUB). Right one shows preliminary setup of the GSOC telescope which will be placed in Sutherland.*

conditions. As the nights are longer in Winter the possible overall observation time is increased. Visibility is not only constrained by the length of the night but also by the weather and other factors. The dependency on good visibility conditions is a major difference between optical and radar observations. The latter can be performed even with clouds in the way and during night and day. Objects in lower orbits can be consequently observed multiple times a day (depending on altitude) even by one sensor. Due to limited resources, i.e. a limited number of observation sites, and the sensitivity to good visibility conditions, observation time is constrained. When maintaining a complete database, every object position must be updated frequently. Hence, the complete orbital domain must be observed on a regular basis. Considering the limited time, only short sequences of measurements, called tracklets, can be collected for each object. Optical tracklets are formed by linking object detections from subsequent frames with each other. Due to the short time gap between the frames, the linking can be achieved with a reasonable confidence level. The information in individual tracklets is insufficient for initial orbit determination. This necessitates a need to associate multiple tracklets and constitutes a key motivation for the research in this thesis.

1.2 Research objective

When surveying the geostationary orbit, each object is observed for short durations (a few seconds up to a few minutes). This provides good knowledge on the line-of-sight and its approximate time derivative, or angles and angular-rates respectively. The latter describes the velocity of the satellite perpendicular to the line-of-sight axis. The angular acceleration of the object, i.e. a change in the rates, can not be accurately observed if only a small fraction of the orbit is covered.

The objects observed in the collected images are unknown, unless they are purposely tracked with prior information on the orbit. Consequently, the observations must be correlated with already existing catalog entries or they must be used to create new catalog objects. The latter task is commonly denoted as initial orbit determination. Initial orbit determination using optical observations is in principle a well known problem and has already been performed centuries ago for celestial objects, e.g. planets and asteroids. The traditional methods developed by Gauss and Laplace can be applied to objects on Earth-bound orbits. But they require a sufficiently long observation arc with well distributed measurements. If that is not fulfilled, finding an orbital solution becomes an ill-conditioned problem, i.e. small errors in the measurements lead to completely different orbits. This makes the classical methods practically unusable for short-arc tracklets.

As one tracklet does not provide enough information, the principle task of this thesis is to combine the information of two tracklets to obtain an orbital solution. This task comprises two steps. The observation arcs are correlated with each other, i.e. tested if they originate from a common object. In case they do, the combined solution must be found.

1.3 Literature review

Many research ideas and approaches for tracklet association originate from the asteroid observation community. They are faced with the same issues, namely short observation arcs. Due to the large revolution periods of asteroids, the collected arcs are short even when observing an object repeatedly for a couple of days. Milani et al., (2004) developed the principle of admissible regions, where they restrict the space of orbital solutions to the physically meaningful ones. An asteroid orbit, for instance, should have a stable elliptic orbit around the Sun. Virtanen et al., (2001), also in the framework of asteroid orbit determination, tackled the issue of short arcs by generating hypotheses directly from two observations. This boundary-value approach, called statistical ranging, uses the angular information at both epochs for a combined orbit solution.

A considerable amount of literature has been published since then on the issue of optical measurement correlation for objects on Earth-bound orbits. Tommei et al., (2007) applied the admissible region theory to orbits of space debris. Maruskin et al., (2009) and Fujimoto et al., (2014b) developed a methodology based on the same principle. Schumacher et al., (2013), similarly to Virtanen et al., (2001), used the boundary-value representation to initiate candidate solutions. The individual approaches originate from using different sensors, observation strategies, and assumptions and consequently each come with advantages and disadvantages. The different methods are categorized and discussed in the following. The discussion follows the presentation in (Siminski et al., 2014a).

The key issue, which comes up when using short-arc tracklets in the association, is the lack of information. The published methods hypothesize free parameters along the undetermined dimensions of the problem. The methods differ, however, in the way the orbit is represented and which testing quantity is used to decide acceptance. The latter quantity is denoted here as the discriminator. As indicated above, the two different representations are the boundary-value and initial-value formulation respectively.

In the initial-value formulation the full information of the first tracklet is used, namely right ascension α and declination δ and their respective rates $\dot{\alpha}$ and $\dot{\delta}$. The angular observations are formally introduced in Chapter 3. The information of the second tracklet is then used as a discriminator. In contrast, the boundary-value formulation takes the line-of-sight information of both epochs to generate the hypothetical states. Both formulations are summarized in Table 1.1, where the discriminators as well as the hypothesized variables are shown.

Table 1.1: *Overview of the current tracklet association approaches adapted from (Siminski et al., 2014a). More details, further approaches, and references are given in the text.*

Orbit representation	Parameters		Approaches	
Initial-value	State	$(\alpha_1, \dot{\alpha}_1, \delta_1, \dot{\delta}_1)$	Grid testing	Tommei et al., (2007)
	Hypotheses	$(\rho_1, \dot{\rho}_1)$	Hyperplane intersection	Maruskin et al., (2009)
	Discriminator	$(\alpha_2, \dot{\alpha}_2, \delta_2, \dot{\delta}_2)$	Optimization	Siminski et al., (2013a,b)
Boundary-value	State	$(\alpha_1, \delta_1, \alpha_2, \delta_2)$	Grid testing	Schumacher et al., (2013)
	Hypotheses	(ρ_1, ρ_2)	Analytical	Taff and Hall, (1977)
	Discriminator	$(\dot{\alpha}_1, \dot{\delta}_1, \dot{\alpha}_2, \dot{\delta}_2)$	Optimization	Siminski et al., (2014a)

1.3.1 Initial-value formulation

As shown in the Table 1.1, three different branches have developed from the initial-value formulation. The regular grid testing and hyperplane intersection are summarized in the following paragraphs. The

approaches and their relation are additionally illustrated in Figure 1.5. The formulating of the association as an optimization problem is one of the research contributions in this work and consequently derived in detail in Chapter 7.

Grid testing

Milani et al., (2004) and Tommei et al., (2007) suggest to restrict the range ρ and range-rate $\dot{\rho}$ space in order to allow only stable bound orbits around the Earth. This is achieved by confining the energy of the orbital solutions to be negative. The hypothesis space (ρ_1, $\dot{\rho}_1$) is sampled with so-called virtual debris particles using a triangulation scheme. The information of the second epoch, i.e. (α_2, $\dot{\alpha}_2$, δ_2, $\dot{\delta}_2$), is used together with a distance measure (attribution penalty) to decide whether the two tracklets are correlated or not. The distance measure compares the modeled measurement of the propagated particle with the actually measured one.

Similarly, DeMars and Jah, (2013) and DeMars et al., (2012) sample the admissible region of the first tracklet with multiple hypotheses on a regular grid to initialize a filter. The region is filled with a set of multivariate normal probability density functions, each consisting of a state, covariance and a weight. The combined density function, a so-called Gaussian mixture, is updated with each incoming measurement. Components which do not represent the new measurements loose weight and are eventually discarded. The surviving hypotheses are improved during the tracking process. Filters are typically used in tracking scenarios. The filters converge after observation arcs with a duration of about 10 minutes in a simulated example in (DeMars et al., 2012). Gadaleta et al., (2012) applied the methodology to the general association problem. They tested different distance measures for the association decision. The filtering approach is furthermore generalized to the multi-object association problem in (DeMars et al., 2015). Instead of representing each object state by a probability density, the whole domain is represented with a hypothesis density function. Consequently, multiple objects are treated simultaneously. This technique could be of special interest when collecting measurements of closely-spaced satellites, such as in clusters of co-located geostationary satellites.

Hyperplane intersection

Maruskin et al., (2009) define bounds on apoapsis and periapsis radii to prevent hypotheses to de-orbit within the next revolutions. Their approach, later refined by Fujimoto et al., (2014b), determines the common solution of two tracklets by computing the overlap of two solutions spaces. Hence, the admissible region is computed for each observation arc and filled with hypotheses. Instead of using the range and range-rate, they sample the Delauney (Maruskin et al., 2009) or Poincaré (Fujimoto et al., 2014b) orbital element space. Usage of orbital elements is advantageous as it provides a natural way of sampling the solution space. The hypotheses are then propagated to a common epoch and compared to each other, which effectively is an intersection of two hyperplanes. Fujimoto et al., (2014b) suggests to affirm or reject the intersecting hypotheses using a least squares fit.

1.3.2 Boundary-value formulation

Two basic approaches have been published which use the information of both observation arcs to define orbital states, namely grid testing and an analytical approach. The approaches are summarized in Figure 1.6. This work contributes with a formulation as an optimization problem. Chapter 7 provides the derivations and details.

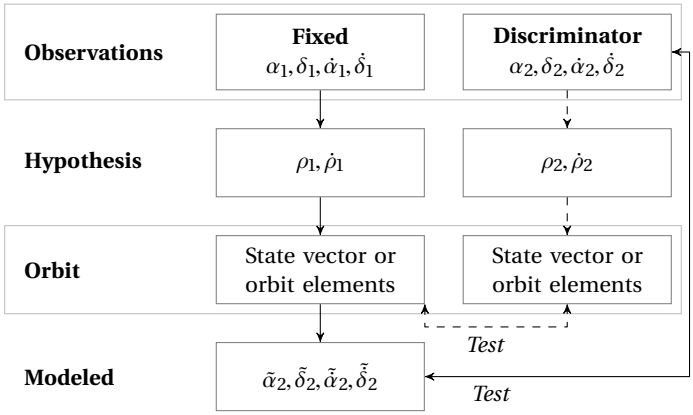


Figure 1.5: *Flow diagram of initial-value approaches: the solid line denotes the approach by DeMars et al., (2012) and Tommei et al., (2007), while the dashed line denotes the path taken by Fujimoto et al., (2014b) and Maruskin et al., (2009). The approach developed in this work follows the solid line as well.*

Grid testing

Schumacher et al., (2013) propose to use only the line-of-sight information if the angular-rates are not available within a sufficient accuracy. The poor accuracy of the angular-rates is caused by the used sensor or due to a short tracklet duration. Very short tracklets (duration in the range of tens of seconds) allow to observe objects multiple times in a night even from the same sensor. This leads to a better coverage of the orbit but also increases the necessary association efforts. If the angular-rate is poorly determined, e.g. for a tracklet duration of a few seconds, at least three tracklets are required to obtain a final orbit. The trade-off between multiple short against less frequent but longer tracklets is discussed in Chapter 3. Nevertheless, the rates can be used to define admissible regions accounting for the large uncertainty (see (Roscoe et al., 2013)). They sample a feasible range-range space (ρ_1, ρ_2) for two measurement arcs. All feasible range combinations create object candidates. These candidates are confirmed and refined with further measurements. Ansalone and Curti, (2013) present a method for an observer in a low-Earth orbit. They sample the feasible range space on a coarse grid and test each hypothesis with the individually measured angles of the two tracklets. The fitness of each hypothesis is then improved iteratively using a genetic algorithm. Schneider, (2012) performs so-called statistical ranging, similar to the approach for asteroids by Virtanen et al., (2001), to generate candidate solutions. He characterizes the error distributions via Markov Chain Monte Carlo and uses them to group tracks originating from the same objects.

Analytical association

Based on the work by Taff and Hall, (1977), Gronchi et al., (2010, 2011) developed an analytical approach to link the measurements. They use the conservation of energy and angular momentum to set up a high-order polynomial (e.g., order 48 degrees in the formulation of Taff and Hall, (1977), and 20 in the one of Gronchi et al., (2011)). The feasible roots of the polynomial then define the orbital solutions. Angles and angular-rates are used simultaneously to define the orbit. Hence, the formulations are also sensitive to errors in the angular-rates and can fail if they are given with insufficient accuracy. It should be noted that the term analytical is used here to categorize the approach, even though the root finding process for high-order polynomials requires numerical tools.

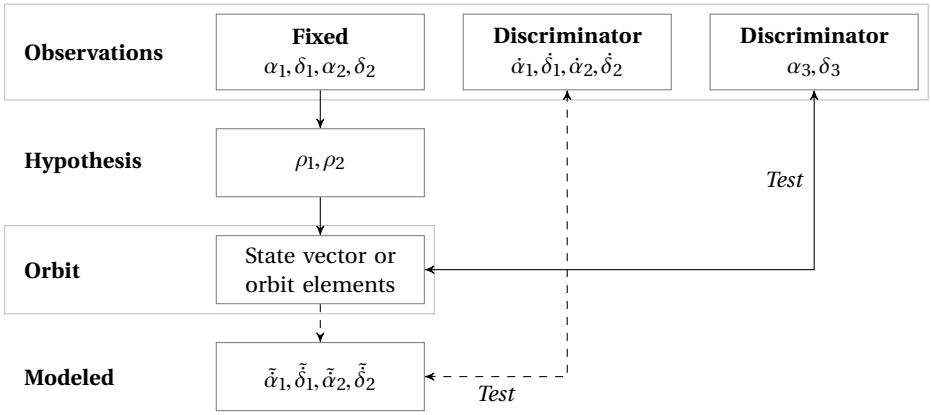


Figure 1.6: *Flow diagram of boundary-value approaches: Solid line denotes approach taken by e.g. Schumacher et al., (2013), while dashed line denotes path taken in this work.*

1.4 Research contribution

The overall goal of this work is to create a robust and fast method to perform the association task assuming tracklets with a duration of about a few minutes. The already published methods either assume a different arc length of the observations (e.g. no usable line-of-sight derivative and thus more than two tracklets are needed for an association decision) or a different observation scenario (e.g. tracking). Methods using a grid evaluation are prone to become computationally intensive or unreliable as the grid size must be very small to guarantee the detection of every target.

Instead of testing the complete admissible regions for feasible combinations, this research contributes by formulating the association as an optimization problem. The methodology is applied to both approaches, i.e., the initial value formulation and the boundary value formulation, and offers great advantages. In contrast to a fixed discretization grid, optimization methods adaptively reduce the step-size to a desired precision. Secondly, a well-posed optimization problem can reduce the computational burden of the association. Considering the large amount of collected tracklets per day in a global telescope network, this advantage becomes crucial when working with limited computational resources. The two formulations are shortly described here and discussed in more detail in Chapter 7.

In the initial-value formulation, the solved-for variables are range ρ_1 and range-rate $\dot{\rho}_1$. The optimized loss function which describes the association probability is computed by comparing modeled and measured observations. Additionally, a new coordinate system is introduced which simplifies the optimization problem. The coordinate system was presented in (Siminski et al., 2013b) and can also be used to efficiently sample the domain in a grid testing approach. As it uses the energy or semi-major axis to represent different solutions, it offers similar advantages as using an orbital element space. In order to find the minimum in a loss function, good starting values are required. Including the conservation of angular momentum into the problem helps to identify such starters.

The boundary-value formulation uses the ranges at both epochs (ρ_1 and ρ_2) as solved-for variables. The angular-rates are used in the loss function to decide whether measurements originate from common objects. Again, starters are obtained by including the angular-momentum conservation.

1.5 Organization of the thesis

The structure of the thesis is illustrated with the diagram in Figure 1.7. The first row of the figure names Chapters 2, 3, and 4 which provide a common terminology and the background theory required for later chapters, i.e. optical observations typically obtained when surveying the geostationary orbit, astrodynamics to predict and determine motion of satellites, and computational optimization methods. The second row continues with theoretical background knowledge based on the previous chapters. The terminology in statistics and data association is generally introduced and then applied to the application case of cataloging space objects. The orbit determination chapter introduces the general methods to find and improve a satellite orbit using the theory derived in the previous parts of the thesis and additionally recapitulates the shortcomings of traditional approaches and thus motivates the major contribution of the thesis in the next row, namely Chapter 7. The developed methods are then tested and assessed with the chapters depicted in the last row. A list with the detailed content of all chapters is provided in the following.

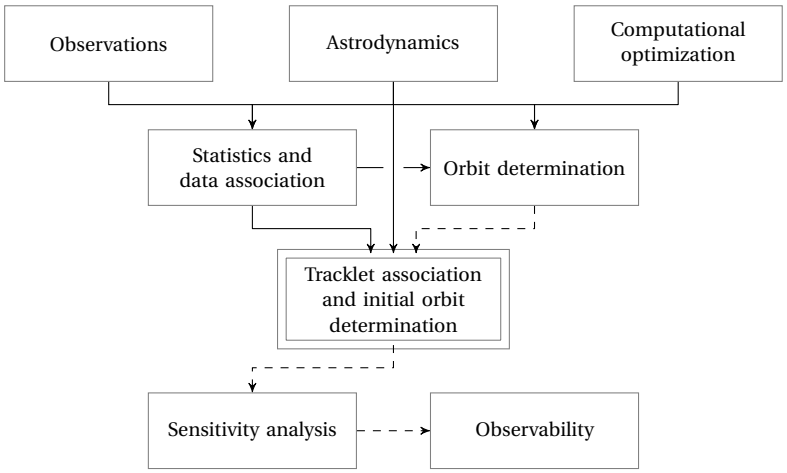


Figure 1.7: *Flow diagram of the thesis. The solid connections denote theory dependence, i.e. one chapter requires the knowledge from another. The dashed connections denote motivational dependence, e.g. a chapter discusses the results and challenges of the other.*

Chapter 2 ***Astrodynamics:*** The initial-value and boundary-value formulation in the tracklet association problem originate from the two corresponding orbital representations. This chapter provides the theoretical astrodynamics background and nomenclature required for the later derivations and discussions.

Chapter 3 ***Observations:*** The selection of an association methodology is coupled with the accuracy and frequency of the observations. The chapter describes visibility conditions and observation strategies. A sensor network simulation is performed in order to analyze the typical measurement distribution, i.e. obtain realistic observation geometries and time gaps between observation epochs. Additionally, the information content of the obtained measurements is analyzed.

Chapter 4 ***Computational optimization:*** The research contribution consists of formulating the correlation as an optimization problem. This chapter introduces a common nomenclature and explains various used computational optimization methods. Additionally, typical difficulties which arise in optimization problems (and also arise in this work), are highlighted and methods to overcome them are presented.

Chapter 5 ***Statistics and data association:*** Association decisions are based on statistical methods and assumptions. The chapter introduces the statistical nomenclature and theory and applies it to the correlation of new observations to catalog objects. It points out difficulties which arise when dealing with uncertainty transformations. The difficulties apply to the tracklet association problem as well and thus become important again in later chapters.

Chapter 6 ***Orbit determination:*** The basics of statistical orbit determination methods are introduced as they form the theoretical foundation of the association method. Classical initial orbit determination methods are explained along with their limitations in the presented observation scenario. Some difficulties are also handled by introducing restricting assumptions, which can be a valid choice in some situations and are presented for completeness.

Chapter 7 ***Tracklet association and initial orbit determination:*** The novel association and initial orbit determination method is described using the initial- and boundary-value formulation. Difficulties appearing in the loss function minimization are discussed and tackled. This includes the proper selection of a coordinate system, starting values, and constraints.

Chapter 8 ***Sensitivity analysis:*** The presented methods are analyzed in terms of association performance and accuracy. For the first quantity, a set of measurements with known associations is used to compute the rate of successful associations. Simulated observations are used to assess the accuracy of the initial orbits depending on observation geometry.

Chapter 9 ***Observability:*** The geometry dependency of the association performance is systematically tested using a simplified case study. The outcome of the study identifies shortages of current observation strategies and suggests possible improvements.

Chapter 10 ***Conclusions:*** The final chapter gives a short summary of the results and a recommendation of how the methods should be implemented in a operational cataloging system. The final section is devoted to an outlook to possible future research tasks.

2. Astrodynamics

This chapter gives an overview on the dynamics in celestial mechanics which describe the motion of satellites and debris around the Earth. It summarizes common textbook knowledge in the field of astrodynamics (see e.g. (Beutler, 2004; Montenbruck and Gill, 2000)) and mainly serves to consistently introduce relevant terms and concepts for the subsequent chapters of this thesis. Readers familiar with the subject may readily skip this part on first reading. Orbital motion models are included in the preliminary orbit estimation process as well as in orbit improvement methods. The underlying physics define conditions which restrict the solution space of orbits for newly detected objects. Classical methods assuming simplified dynamics (such as two-body motion) help to reduce the complexity of otherwise challenging estimation problems.

The orbital motion of satellites is described in terms of different accelerating forces. The accelerations acting on satellites in the vicinity of the Earth are listed in the following equation

$$\ddot{\boldsymbol{r}} = \boldsymbol{F}(t, \boldsymbol{r}, \dot{\boldsymbol{r}}) = \boldsymbol{F}_{\oplus} + \boldsymbol{F}_{\odot} + \boldsymbol{F}_{M} + \boldsymbol{F}_{D} + \boldsymbol{F}_{S} \ \ldots \,, \tag{2.1}$$

where $\boldsymbol{F}_{\oplus}$ denotes the acceleration due to the gravitational potential of the Earth, $\boldsymbol{F}_{\odot}$ and $\boldsymbol{F}_{M}$ the tidal accelerations by Sun and Moon, $\boldsymbol{F}_{D}$ is caused by the atmospheric drag, and $\boldsymbol{F}_{S}$ due to solar radiation pressure. Depending on the required precision, further perturbing terms can be included (denoted by the dots in the equation).

If only two-body dynamics are modeled, i.e. only the acceleration of a point-mass Earth is considered, the acceleration reduces to

$$\ddot{\boldsymbol{r}} = -\frac{\mu_{\oplus}}{\|\boldsymbol{r}\|^3}\boldsymbol{r} \,, \tag{2.2}$$

where $\mu_{\oplus}$ is the product of Earth's mass and the gravitational constant. Equation (2.1) (equivalently (2.2)) forms a set of three second order, non-linear, differential equations and requires double integration for its solution. It is commonly transformed into a set of six ordinary, first order, differential equations by introducing a new state variable

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{r} \\ \dot{\boldsymbol{r}} \end{pmatrix} , \tag{2.3}$$

which consists of the position $\boldsymbol{r}$ and velocity $\dot{\boldsymbol{r}}$. The new system of equations is then given by

$$\dot{\boldsymbol{y}} = \begin{pmatrix} \dot{\boldsymbol{r}} \\ \boldsymbol{F}(t, \boldsymbol{r}, \dot{\boldsymbol{r}}) \end{pmatrix} = \boldsymbol{f}(\boldsymbol{y}, t) \,. \tag{2.4}$$

In order to solve these equations, they need to be restricted to a certain orbit, i.e. to the orbit which represents the motion of a selected satellite or debris particle. The solution of the system is uniquely

defined by six constraints, thus an orbit is e.g. defined by one initial-value

$$\boldsymbol{y}(t_1) = \begin{pmatrix} \boldsymbol{r}_1 \\ \dot{\boldsymbol{r}}_1 \end{pmatrix} \tag{2.5}$$

or by two boundary-values

$$\boldsymbol{y}_{1:3}(t_1) = \boldsymbol{r}_1 \quad \text{and} \quad \boldsymbol{y}_{1:3}(t_2) = \boldsymbol{r}_2\,, \tag{2.6}$$

where the subscript denotes the constrained elements of the state vector. The two formulations inherit certain advantages and difficulties when working with them and will be used throughout this thesis. Additionally to the six independent parameters, force model parameters have to be defined. Depending on the complexity of the model, the mass and shape of satellite, drag and solar radiation pressure coefficients and others might be needed to compute the respective acceleration terms.

The full state required for the initial-value representation can alternatively be expressed in terms of six orbital elements. Most orbital elements describe the state as a location on an ellipse with its focal point in the center of the Earth. Classical Keplerian elements use the eccentricity e and semi-major axis a to define the size and shape of the ellipse. The orientation of the orbital plane is described with the inclination i and the ascending node Ω. The orientation of the ellipse within the plane (the orientation of the major axis) is fixed with the argument of periapsis ω. Finally, the true ν, the mean M, or the eccentric anomaly E define the location on the ellipse at a specific epoch. The true anomaly describes the angle between the current location and the periapsis. The mean anomaly describes the location of the object if it were on a circular orbit with equal major axis. The auxiliary angles (mean and eccentric anomaly) are calculated from the true anomaly and vice versa. This representation is useful, amongst other benefits, when propagating unperturbed orbits, as the solution to the equation of motion reduces to linear motion in the mean anomaly (see next section). Keplerian elements run into singularities when working with low inclination, circular orbits. The node Ω becomes undefined as it lies on the same plane as ω and ν, while the periapsis can not be determined if the radius of the orbit is constant. This difficulty arises, most importantly in the context of this research, when dealing with geostationary objects. Alternative orbital elements, which avoid these difficulties, can be used instead, e.g. equinoctial elements (see (Broucke and Cefola, 1972)). Soop, (1994) suggests to use special geosynchronous elements when studying geostationary objects. The transformation between the different orbital representations and further details can be obtained from (Montenbruck and Gill, 2000). The following two sections provide the methods to obtain an orbital solution from either two boundary-values or one initial-value.

2.1 Initial-value solution

Given the initial state, the state at any other epoch can be predicted by single integration. The following equation shows the integration for the cartesian state $\boldsymbol{y}$:

$$\boldsymbol{y}(t_2) = \boldsymbol{y}(t_1) + \int_{t_1}^{t_2} \boldsymbol{f}(\boldsymbol{y}(t),\, t)\,\mathrm{d}t \tag{2.7}$$

The numerical integration is performed with e.g. multistep Adam-Bashforth methods (practical implementation given by Shampine and Gordon, (1975)). Alternatively, it is solved analytically or semi-analytically using other orbital state representations (cf. Brouwer, (1959) and Kozai, (1959)).

A prominent analytical solution is given when restricting the dynamics to a two-body model. Important equations and properties of this orbital solution are recapitulated in the following from the book by Battin, (1999) in order to develop a common terminology. Details and intermediate steps can be obtained from the book.

Two conserved quantities are used in this work to restrict the solution space, namely the energy and the angular momentum. Starting with the latter, taking the time derivative of it, and inserting (2.2) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}\times\dot{\boldsymbol{r}}) = \dot{\boldsymbol{r}}\times\dot{\boldsymbol{r}} + \boldsymbol{r}\times\ddot{\boldsymbol{r}} = 0 \tag{2.8}$$

which demonstrates the conservation of angular momentum

$$\boldsymbol{h} = \boldsymbol{r}\times\dot{\boldsymbol{r}} = \text{const.} \tag{2.9}$$

The pointing of $\boldsymbol{h}$ describes the orientation of the plane and it is therefore also called inclination vector. Describing the position in spherical coordinates, gives Kepler's second law

$$\|\boldsymbol{h}\| = r^2\frac{\mathrm{d}\nu}{\mathrm{d}t} \tag{2.10}$$

which states that objects sweep out equal areas during equal intervals of time. Another conserved quantity, the eccentricity vector, is derived by taking the following derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}(\dot{\boldsymbol{r}}\times\boldsymbol{h}) = \ddot{\boldsymbol{r}}\times\boldsymbol{h} + \dot{\boldsymbol{r}}\times\boldsymbol{0} \tag{2.11}$$

which is again integrated and rearranged, using spherical coordinates, to

$$\boldsymbol{e} = \frac{1}{\mu_\oplus}(\dot{\boldsymbol{r}}\times\boldsymbol{h}) - \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|} = \text{const.} \tag{2.12}$$

and after squaring it with itself gives

$$1 - \|\boldsymbol{e}\|^2 = \frac{\|\boldsymbol{h}\|^2}{\mu_\oplus}\left(\frac{2}{\|\boldsymbol{r}\|} - \frac{\|\dot{\boldsymbol{r}}\|^2}{\mu_\oplus}\right). \tag{2.13}$$

The equation exposes the conservation of energy. The bracketed term on the right side is the reciprocal of the semi-major axis

$$a = \left(\frac{2}{\|\boldsymbol{r}\|} - \frac{\|\dot{\boldsymbol{r}}\|^2}{\mu_\oplus}\right)^{-1}. \tag{2.14}$$

Using the expressions for kinetic ($\frac{\|\dot{\boldsymbol{r}}\|^2}{2}$) and potential energy ($\frac{\mu_\oplus}{\|\boldsymbol{r}\|}$), a relation to the total energy per unit mass is derived as

$$\mathcal{E} = \frac{\|\dot{\boldsymbol{r}}\|^2}{2} + \frac{\mu_\oplus}{\|\boldsymbol{r}\|} = \text{const.} = -\frac{\mu_\oplus}{2a} \tag{2.15}$$

The geometric interpretation of a as the semi-major axis of the ellipse allows dividing the orbits into two categories: elliptic or circular bound orbits around Earth where $a > 0$ and $\mathcal{E} < 0$, and unbounded hyperbolic or parabolic orbits with $a \leq 0$ and $\mathcal{E} \geq 0$. Objects on latter named orbits will eventually escape Earth's sphere of influence and play no role when determining orbits of space debris particles. The energy

thus serves to discriminate between feasible Earth-bound orbits and unfeasible ones. Lastly, the rotation period is computed from the overall area covered by the orbit using (2.10) and inserting the semi-major axis and eccentricity

$$P = 2\pi\sqrt{\frac{a^3}{\mu_\oplus}} \tag{2.16}$$

which is then used to determine the rate of change $\sqrt{\frac{\mu_\oplus}{a^3}}$ of the mean anomaly M. Assuming two-body dynamics a linear equation to propagate the mean anomaly is derived:

$$M(t_2) = \sqrt{\frac{\mu_\oplus}{a^3}}(t_2 - t_1) + M_1\,. \tag{2.17}$$

The elements are then transformed back to the cartesian position and velocity vector if needed. Typically, the two-body solution deviates largely from the one using high-fidelity force models (especially for longer propagation periods). However, it can be used to determine initial-values for more complex methods, as done in the next section.

2.2 Boundary-value solution

The orbital boundary-value problem, also known as Lambert's problem, requires a different approach and cannot simply be obtained by previously mentioned integration methods. One possibility to solve the perturbed Lambert's problem is, to use so-called shooting-methods. The initial-value integration is used to propagate a first guess of the initial state $\tilde{\boldsymbol{y}}_1$

$$\tilde{\boldsymbol{y}}(t_2) = \tilde{\boldsymbol{y}}(t_1) + \int_{t_1}^{t_2} \boldsymbol{f}(\tilde{\boldsymbol{y}}(t),\, t)\,\mathrm{d}t\,, \tag{2.18}$$

where the known position $\boldsymbol{r}_1$ at the initial epoch is augmented with a first guess of the velocity $\tilde{\dot{\boldsymbol{r}}}_1$

$$\tilde{\boldsymbol{y}}(t_1) = \begin{pmatrix} \boldsymbol{r}_1 \\ \tilde{\dot{\boldsymbol{r}}}_1 \end{pmatrix}. \tag{2.19}$$

Now, a three dimensional root-finding problem is obtained, i.e. a velocity is searched which solves

$$0 \stackrel{!}{=} \boldsymbol{r}_2 - \tilde{\boldsymbol{r}}_2(t_2,\, \dot{\boldsymbol{r}}_1) \quad \text{where} \quad \tilde{\boldsymbol{r}}_2 = \tilde{\boldsymbol{y}}_{1:3}(t_2)\,. \tag{2.20}$$

The exclamation mark notation is used to indicate root-finding problems throughout this thesis. Numerical root-finding methods, e.g. the Newton-Raphson one, are used for that purpose. They can be obtained e.g. from (Press et al., 2007). The shooting-method requires a good guess for the velocity $\tilde{\dot{\boldsymbol{r}}}_1$ at the initial epoch. Otherwise, the numerical methods might not converge towards the wanted solution. The situation deteriorates with larger time intervals between the epochs as even small deviations from the true solution can lead to large differences in the position vector and hence divergence. More advanced methods are therefore required which robustly solve the boundary-value problem (e.g. multiple-shooting approaches or numerical continuation methods as used for initial orbit determination by Lenz et al., (2010) and Smith and Huang, (1986)).

A practical approach to determine a good initial solution is to consider only two-body dynamics. Equation (2.17) can be rearranged for the time-of-flight, i.e. the time interval between the two epochs

$$t_2 - t_1 = \sqrt{\frac{a^3}{\mu_\oplus}}(M_2 - M_1 + 2\pi n_{\text{rev}})\,, \tag{2.21}$$

which is dependent on the number of completed revolutions n_{rev}. The two-body boundary-value problem has potentially multiple solutions due to the ambiguity in the anomalies. Additionally, the direction of flight must be provided, i.e. whether the central angle between the two positions (the difference in true anomaly) is defined by

$$\Delta\nu = \arccos\frac{\boldsymbol{r}_1\boldsymbol{r}_2}{\|\boldsymbol{r}_1\|\|\boldsymbol{r}_2\|} \quad \text{or} \quad \Delta\nu = 2\pi - \arccos\frac{\boldsymbol{r}_1\boldsymbol{r}_2}{\|\boldsymbol{r}_1\|\|\boldsymbol{r}_2\|}\,. \tag{2.22}$$

In the first case $\Delta\nu$ lies between 0 and π. Consequently, the short arc around the orbital ellipse is taken between the epochs. The other direction is selected along the longer arc where the central angle is between π and 2π. In case of $\Delta\nu = \pi$ or $\Delta\nu = 2\pi$, the orbital plane cannot be determined. Additionally, the eccentricity is unobservable for $\Delta\nu = 2\pi$. It is convenient to use the number of completed half orbital periods k instead of n_{rev}. If k is odd, the central angle is between 0 and π, whereas if even, the angle is between π and 2π. The situation is illustrated in Figure 2.1. In order to determine all possible solutions, the problem must be solved for all feasible k integer values. The feasible interval $k = [1, \ldots, n_k]$ for a specific problem will be derived in Chapter 7 where the Lambert's problem solver discussed here will be used.

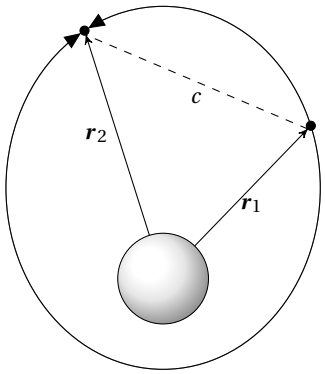


Figure 2.1: *Geometry of Lambert's problem. The short ($k = 0$) and long paths ($k = 1$) are illustrated. The dashed line is the chord joining both positions.*

The time-of-flight is expressed in terms of the semi-major axis a, the length of the chord joining the two position vectors

$$c = \|\boldsymbol{r}_2 - \boldsymbol{r}_1\| \tag{2.23}$$

and the sum of geocentric distances and chord length

$$2\,s = \|\boldsymbol{r}_1\| + \|\boldsymbol{r}_2\| + c\,. \tag{2.24}$$

The time-of-flight equation (2.21) essentially states a one-dimensional root-finding problem in the variable a. Multiple variations of expressing (2.21) using the aforementioned parameters have been developed (cf. Battin, (1999)). Typically, a change of variables is introduced for more efficient root finding, i.e. substituting a with better behaving variables. One advantage of other variables can be a bounded solution space and well-defined derivatives for Newton-Raphson or Halley's root finding. In this work, the implementation by Gooding, (1990) is used, which takes the formulation (variables) from Lancaster and Blanchard, (1969) and

includes some improvements (e.g. the selection of suitable starting values for the root-finding process). Following the derivations in (Lancaster and Blanchard, 1969), the time-of-flight function is given by

$$t_2 - t_1 = \sqrt{\frac{a^3}{\mu_\oplus}}(\alpha - \beta - (\sin\alpha - \sin\beta) + 2\pi n_{\text{rev}}), \tag{2.25}$$

where

$$\sin^2\tfrac{\alpha}{2} = \frac{s}{2a} \quad \text{and} \quad \sin^2\tfrac{\beta}{2} = \frac{s-c}{2a}. \tag{2.26}$$

Instead of using the semi-major axis, the alternative variable x with

$$x^2 = 1 - \frac{s}{2a} \quad \text{or} \quad x = \cos\tfrac{\alpha}{2} \tag{2.27}$$

is used as a parameter to find the root. It is bounded for elliptical orbits between −1 and 1.

Figure 2.2 shows the time-of-flight function dependent on x for an arbitrary geostationary satellite. As can be seen in the figure, Lambert's problem has exactly one solution when the satellite has not completed a full orbital revolution yet (i.e. $k = \{0,1\}$). However, after one completed revolution ($k \geq 2$), the measured time interval between the epochs (illustrated with the dotted line) can intersect the time-of-flight function in two x values. Depending on the geometry and the number of allowed revolutions, the two-body boundary-value problem can thus have different numbers of solutions.

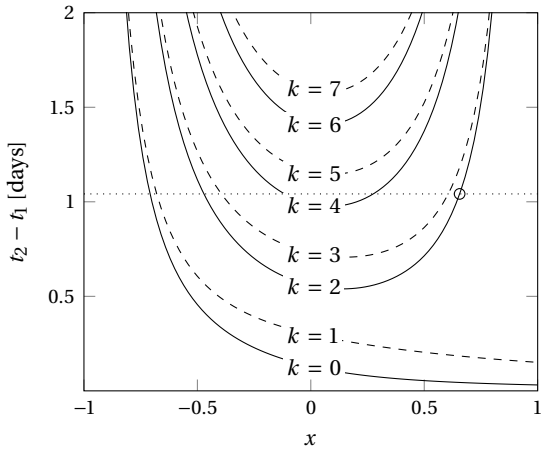


Figure 2.2: *Time of flight as a function of x for an arbitrary near-geostationary orbit. Dotted line shows the true time difference (25 hours). All intersections between the function and the dotted line mark possible solutions. The circle marks the true solution. Figure is based on (Gooding, 1990).*

3. Observations

Space surveillance tasks are performed using either active or passive observation methods. Active ones, e.g. radar or laser ranging, use two-way signal travel times to determine the topocentric distance along with optional pointing angles. Additionally, the range-rate, i.e. the velocity along the line-of-sight relative to the observer, can be measured with some of these techniques. As the signal has to travel the distance twice, the sensitivity of active sensors degrades with the fourth power of the distance. Consequently, when observing objects in high-altitude orbits, only large objects are observed or a strong transmitting signal is required. Phased-array radars are capable of simultaneously tracking multiple satellites by repeatedly scanning a wide field of view with short revisit times. They are hence used for surveillance tasks in the lower orbits while mechanical parabolic dish antennas, such as TIRA operated by Fraunhofer-FHR in Wachtberg, are used for individual object tracking. Radars have the advantage that they can be used during night and day and depend less on weather conditions. Laser ranging can only be performed under good visibility conditions and is commonly used to validate other measurement techniques. Nevertheless, due to their high precision, recent efforts have been made to utilize them in space surveillance (cf. Kirchner et al., (2013)).

Passive observation methods, in contrast, use the energy of external sources. Optical sensors capture the light which is emitted by the Sun and reflected by the observed objects. Considering a constant solar irradiance in the vicinity of the Earth, the sensitivity of optical sensors is only decreased by the distance squared. This makes optical telescopes suitable for detecting objects in high-altitude orbits. However, if no information about the objects shape and attitude is known, the sensors only provide a line-of-sight. Furthermore, optical observations require certain visibility conditions, i.e. objects must be in sunlight while the background must be dark. As a result, ground-based optical telescopes are only operated in the night and cannot observe objects inside the Earth's shadow cone. Moreover, the visibility is limited by clouds or other light sources in the field of view. But, telescopes are considerably less expensive than radars in installation and operation. In consequence, the near-geostationary orbit domain is typically scanned using ground-based telescopes. Space-based optical sensors have been proposed and first missions have already been launched or are currently planned (Norris, 2015, pp. 770-774). T. Flohrer et al., (2011) analyze

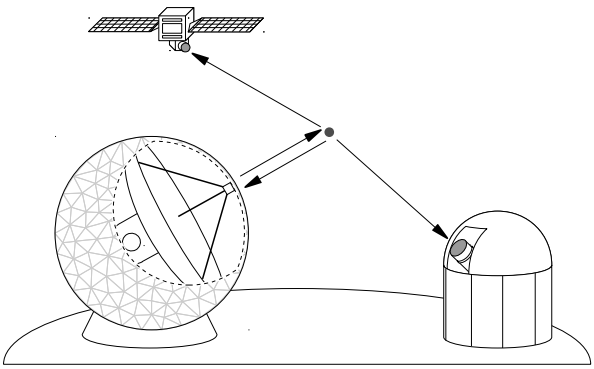


Figure 3.1: *Observation methods used for space surveillance tasks: a typical tracking radar (e.g. TIRA) with radome, an optical ground-based telescope (e.g. ZimSMART) and space-based optical sensors.*

the feasibility of such payloads and their contribution to a frequently updated catalog of geostationary objects. The different sensor systems are summarized and illustrated in Figure 3.1.

In the framework of this research, measurements obtained from ground-based sensors are discussed. As this work is motivated by the fact that GSOC builds up a telescope network (Fiedler et al., 2015), the expected sensor properties are used as a reference. The algorithms and derivations can be applied to other telescopes and also space-based observations with moderate adjustment. A discussion of optical observations and their information content is presented in the following sections. Different observation strategies and survey scenarios result in a diverse set of consequences. These are, the distribution of measurements along the orbit and in time, the accuracy, and the duration of the individual measurements arcs. It is important to understand the properties and limits of optical measurements in order to develop effective methods for orbit determination and measurement association. A more thorough discussion about the measurement generation is given by Schildknecht, (1994).

3.1 Telescope setup

The principle setup of an optical sensor system consists of a telescope and an attached detector. The detection is either performed by an integrating photon counting array (e.g. a charge-coupled device sensor) or an active-pixel sensor (e.g. using complementary metal-oxide-semiconductors) (Schildknecht, 2007). The aperture and focal length of the telescope define the field of view and the angular resolution which can be achieved by the detector pixels. Additionally, the pixel size, focal ratio, and the other parameters have a direct impact on the telescope performance, namely the signal-to-noise ratio and accuracy, and are all connected to each other. The overall design of a sensor is thus a multivariate optimization process and returns different results depending on the planned observation scenario and strategy. T. Flohrer, (2012) provides a detailed description of such an optimization process considering user requirements.

For simplicity, the following discussion focuses on aperture and field of view when designed the telescope for its specific purpose. Two scenarios occur during catalog build-up and maintenance: tracking and surveying. Firstly, if the position of an object is approximately known, the aim of the sensor is to improve this current estimate. Thus, only a limited field of view is required, i.e. as small as needed to find the object. A small field of view allows a greater angular resolution and is achieved by increasing the focal length of the telescope. This leads to a better accuracy in the measurements and ultimately in better estimates of the object's orbits. The second scenario uses no prior information of any object and scans the complete orbital region of interest. This work focuses on near-geostationary objects, but other domains are likewise scanned. If the entire region is covered by a small number of telescopes, a large field of view is essential to guarantee a frequently updated object database. It is achieved by a shorter focal length and comes with the drawback of a lower angular resolution.

The telescope system planned by GSOC will consist of two telescopes on one mount, where each telescope is used for a different operational mode: surveillance or tracking. The surveillance telescope will be equipped with a primary mirror with a diameter of about 20 cm and a focal length of 50 cm. The corresponding field of view is $2^\circ \times 2^\circ$. The tracking telescope has a primary mirror diameter of 50 cm and focal length of 300 cm, which results in a $42' \times 42'$ field of view (Fiedler et al., 2014, 2015). It will be mostly used to chase cataloged objects. Since the main focus of this work lies in the initial orbit determination without prior knowledge the surveillance telescope is used as a reference for the following simulations and discussions.

3.2 Astrometric reduction and errors

The images taken by the sensor contain information about the celestial position of an observed object. The topocentric astrometric place, i.e. the location on the sphere w.r.t an observer in the coordinate system origin, is illustrated in Figure 3.2. The right ascension α describes the horizontal angle between the observed position and the vernal equinox in the equatorial plane. The declination δ describes the vertical angle between the above noted plane and the position.

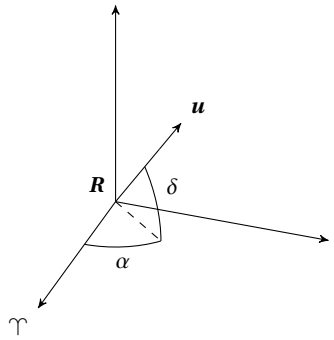


Figure 3.2: *Topocentric coordinate system: line-of-sight* $\boldsymbol{u}$*, sensor location* $\boldsymbol{R}$ *and astrometric place* (α,δ) *w.r.t. to vernal equinox.*

Stars and space objects moving relative to an observer appear as stripes on the image assuming a fixed telescope pointing, whereas geostationary objects appear as a bright spot. The coordinates of objects in the image reference system are obtained by using image processing methods. First, the images are segmented using e.g. thresholds to identify object pixels standing out of the image background. Then, so-called centroiding methods determine the center of objects in the image reference system (x,y), e.g. by fitting point spread functions to the segmented object pixels. The path of the photons taken from their celestial location to the detector is described by a transformation. In order to obtain the astrometric position (α,δ) of an observed object at (x,y), the transformation parameters must be determined (see Figure 3.3 for an illustration of the transformation).

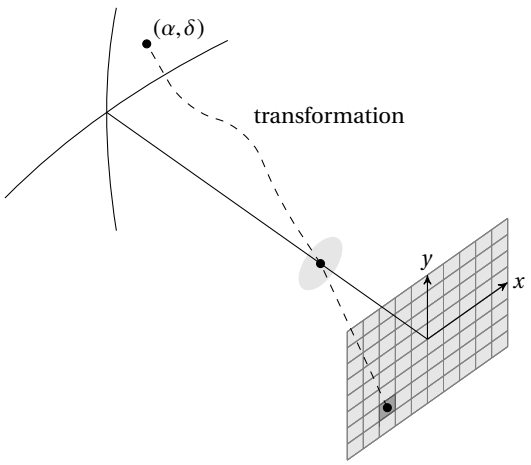


Figure 3.3: *Illustration of the astrometric reduction. The celestial position* (α,δ) *is projected on the image frame* (x,y)*.*

In principle, the transformation can be described as a projection. However, the path of the photons is distorted and the sensor model is not perfectly known or changing with time. Atmospheric refraction (wavelength dependent change in direction of the photons path) has to be considered as well as aberration. Latter is caused by the motion of the telescope (w.r.t. the geocenter and to the barycenter of solar system). The transformation parameters have to cover these distortions as well. Stable (or long-term varying)

parameters are calibrated, e.g. once every night. On the other side, short term distortions need to be modeled for each image or for small series of images. Known star positions are used as a reference to determine the remaining parameters of transformation through a least squares adjustment. The places of the unknown light sources are then obtained by inverting the transformation. The uncertainty of this mapping function depends on the distribution of reference stars around the objects of interest. Consequently, the uncertainty can vary for objects on the same image but different location. If a series of images of an object is collected with the same star background, similar errors are expected. Reference star catalogs can introduce a bias if the positions are erroneous (C. Flohrer, 2008). Usually, the contribution is neglected as modern catalogs promise very good accuracies (Schildknecht, 2007).

The quality of each individual measurement is mostly limited by the atmospheric seeing. The seeing is caused by turbulence in the air of the atmosphere. The image of the observed objects is blurred and possibly shifted. The magnitude of this blurring (typically described with the so-called seeing disc diameter) differs for each location and observation time, and is mostly influenced by the observer location and telescope pointing.

An additional error source is the erroneous registration of the observation epoch. In case of a photon counting array, the collection of photons (typically called integration) is started and stopped by opening and closing of a shutter. Afterward, the registers are read out to obtain the image. The movement of the shutter blades is detected and used to obtain a mean epoch. As the mechanical motion can vary due to changing temperature, humidity, or orientation, an exact timing cannot be guaranteed. C. Flohrer, (2008) computed corrections for these errors, which can be used for routine calibration. Schildknecht, (1994) proposes to read out the detector register after opening the shutter in order to detect the first collected photon. In case of active-pixel sensors, the problem is avoided as each instance of an arriving photon can be timed (Silha et al., 2014).

The overall accuracy is typically estimated using precisely known positions of e.g. navigation satellites (C. Flohrer, 2008). For this research, the error of each measurement noise is assumed to be caused primarily by the seeing and in the order of 1″, while a bias affecting all measurements of a series is caused by the shutter delay and expected to be in the range of a few arc seconds. The performance is consistent with the observations of the ZimSMART telescope, which has been analyzed by Herzog et al., (2013). These assumptions do not fully represented the expected performance of GSOC's network, but rather a conservative approximation.

3.3 Characteristics of observed domain

This work focuses on objects in near-geostationary orbits, i.e. objects that lie on or intersect with the geostationary ring. Common orbits in this region are the geostationary orbit itself, inclined geosynchronous orbit, graveyard orbit, and the geostationary transfer orbit. The latter is typically used by satellite launchers to place objects into the geostationary orbit. The graveyard orbit (also called super-geosynchronous orbit) lies approximately 200–300 km above the geostationary ring and serves as a disposal place for disused satellites (Flury, 1991). Inclined geosynchronous orbits share the same altitude with geostationary objects, but a different orbital plane orientation. They are used for navigation satellites as e.g. in Japan's Quasi-Zenith Satellite System or some of China's Compass satellites. The dominant perturbing forces and long-term dynamics are comparable for all different geosynchronous objects. Thus, this section covers the behavior of the two different regimes: geosynchronous (including geostationary, inclined geosynchronous and super-geosynchronous) and transfer orbits.

The most dominant perturbing forces from (2.1) for both orbit types are summarized in Table 3.1. The perturbing forces have an impact on the long-term evolution of the orbital domain, and consequently the

distribution of objects. The effects are described in the following for each type.

3.3.1 Geosynchronous orbits

The ideal geostationary orbit is defined by the orbital parameters

$$a = 42{,}164\text{km}, \quad e = 0, \quad i = 0° . \tag{3.1}$$

However, the actual distribution of objects along the ring differs from the perfect circular ring.

The tesseral terms of Earth's gravitational potential are longitude dependent and cause librational motion along the orbit around stable points. Controlled satellites perform East-West maneuvers to stay within their assigned longitude slot. The zonal terms affect the orientation of the orbital plane. The orientation is described with the normalized inclination vector from (2.9) in terms of orbital parameters

$$\boldsymbol{h} = \frac{\boldsymbol{r} \times \dot{\boldsymbol{r}}}{\|\boldsymbol{r} \times \dot{\boldsymbol{r}}\|} = \begin{pmatrix} \sin i \sin \Omega \\ -\sin i \cos \Omega \\ \cos i \end{pmatrix} . \tag{3.2}$$

The zonal terms (namely the oblateness) cause this vector to precess around the Earth's rotation axis. Additionally, the gravitational pull from Sun and Moon cause a similar precession around the ecliptic pole. The combined effect of these three perturbations then creates a precession around the normal $\boldsymbol{h}_L$ of the so-called Laplace-plane with a period of around 53 years. The normal is inclined to the Earth rotation axis by $\sim$ 7.5° in direction of the ecliptic pole. The Laplace-plane normal is approximately constant, however, dependent on the area-to-mass ratio the solar radiation pressure has a similar effect and pulls the Laplace-plane normal even further towards the ecliptic pole.

Thus, controlled satellites have to perform out-of-plane (North-South) maneuvers to maintain their orbital plane close to the Earth's equator. As these satellites are regularly maneuvered, they typically stay very close ($\sim$ 0.1°) to the equatorial plane. Maneuvers have to be performed to correct the orbits which would otherwise drift away from their assigned location due to perturbations. Active objects on purposely inclined orbits will equally stay close to their respective plane.

However, as the orbital plane of disused satellites and debris is constantly changing, the overall distribution of objects close to the geostationary ring can be covered by the constraints

$$30{,}000 \leq a \leq 50{,}000\text{km}, \quad e \leq 0.3, \quad i \leq 30° . \tag{3.3}$$

Table 3.1: *Order of magnitude of dominant perturbing accelerations, as defined in (2.1), acting on satellites in geostationary and geostationary transfer orbits. The magnitude for the transfer orbit is only given for the perigee passage (apogee similar to geostationary case).*

	Acceleration [m s^{-2}]						
	$\boldsymbol{F}_\oplus$			$\boldsymbol{F}_\odot$	$\boldsymbol{F}_M$	$\boldsymbol{F}_D$	$\boldsymbol{F}_S$
	$\mu_\oplus$	Zonal	Tesseral				
Geostationary orbit	10^0	10^{-5}	10^{-7}	10^{-5}	10^{-5}		10^{-7}
Geostationary transfer orbit (perigee)	10^1	10^{-2}	10^{-4}	10^{-6}	10^{-6}	10^{-2}–10^{-7}	10^{-7}

Using these bounds and the publicly available catalog, the distribution of objects is plotted. Figure 3.4 visualizes the precession of the orbital plane normal around $\boldsymbol{h}_\mathrm{L}$.

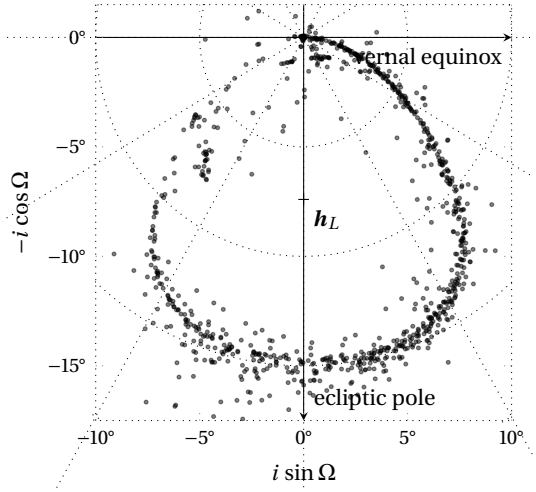


Figure 3.4: *The orbital plane orientation of cataloged geostationary objects (status January 2015).* $(i\sin\Omega, -i\cos\Omega)$ *is the approximate projection of* $\boldsymbol{h}$ *on the equatorial plane. The precession around* $\boldsymbol{h}_L$ *is indicated. The figure adapted from (Soop, 1994).*

When placing circular orbits on these orbital planes with $\sim 7.5°$ angular separation between each $\boldsymbol{h}$ and $\boldsymbol{h}_\mathrm{L}$, a ring of orbits with a diameter about 15° is obtained. Figure 3.5 illustrates the resulting ring. The gray-scale color of the orbits lightens with the progress around the precession axis, i.e. the black thick line represents the starting geostationary orbit while the light gray orbits represents the status at the end of the cycle (after approximately 50 years).

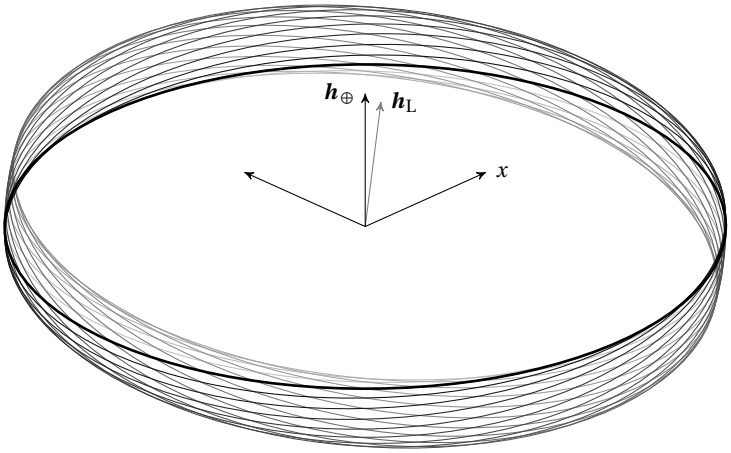


Figure 3.5: *Evolution of orbital planes around the precision axis* $\boldsymbol{h}_L$. $\boldsymbol{h}_\oplus$ *denotes the Earth rotation axis (and consequently the axis of geostationary orbits). Further details in the text.*

The ring with the expected density of objects is mapped into the geocentric α-δ-space. Instead of showing a set of virtual orbits, the real distribution of objects from the publicly available catalog is plotted. Figure 3.6 shows the density of objects in the observed domain. The density of orbital planes around the Laplace-plane (emphasized with the white line in the figure) has consequences on the planning of optical observations, which will be discussed in section 3.5.

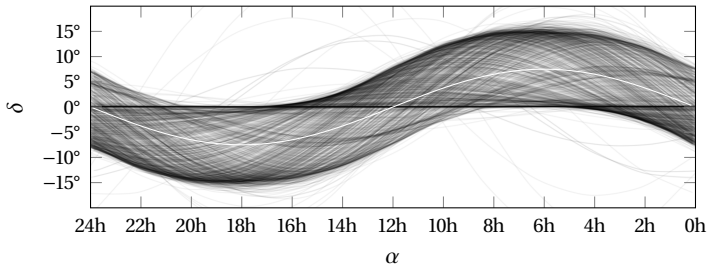


Figure 3.6: *Distribution of cataloged geostationary objects (status January 2015) in right ascension (α) and declination (δ) space. Figure adapted from (Schildknecht, 2007).*

3.3.2 Geostationary transfer orbits

When placing a satellite into the geostationary orbit, an intermediate Hohmann transfer orbit is typically used. The apogee of the orbit corresponds to the geostationary altitude, whereas the perigee stays close to the Earth's atmosphere. Satellites are released there by the launchers from where they perform apogee kick thruster burns to circularize their orbits. The carrier rocket stages remain in the transfer orbit. The long-term evolution of objects in this domain is dependent on their area-to-mass ratio and launch date. In theory, the atmospheric drag slows down the satellite at the perigee and consequently reduces the apogee height (with approximately constant perigee). This circularizes the orbit and will eventually lead to a decay. The effect on the distribution of objects can be observed in Figure 3.7, where the semi-major axis is plotted against the eccentricity.

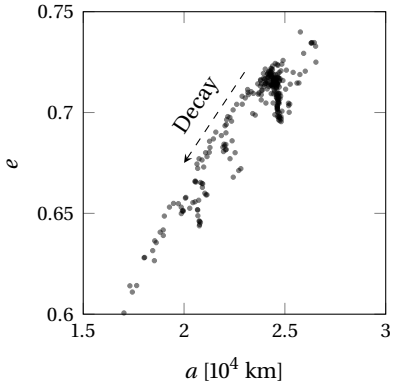


Figure 3.7: *Distribution of cataloged geostationary transfer orbit objects (status January 2015) in semi-major axis (a) and eccentricity (e) space. Orbital decay of objects is illustrated with the arrow. The figure is adapted from (Musci et al., 2007).*

The overall distribution of objects that started on geostationary transfer orbits is covered by the constraints

$$300 \leq r_{\mathrm{P}} \leq 1,200\mathrm{km}, \quad 0.3 \leq e \leq 0.75, \quad i \leq 15°, \tag{3.4}$$

where $r_{\mathrm{P}} = a(1-e)$ is the radius at the perigee. Due to large apparent velocity relative to the observer, the objects are mostly observed at apogee.

3.4 Observation strategy and constraints

In order to distinguish illuminated objects from the sky background, enough reflected photons must be collected by the detector. The high angular velocity of some satellites w.r.t. an observer, increases this difficulty. Fortunately, objects in the neighborhood of the geostationary ring typically stay at a nearly fixed location on the sky within reasonable time. Consequently, the pointing direction of a telescope can be kept constant.

Exposure and read-out

The minimum exposure time is dependent on the brightness of the observed objects and the topocentric distance. The first depends on the Sun illumination, object size, shape, and orientation while the latter stays approximately constant for objects in the region of interest. Optimally, objects are observed when the angle between the line-of-sight and the Sun-object vector, the so-called solar phase angle, is small. This increases the chances of detection, as parts of the surface pointing to the observer are illuminated. It is achieved by observing the areas in the sky opposite to the Sun. Geostationary objects theoretically allow arbitrary exposure times. However, star trails appear that cut out large parts of the image and reduce the usable area. The selection of a proper integration time is therefore an optimization task depending on hardware and observed orbital domain. Objects with larger apparent velocities, e.g. on geostationary transfer orbits, require tracking unless observed at their apogee. At this location they exhibit a smaller apparent velocity and can be observed within the same survey as near-geostationary objects. After the exposure, the information is read out and simultaneously the telescope is pointed to a new direction. The individual tasks accumulate to one characteristic interval, determining the minimum spacing between individual measurements.

Orbital coverage with limited resources

Besides collecting enough light from the objects, the sky background has to be dark enough. Any bright light sources in the vicinity of the observed objects decrease the signal-to-noise ratio. Thus, the viewing conditions are only acceptable in certain parts of the sky. Figure 3.8 shows the effective visible regions on two dates. The hatched areas cannot be observed due to the high density of stars close to the Milky Way or due to the Moon. Additionally, the area covered by the Earth's shadow at geostationary altitude is hatched. The figure depicts a major difficulty when dealing with optical measurements, i.e. the uneven

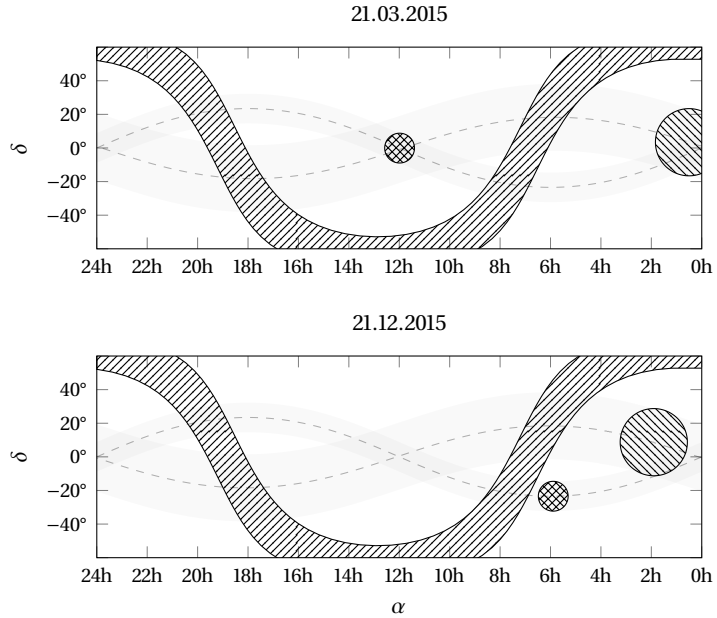


Figure 3.8: *Geocentric right ascension and declination observation constraints for two nights (21. March 2015 upper and 21. December 2015 lower part). Shaded area cannot be observed: no visibility inside Penumbra cone (▨) , 20° angular distance to Moon (▧) , and* 10° *to Galactic plane (▨) . The light gray area illustrates the changing position of the shadow and Moon over time.*

distribution of measurements along the orbits of the observed objects. At the vernal equinox (upper part of figure), the viewing conditions are favorable in the selected year, as the area close to the shadow is free of any other disturbing light sources. However, the situation at the winter solstice (lower part of the figure) is different. It is the day in the year with most night hours in the northern hemisphere, which allows the observation of a large arc. However, the objects can only be observed at a few locations of their orbit.

In order to keep an object database as updated as possible, as many objects as possible have to be observed within one night. This is achieved by collecting only one angular measurement pair (α, δ) per object. However, the information content of such an individual measurement is limited, i.e. it allows no initial orbit determination and complicates an association to known objects. To overcome this difficulty, short observation arcs, called tracklets, are collected instead. This strategy increases the information content by a measure for the apparent velocity $(\dot{\alpha}, \dot{\delta})$ on the sky. Longer tracklets provide more accurate information on the velocity, but reduce the overall number of covered objects per night.

Alternatively, individual angular measurements at multiple epochs distributed throughout the night can be used. A larger arc is then covered by the same number of measurements than when using short-arc tracklets. However, it increases the efforts of combining the individual measurements as three pairs must be tested against each other to identify one candidate object (see section 6.2). If a set of m measurements is collected (e.g. within one or multiple nights), in total

$$\binom{m}{3} = \tfrac{1}{6} m(m-2)(m-1) \tag{3.5}$$

combinations must be tested. The test effort thus increases with the third power of the total number of tracklets ($\mathcal{O}(m^3)$). Tracklets, however, allow for an initial orbit determination under certain assumptions, e.g. for circular orbits (cf. section 6.3). Additionally, two tracklets provide sufficient information to perform initial orbit determination (see Chapter 7). Thus, when testing all possible orbits of a set of m measurements, (3.5) is reduced to

$$\binom{m}{2} = \tfrac{1}{2}(m-1)\ m \tag{3.6}$$

correlation tests must be performed, i.e., the total effort increases with the square of the number of tracklets ($\mathcal{O}(m^2)$).

The latter approach of collecting tracklets is taken as a reference for this work. Considering a limited number of telescopes surveying the near-geostationary domain, typically 5-10 measurements are collected per tracklet to still guarantee a good coverage of the domain.

3.5 Scheduling

The planning of observations should be optimized for two different goals: firstly, to assure good visibility of the objects and secondly to maximize the information gain. A schedule (or observation plan) contains pointing directions for each epoch in the night. If an object catalog exists, the scheduler can include the knowledge into the optimization process. DeMars and Jah, (2011) propose to use the expected information increase to decide which object is observed when and from which telescope. The latter plays a role when the observation geometry from one location is more beneficial than from another. When building up a catalog, the expected density of objects can be used to define a schedule. The density of objects in the near-geostationary region can be approximated with the ring around the Laplace-plane (see Figure 3.6).

A basic planning tool has been developed to simulate realistic observation conditions for a survey telescope. For simplicity, the tool does not optimize a complete timeline of tasks but fills up a schedule

considering three main constraints. As written above, good visibility is typically expected at small solar phase angles. Thus, objects are observed when they are close to the Earth's shadow. In principle, when collecting multiple measurements of the same object, the information increases with larger arcs covered. It is therefore beneficial to avoid observing an object at the same location on their orbits in consecutive nights. Lastly, a survey should find every object within a reasonable time in order to update the complete database frequently.

Schildknecht et al., (1999) and T. Flohrer et al., (2005) describe a strategy where a declination stripe is repetitively scanned with a fixed right ascension. Each object passes through this space-fixed direction once, and if observed continuously, every object will be observed. Figure 3.9 (left side) illustrates the selection of the fixed right ascension α_F (in geocentric reference system) close to the shadow boundary. The boundaries are assumed to be constant for one (or a few) nights. The right side of the figure shows the declination scanning. Instead of testing the complete declination range of this stripe, the actual distribution of objects from Figure 3.6 is considered. The relation between i and Ω is used to ignore all improbable declination values. A declination interval for each right ascension covering most of the objects is approximately given by the Laplace-plane ±7.5°.

Herzog et al., (2010) suggest to scan not one but multiple of these stripes in each night. The planning tool then iterates through a set of fixed directions. This increases the time until complete orbital coverage is obtained, but instead provides better options to re-observe objects (even in the same night). By placing the additional stripes α_R in a predefined angular distance, the same longitude slot of the geostationary orbit is re-observed after a certain time interval. Furthermore, different stripes allow a better distribution of observations along the orbits.

Reference survey algorithm

The basic algorithm to select these stripes is presented in the following. The directions are fixed for a certain time-interval (3 nights), where celestial bodies stay approximately within reasonable boundaries. As proposed in (Herzog et al., 2010) two fixed stripe locations are selected $\alpha_{F,1}$ and $\alpha_{F,2}$, and each one is accompanied with a re-observation stripe $\alpha_{R,1}$ and $\alpha_{R,2}$. In order to compute the places for the re-observation fields, the stripe duration must be considered. The duration comprises the exposure and read-out time of the detector for each image and is consequently different for each telescope setup. The time is also dependent on the field of view of each sensor and the corresponding different number of images per stripe. For the sake of simplicity, the telescopes are assumed to be equal in the network. However, the strategy can be adjusted to different sensors which then requires careful planning of re-observation epochs. The selection of stripes consists of the following steps:

1. Find the two right ascension values $\alpha_{F,1}$ and $\alpha_{F,2}$ bounding the Earth shadow at geostationary altitude

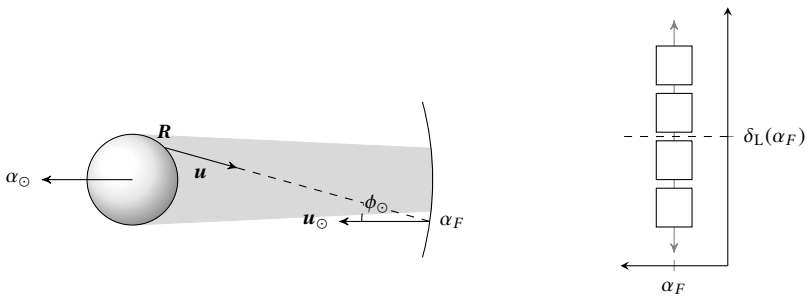


Figure 3.9: *Fixed geocentric right ascension α_F with favorable solar phase angle throughout the night.*

throughout the time interval. If the stripes are disturbed by other light sources, shift them until a visible value is found.

2. Select further two right ascension values $\alpha_{R,i}$ with a specific offset to each $\alpha_{F,i}$, $i \in 1,2$. The offset is chosen as an integer number of stripe durations $j\ \Delta t_{\text{str}}$. As objects on the geostationary orbit rotate around Earth with sidereal rate, the objects appearing at $\alpha_{F,i}$ are re-observed if $\alpha_{R,i}$ is observed exactly after j stripe iterations. The number j is selected accounting for the rate and visibility constraints.

Figure 3.10 shows an example stripe selection for an unfavorable case, i.e. a large fraction of the α-δ space is masked by other light sources.

After finding the suitable set of stripes, the planning continues by creating a timeline of observation tasks, i.e. topocentric telescope pointing directions. So far, the direction computation was independent of the sensor location. At first, the nights are divided into Δt_{str} intervals. In each interval, one of the visible stripes is scanned. Visibility is tested by computing the elevation of the virtual objects along the stripe at geostationary altitude. If the elevation is above the nautical horizon, the objects are assumed to be observable. Typically, in the beginning of a night, only one stripe in eastern direction is visible. Then later, all stripes can be observed. At the end of the night, the sensor points towards western direction and sees only one stripe. Another station dependent constraint is the actual solar phase angle. Stripes, with an angle larger than 120° are ignored. Other visibility constraints are avoided by the previous stripe selection. The constraints are summarized in the following Table 3.2.

Table 3.2: *Visibility constraints for optical observations*

Constraint	
Min. Moon angular distance to observed direction	20°
Min. galactic latitude of observed direction	20°
Max. Sun elevation	−12°
Max. solar phase angle	120°

After selecting a stripe direction, the planning tool continues with the next stripe observation epoch. The re-observation stripes $\alpha_{R,i}$ are scanned j epochs after their respective $\alpha_{F,i}$ partners. An additional criterion for the exclusion of stripes is included: if the geostationary longitude of the stripe has been observed in the previous nights, similar observation geometries must be avoided (see Chapter 8 and Chapter 9 for the effects). The same position on the orbit is observed if the geostationary longitude is

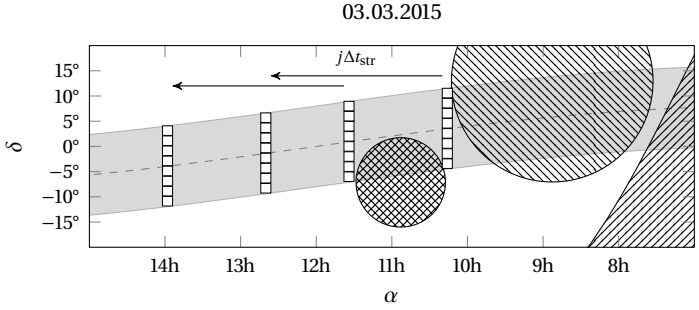


Figure 3.10: *Selection of four declination stripes. Hatched areas denote visibility constraints (same as in Figure 3.8). Gray area illustrates the likely distribution of objects around Laplace plane (dashed line).*

observed after integer numbers of sidereal days. Thus, stripes are eliminated which come close to this re-observation time (e.g. one hour). The described algorithm serves as a reference but does not promise to obtain optimal results. Its performance is estimated using a telescope network simulation.

Simulation

The reference planning algorithm is used to simulate a small telescope network. The simulation is carried out to obtain a likely distribution of measurements and geometries. As stated in the previous section, the schedule is not optimal in terms of coverage or visibility but serves as a reference. Similarly, the simulation does not cover all physical effects and challenges. That is, it does not include weather (cloudy skies), or the intersection of star trails with observed objects. Both would cause a loss of observations. Of course, several weeks of bad viewing conditions will alter the distribution. However, the simulation suffices to give an approximate value for the expected (nominal case) re-observation times and re-observation geometries.

In order to cover a specific longitude range continuously, it is beneficial to place telescopes on the northern and southern hemisphere. This reduces seasonal effects and increases chances of good visibility for at least one site. Thus, in accordance with GSOC's plans, a telescope is placed in Zimmerwald (Switzerland), and one in Sutherland (South Africa). The location of both simulated telescopes is given in Table 3.3.

Table 3.3: *Simulated telescope locations*

	λ	ϕ	h
Zimmerwald	7.465	46.877	970
Surtherland	20.813	−32.937	1700

The sensors are modeled with a $2° \times 2°$ field of view. The simulation is performed using a set of near-geostationary satellites from the publicly available catalog *space-track.org*. The objects were extracted according to the previously defined region constraints (3.3) and (3.4). In total, 661 objects which are visible from at least one sensor are taken into account for the simulation.

Using the previously described schedule, the telescope pointing for each simulation step is computed. Each object that appears in the field of view at least three times in a series of five images, creates a tracklet. A tracklet consists therefore of at most 5 individual right ascension and declination measurements. The objects' states are propagated to the observation epochs accounting for the simplified perturbation model (SGP4). Observations are generated for the whole year 2015. In total about 80,000 tracklets are collected.

The performance of the scheduling algorithm and the implications for orbit determination tasks are analyzed with the following figures. The first (Figure 3.11) shows the distribution of the solar phase angle $\phi_\odot$ and serves to check for the plausibility of the plan. Most objects are observed at favorable angles (around 10°–30°). The objects are likely to be successfully observed at these geometries. Thus, the simulation is plausible in terms of realistic geometries.

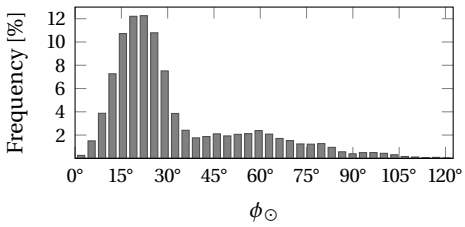


Figure 3.11: *Distribution of solar phase angles for simulated observations.*

The other two histograms in (Figure 3.12) set first requirements for the orbit determination and association process. The left side shows the time intervals between observations for each object. Typically, the objects are observed in the same or consecutive nights. However, due to visibility and time constraints, it is very likely to observe objects only after three or four days. The developed algorithms must deal with the expected numbers of days (1-5) in between measurements. The right side of the figure shows the distribution of angular separations $\Delta\nu$ on the respective orbits between two observation epochs (the difference in true anomaly). The distinct peaks are caused by the definition of the re-observation stripes in the planning algorithm. Even though the algorithm tries to avoid observing objects at the same place of their orbits, the resulting distribution shows a large fraction of objects with small $\Delta\nu$ values. This is caused by the visibility constraints and that the algorithm only accounts for objects moving with sidereal rate. Observing an object at the same location of its orbit returns less new information than when observing it somewhere else. The consequences will be discussed in Chapter 8. Unless more advanced planning algorithms are able to avoid these geometries completely, a catalog software should be able to work with these geometries e.g. by including restrictions on the orbit (as presented in section 6.3).

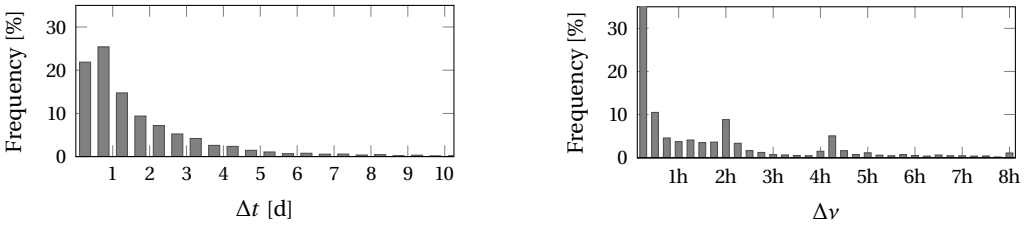


Figure 3.12: *Distribution of Δt (time intervals between tracklets) for simulated objects in near-geostationary domain on left side. Distribution of angular separations $\Delta\nu$ for same tracklets.*

3.6 Information content

As described in the previous sections, optical telescopes surveying the geostationary domain capture short series of astrometric observations for each object, i.e. right ascension α and declination δ in the topocentric reference frame

$$\alpha_i, \delta_i \quad \text{at epochs} \quad t_i \quad \text{for} \quad i = 1, \ldots, n\,. \tag{3.7}$$

The information of this sequence is compressed in order to use it conveniently later. Thus, a polynomial motion model is fitted to the angular measurements, where the estimated coefficients then contain the information. This compression comes with a loss of information, but simplifies the orbit determination and association. The order of the polynomial must be carefully selected. A higher order polynomial increases the risk of over-fitting the measurement arc, i.e. instead of finding a model that describes the information of the tracklet, a model fitting the random measurement noise is found. On the other hand, if

the degree of the polynomial is too small, possibly usable information is lost and systematic errors can be introduced. This problem is commonly referred to as the variance-bias tradeoff. In case of short-arc tracklets (spanning a few minutes) of high altitude objects, a linear models is sufficient. When using different observation strategies (e.g. longer tracklet duration) or when observing low altitude objects, a quadratic motion model becomes feasible. Maruskin et al., (2009) derive in detail how to fit a quadratic motion model to the measurements of one observation arc and what errors to expect in the estimated parameters. The derivation here follows the work by Maruskin et al., (2009), but adapts it to the linear model

$$\alpha(t) = \bar{\alpha} + (t - \bar{t})\,\dot{\alpha} + b_\alpha, \quad \delta(t) = \bar{\delta} + (t - \bar{t})\,\dot{\delta} + b_\delta, \tag{3.8}$$

including the bias terms b_α and b_δ. These terms are assumed to be constant for each tracklet. The model parameters are the mean angles $\bar{\alpha}$ and $\bar{\delta}$ and the respective angular rates $\dot{\alpha}$ and $\dot{\delta}$. They are estimated around a mean epoch $\bar{t}$. The right ascension and declination values are assumed to be uncorrelated to each other and are individually estimated. For the sake of notational simplicity, the generic angular variable θ is used in the following as a replacement for α and δ. The linear motion model function is then

$$\theta(t) = \bar{\theta} + (t - \bar{t})\dot{\theta} + b_\theta\,. \tag{3.9}$$

Inserting the angular measurements $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\mathsf{T}$, a set of linear equations is obtained

$$\boldsymbol{\theta} = \boldsymbol{J}\boldsymbol{x} + \boldsymbol{\varepsilon} \quad \text{where} \quad \boldsymbol{J} = \begin{pmatrix} \left(\frac{\partial\theta(t)}{\partial \boldsymbol{x}}\Big|_{t_1}\right)^\mathsf{T} \\ \vdots \\ \left(\frac{\partial\theta(t)}{\partial \boldsymbol{x}}\Big|_{t_n}\right)^\mathsf{T} \end{pmatrix} = \begin{pmatrix} 1 & t_1 - \bar{t} \\ \vdots & \vdots \\ 1 & t_n - \bar{t} \end{pmatrix}, \tag{3.10}$$

where $\boldsymbol{x} = (\bar{\theta} + b_\theta, \dot{\theta})^\mathsf{T}$ denotes the estimated parameters, and $\boldsymbol{\varepsilon}$ the measurement noise. As the bias is assumed to be constant, it is estimated together with the mean angle. The noise is assumed to be zero mean and normally distributed with the diagonal covariance matrix $\boldsymbol{C}_{\boldsymbol{\varepsilon}} = \sigma_\mathrm{N}^2\,\mathbb{1}$. The least squares solution is computed with

$$\boldsymbol{x} = \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{J}^\mathsf{T}\boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\theta} \tag{3.11}$$

where the covariance

$$\boldsymbol{C}_{\boldsymbol{x}} = (\boldsymbol{J}^\mathsf{T}\boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{J})^{-1} \tag{3.12}$$

describes the uncertainty of the estimated vector $\boldsymbol{x}$. The so-called innovation matrix

$$\boldsymbol{J}^\mathsf{T}\boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{J} = \frac{1}{\sigma_\mathrm{N}^2}\begin{pmatrix} n & 0 \\ 0 & \frac{1}{3}\Delta t^2(n+1)(n-1)n \end{pmatrix} \tag{3.13}$$

is derived analogously to (Maruskin et al., 2009). The n measurement epochs are assumed to be distributed on an equally spaced grid around the mean epoch, thus

$$t_i = \bar{t} + i\,\Delta t \quad \text{where} \quad i = -\tfrac{n-1}{2}, \ldots, 0, \ldots, \tfrac{n-1}{2}\,. \tag{3.14}$$

The variance of the mean angle is determined by inverting the innovation matrix and including the variance of the bias. The bias is assumed to be normally distributed with $\mathrm{E}[b_\theta] = 0$ and $\mathrm{E}[b_\theta^2] = \sigma_\mathrm{B}^2$. The variance of the angle is then

$$\sigma_{\bar{\theta}}^2 \approx \sigma_\mathrm{B}^2 + \tfrac{1}{n}\sigma_\mathrm{N}^2 \,. \tag{3.15}$$

The estimated angular rate is independent from the bias and its variance and given by

$$\sigma_{\dot{\theta}}^2 \approx \frac{3\,\sigma_\mathrm{N}^2}{(\Delta t)^2\, n\,(n^2-1)} \,. \tag{3.16}$$

The last two equations illustrate an advantage of using the compressed information in terms of mean angle and angular rates: the model smoothes out random noise. The mean angle has a smaller variance when compared to one of the measured ones. Furthermore, the angular rate accuracy $\sigma_{\dot{\theta}}$ is approximately inversely proportional to the tracklet length ($\propto \frac{1}{n\,\Delta t}$) or to the number of individual measurements ($\propto \frac{1}{n}$). Hence, observation strategies can be adapted to guarantee a sufficient accuracy of the new parameters.

For completeness, the derived variances from (Maruskin et al., 2009) for the motion model

$$\theta(t) = \bar{\theta} + (t-\bar{t})\dot{\theta} + (t-\bar{t})^2\ddot{\theta}\,, \tag{3.17}$$

are given by

$$\sigma_{\bar{\theta}} \approx \tfrac{3}{2\sqrt{n}}\sigma_\mathrm{N}\,, \quad \sigma_{\dot{\theta}} \approx \tfrac{2\sqrt{3}}{(n-1)\Delta t\sqrt{n}}\sigma_\mathrm{N}\,, \text{ and } \quad \sigma_{\ddot{\theta}} \approx \tfrac{12\sqrt{5}}{(n-1)^2\Delta t^2\sqrt{n}}\sigma_\mathrm{N}\,, \tag{3.18}$$

where simplifications have been included assuming that the number of individual measurements is large, i.e. $n \gg 1$, and no bias term is added. In order to improve readability, the bar above the mean epoch and mean angle is dropped in the following derivations. The final measurement vector is then

$$\boldsymbol{z} = (\alpha,\, \dot{\alpha}, \delta, \dot{\delta})^\mathsf{T} \quad \text{or} \quad \boldsymbol{z} = (\alpha,\, \dot{\alpha},\, \ddot{\alpha}, \delta, \dot{\delta}, \ddot{\delta})^\mathsf{T} \tag{3.19}$$

at the epoch t.

Line-of-sight and derivatives

An alternative representation of the angles, angular rates, and angular acceleration is given by the line-of-sight and its time-derivatives: $\boldsymbol{u}$, $\dot{\boldsymbol{u}}$, $\ddot{\boldsymbol{u}}$. The topocentric line-of-sight is illustrated in Figure 3.2 and is computed from the right ascension and declination with

$$\boldsymbol{u} = \begin{pmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{pmatrix} . \tag{3.20}$$

The first time-derivative of the line-of-sight is given in terms of angles and rates by

$$\dot{\boldsymbol{u}} = \frac{\partial \boldsymbol{u}}{\partial \alpha}\,\dot{\alpha} + \frac{\partial \boldsymbol{u}}{\partial \delta}\,\dot{\delta} \quad \text{with} \quad \frac{\partial \boldsymbol{u}}{\partial \alpha} = \begin{pmatrix} -\sin\alpha\cos\delta \\ \cos\alpha\cos\delta \\ 0 \end{pmatrix} \quad \text{and} \quad \frac{\partial \boldsymbol{u}}{\partial \delta} = \begin{pmatrix} -\cos\alpha\sin\delta \\ -\sin\alpha\sin\delta \\ \cos\delta \end{pmatrix} . \tag{3.21}$$

The second derivative of the line-of-sight is likewise derived using the angular accelerations $\ddot{\alpha}$ and $\ddot{\delta}$.

Modeled observation from a satellite state

The geocentric position in terms of the line-of-sight is expressed by

$$\boldsymbol{r} = \boldsymbol{R} + \rho\,\boldsymbol{u}\,, \tag{3.22}$$

where ρ is the distance between the observer and the satellite, and $\boldsymbol{R}$ is the position of observer. Consequently, the time derivatives of the satellite position are given by

$$\dot{\boldsymbol{r}} = \dot{\boldsymbol{R}} + \rho\,\dot{\boldsymbol{u}} + \dot{\rho}\,\boldsymbol{u}\,, \tag{3.23}$$

$$\ddot{\boldsymbol{r}} = \ddot{\boldsymbol{R}} + \rho\,\ddot{\boldsymbol{u}} + 2\dot{\rho}\,\dot{\boldsymbol{u}} + \ddot{\rho}\,\boldsymbol{u}\,. \tag{3.24}$$

Accordingly, the modeled line-of-sight and the first derivative from a state are given by

$$\boldsymbol{u} = \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|} \quad \text{and} \quad \dot{\boldsymbol{u}} = \frac{\dot{\boldsymbol{s}}}{\|\boldsymbol{s}\|} - \boldsymbol{u}\frac{\boldsymbol{s}^{\mathsf{T}}\dot{\boldsymbol{s}}}{\|\boldsymbol{s}\|^2} \tag{3.25}$$

where

$$\boldsymbol{s} = \boldsymbol{r} - \boldsymbol{R} \quad \text{and} \quad \dot{\boldsymbol{s}} = \dot{\boldsymbol{r}} - \dot{\boldsymbol{R}}\,. \tag{3.26}$$

The modeled angular rates can be derived from these.

4. Computational optimization

If a function $L(\boldsymbol{x})$ describes the cost or loss of a problem depending on some variable $\boldsymbol{x}$, then optimization algorithms are used to determine the optimal value of $\boldsymbol{x}$, i.e. where the loss becomes as small as possible

$$\arg\min_{\boldsymbol{x}} L(\boldsymbol{x}) := \{\boldsymbol{x} \mid \forall \tilde{\boldsymbol{x}} \colon L(\tilde{\boldsymbol{x}}) \geq L(\boldsymbol{x})\}\,. \tag{4.1}$$

This chapter recapitulates some basics of optimization and provides the required knowledge for later discussions. It follows the explanations in (Press et al., 2007). An example loss function is illustrated in Figure 4.1. The right plot shows the topography as seen from above. This representation is also used later for the discussions of loss functions in the tracklet association problem.

Several computational optimization methods have been developed, each one more or less suitable for specific loss functions and problem statements. If $L(\boldsymbol{x})$ is smooth and has one distinct minimum, iterative algorithms can be used. These methods start at an initial guessed value $\boldsymbol{x}^{[0]}$, and gradually improve the state until convergence is reached. The square brackets denote the current iterate (starting from 0). The state is either improved along some specified directions or using the gradient information. Two different methods are discussed in this work, namely the *quasi-Newton* approach and *Powell's line search*.

Often, the variable space of an optimization problem is bounded, i.e. an optimal point is forced to be in a specified region

$$\mathcal{C}(\boldsymbol{x}) = \{\boldsymbol{x} : c_1(\boldsymbol{x}) \leq 0,\, c_2(\boldsymbol{x}) \leq 0,\, \ldots\, c_{n_{\text{con}}}(\boldsymbol{x}) \leq 0\}\,, \tag{4.2}$$

where c_i for $i = 1, \ldots, n_{\text{con}}$ define inequality constraints. Equality constraints are included similarly. The region is typically either bounded due to physical constraints or, alternatively, to restrict the solution space to a subspace of interesting solutions.

A loss function may contain multiple local minima inside the region (as in the example from Figure 4.1). Therefore, a difficulty arises that the global minimum (out of many local ones) needs to be identified. The

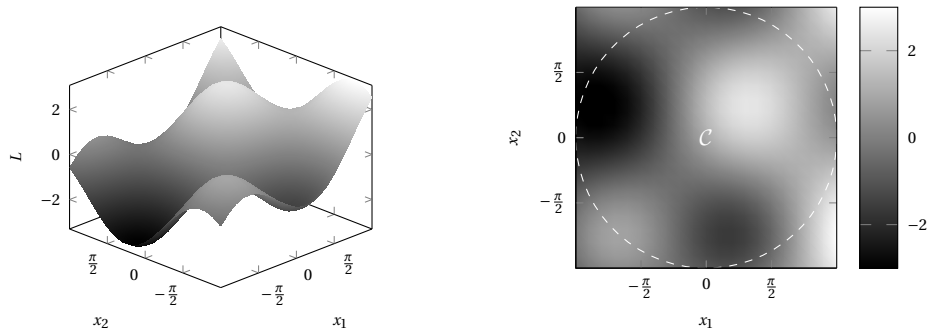


Figure 4.1: *Two illustrations of an example loss function* $L(x_1, x_2) = \frac{x_1}{2} + \sin(x_1 + x_2) + \cos(x_1 - x_2)$. *The plot on the right shows an intensity-coded representation of the loss function over the two-dimensional variable space.*

aforementioned iterative methods search in the neighborhood of an initial point and consequently find the next local minimum. Different so-called heuristics have been developed to overcome this difficulty. Typically, the constrained domain is sampled with a set of initial points. Then, the points are altered considering certain rules until local minima are identified. The minimum with the smallest loss is selected as the global one. Common examples for these multimodal optimization methods are *simulated annealing* or *differential evolution*. The latter is used in this work and is described in section 4.3.

In principle, the global minimum is also found, if the complete search space is evaluated. This brute-force approach requires the highest computational resources, but can be a safe backup solution if the right search grid is selected and other methods fail.

4.1 Quasi-Newton methods

If the function is smooth enough, algorithms can use the gradient information to efficiently reach the minimum. Prominent examples are conjugate gradient or quasi-Newton methods. The latter has been used in this work and will be explained in the following. The quasi-Newton methods approximate the loss function locally with a quadratic function using a second order Taylor expansion around the current value $\boldsymbol{x}^{[i]}$ (e.g. the initial value $\boldsymbol{x}^{[0]}$)

$$L(\boldsymbol{x}) \approx L(\boldsymbol{x}^{[i]}) + \boldsymbol{\nabla} L \cdot (\boldsymbol{x}^{[i]})(\boldsymbol{x} - \boldsymbol{x}^{[i]}) + \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{[i]})^{\mathsf{T}} \boldsymbol{H}(\boldsymbol{x}^{[i]})(\boldsymbol{x} - \boldsymbol{x}^{[i]}), \tag{4.3}$$

where $\boldsymbol{\nabla} L(\boldsymbol{x}^{[i]})$ is the gradient at $\boldsymbol{x}^{[i]}$ and

$$\boldsymbol{H}(\boldsymbol{x}^{[i]}) = \left.\frac{\mathrm{d}^2 L}{\mathrm{d}x^2}\right|_{\boldsymbol{x}=\boldsymbol{x}^{[i]}} \tag{4.4}$$

the corresponding Hessian. The minimum of the approximated function is then determined by finding the root of the first derivative

$$0 \overset{!}{=} \boldsymbol{\nabla} L(\boldsymbol{x}^{[i]}) + \boldsymbol{H}(\boldsymbol{x}^{[i]})(\boldsymbol{x} - \boldsymbol{x}^{[i]}), \tag{4.5}$$

where the Hessian must be a positive-definite matrix. If no analytical expression for the gradient is known, it can be approximated using a finite difference scheme. Additionally, the inverse of the Hessian matrix $\boldsymbol{H}^{-1}$ is required to solve equation (4.5) with

$$\boldsymbol{x}^{[i+1]} = \boldsymbol{x}^{[i]} - \boldsymbol{H}^{-1}(\boldsymbol{x}^{[i]}) \boldsymbol{\nabla} L(\boldsymbol{x}^{[i]}), \tag{4.6}$$

which gives the improved solution $\boldsymbol{x}^{[i+1]}$. Such an iterative improvement is denoted as a Newton method. Quasi-Newton methods do not require the Hessian or its inversion. Instead of computing $\boldsymbol{H}^{-1}$ directly, it is approximated and iteratively improved during the minimization process. Different variants of the quasi-Newton method exist, resulting in different schemes on how to improve the matrix. An implementation for the popular Broyden-Fletcher-Goldfarb-Shanno (BGFS) scheme can be obtained from (Press et al., 2007). For a detailed derivation of the algorithm refer to the book of Fletcher, (1987).

The procedure is iterated until convergence. Convergence is reached if the correction between iterations is small enough

$$\boldsymbol{x}^{[i]} - \boldsymbol{x}^{[i+1]} < t_{\boldsymbol{x}}\,, \tag{4.7}$$

where $t_{\boldsymbol{x}}$ is a predefined tolerance. Alternatively, the iterations can be stopped if the right hand side of (4.5) falls below a predefined threshold.

Advantages and disadvantages

The quasi-Newton approaches converge fast if the loss function is sufficiently similar to a quadratic function. The approximation is good when the initial value is close to the actual minimum. Good initial values therefore reduce the number of required steps. In case of a true quadratic function, the search stops after $n_{\boldsymbol{x}}$ iterations, i.e. the number of dimensions.

When using a finite difference scheme to obtain the gradient $\boldsymbol{\nabla} L$ in very flat valleys, the numerical error can exceed the actual value and therefore cause divergence or false results. If no analytical expression for this gradient is available and flat valleys are expected, other approaches which do not use any gradient information are favorable.

4.2 Powell's line search

Unlike the previously explained quasi-Newton methods, Powell's line search does not use any gradient information. Instead, a set of directions is used, represented with unit vectors $\boldsymbol{u}_j$ for $j = 1, \ldots n_{\boldsymbol{x}}$. The directions are iteratively improved during the minimization process. The principle of the line search method is illustrated in Figure 4.2. Commonly, the directions aligned with the coordinate axes are used as initial values, i.e.

$$\boldsymbol{u}_1^{[0]} = (1,\,0)^{\mathsf{T}} \quad \text{and} \quad \boldsymbol{u}_2^{[0]} = (0,\,1)^{\mathsf{T}} \tag{4.8}$$

for the two-dimensional case. Known more promising directions can be used to speed up the process. These are directions where a decrease in the loss function is expected. The algorithm then proceeds as follows: starting from the current value $\boldsymbol{x}^{[i]}$, for each direction $\boldsymbol{u}_j^{[i]}$ a one-dimensional minimization problem is obtained, i.e. for the first one

$$\lambda_1 = \arg\min_{\lambda} L\left(\boldsymbol{x}^{[i]} + \lambda\, \boldsymbol{u}_j^{[i]}\right) \tag{4.9}$$

and for $j > 1$

$$\lambda_j = \arg\min_{\lambda} L\Big(\boldsymbol{x}^{[i]} + \sum_{k=1}^{j-1} \lambda_k\, \boldsymbol{u}_k^{[i]} + \lambda\, \boldsymbol{u}_j^{[i]}\Big)\,. \tag{4.10}$$

The superscript index $[i]$ is also applied to the direction vectors as they are iteratively updated during the optimization process. A commonly used approach to solve the problem in (4.9) and (4.10) is Brent's algorithm, which can be obtained from (Press et al., 2007) and is described in (Brent, 1973).

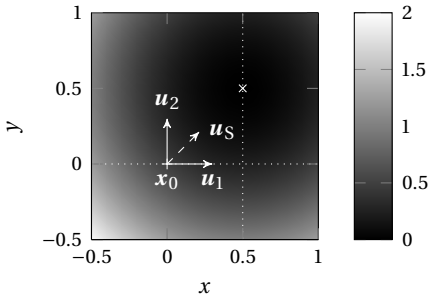


Figure 4.2: *Illustration of Powell's line search. Dotted lines indicate the one-dimensional search space. In this quadratic example, the minimum is found after one iteration.*

The auxiliary scalar variables λ_j, which minimize the loss function along the lines, are then used to obtain an improved value

$$\boldsymbol{x}^{[i+1]} = \boldsymbol{x}^{[i]} + \sum_{k=1}^{n_{\boldsymbol{x}}} \lambda_k \, \boldsymbol{u}_k^{[i]}. \tag{4.11}$$

The improved $\boldsymbol{x}^{[i+1]}$ is used together with the previous value $\boldsymbol{x}^{[i]}$ to update the most influential direction, i.e. the direction with the largest correction in one iteration

$$k^* = \arg\max_k \lambda_k \,, \tag{4.12}$$

with

$$\boldsymbol{u}_{k^*}^{[i+1]} = \frac{\boldsymbol{x}^{[i+1]} - \boldsymbol{x}^{[i]}}{\|\boldsymbol{x}^{[i+1]} - \boldsymbol{x}^{[i]}\|} \,. \tag{4.13}$$

Additional rules when to drop direction updates and modified ways to update the set of directions can be implemented to improve convergence of the method.

Advantages and disadvantages

The line search typically requires more loss function evaluations than the quasi-Newton method. Of course, this is depending on the shape of the loss function and the quality of the initial value and the initial directions. However, it does not use any information on the gradient. Especially when considering loss functions with no analytical representation of the gradient and narrow and flat valleys, it can outperform the gradient-based approaches.

4.3 Differential evolution

A multimodal problem is minimized with heuristic methods, if no prior information on the location of the local minima is available (e.g. constraints which bound the solution space to one valley). In this work, differential evolution has been implemented to solve these problems. A detailed description and code samples can be obtained from the book by Price et al., (2005). The complete details go beyond the scope of this thesis. As the approach is used in Chapter 7, it is sketched here to illustrate its basic principle and advantages/disadvantages. Figure 4.3 shows the principle of it with the example loss function from the beginning of this chapter.

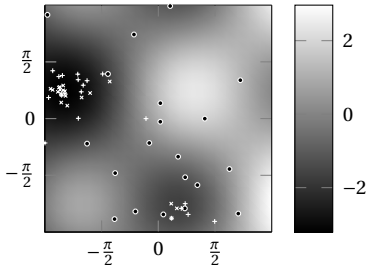


Figure 4.3: *Illustration of differential evolution optimization scheme. Circles show the first generation of 20 candidates (randomly selected inside domain), + after 25 iterations, and × after 50 iterations.*

To initialize the method, the complete domain is randomly sampled with a candidate population

$$\boldsymbol{x}_l^{[i]} \quad \text{for} \quad l = 1, \ldots, n_{\text{pop}}\,, \tag{4.14}$$

where n_{pop} is the potentially large number of candidates. In every iteration of the algorithm, each candidate solution is updated using randomly picked other candidates $\boldsymbol{x}_{\text{a}}^{[i]}$, $\boldsymbol{x}_{\text{b}}^{[i]}$, and $\boldsymbol{x}_{\text{c}}^{[i]}$ with a so-called crossover strategy

$$\boldsymbol{x}_l^{[i+1]} = c\left(\boldsymbol{x}_l^{[i]}, \boldsymbol{x}_{\text{a}}^{[i]}, \boldsymbol{x}_{\text{b}}^{[i]}, \boldsymbol{x}_{\text{c}}^{[i]}\right)\,, \tag{4.15}$$

where different c-functions exist. Alternatively, specific crossover candidates can be selected deterministically, e.g. the best candidate of the population. The strategies are tuned for the individual problem using a crossover probability threshold and a differential weight. The candidate $\boldsymbol{x}_l^{[i]}$ is only updated if the crossover result $\boldsymbol{x}_l^{[i+1]}$ is an improvement, i.e. the loss falls below the initial loss

$$L(\boldsymbol{x}_l^{[i+1]}) < L(\boldsymbol{x}_l^{[i]})\,. \tag{4.16}$$

After a previously defined maximum number of iterations or convergence, the best candidate is selected. The exponential crossover method is used along with random crossover candidate selection (cf. the book of Price et al., (2005) for details). More advanced variations of the method, e.g. from Takahama and Sakai, (2006), use the gradient information locally for each candidate.

Advantages and disadvantages

As any type of heuristic, the differential evolution cannot guarantee convergence towards the global minimum. It can randomly hit the desired valley, but it might also miss it. It is therefore favorable to split the problem into single valleys and then proceed with the above described iterative methods. However, if no separation is possible, heuristics can become a valid choice to minimize the loss function. They can be also used after small modification to not only find the global, but also multiple local minima.

4.4 Change of variables

An optimization problem can be simplified by substituting the search variable $\boldsymbol{x}$ with an other expression $\boldsymbol{x}'$. The loss function L is rewritten in terms of the new variable with

$$L'(\boldsymbol{x}') = L(\boldsymbol{g}(\boldsymbol{x}')) = L(\boldsymbol{x})\,, \tag{4.17}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ satisfy the coordinate map

$$\boldsymbol{x}' = \boldsymbol{g}^{-1}(\boldsymbol{x}) \quad \text{and} \quad \boldsymbol{x} = \boldsymbol{g}(\boldsymbol{x}') \,. \tag{4.18}$$

Three different example substitutions are shown in the following, which illustrate how a new coordinate system can be beneficial in terms of stability and solvability for the optimization.

Simplified constraints

A possible improvement could be, that the solution space is easier to bound, e.g. a rectangular region

$$\mathcal{C}(\boldsymbol{x}') = \{\boldsymbol{x}' : \boldsymbol{x}'_l \leq \boldsymbol{x}' \leq \boldsymbol{x}'_u\} \,. \tag{4.19}$$

where $\boldsymbol{x}_l$ and $\boldsymbol{x}_u$ are the lower and upper constraint respectively. The constraint boundary (and the solution space) is then rectified. A non-linear constraint, especially around narrow and curved solution spaces, typically cause the iterative algorithms to step outside the feasible area more often, which increases the number of required steps. Consequently, the iterations might not converge within a predefined number of steps.

Scaling

Commonly, the elements x_i of $\boldsymbol{x}$ are scaled using an expected value for each dimension $x_{\mathrm{e},i}$, i.e.

$$x'_i = \frac{x_i}{x_{\mathrm{e},i}} \,. \tag{4.20}$$

Alternatively, if the solution space is rectangular, each element can be scaled to stay in a specific interval (e.g. $x'_i \in [-1,1]$). Differently scaled elements can lead to numerical difficulties, as the algorithm reacts more sensitive to changes in one dimension than in the other, especially when using gradient information.

Separation of local minima

A change in coordinates can also help to separate the different local minima. Let L' be the loss function dependent of $\boldsymbol{x}'$ and some additional discrete parameter k constraining the function to each known local valley (containing only one minimum). Then, finding the global minimum reduces to finding the smallest of all local minima

$$\hat{\boldsymbol{x}}'_k = \arg\min_{\boldsymbol{x}'} L'(\boldsymbol{x}', k) \,, \tag{4.21}$$

where $k = 1, \ldots, n_k$ and n_k is the total number of local minima. The global minimum is then

$$\hat{\boldsymbol{x}} = \arg\min_{k} L(\boldsymbol{g}(\hat{\boldsymbol{x}}'_k)) \,. \tag{4.22}$$

4.5 Constrained optimization

Constraints (as defined in equation (4.2)) can be included into the optimization problem by adding penalty terms to the loss

$$L_c(\boldsymbol{x}) = L(\boldsymbol{x}) + \sum_{j=1}^{n_{\text{con}}} \phi(c_j(\boldsymbol{x}), \lambda)\,, \tag{4.23}$$

where ϕ denotes the penalty function, λ is a tuning parameters and scales the penalty, and L_c is the modified loss function. The function ϕ should penalize all values of $\boldsymbol{x}$ outside the feasible region. A common practice for a penalty is to use a quadratic function q, which increases when exceeding the constraint boundaries:

$$\phi(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in \mathcal{C} \\ \lambda q(\boldsymbol{x}) & \text{otherwise.} \end{cases} \tag{4.24}$$

Consequently, when the iterations step outside the feasible domain, they are pushed back by the slope. The function must be carefully scaled using λ in order to superpose the loss function and the latter must be defined outside $\mathcal{C}$. Alternatively, so-called barrier methods, prevent the iteration to step outside by superimposing a function which increases inside the region when approaching the constraint boundary. This is achieved, for example, with a logarithmic function and a singularity at the boundary.

Both methods alter the loss function and therefore potentially find a different minimum than when using no additional terms. Barrier methods work well with loss functions that are undefined outside the feasible domain. More advanced algorithms to solve constraint optimization problems exist. However, in this work, the above mentioned approaches were implemented due to their simplicity and transparency. The iterative optimizers (quasi-Newton or line search) can be used together with the penalty function without any further modification.

5. Statistics and data association

An association decision relies on statistical assumptions and methods. A statistical decision is taken when a new measurement is linked to other measurements or to an existing catalog object. The first three sections of this chapter summarize the relevant statistical theory. The content is common textbook knowledge (e.g. by Bar-Shalom et al., (2004) and Sivia, (2006)) and serves to introduce the nomenclature.

Section 5.4 addresses a typical problem which arises when using statistical methods, i.e. the proper quantification and transformation of uncertainties. The last section describes current methods and possible improvements for the task of associating new tracklets to catalog objects. The association framework builds the theoretical foundation for the later derived tracklet-to-tracklet association.

5.1 Probability density functions

A continuous probability distribution of a random variable is described by a density function $f(\boldsymbol{x})$. It is defined as the derivative of the cumulative distribution

$$f(\boldsymbol{x}) = \frac{d}{dx} F(\boldsymbol{x}) \tag{5.1}$$

where F computes the probability that the random variable is less than a given upper value $\boldsymbol{x}_u$, i.e.

$$P(\boldsymbol{x} \leq \boldsymbol{x}_u) = F(\boldsymbol{x}_u) = \int_{-\infty}^{\boldsymbol{x}_u} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\,. \tag{5.2}$$

The equations use vector notation to describe multivariate random variables, but they apply to scalar variables as well.

Two probability density functions are used within this thesis. The first one is the *multivariate normal distribution*

$$\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{C}_{\boldsymbol{x}}) \coloneqq f(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{C}_{\boldsymbol{x}}) = \frac{1}{\sqrt{\det(2\pi \boldsymbol{C}_{\boldsymbol{x}})}} \exp\{-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{x}}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}})\}\,, \tag{5.3}$$

which describes the probability of a $n_{\boldsymbol{x}}$-dimensional $\boldsymbol{x}$ given the mean $\bar{\boldsymbol{x}}$, and the covariance matrix $\boldsymbol{C}_{\boldsymbol{x}}$. The det operator computes the determinant of the matrix. The function is used, e.g., to describe the orbital state estimates and measurement noise. The mean and covariance of the normal distribution are the first and second moments of the density function, i.e. $\bar{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}]$ and $\boldsymbol{C}_{\boldsymbol{x}} = \mathrm{E}[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}]$, where the expectation for continuous random variables is defined as

$$\mathrm{E}[\boldsymbol{x}] = \int_{-\infty}^{\infty} \boldsymbol{x}\, f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\,. \tag{5.4}$$

In the univariate case, the function is described by a standard deviation σ and mean μ

$$f(x, \bar{x}, \sigma_x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\{-\tfrac{1}{2} \frac{(x-\bar{x})^2}{\sigma_x^2}\} . \tag{5.5}$$

The other used function is the *chi-squared distribution* for a scalar random variable x

$$f(x, n_{\boldsymbol{x}}) = \frac{1}{2^{\frac{n_{\boldsymbol{x}}}{2}} \Gamma(\frac{n_{\boldsymbol{x}}}{2})} x^{\frac{n_{\boldsymbol{x}}}{2}-1} e^{-\frac{x}{2}} \quad \text{for} \quad x \geq 0 , \tag{5.6}$$

where Γ denotes the Gamma function and $n_{\boldsymbol{x}}$ the degrees of freedom. The distribution can be related to the multivariate normal distribution, i.e. if an $n_{\boldsymbol{x}}$-dimensional random variable is normally distributed as described by (5.3), the scaled quadratic distance

$$d^2(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{x}}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) \sim \chi^2(n_{\boldsymbol{x}}) \tag{5.7}$$

is distributed ($\sim$) according to the chi-squared distribution (χ^2) with $n_{\boldsymbol{x}}$ degrees of freedom (Bar-Shalom et al., 2004, pp. 57–58). The quantity d^2 is used to describe the statistical distance to or from a multivariate normal distribution and its square root is commonly denoted as the Mahalanobis distance (Mahalanobis, 1936). The two density functions are illustrated in Figure 5.1 and Figure 5.2 along with their relationship. The first figure shows a bivariate normal distribution for $\boldsymbol{x} = (x_1, x_2)^{\mathsf{T}}$. Three iso-lines indicate a certain constant distance d^2 from the distribution and contain a certain fraction of the overall volume under the function. The most inner one comprises $\sim$ 68.27 % of the volume, the middle one $\sim$ 95.45 %, and the most outer one $\sim$ 99.73 %. The chi-squared distribution with $n_{\boldsymbol{x}} = 2$ is presented in Figure 5.2. The distance values used for the iso-lines in Figure 5.1 are depicted with the vertical lines. The selected percentages correspond to the σ-boundaries commonly used with the univariate normal distribution, i.e. the interval $[-1\sigma, 1\sigma]$ covers $\sim$ 68.27 % of the distribution ($2\sigma \approx 95.45$ %, $3\sigma \approx 99.73$ %).

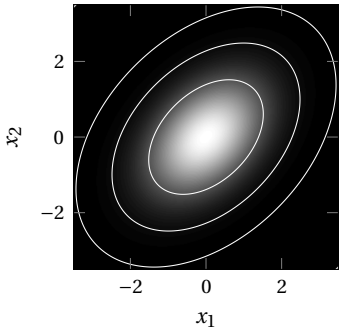


Figure 5.1: *Bivariate normal distribution of $\boldsymbol{x} = (x_1, x_2)^{\mathsf{T}}$. The contours (ellipses) depict lines of constant density (equivalently constant d^2 distance).*

The bounding distances, from here on after called gates, are typically defined accounting for a certain significance. With the known density function (5.6), gates are selected to cover a desired percentile, e.g. P = 95%, using the cumulative distribution function $P(x \leq x_u) = F(x_u, n)$. The upper bound is computed by inversion

$$x_u = F^{-1}(P, n) . \tag{5.8}$$

Common gates for different significance values are listed in the table below. The values are listed for $n_{\boldsymbol{x}} = 2$ in accordance with Figure 5.1 and Figure 5.2. Additionally, the values for $n_{\boldsymbol{x}} = 4$ are shown as these are mostly used for the association tasks in this work.

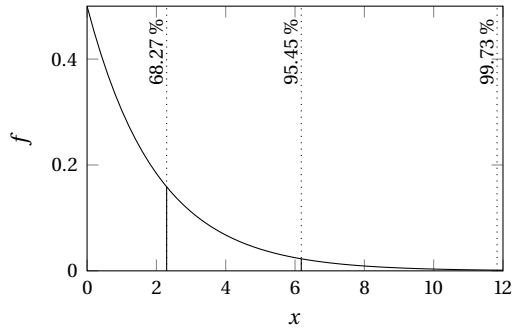


Figure 5.2: *Chi-square distribution f of x with 2 degrees of freedom. The vertical lines bound the shown percentage of the distribution from the right (corresponding to the contour lines in Figure 5.1).*

Table 5.1: *Common gates for the chi-squared distribution with 2 and 4 degrees of freedom.*

P[%]	10	20	30	40	50	60	70	80	90	95	99
$x_u(n_{\boldsymbol{x}}=2)$	0.21	0.45	0.71	1.02	1.39	1.83	2.41	3.22	4.60	5.99	9.21
$x_u(n_{\boldsymbol{x}}=4)$	1.06	1.65	2.20	2.75	3.36	4.04	4.88	5.99	7.78	9.49	13.28

5.2 Bayesian estimation

The joint probability density describing the distribution of two variables $f(\boldsymbol{x}, \boldsymbol{z})$ can be described in terms of the conditional probability (likelihood of one variable given the other)

$$f(\boldsymbol{x}, \boldsymbol{z}) = f(\boldsymbol{x}|\boldsymbol{z})\, f(\boldsymbol{z}) = f(\boldsymbol{z}|\boldsymbol{x})\, f(\boldsymbol{x})\,. \tag{5.9}$$

In order to introduce a common nomenclature, $\boldsymbol{x}$ denotes now the estimated parameter and $\boldsymbol{z}$ the measurements or data. Starting from (5.9), Bayes' theorem is derived accounting additionally for the underlying model M (background information such as a dynamical and/or measurement model)

$$f(\boldsymbol{x}|\boldsymbol{z}, M) = \frac{f(\boldsymbol{z}|\boldsymbol{x}, M)\, f(\boldsymbol{x}, M)}{f(\boldsymbol{z},\, M)}\,. \tag{5.10}$$

The *posterior density* $f(\boldsymbol{x}|\boldsymbol{z}, M)$ then describes the probability of the parameter $\boldsymbol{x}$ given the measurements and the model. The *likelihood* of the measurements $f(\boldsymbol{z}|\boldsymbol{x}, M)$ accounts for the data noise. If *prior* knowledge on the parameters is available, it is incorporated in $f(\boldsymbol{x}, M)$. Lastly, $f(\boldsymbol{z},\, M)$ is referred to as the *evidence* and can help to select an appropriate model (Sivia, 2006). When estimating the most probable $\hat{\boldsymbol{x}}$, the evidence can be neglected as it is independent of the parameters. Hence, it is omitted from the following derivations along with the model M.

The mode which maximizes the posterior density function is denoted as the maximum a-posteriori (MAP) estimate

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{z}|\boldsymbol{x})\, f(\boldsymbol{x})\,. \tag{5.11}$$

If no prior knowledge is available, the maximum likelihood estimate (MLE) is searched for

$$\hat{\boldsymbol{x}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{z}|\boldsymbol{x})\,. \tag{5.12}$$

In both cases, the probability functions are maximized to get the most likely state estimate. Chapter 4 provides the necessary tools for such optimization problems. The estimation process in case of univariate density functions is illustrated in Figure 5.3. The unscaled posterior function (depicted with the dashed line) is obtained by multiplication of prior and likelihood. The scaling factor (evidence) is unknown but the $\hat{x}_{\text{MAP}}$ estimate can still be determined. A first practical estimation problem is provided in the next section. The presented methodology is repeatedly used in the following chapters.

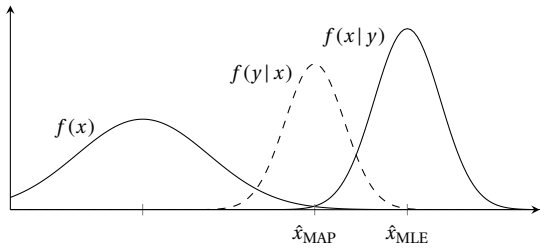


Figure 5.3: *Illustration of example estimation results for a given scalar prior $f(x)$, likelihood $f(x|y)$, and posterior density function $f(y|x)$. The maximum a posteriori $\hat{\boldsymbol{x}}_{MAP}$ and maximum likelihood estimate $\hat{\boldsymbol{x}}_{MLE}$ are indicated on the bottom axis.*

5.3 Classification

The association of a new observation $\boldsymbol{z}$ is interpreted as a classification based on Bayes' theorem. Each object is represented with a class name c_i, where the index identifies the object number. In principle, prior knowledge $f(c_i)$ on object probability can be included, e.g. when performing a survey for geostationary objects, an observation of an object in a low-Earth orbit is unlikely. Following (5.10), the probability of each object candidate given the new measurement is expressed proportionally with

$$f(c_i|\boldsymbol{z}) \propto f(\boldsymbol{z}|c_i)f(c_i)\,. \tag{5.13}$$

In the simple case, each object is equally likely and modeled with the same probability. Consequently, the prior is neglected in the following.

Each object is attributed with a state vector at an epoch and additional model parameters. The parameters and their uncertainty are described with the multivariate normal distribution, i.e. the mean vector $\boldsymbol{x}_i$ and the covariance matrix $\boldsymbol{C}_{\boldsymbol{x},i}$. The measurement noise is assumed to be normally distributed as well with the mean $\boldsymbol{z}$ and covariance $\boldsymbol{C}_{\boldsymbol{z}}$ (see section 3.6). When comparing the two different functions, first a common frame must be determined. One possible frame is the observation space, hence, the object state density functions are transformed with a model function $\boldsymbol{h}$. The outcome of this function is denoted as the modeled observation. The modeled observations are random variables as well and can be described again with a normal distribution (mean $\tilde{\boldsymbol{z}}_i$ and covariance $\boldsymbol{C}_{\tilde{\boldsymbol{z}},i}$). This representation is not always a valid choice. The transformation of density functions and resulting limitations are explained in the following section 5.4.

The probability density function of the difference between two independent normally distributed random variables $\Delta\boldsymbol{z}_i = \boldsymbol{z} - \tilde{\boldsymbol{z}}_i$ is given by the cross-correlation (equivalently convolution) between both

describing density functions (Sivia, 2006, pp. 72–73)

$$f(\Delta \boldsymbol{z}_i) = \int_{-\infty}^{\infty} f(\boldsymbol{z}') f(\tilde{\boldsymbol{z}}_i' + \Delta \boldsymbol{z}_i)\, \mathrm{d}\boldsymbol{z}' \,. \tag{5.14}$$

In case of two normally distributed variables (as assumed here), a closed expression for the distance distribution can be derived as

$$f(\Delta \boldsymbol{z}_i) = \mathcal{N}(\Delta \boldsymbol{z}_i, \boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i}) \,. \tag{5.15}$$

The likelihood of the measurements given an object c_i is modeled using this function, thus, $f(\boldsymbol{z}|c_i) = f(\Delta \boldsymbol{z}_i)$. The goal of the classification process is to find the object which maximizes the likelihood.

$$\hat{c}_i = \arg\max_{c_i} \mathcal{N}(\Delta \boldsymbol{z}_i, \boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i}) \,. \tag{5.16}$$

When dealing with normal distributions, often the negative logarithm is minimized instead, i.e.

$$\hat{c}_i = \arg\min_{c_i} -\ln f(\boldsymbol{z}|c_i) \tag{5.17}$$

$$= \arg\min_{c_i} \tfrac{1}{2} \ln\left\{\det(2\pi(\boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i}))\right\} + (\boldsymbol{z} - \tilde{\boldsymbol{z}}_i)^{\mathsf{T}} (\boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i})^{-1} (\boldsymbol{z} - \tilde{\boldsymbol{z}}_i) \,, \tag{5.18}$$

where the first term prunes large uncertainties of modeled measurements and the second term is the statistical distance from (5.7).

All objects must be tested and the one which comes closest, i.e. will produce the smallest value in (5.17), is associated. However, if catalog objects are closely-spaced and have large uncertainties in their states, the association decision becomes sensitive to small errors and is ambiguous. An ambiguous association is illustrated in Figure 5.4, where three modeled observations $\tilde{\boldsymbol{z}}_i = (\tilde{\alpha}_i, \tilde{\delta}_i)^{\mathsf{T}}$ corresponding to the objects c_i are shown along with difference covariances $\boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i}$. The covariances are depicted with lines of constant probability (as in Figure 5.1). The new observation $\boldsymbol{z}$ (white square) is associated to the second object c_2 in this example. However, it is also inside the significant domain of object c_1 (indicated by the $3\,\sigma$-line around $\tilde{\boldsymbol{z}}_1$ covering 99.73 % of the distribution).

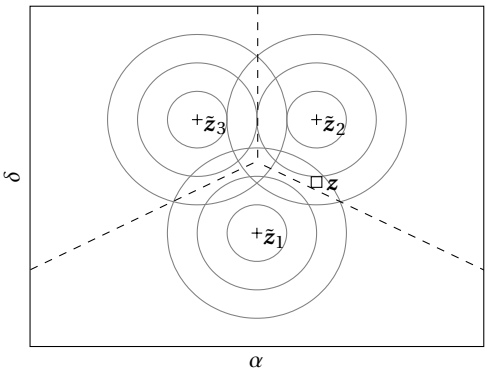


Figure 5.4: *Bayesian classification for a new observation $\boldsymbol{z}$ (white square). Three classes (catalog objects) with modeled observations $\tilde{\boldsymbol{z}}_i$. Each likelihood function is represented with the σ-lines (explained in the text). Dashed lines show the boundaries between the different classifiers.*

Thus, the general association decision should be based on whether the difference $\Delta \boldsymbol{z}_i$ is in a statistically significant part of the distribution. As described in section 5.1, certain significance intervals can be selected for the multivariate distribution with the d^2-distance. Using (5.7) and (5.15), the test quantity is computed

by

$$d^2 = (\boldsymbol{z} - \tilde{\boldsymbol{z}}_i)^\mathsf{T} (\boldsymbol{C}_{\boldsymbol{z}} + \boldsymbol{C}_{\tilde{\boldsymbol{z}},i})^{-1} (\boldsymbol{z} - \tilde{\boldsymbol{z}}_i) \,. \tag{5.19}$$

Other distance measures between probability density functions exist, e.g. *Kullback-Leibler information* or *Bhattacharyya's distance* (Everitt and Skrondal, 2010, pp. 44,239). DeMars, (2010) and DeMars and Jah, (2011) provide an overview on similarity measures used for space object cataloging tasks (e.g. sensor management) and derive closed expressions for the multivariate normal density function. Gadaleta et al., (2012) test different measures for the special case of tracklet-to-tracklet association.

If the statistic is below a predefined threshold (see Table 5.1), the measurement is considered likely to be generated by the object c_i. Hence, the decision is now based on an individual assessment of each object. However, as also indicated in the figure, multiple objects may pass the gate leading to multiple associations of the same observation. The ambiguous association is particularly challenging for observations of satellites flying in formations. One way of overcoming the ambiguities is to associate loosely with an association weight. This follows the multi-hypothesis filtering approach as described e.g. in (Bar-Shalom et al., 2004). The individual likelihood is scaled with the overall association likelihood and then used as the weight for each hypothesis. If a measurement is erroneously mistagged, it results in a false posterior state after an orbit improvement step. Thus the ambiguity might be resolved after further measurements are obtained which cannot be associated anymore to the false candidates. The focus of the thesis lies on the individual association. The efficient handling of a multi-hypotheses association problem is shortly discussed Chapter 10.

5.4 Uncertainty transformation

The realistic representation of uncertainties is required in various space object catalog applications: the association of tracklets during the catalog build-up, but also, for instance, when computing collision probabilities between two satellites. The focus here lies on the first problem, nevertheless, the methods are generally used wherever density functions must be transformed.

When transforming a random variable through a function $\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x})$, the result is also a random variable and thus also described by some density function. The mean $\bar{\boldsymbol{x}}$ and covariance $\boldsymbol{C}_{\boldsymbol{x}}$ are assumed to properly describe the initial distribution. Indices and accents from previous sections are omitted to reduce notational clutter. The distribution of the transformed variable can be found by applying a linear approximation of the transformation, i.e. using a first order expansion around the mean value

$$\boldsymbol{h}(\boldsymbol{x}) \approx \boldsymbol{h}(\bar{\boldsymbol{x}}) + \boldsymbol{J}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad \text{,where} \quad \boldsymbol{J} = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}} . \tag{5.20}$$

and inserting it into the definition of mean and covariance

$$\bar{\boldsymbol{z}} = \mathrm{E}[\boldsymbol{z}] = \boldsymbol{h}(\bar{\boldsymbol{x}}) \quad \text{and} \quad \boldsymbol{C}_{\boldsymbol{z}} = \mathrm{E}[(\boldsymbol{z} - \bar{\boldsymbol{z}})(\boldsymbol{z} - \bar{\boldsymbol{z}})^\mathsf{T}] = \boldsymbol{J}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{J}^\mathsf{T} . \tag{5.21}$$

The distribution of the transformed variable is also described with the normal distribution under linear assumptions. If $\boldsymbol{h}$ is nonlinear around $\bar{\boldsymbol{x}}$, the transformation can be a poor approximation.

Alternatively, Julier, (2002) proposes to approximate the resulting density directly instead of approximating the transformation. The so-called *unscented transform* samples the initial distribution with sigma-points $\boldsymbol{\mathcal{X}}_i$, where $i = 1, \ldots, n_{\boldsymbol{\mathcal{X}}}$. Each point is individually transformed with

$$\boldsymbol{\mathcal{Z}}_i = \boldsymbol{h}(\boldsymbol{\mathcal{X}}_i)\,. \tag{5.22}$$

The sample mean and covariance are computed with

$$\bar{\boldsymbol{z}} = \sum_{i=1}^{n_{\boldsymbol{\mathcal{X}}}} w_i \boldsymbol{\mathcal{Z}}_i \quad \text{and} \quad \boldsymbol{C}_{\boldsymbol{z}} = \sum_{i=1}^{n_{\boldsymbol{\mathcal{X}}}} w_i (\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})(\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})^{\mathsf{T}}\,, \tag{5.23}$$

where w_i denote the weights for each sigma-point and $\sum_{i=1}^{n_{\boldsymbol{\mathcal{X}}}} w_i = 1$. Different schemes to define the sigma-points and weights have been developed (Julier, 2002). Typically, sigma-points are selected along the eigenvectors of the covariance matrix or the column vectors of its Cholesky decomposition. More details can be obtained from (Julier, 2002; Julier and Uhlmann, 1997). The unscented transform will be referred to as a non-linear transformation in the following to distinguish it from the one using the linear transformation.

The mean and covariance of the transformation result can be also estimated by randomly sampling the distribution. They are computed as well with (5.23), where all samples are equally weighted ($1/n_{\boldsymbol{\mathcal{X}}}$ for large samples). A considerable amount of methods has been published on the analysis of density samples. For instance, the samples can also be used to estimate higher order moments of the distribution such as skewness and kurtosis (Everitt and Skrondal, 2010, pp. 397, 239). These moments provide information on the geometry of the transformation, e.g. they can indicate some deformation w.r.t. to the normal distribution. A large skewness indicates a lack of symmetry, while the kurtosis describes the peakiness or flatness of the density function. Mardia, (1970) developed a formulation for the multivariate skewness

$$\beta_1 = \frac{1}{n_{\boldsymbol{\mathcal{X}}}^2} \sum_{i=1}^{n_{\boldsymbol{\mathcal{X}}}} \{(\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{z}}^{-1} (\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})\}^3 \tag{5.24}$$

and kurtosis

$$\beta_2 = \frac{1}{n_{\boldsymbol{\mathcal{X}}}} \sum_{i=1}^{n_{\boldsymbol{\mathcal{X}}}} \{(\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{z}}^{-1} (\boldsymbol{\mathcal{Z}}_i - \bar{\boldsymbol{z}})\}^2 \tag{5.25}$$

He used the terms to test if a set of samples originate from a normal distribution. As the multivariate normal distribution is symmetric w.r.t. its major axes, the skewness β_1 is 0. Mardia's kurtosis β_2 is $n_{\boldsymbol{z}}(n_{\boldsymbol{z}}+2)$ for a multivariate normal distribution, where $n_{\boldsymbol{z}}$ is the dimension of $\boldsymbol{z}$. The kurtosis is normalized to β_2' in the following by subtracting the expected value for a normal distribution from β_2. A deviation from this value thus indicates that the function is more peaky or flatter than the normal distribution.

In order to illustrate the transformation methods and also their shortcomings, the deformation of a satellite state uncertainty through orbit propagation is demonstrated. The satellite motion is reduced to the planar case, i.e. observing just along-track and in-radial uncertainty. This has the advantage that unfavorable effects of the transformation are easier to visualize. Starting from the apogee, a satellite state located on a geostationary transfer orbit ($a = 24{,}500$ km, $e = 0.7$, $i = 0°$), is propagated for 1.5 orbital periods assuming two-body dynamics. The state is then located at the perigee of the orbit, which is intentionally selected to increase the deformation effect. State and covariance are defined in terms of

satellite-centered coordinates

$$\boldsymbol{y} = (R, T, N, \dot{R}, \dot{T}, \dot{N})^{\mathsf{T}} \quad \text{and} \quad \boldsymbol{C}_{\boldsymbol{y}} = \operatorname{diag}\left(\sigma_R^2, \sigma_T^2, \sigma_N^2, \sigma_{\dot{R}}^2, \sigma_{\dot{T}}^2, \sigma_{\dot{N}}^2\right). \tag{5.26}$$

The local satellite-centered reference frame is defined as follows: the main axis is tangential to the orbit (along the velocity-vector), the second one is parallel to orbit normal N, and the last is in-radial R direction and forms the orthonormal basis. The initial distribution is described with the mean $\boldsymbol{y}_1$ and the diagonal terms $\sigma_{R,1} = \sigma_{T,1} = \sigma_{T,1} = 10\,\text{km}$, $\sigma_{\dot{R},1} = 1\,\text{m/s}$, $\sigma_{\dot{T},1} = 3\,\text{m/s}$, $\sigma_{\dot{N},1} = 1\,\text{m/s}$. The distribution defined by mean and covariance is randomly sampled with $n_{\boldsymbol{\mathcal{X}}} = 5000$. The samples along with the σ-lines are shown in Figure 5.5. The distribution is projected onto the $R \times T$ plane for better visualization.

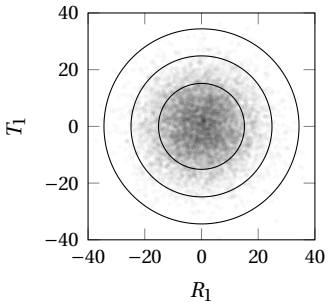


Figure 5.5: *Initial distribution of satellite state samples and σ-lines of the describing normal distribution projected on the $R \times T$-frame.*

After propagating each sample point for the specified time interval, the deformed distribution can be observed in Figure 5.6. The σ-lines of the projected distribution are again shown, once for the linear transformation (left side) and once for the sigma-point transformation (right side). Both distributions are poor approximations of the actual one. The distribution obtained from the linear transformation properly describes the actual samples around the mean, but deteriorates as soon as the distribution bends. The one obtained from the unscented transform covers most sample points, but overestimates the size of the distribution.

Instead of visually comparing the samples and distribution in state space, the distance measure from (5.7) can be used to see how much the sample distribution deviates from the theoretical chi-squared distribution. Figure 5.7 shows the histogram of the d^2-values calculated from the samples using the covariance and mean from each of the two transformation methods. The figures show the same result as before. The d^2-values from the linear approximation show a good match to the theoretical function close to zero (i.e. close to the mean). However, a the amount of samples with large d^2-values is larger than

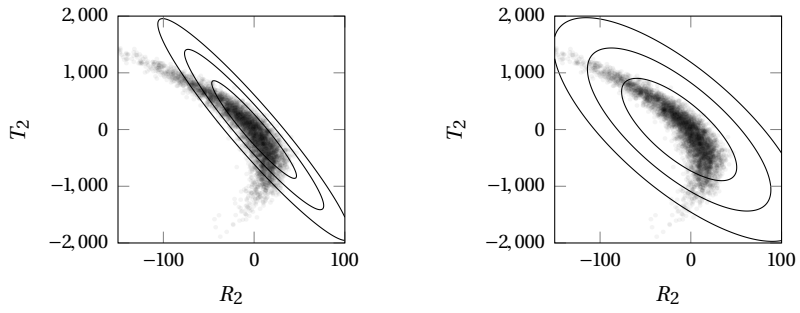


Figure 5.6: *Propagated satellite state samples propagated for 1.5 orbital periods in cartesian object-centered coordinates. Left side shows the σ-isolines of the linearly transformed function, while right side shows the result of the non-linear transformation.*

expected. The dashed line represents the gate where theoretically 95% of the samples should be covered ($P = 95\%$ and $d_u^2 \approx 5.99$ according to Table 5.1, where the upper boundary x_u is denoted with d_u^2 in the following for consistency). In this example, $\sim 24\%$ of the samples fall outside of the gate, instead of 5% that should be. This effectively causes missed associations.

The sample distance values, obtained when using the non-linear transform, also resemble the impression from Figure 5.6. More samples lie within the region close to the mean value than they should (small d^2 values). However, for larger d^2-values the distribution matches the theoretical one well. The number of samples above the previously defined gate accumulate to 4.9%. The overestimation of the significant region size comes with certain disadvantages. In the presence of closely-spaced objects, this covariance expansion can lead to multiple or false associations (as illustrated in Figure 5.4). The histograms are again used in Chapter 8 to identify problems with the uncertainty transformations of the two proposed formulations for the tracklet-to-tracklet association.

While some transformation can be properly transformed using the two described methods, others cannot. If the satellite is propagated for a short duration, its uncertainty is mostly representable by the transformation, while when propagating for larger durations it is not necessarily. Additionally, the success of a simplified transformation depends on the initial covariance. A small covariance allows larger propagation periods. The histogram provides an assessment of this validity for a given transformation and propagation time. The time dependency of the covariance deformation is visualized using the 95% association gate and the multivariate skewness and kurtosis as defined in (5.24) and (5.25). Figure 5.8 shows the time series of accepted samples points, i.e. $d_i^2 < d_u^2$ ($P = 95\%$) for $i = 1, \ldots, n_{\boldsymbol{\mathcal{X}}}$, once for the linear and for the unscented transform. The state uncertainty is properly transformed for most time intervals, but around the perigees. The situation deteriorates with larger propagation periods. The unscented transform captures the right percentage of the distribution even around the perigee. The skewness and kurtosis on the right side of the figure again illustrate the deformation of the distribution.

To overcome these issues, various researchers propose alternative uncertainty representations, such as *Gaussian mixtures* as done by DeMars, (2010), calculating higher order moments using *state-transition tensors* (Fujimoto et al., 2012), using the so-called *Gauss-von-Mises* distribution as in (Horwood and Poore, 2014), or using polynomial approximations (Jones et al., 2013). The different approaches typically require additional computational resources and/or a rewriting of well-established orbit propagation methods. The Gaussian mixture approximation as proposed by DeMars, (2010) detects the non-linearity of the transformation from the dynamics, i.e., since orbit propagation is approximately volume-preserving, the non-linearity is detected when the volume of the covariance changes after using the unscented transform. Alternatively, Faubel et al., (2009) detects the non-linearity by determining an asymmetry in the transformed sigma-points.

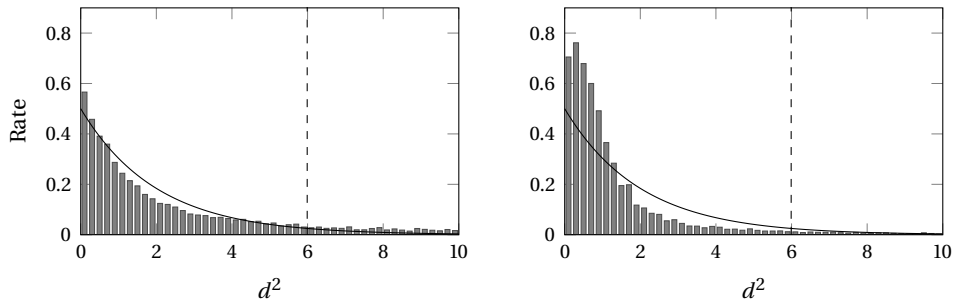


Figure 5.7: *Theoretical chi-squared distribution (solid line) and actual distribution of d^2-values for the linear (left) and non-linear transform (right) using cartesian coordinates. The vertical dashed line represents the $P = 95\%$ gate.*

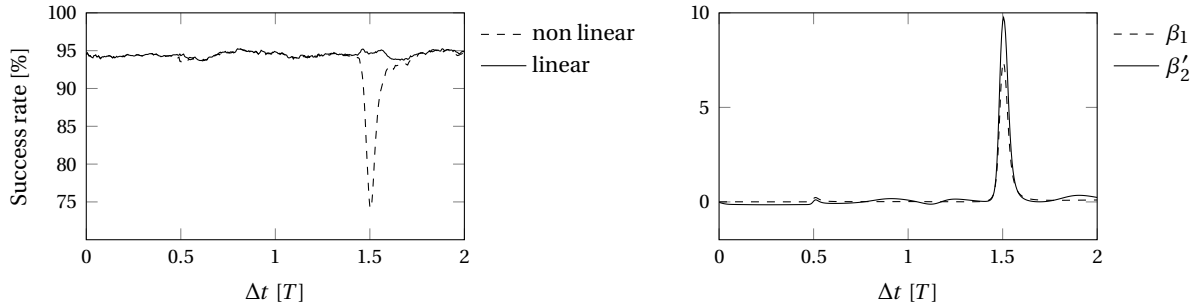


Figure 5.8: *Change of uncertainty quantification errors in cartesian state space over two revolutions. The parameters of the tested orbit are provided in the text along with further details. Left side shows percentage of samples passing the gate ($P = 0.95\%$) for the transformed covariance from the linear and non-linear transform. The right side shows Mardia's skewness (β_1) and the normalized kurtosis β_2' over the same time period.*

The initial multivariate normal distribution is split whenever a non-linearity is detected during the transformation. The initial function is thus, after dividing it multiple times, represented as a sum of weighted normal distributions. The weights and components are determined by an optimization scheme guaranteeing good agreement with the single initial multivariate distribution. Each component is then individually transformed using either the non-linear or linear transform. As the covariance volumes reduce with each subdivision, the linear assumptions become valid at some point. Arbitrary shapes and deformations can be covered with this formulation if enough components are used. Figure 5.9 shows the transformed distribution of the example problem using 100 components. The initial distribution is adaptively split using the approach by Faubel et al., (2009).

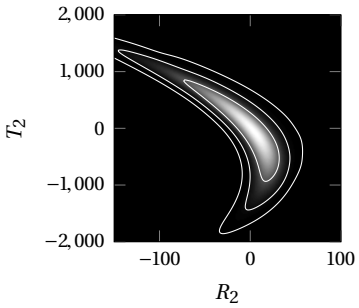


Figure 5.9: *Transformed probability density function using an adaptively split Gaussian-mixture distribution.*

Working with single normal distributions comes with advantages such as computational efficiency, simplicity, and consequently transparency and implementation robustness. Instead of searching for an appropriate representation, a different coordinate system can be used in which the transformation does not become non-linear. Coordinates are favorable if the uncertainty remains almost normally distributed over longer time-spans. Hill et al., (2012) compare different coordinate systems for their effectiveness in radar-track association. Hill et al., (2008) propose to use *object-centered curvilinear coordinates* when correlating radar-tracks to each other and to catalog objects. They use the Mahalanobis distance as in (5.7), to compare two states. Their analysis contains also orbital elements, which show promising performance as well. The mentioned curvilinear coordinates, however, allow a separation of position and velocity, which helps to visualize the covariance. A similar formulation of the coordinates, as described by Vallado and Alfando, (2014), is illustrated in Figure 5.10. They use the coordinates to describe relative motion of an interceptor and target. The coordinate system is defined by a reference position $\tilde{\boldsymbol{r}}$ and velocity $\dot{\tilde{\boldsymbol{r}}}$. When describing the position $\boldsymbol{r}$ in terms of the reference, the along-track arc length s_T, cross-track arc length s_N,

and radial distance s_R are used. The values are computed using auxiliary positions $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$. The vector $\boldsymbol{r}_a$ is obtained by projecting $\boldsymbol{r}$ onto the reference orbit. $\boldsymbol{r}_b$ is calculated from by rotating $\boldsymbol{r}_a$ to be aligned with $\boldsymbol{r}$.

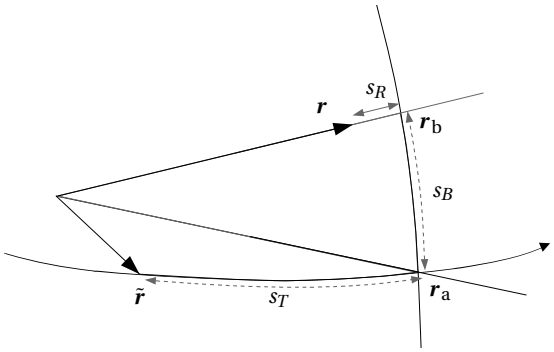


Figure 5.10: *Individual components of curvilinear coordinates s_T (tangential), s_B (cross-track or binormal), and s_N (normal or in-radial). Illustration adapted from (Vallado and Alfando, 2014).*

Figure 5.11 shows the propagated random states in the curvilinear coordinate system along with σ-lines of the linear transformed covariance matrix. Additionally, the histogram of the d^2-values, the time-series of samples passing the 95%-gate, and the skewness and kurtosis are shown in Figure 5.12 and Figure 5.13. The gated samples now perfectly match the theoretical chi-squared distribution. The good representation also holds for other epochs of the propagation time interval as illustrated by β_1 and β_2'.

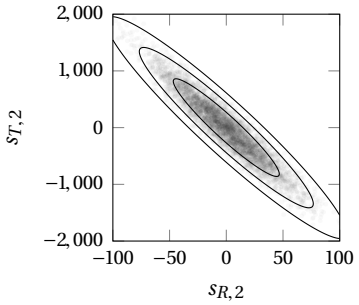


Figure 5.11: *Transformed samples in curvilinear frame and $3\,\sigma$ lines illustrating the normal density function.*

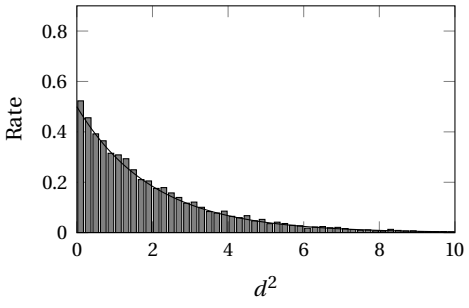


Figure 5.12: *Theoretical chi-squared distribution (solid line) and actual distribution of d^2-values for the linear transform using object-centered curvilinear coordinates.*

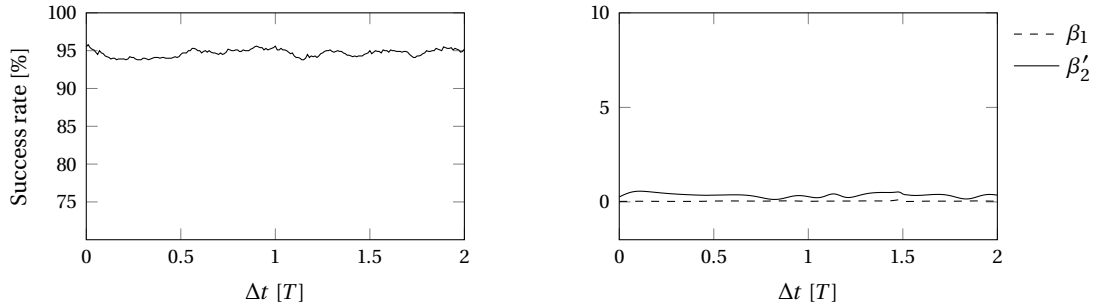


Figure 5.13: *Change of uncertainty quantification errors in curvilinear coordinates over two revolutions. The same tested orbit from Figure 5.8 is propagated for the analysis. The left side shows the percentage of samples passing the gate ($P = 0.95\%$) for a covariance from a linear transform. The right side shows Mardia's skewness (β_1) and the normalized kurtosis (β_2').*

5.5 Association of tracklets to catalog objects

The previous section discusses the difficult representation of state uncertainty propagation. An association of state vectors to each other becomes relevant if the new observation provides enough information to determine an orbital state, e.g. when correlating radar-tracks or long optical observation arcs. In this work, the focus lies on short-arc measurement association. The tracklet provides only information on the angles and angular-rates (cf. section 3.6). As proposed in section 5.4, the catalog state is propagated to the observation epoch, where it is transformed into the observation space. Thus, the function $\boldsymbol{h}$ returns the angles and angular rates at the observation epoch. The association decision is then based on the distance metric (5.19). In principle, angles and angular rates form a curvilinear frame as well, i.e. they should better represent the along-track deviation than cartesian coordinates. However, the transformation can nevertheless introduce deformations as illustrated in the following.

The analysis from the previous section is repeated for the association of tracklets, i.e. $n_{\boldsymbol{\mathcal{X}}} = 5000$ samples are randomly drawn from the initial catalog state distribution (same covariance as before in satellite-centered reference frame) and propagated for a certain time interval. Instead of propagating a satellite on the transfer orbit, the more likely case of a geostationary one is selected. This assures a feasible observation geometry for the whole propagation period (neglecting other visibility conditions). The time-series of the association performance using the 95%-gate is shown in Figure 5.14. The right side shows the time-series of the β_1 and β_2' quantities. Similarly as for the cartesian state coordinates, the distribution deforms significantly after a certain propagation time. The conclusion is drawn from the difference between linear and non-linear transformation as well as from the increase in kurtosis and skewness.

The distribution of the samples for a critical epoch, i.e. local maximum of the deformation, at $\Delta t = 5.48$ days is shown in Figure 5.15. The figure visualizes the deformation in the angular projection of the uncertainty distribution, while the time-series in Figure 5.14 uses the full angles and angular rates covariance in the calculation of the quantities. Thus, the same limitations of the approximate transformation apply to angles and angular rates as for the cartesian coordinate system. This difficulty needs to be considered when developing a tracklet association method. The implications of an erroneous representation can be observed in Chapter 8.

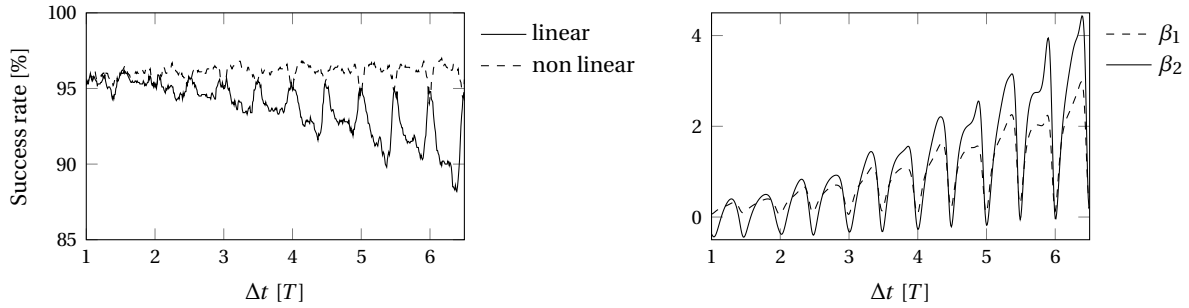


Figure 5.14: *Change of uncertainty quantification errors in angles and angular rates over six revolutions. The parameters of the tested orbit are provided in the text along with further details. Left side shows percentage of samples passing the gate ($P = 0.95\%$ corresponding to $d_u^2 = 9.49$) for the transformed covariance from the linear and non-linear transform. The right side shows Mardia's skewness (β_1) and the normalized kurtosis β_2'.*

As in the example from section 5.4, another coordinate system can be beneficial when comparing the catalog state with the observation. Früh and Schildknecht, (2009) suggest to transform the observation into state space by assuming an equal geocentric distance of catalog state and observed state. A comparison in the state space can improve the performance, as the uncertainty in the observed quantities is usually smaller than the one from the state. Consequently, a transformation of the observation covariance can be less lossy as the other way around. If the covariance in state space describes the uncertainty in a realistic way (e.g. when using orbital elements or object-centered curvilinear coordinates), the association can thus become more robust. Robustness implies not only that most observations are successfully associated to the respective catalog object, but also that the association performance can be properly calibrated.

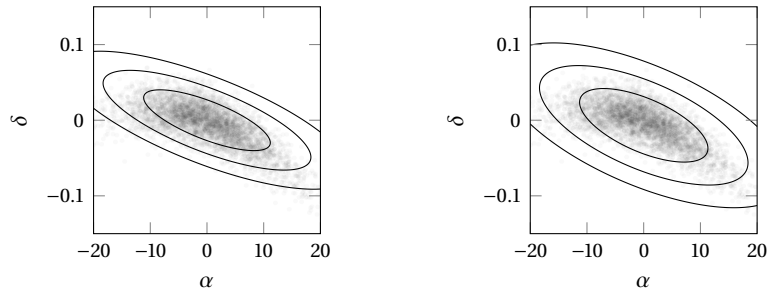


Figure 5.15: *Satellite state samples propagated for* 5.48 *days in angular coordinates. Left side shows the σ-isolines of the linearly transformed function, while right side shows the result of the non-linear transformation.*

6. Orbit determination

Determining unknown parameters describing an orbit from a set of measurements is an inversion problem. Given a set of measurements, which constrain at least as many degrees of freedom as there are unknown parameters, an orbit can be determined. The different orbit determination tasks which have to be performed for cataloging space debris are: initial orbit determination and orbit improvement. The latter is explained first in this chapter as it motivates the need for good initial states. Afterward, different preliminary orbit determination methods and their difficulties with short optical arcs are presented. Finally, possible ways to overcome the difficulties by including restricting assumptions are provided.

6.1 Orbit improvement

As outlined in the previous sections, an orbital state $\boldsymbol{y}$ is defined using six parameters (here a cartesian state or equivalently orbital elements). Typically, force model parameters $\boldsymbol{p}$ are unknown (unless provided by satellite operators) and are estimated from the data as well. Additional parameters $\boldsymbol{q}$ (e.g. biases) are introduced in the estimation process to overcome measurement model deficiencies. The estimated orbital parameters are then summarized in the variable

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix}. \tag{6.1}$$

The observations $\boldsymbol{z}$ are described with the function $\boldsymbol{h}$ by

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{6.2}$$

given the orbit parameters and accounting for the presence of noise $\boldsymbol{\varepsilon}$. Measurements in the context of this thesis are angles only (α, δ), or angles and angular-rates $(\dot{\alpha}, \dot{\delta})$.

Similar to the derivations in Chapter 5 and Chapter 7, the inversion of equation (6.2) is derived as a Bayesian estimation process. The explanation here follows the book of Bar-Shalom et al., (2004), but also borrows nomenclature from (Beutler, 2004) and (Montenbruck and Gill, 2000). Starting with Bayes' theorem (5.10), the posteriori probability density function of the orbit $f(\boldsymbol{x}|\boldsymbol{z})$ can be computed from

$$f(\boldsymbol{x}|\boldsymbol{z}) = \frac{f(\boldsymbol{z}|\boldsymbol{x})f(\boldsymbol{x})}{f(\boldsymbol{z})}, \tag{6.3}$$

where $f(\boldsymbol{z}|\boldsymbol{x})$ is the likelihood of the measurements, $f(\boldsymbol{x})$ the prior knowledge, and $f(\boldsymbol{z})$ the evidence. Analogously to the derivations in section 5.2, the maximum a-posteriori estimate is given by

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{z}|\boldsymbol{x})f(\boldsymbol{x}), \tag{6.4}$$

assuming prior information on the orbit is available (e.g. from the satellite catalog). If no a-priori density function is known, the maximum likelihood is estimated instead, i.e.

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{z}|\boldsymbol{x}) \,. \tag{6.5}$$

It is now assumed that $\boldsymbol{\varepsilon}$ and $\boldsymbol{x}$ are normally distributed according to (5.3) with zero-mean noise ($\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and covariance $\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}}] = \boldsymbol{C}_{\boldsymbol{\varepsilon}}$)

$$f(\boldsymbol{\varepsilon}) = \mathcal{N}(\boldsymbol{\varepsilon}; 0, \boldsymbol{C}_{\boldsymbol{\varepsilon}}) \,. \tag{6.6}$$

The likelihood function is then computed with

$$f(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{C}_{\boldsymbol{\varepsilon}}) \tag{6.7}$$

If the orbital state $\boldsymbol{x}_a$ of a catalog object is known along with its corresponding covariance $\boldsymbol{C}_{\boldsymbol{x},a}$, then the prior is written as

$$f(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}_a, \boldsymbol{C}_{\boldsymbol{x},a}) \,. \tag{6.8}$$

The two last equations are inserted into (6.4) and (6.5). Instead of searching for the parameter which maximizes the maximum a-posteriori or likelihood respectively, the following two loss functions are used. They minimize the logarithm of the corresponding density function and consequently maximize the functions. For the maximum a-posteriori, the following loss function

$$L(\boldsymbol{x}) = (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}))^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1} (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})) + (\boldsymbol{x} - \boldsymbol{x}_a)^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{x},a}^{-1} (\boldsymbol{x} - \boldsymbol{x}_a) \tag{6.9}$$

whereas for the maximum likelihood this function

$$L(\boldsymbol{x}) = (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}))^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1} (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})) \,, \tag{6.10}$$

is minimized. The different cases will not be explicitly separated in the following as the principle remains the same. The respective loss function is minimized and the orbit estimate is finally given by

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}) \,. \tag{6.11}$$

Several approaches have been developed to determine this minimum. One option is to sequentially go through the measurements and update the a-priori state. An alternative option is to use the batch least squares formulation. The latter will be shortly explained here.

The loss function can be, in principle, minimized using the optimization approaches from Chapter 4, e.g. by applying the Newton method. However, the structure of the function can be exploited using the so-called Gauss-Newton method (effectively a simplification of the Newton method). The non-linear measurement function $\boldsymbol{h}(\boldsymbol{x})$ is linearized about a current estimate $\boldsymbol{x}^{[i]}$ (starting with a guessed initial point $\boldsymbol{x}^{[0]}$) with the first order Taylor expansion

$$h(\boldsymbol{x}) \approx \boldsymbol{h}(\boldsymbol{x}^{[i]}) + \boldsymbol{J}(\boldsymbol{x}^{[i]})(\boldsymbol{x} - \boldsymbol{x}^{[i]}) \,, \tag{6.12}$$

Finding the minimum is equivalent to finding the root in the derivative of L (similar to (4.5)), which defines the following system of linear equations after the insertion of (6.12) into the function

$$0 \stackrel{!}{=} \boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1}\left(\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x}^{[i]})\right)+\boldsymbol{C}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{J}(\boldsymbol{x}^{[i]})(\boldsymbol{x}-\boldsymbol{x}^{[i]}), \tag{6.13}$$

where $\boldsymbol{x}$ is unknown and can be solved for, and the Jacobian at $\boldsymbol{x}^{[i]}$ is

$$\boldsymbol{J}(\boldsymbol{x}^{[i]}) = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{[i]}}. \tag{6.14}$$

The solution to (6.13) is denoted as $\boldsymbol{x}^{[i+1]}$ and provides an improvement w.r.t. the previous iterate. Consequently, the process is iterated until convergence, i.e. the correction between two consecutive improvements is negligible or the loss function value at the estimate is not changing anymore. The presented process is commonly called differential correction and describes only one of the possible ways how to minimize the function. Other methods use damping factors to scale the improvement step or so-called trust-region approaches (e.g. Levenberg-Marquardt) to increase the convergence radius.

Still, the approach serves the purpose to illustrate the need for a good initial state guess $\boldsymbol{x}^{[0]}$. If the guessed point largely differs from the true solution, the iteration might diverge. It could either find another local minimum of the loss function or not find any at all. The convergence radius is larger with more advanced update schemes. Nevertheless, good starters are still required to find the right valley of the loss function. The dependency on good initial values motivates the next section about initial orbit determination.

6.2 Preliminary orbit determination

As discussed in the previous section, a sufficiently good estimate of an orbital state is required to perform orbit improvement. Sufficiently good means that the estimate is close enough to the final solution (the minimum of the loss function), so that the iteration converges. Chapter 2 shows that at least six independent parameters are required to describe a two-body dynamics orbit. Consequently, at least six independent observations must be provided to determine a first orbit. For high-altitude orbits, a set of optical observations as described in Chapter 3 is taken to compute the first estimate. A line-of-sight direction is described by the two angles: right ascension and declination. Thus, at least 3 pairs, or three line-of-sights $\boldsymbol{u}_i$ at three observations epochs t_i where $i = \{1,2,3\}$, are required to perform the initial orbit determination.

Classical methods have been derived by Gauss and Laplace. Gibbs, Herrick and Gooding, (1996) published more advanced versions of it. For more detailed derivations, the reader is referred to (Beutler, 2004; Escobal, 1965; Montenbruck and Gill, 2000). The different approaches can be categorized according to the formulation used to represent the orbit, i.e. the initial-value formulation and the boundary-value formulation. As described in Chapter 2 both representations can be used to uniquely define an orbital state. Here, one approach of each formulation is outlined. The different formulations will be likewise used later in the tracklet association method. The methods are outlined in order to illustrate the deficiencies when dealing with real observations.

6.2.1 Initial-value formulation

Laplace's solution is selected as a representative initial-value approach due to its simple derivation and transparency. Laplace merges the information of the three line-of-sights and the time difference between

them, to determine the line-of-sight $\boldsymbol{u}$ at a mean epoch and its first and second time derivative $\dot{\boldsymbol{u}}$ and $\ddot{\boldsymbol{u}}$. Inserting the acceleration from (3.24) into (2.2) gives a non-linear equation in three unknowns (topocentric range ρ and its time derivatives $\dot{\rho}$ and $\ddot{\rho}$)

$$\ddot{\boldsymbol{R}} + \rho\ddot{\boldsymbol{u}} + 2\dot{\rho}\dot{\boldsymbol{u}} + \ddot{\rho}\boldsymbol{u} = -\frac{\mu_\oplus}{\|\boldsymbol{r}\|^3}\left(\boldsymbol{R} + \rho\boldsymbol{u}\right) . \tag{6.15}$$

Substituting the geocentric distance $\|\boldsymbol{r}\| = r$ and taking the time derivative on both sides of the equation with $(\boldsymbol{u} \times \dot{\boldsymbol{u}})$ gives the topocentric distance as a function of r

$$\rho(r) = \frac{1}{(\boldsymbol{u} \times \dot{\boldsymbol{u}}) \cdot \ddot{\boldsymbol{u}}} \left(-\frac{\mu_\oplus}{r^3}\boldsymbol{R} \cdot (\boldsymbol{u} \times \dot{\boldsymbol{u}}) - \ddot{\boldsymbol{R}} \cdot (\boldsymbol{u} \times \dot{\boldsymbol{u}}) \right) \tag{6.16}$$

The initial orbit determination reduces to finding the roots in

$$0 \stackrel{!}{=} r - \|\boldsymbol{R} + \rho(r)\,\boldsymbol{u}\| . \tag{6.17}$$

With the known root r, the range-rate

$$\dot{\rho}(r) = \frac{1}{(\boldsymbol{u} \times \ddot{\boldsymbol{u}}) \cdot \dot{\boldsymbol{u}}} \left(-\frac{\mu_\oplus}{r^3}\boldsymbol{R} \cdot (\boldsymbol{u} \times \ddot{\boldsymbol{u}}) - \ddot{\boldsymbol{R}} \cdot (\boldsymbol{u} \times \ddot{\boldsymbol{u}}) \right) \tag{6.18}$$

is computed. Range ρ and its rate are then inserted into equations (3.22) and (3.23) to determine the position $\boldsymbol{r}$ and velocity $\dot{\boldsymbol{r}}$, i.e. the state $\boldsymbol{y}$ at the mean epoch.

6.2.2 Boundary-value formulation

Gauss proposed a method to determine the positions of celestial objects based on three line-of-sights. A detailed derivation of Gauss' method can be obtained from (Escobal, 1965). Gibbs, (1889) and Herrick, (1971) extended the method by additionally computing the velocities based on three position vectors. Gauss uses the principle that the two outer line-of-sights and the corresponding topocentric ranges (if they were known) fully describe the orbital state, i.e.

$$\boldsymbol{r}_1 = \boldsymbol{R}_1 + \rho_1\,\boldsymbol{u}_1 \quad \text{and} \quad \boldsymbol{r}_3 = \boldsymbol{R}_3 + \rho_3\,\boldsymbol{u}_3 . \tag{6.19}$$

The position at the central observation is then described as a linear combination of the outer positions, assuming coplanar motion

$$\boldsymbol{r}_2 = n_1\boldsymbol{r}_1 + n_2\boldsymbol{r}_2 \tag{6.20}$$

It is furthermore possible to express the range values at the outer epochs ρ_1 and ρ_2 in terms of n_1 and n_2 Starting with approximate values for these two quantities, the orbit is improved using fixed-point iterations.

The situation for an asteroid orbit determination is illustrated in Figure 6.1. The observation geometry in the figure is taken from the example in (Bauschinger, 1906, pp. 316–320). Bauschinger describes a method named *variation of geocentric distances*, where he alters the ranges ρ_1 and ρ_3 to obtain the astrometric place at the center epoch (namely the line-of-sight $\tilde{\boldsymbol{u}}_2$). He uses a finite difference scheme to update the guessed range values and an approximate motion model. He iteratively improves the bounding range

values until the computed line-of-sight $\tilde{\boldsymbol{u}}_2$ approximates the actually measured $\boldsymbol{u}_2$ as good as possible, i.e.

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}) \quad \text{where} \quad L(\boldsymbol{x}) = \|\tilde{\boldsymbol{u}}_2(\boldsymbol{x}) - \boldsymbol{u}_2\| \quad \text{and} \quad \boldsymbol{x} = (\rho_1, \rho_2)\,. \tag{6.21}$$

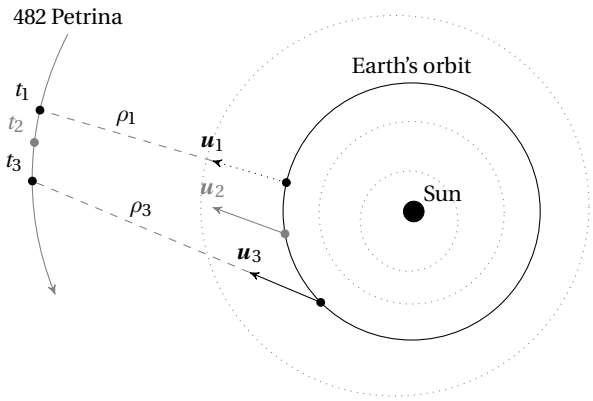


Figure 6.1: *Initial orbit determination of the asteroid 482 Petrina. Measurements taken March 7, March 30, and May 7 1902. The orbits of Mars, Venus, and Mercury are denoted with dotted lines.*

In fact, he used angular differences at the middle epoch, but the line-of-sight serves to show the principle. In his book he notes, that the method can be easily extended to more than three observations. Herget, (1965) later published an algorithm to perform this distance variation on a computer using as many observations as wanted. Gooding, (1996) developed a method very similar to this approach to determine the orbit from three line of sights. However, he does not use an approximate motion model, but instead the solution to the two-body boundary-value problem (cf. section 2.2). For the solution he uses a Newton-Raphson multivariate root-finding procedure. The boundary-value problem can be solved for multiple revolutions and consequently Gooding's approach allows the combination of optical observations from different revolutions as well. It can be therefore used to determine initial orbits using tracklets of geostationary objects (only the angular information) from three different observation nights.

6.2.3 Sensitivity to errors in very short arcs

Equation (6.16) and (6.18) reveal a major challenge, i.e. the line-of-sight acceleration comes up in the denominator of the function. Short arcs will lead to small estimates of this quantity with relatively large errors due to noise, which are effectively amplified when computing the unknown quantities. The preliminary orbit determination methods using the boundary-value formulation suffer from the same fact as the initial-value methods when dealing with very short arcs. That is, the curvature of the observed arc is used in the determination process. For Gauss' method, it appears in the denominator when computing the ranges ρ_1 and ρ_2 from n_1 and n_2 (Gronchi, 2004). Gooding's approach on the other hand suffers from the sensitivity of Lambert's problem to short arcs. The computation of the semi-major axis is sensitive to small angular errors for very short time intervals between the epochs. Additionally, a small central angle between two position vectors will make the orbital plane orientation more sensitive to errors in the pointing of the position vectors.

The sensitivity of the classical preliminary orbit determination methods is illustrated in Figure 6.2. The orbits defined by the erroneous observations (gray circles) are completely different from the one connecting the three true observations (black circles).

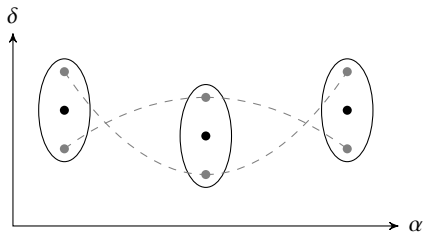


Figure 6.2: *Visualization of the short-arc problem: small errors in observations can result in large orbital errors. Dashed gray lines show erroneous curvature of orbit solutions using gray circles, while black circles denote the truth. Figure adapted from the sketch in (Gronchi, 2004).*

6.3 Preliminary orbit determination with restricting assumptions

Considering a special orbital region and observation constraints, certain geometric simplifications can be made to obtain first initial orbit states. This helps to exclude the possibly corrupted angular accelerations from the orbit determination process, with errors expected to be distributed as in (3.18). However, the simplifications also restrict the solutions to the selected orbital classes. Typically, geostationary and also transfer orbits are located at low inclinations. One simplification could thus be to constrain the orbital plane. In the following, methods which use this or other assumptions are presented. In contrast to most traditional approaches which use a set of line-of-sights, the methods here are formulated around the measured line-of-sight $\boldsymbol{u}$ and its derivative $\dot{\boldsymbol{u}}$. As the derivative is obtained from a line-fitting process (see section 3.6), it can smooth out errors if more than two line-of-sights are used for the fitting.

6.3.1 Fixed orbital plane

The orientation of the orbital plane is fixed by either specifying the right ascension of the ascending node Ω and inclination i or the normal vector $\boldsymbol{h}$. This reduces the degrees of freedom to 4 and hence requires only 4 independent observations, e.g. two angle pairs or one angle pair together with the rates.

In case of a zero inclination orbit, the normal vector is parallel to the Earth rotation axis, i.e. $\boldsymbol{h} = (0,0,1)^{\mathsf{T}}$. When the topocentric line-of-sight $\boldsymbol{u}$ is known, the geocentric position is determined by computing the intersection between the geocentric line-of-sight ($\boldsymbol{R} + \rho\,\boldsymbol{u}$) and the orbital plane. The intersection is then found by

$$\boldsymbol{h} \cdot \boldsymbol{r} = 0 \rightarrow \rho = -\frac{\boldsymbol{h} \cdot \boldsymbol{R}}{\boldsymbol{h} \cdot \boldsymbol{u}}\,. \tag{6.22}$$

Analogously, the geocentric velocity is obtained using

$$\boldsymbol{h} \cdot \dot{\boldsymbol{r}} = 0 \rightarrow \dot{\rho} = -\frac{\boldsymbol{h} \cdot \dot{\boldsymbol{R}} + \boldsymbol{h} \cdot \rho\,\dot{\boldsymbol{u}}}{\boldsymbol{h} \cdot \boldsymbol{u}}\,. \tag{6.23}$$

Inserting the intersecting range ρ and range-rate $\dot{\rho}$ into (3.22) and (3.23) yields the state on the fixed orbital plane.

6.3.2 Circular orbits

Another good simplification for geostationary orbits is the circular orbit. Circular orbits have four degrees of freedom, i.e. the orientation of orbital plane, the radius, and the location on the ring. A measured

line-of-sight and its derivative constrain the motion in four degrees, thus, provide sufficient information to obtain an orbital estimate. Beutler, (2004) uses two line-of-sights (each constraining two degrees of freedom) to perform the preliminary orbit determination for a circular orbit (similar also to the approach by Dubyago, (1961)). He sets up a root-finding problem to find the semi-major axis which zeros the difference between a computed angular separation with the measured one.

Two properties of circular orbits are used here to obtain the initial estimate: firstly, the velocity $\dot{\boldsymbol{r}}$ is always perpendicular to the position vectors, secondly, the magnitude of the velocity $\dot{r}$ is constant and only depends on the semi-major axis (equal to the geocentric distance r for circular orbits). The here

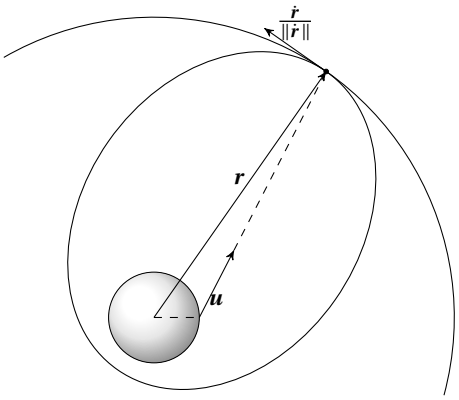


Figure 6.3: *Observation geometry for object on geostationary orbit and at the apogee of a geostationary transfer orbit. The normalized velocity vector is shown to illustrate the orthogonality for both orbits*

presented approach is equivalent to one by Fujimoto et al., (2010) with minor modifications. The two used properties, effectively constraints, are mathematically written as

$$\boldsymbol{r} \cdot \dot{\boldsymbol{r}} = 0 \quad \text{and} \quad \dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}} = \dot{r}^2 = \frac{\mu_\oplus}{r} \,. \tag{6.24}$$

Expressing the position and velocity in terms of observations (equation (3.22) and (3.23)) and inserting it to the first constraint yields

$$\boldsymbol{r} \cdot \dot{\boldsymbol{r}} = \left(\boldsymbol{R} \cdot \dot{\boldsymbol{u}} + \dot{\boldsymbol{R}} \cdot \boldsymbol{u}\right) \rho + \boldsymbol{R} \cdot \boldsymbol{u} \dot{\rho} + \rho \dot{\rho} = 0 \,, \tag{6.25}$$

which is rearranged for the range-rate $\dot{\rho}$. The rate is thus defined only by the topocentric range

$$\dot{\rho}(\rho) = \frac{\left(\boldsymbol{R} \cdot \dot{\boldsymbol{u}} + \dot{\boldsymbol{R}} \cdot \boldsymbol{u}\right) \rho}{\boldsymbol{R} \cdot \boldsymbol{u} + \rho} \,. \tag{6.26}$$

Using (3.22) and simplifying with $\boldsymbol{r} \cdot \boldsymbol{r} = r^2$ gives a function of the range depending on the geocentric distance r:

$$\rho(r) = -\boldsymbol{R} \cdot \boldsymbol{u} + \sqrt{(\boldsymbol{R} \cdot \boldsymbol{u})^2 - \boldsymbol{R} \cdot \boldsymbol{R} + r^2} \tag{6.27}$$

Consequently, the range-rate can also be defined in terms of the geocentric distance, i.e. $\dot{\rho}(r)$. Inserting $\dot{\rho}(r)$ and $\rho(r)$ into second constraint from (6.24) gives a root finding problem

$$0 \stackrel{!}{=} \dot{r}^2(r) - \frac{\mu_\oplus}{r} \,, \tag{6.28}$$

which is numerically solved. An example of the function in (6.28) is shown in Figure 6.28. The resulting root r is then used to obtain the preliminary orbit estimate. The derivative w.r.t. r is easily derived from

the equations above and allows the efficient usage of Newton-type root finding methods.

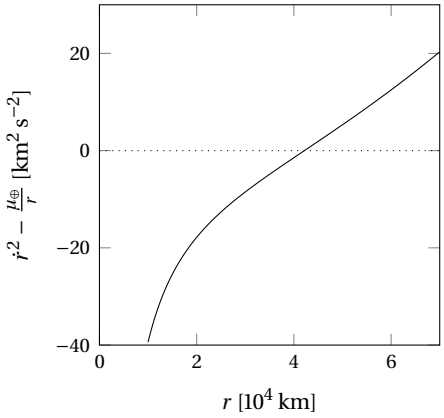


Figure 6.4: *Root-finding problem for preliminary orbit determination assuming a circular orbit. Observation taken from a geostationary object (sub-satellite longitude $\lambda = 100°$ and station located at $\lambda = 30°$ and latitude $\phi = 40°$). Solution is the intersection of dotted line and solid path. Function similar for the apogee orbit determination in section 6.3.3.*

6.3.3 Apogee orbit determination

Objects on geostationary transfer orbits are typically observed at their apogee. Väisälä, (1939) computed orbits with two line-of-sights assuming the objects to reside at their apsides during observation. Here, similarly to the approach for circular orbits, the apogee cases are constrained by

$$\boldsymbol{r} \cdot \dot{\boldsymbol{r}} = 0 \quad \text{and} \quad \dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}} = \dot{r}^2 = \frac{(1-e)\,\mu_\oplus}{r}, \tag{6.29}$$

where the velocity is again perpendicular to the position. The orthogonality assumption (left side) constrains one degree of freedom, which leaves five independent unknown parameters. Therefore another parameter must be fixed in order to obtain the full state. Väisälä, (1939) proposed to guess either the geocentric distance r or the eccentricity (as in the right side of (6.29)). If the distance r is known, the topocentric range ρ is calculated from (6.27) and consequently the range-rate $\dot{\rho}$ from (6.26). After calculating both variables, the orbit is ultimately known (cf. equations (3.22) and (3.23)). If the eccentricity is assumed to be known, then, analogously to the circular case, a root finding problem is obtained

$$0 \overset{!}{=} \dot{r}^2(r) - \frac{(1-e)\,\mu_\oplus}{r}, \tag{6.30}$$

which must be again solved using numerical methods. As transfer orbits typically circularize during their life-time, different eccentricity values should be tested in order to compute at least one feasible preliminary orbit candidate. However, as long as no further information on the eccentricity of the orbit is available, each candidate is equally likely.

7. Tracklet association and initial orbit determination

Only the reliable information of the tracklets, i.e. the line-of-sight $\boldsymbol{u}$ and its first time-derivative $\dot{\boldsymbol{u}}$, is used for the association and initial orbit determination. This information constrains the equations of motion in four degrees of freedom. As an orbital state is uniquely defined by six independent constraints (see Chapter 2) the orbit determination problem becomes under-determined. The missing parameters for a full orbital state in the topocentric reference system are the range ρ and range-rate $\dot{\rho}$ (with the position defined as in (3.22) and the velocity as in (3.23)). They are called *free* or *hypothesized parameters* in the following as they can be altered in order to derive a 6-dimensional orbital state from the 4 independent tracklet parameters. In principle, three different strategies can be used to find an orbit fitting the tracklet:

1. preliminary orbit determination using restricting assumptions (cf. section 6.3),
2. association to already cataloged objects (cf. section 5.5),
3. association to other tracklets (derived here).

The first strategy is only successful as long as the assumptions are valid. Assuming circular orbits, is a reasonable and mostly valid assumption for objects on geostationary orbits. However, it is not a good approximation for other objects close to or intersecting the geostationary ring (e.g. on geostationary transfer orbits).

An association to already cataloged objects is only possible if the object's location is well known. Of course, new objects cannot be linked to known cataloged objects. However, re-observed old objects might not be successfully associated as well due to unpredictable dynamics or orbital maneuvers. The cases where both previous strategies fail, motivate the tracklet-to-tracklet association as developed in this work. A method is derived, which uses the information from two observations to obtain an orbit estimate and an association probability. The latter is used to decide, whether two observation arcs belong to the same object. The used information for the orbit and the decision is $\boldsymbol{u}_1$, $\dot{\boldsymbol{u}}_1$, $\boldsymbol{u}_2$, and $\dot{\boldsymbol{u}}_2$ at the epochs t_1 and t_2 or alternatively

$$\alpha_1, \dot{\alpha}_1, \delta_1, \dot{\delta}_1 \quad \text{and} \quad \alpha_2, \dot{\alpha}_2, \delta_2, \dot{\delta}_2 \tag{7.1}$$

along with the respective uncertainties. The observations from two epochs constrain the equations of motion in eight degrees of freedom, which results in an over-determined system to solve.

Similarly to the presented preliminary orbit determination cases in section 6.2, the methods employ two different representations, namely the initial-value and boundary-value formulation. Both methods share the same working principle. The observations are divided into two groups: one is augmented with a hypothesis to define a state, while the second group is used to discriminate between good and bad solutions. Then, irrespective of the formulation, a optimization scheme is used to find the best fitting orbit.

7.1 Initial-value approach

The observation vector from the first epoch t_1

$$\boldsymbol{z}_1 = (\alpha_1, \dot{\alpha}_1, \delta_1, \dot{\delta}_1)^\mathsf{T} \tag{7.2}$$

is augmented with the hypothesized topocentric range and range-rate

$$\boldsymbol{x} = (\rho,\, \dot{\rho})^\mathsf{T} \tag{7.3}$$

which then defines the orbital state with

$$\tilde{\boldsymbol{y}}_1(\boldsymbol{x}) = \begin{pmatrix} \tilde{\boldsymbol{r}}(\rho_1) \\ \tilde{\dot{\boldsymbol{r}}}(\rho_1,\, \dot{\rho}_1) \end{pmatrix} . \tag{7.4}$$

The tilde notation is used to indicate the modeled variables, which are intermediate hypotheses. The test state is propagated to the second observation epoch t_2 as described in section 2.1. Subsequently, the resulting $\tilde{\boldsymbol{y}}_2(\boldsymbol{x})$ is transformed back to observation space, which gives a modeled observation $\tilde{\boldsymbol{z}}_2$. Additionally, together with the observation, the uncertainty of the first measurement $\boldsymbol{C}_{z_1}$ is transformed to the second epoch ($\boldsymbol{C}_{\tilde{z}_2}$) using $\boldsymbol{x}$ and the non-linear approximation as described in section 5.4.

The measurement association is analogously derived to the classification in section 5.3. The classification test as proposed in the latter named section, cannot be directly applied but requires first an orbital estimate defined by $\boldsymbol{x}$. The proportional probability of the variable $\boldsymbol{x}$ given the two observations $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ is computed according to Bayes' theorem

$$f(\boldsymbol{x}|\boldsymbol{z}_2,\, \boldsymbol{z}_1) \propto f(\boldsymbol{z}_2|\boldsymbol{x}, \boldsymbol{z}_1) f(\boldsymbol{x}|\boldsymbol{z}_1) . \tag{7.5}$$

The value $\boldsymbol{x}$ is sought which maximizes this function and consequently gives the best estimate. The prior $f(\boldsymbol{x}|\boldsymbol{z}_1)$ is assumed to be constant inside a feasible domain $\mathcal{C}$ and is therefore neglected. This assumption is not always valid. More details about this admissible region and about the possible invalidity is provided in the following section 7.1.2. The remaining likelihood is modeled similarly to (5.15) with the difference between the modeled and the observed measurement

$$f(\boldsymbol{z}_2|\boldsymbol{x}, \boldsymbol{z}_1) = \mathcal{N}(\tilde{\boldsymbol{z}}_2 - \boldsymbol{z}_2; \boldsymbol{0},\, \boldsymbol{C}_{\tilde{z}_2} + \boldsymbol{C}_{z_2}) . \tag{7.6}$$

The maximum likelihood estimate for the hypothesis $\boldsymbol{x}$ is then

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathcal{C}}{\arg\max}\; f(\boldsymbol{z}_2|\boldsymbol{x}, \boldsymbol{z}_1) \tag{7.7}$$

where $\mathcal{C}$ denotes the admissible region in which the prior is assumed to be constant. The function can be maximized using the optimization methods as described in Chapter 4. Instead of maximizing (7.6), the negative logarithm of the same function is minimized to find the identical point

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathcal{C}}{\arg\min} \left\{ \tfrac{1}{2} \ln\det(2\pi(\boldsymbol{C}_{\tilde{z}_2} + \boldsymbol{C}_{z_2})) + (\boldsymbol{z}_2 - \tilde{\boldsymbol{z}}_2)^\mathsf{T} (\boldsymbol{C}_{\tilde{z}_2} + \boldsymbol{C}_{z_2})^{-1} (\boldsymbol{z}_2 - \tilde{\boldsymbol{z}}_2) \right\} , \tag{7.8}$$

where the first term prunes large uncertainties $\boldsymbol{C}_{\tilde{z}_2}$ and the second term gives a measure how close the modeled $\tilde{\boldsymbol{z}}_2$ is to the actually observed one.

When deciding whether two tracklets belong to a common object, the second term becomes relevant

as it can be directly used to gate observation pairs (cf. with the Mahalanobis distance as provided in (5.7)). The location of the minimum of the loss function

$$L(\boldsymbol{x}) = (\boldsymbol{z}_2 - \tilde{\boldsymbol{z}}_2)^{\mathsf{T}} \left(\boldsymbol{C}_{\tilde{\boldsymbol{z}}_2} + \boldsymbol{C}_{\boldsymbol{z}_2}\right)^{-1} (\boldsymbol{z}_2 - \tilde{\boldsymbol{z}}_2) \tag{7.9}$$

is denoted with

$$\hat{\boldsymbol{x}}_d = \underset{\boldsymbol{x}\in\mathcal{C}}{\arg\min}\, L(\boldsymbol{x})\,. \tag{7.10}$$

and the loss function value at it

$$L(\hat{\boldsymbol{x}}_d) = d^2 \leq L(\hat{\boldsymbol{x}}) \tag{7.11}$$

can be gated using the thresholds in Table 5.1 for the chi-squared distribution with 4 degrees of freedom. The assumption is valid if the probability density of $\boldsymbol{z}_2 - \tilde{\boldsymbol{z}}_2$ is assumed to be normally distributed with zero mean and covariance $\boldsymbol{C}_{\tilde{\boldsymbol{z}}_2} + \boldsymbol{C}_{\boldsymbol{z}_2}$. Neglecting any bias or transformation errors, any value within the significant region of the distribution indicates that the observations originate from the same object. Otherwise, the mean will be shifted, leading to a non-central chi-squared distribution. In practice, unmodeled dynamics or non-linearities affecting the density transformation lead to a decentralization of the distribution. Therefore, thresholds can be selected considering the actual distribution.

The estimate $\hat{\boldsymbol{x}}_\mathrm{d}$ describes the orbit which approximates both measurements best but not the one which maximizes the likelihood. The difference between the maximum likelihood estimate $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_d$ is illustrated in Figure 7.1. The figures shows the two modeled density functions (simplified to one dimension) of each estimate. In this example, the statistical distance d^2 is equal for both estimates, however, the likelihood of $\hat{\boldsymbol{x}}$ is larger due to the smaller modeled covariance.

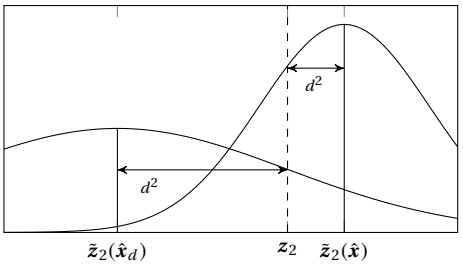


Figure 7.1: *Comparison between maximum likelihood estimate $\hat{\boldsymbol{x}}$ and minimum of loss function $\hat{\boldsymbol{x}}_d$. The scaled distance d^2 is the same for both points for the illustrated example (1 σ-distance of the univariate normal distribution).*

If the covariance pruning term is approximately constant in the vicinity of the local minimum, then $\hat{\boldsymbol{x}}_d \approx \hat{\boldsymbol{x}}$ gives a good initial estimate. The subscript d is dropped in the following derivations for notational simplicity. Afterward, an orbit improvement can be performed, either using the log-likelihood from (7.8) or directly a least squares approach with the angular data. The latter has the advantage of using the full information from two tracklets, rather than using the derived mean angle and angular rate quantities.

7.1.1 Loss function topography

The loss function defined above, evaluated at feasible $\boldsymbol{x}$ values, is illustrated in Figure 7.2. The feasible range is defined in section 7.1.2. A geostationary object at longitude $\lambda = 100°$ is observed from a ground-based telescope with coordinates $\lambda = 30°$ and $\phi = 40°$ (longitude and latitude). This example will be used throughout this chapter. The satellite is once re-observed in the same night after three hours and in the consecutive night after 27 hours. Measurement uncertainty is for simplicity modeled with $\sigma_{\alpha,\delta} = 2\,''$ and

$\sigma_{\dot{\alpha},\dot{\delta}} = 0.004\,''/\mathrm{s}$ which corresponds approximately to 5 measurements within one tracklet, a spacing of about 20 seconds, $\sigma_\mathrm{N} \approx 0.5\,''$, and $\sigma_\mathrm{B} \approx 2\,''$. The logarithm to base 10 of the loss function is plotted in order to emphasize the shape of the topography. As can be seen in the figure, the function contains one or several minima located within narrow valleys. The number of minima increases with the time interval between the measurements.

In order to interpret the topography, the loss function is decomposed into two functions. The loss is effectively a linear combination of the scaled angular and angular rate differences (between modeled and observed ones). Both parts are illustrated in Figure 7.3 for the 27 hour case. The pattern from the difference between angular positions is dominating in the combined loss function. This is caused by the large uncertainty of the angular rates (the measurement uncertainty as well as the propagated modeled uncertainty). A large uncertainty scales the difference to a small value and the better known angular position becomes more influential. Similar to Lambert's orbital boundary-value problem (see section 2.2), an angular position can be approximated for different numbers of completed full and half revolutions. This number is conveniently described with the variable k as in the previously mentioned section. This property is illustrated by overlaying the loss function from above with contours of equal k-values. It can be observed in Figure 7.4 that the valleys are distributed between the contour lines.

The difference in angular rates shows a different pattern. It resembles the property from the angular difference, i.e. the angular rates differences are locally smaller for different numbers of completed revolutions (and thus different semi-major axis values). The function can be further divided by geometrical considerations. Neglecting orbital constraints, the modeled angular rates at the second epoch can approximate the observed ones in two different geometrical ways: with the modeled range-rate pointing outwards $\dot{\rho}_2 > 0$ or inwards $\dot{\rho}_2 < 0$. Figure 7.5 illustrates two possible approximations to the line-of-sight derivative for different orbits. Figure 7.6 shows the $\dot{\rho} = 0$ boundaries on top of the angular rate loss function. In combination with the k-boundaries, the local minima appear in separated regions. The geometrical separation does not, however, guarantee that all minima are logically separated.

Summarizing the discussion about the loss function topography: the initial-value loss function in its standard form (using range and range-rate as free parameters) is a difficult multi-modal optimization problem. Restrictions (e.g. k-intervals or $\dot{\rho}$ boundaries) and coordinate transformation can improve the stability and convergence of its solution. The latter will be explained in the following sections. Alternatively, the angular rates can be dropped from the discriminating loss function. With typical measurement properties (uncertainty as described above), the angular rates at the second observation epoch play a minor role in the loss function. The dominating discriminators, namely the angular variables at the second

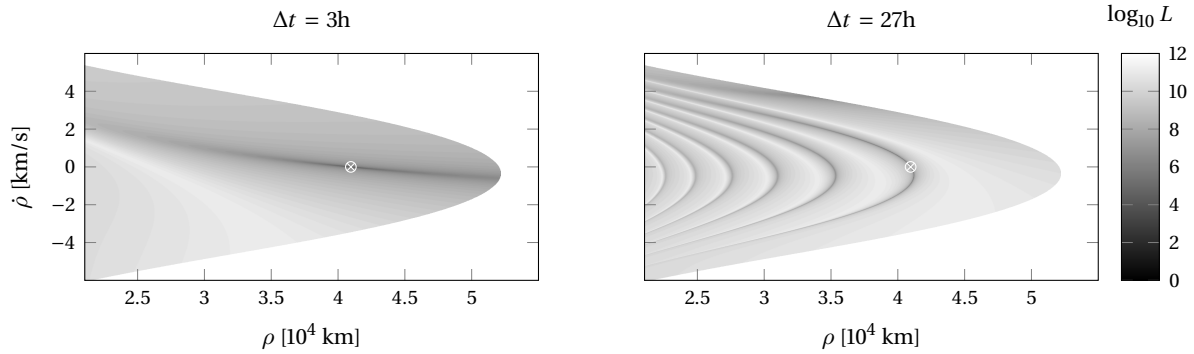


Figure 7.2: *The loss function topography for an example geostationary object (explained in text). The left figure shows the topography for a re-observation time of 3 hours and the right one shows it for 27 hours.*

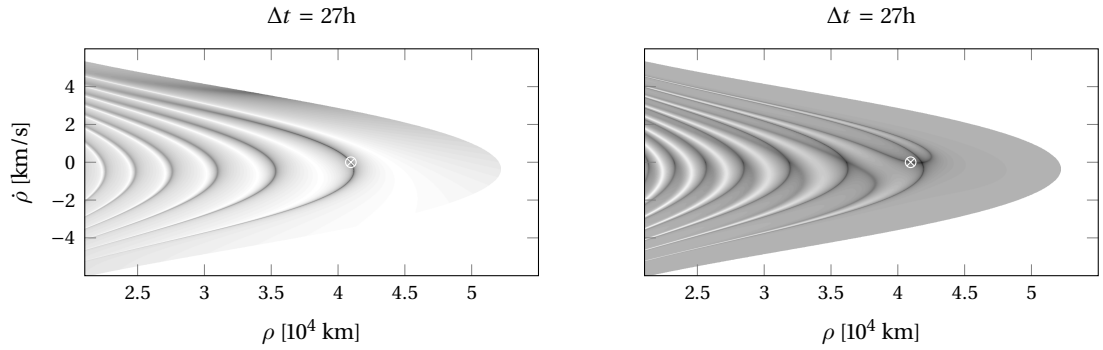


Figure 7.3: *Decomposition of the loss function. The left side shows difference between measured and modeled angles. The right side illustrates the influence of angular rates, i.e. angular rate difference between modeled and observed ones. The differences are not scaled with the uncertainty. The gray color ranges between the maximum and the minimum value of the respective function.*

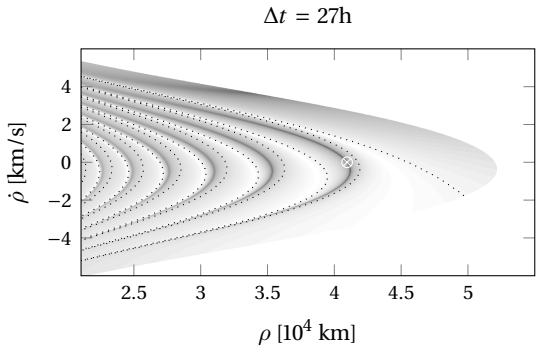


Figure 7.4: *Difference between modeled and observed angles for complete feasible domain. The local valleys are bounded by the dotted lines denoting the k-intervals. Each bounded subdomain corresponds to one number of completed full and half revolutions.*

epoch, are mostly affected by the angular rates at the first epoch. This can be shown with

$$\frac{\partial \tilde{\theta}_2}{\partial \dot{\theta}_1}\sigma_{\dot{\theta}_1} \gg \frac{\partial \tilde{\theta}_2}{\partial \theta_1}\sigma_{\theta_1}, \tag{7.12}$$

where θ represents both angles. Consequently, the loss function is not very sensitive to errors in the angular variables, but to ones in the angular rates. Additionally, the uncertainty in the modeled angles due to the angular rate uncertainty is typically larger than the one of the measured angles, i.e.

$$\sigma_{\tilde{\theta}_2}(\boldsymbol{x}, \boldsymbol{z}_1) \gg \sigma_{\theta_2}\,. \tag{7.13}$$

Considering that a bias affects only the mean angles, its effect on a practical loss function is negligible. Furthermore, the orbit is fully described by six independent quantities, i.e. angular position at both epochs and rates at first. Instead of finding the minimum, a non-linear system of equations can then be solved, e.g. the difference between modeled line-of-sight and observed one at the second epoch. This can be achieved with numerical root-finding algorithms as discussed for the Lambert's problem in section 2.2. The reduction, however, comes with the cost of losing information. In this work, the reduction is not applied in order to develop general methods independent of the specific telescope setup.

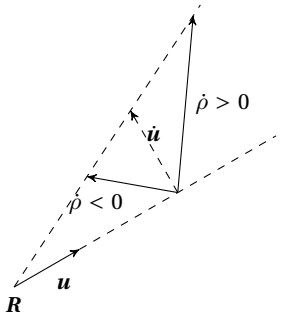


Figure 7.5: *Illustration of two possible orbits with the same line-of-sight time-derivative $\dot{\boldsymbol{u}}$ with equal line-of-sight $\boldsymbol{u}$. The telescope is located at $\boldsymbol{R}$.*

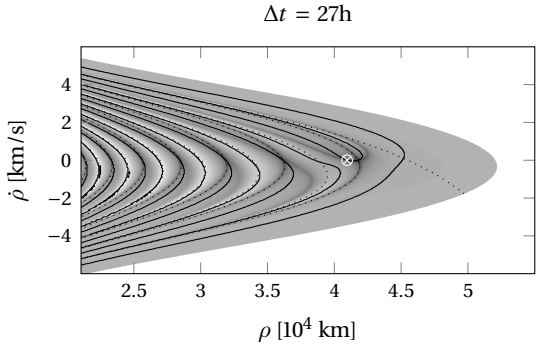


Figure 7.6: *Difference between modeled and observed angular rates for complete feasible domain. The local valleys are bounded by the dotted and the solid lines. The first denotes the k-intervals as in the Figure 7.4. The solid lines denote the range-rate boundaries as explained in the text.*

7.1.2 Admissible region

The two dimensional space of hypotheses $\boldsymbol{x}$, essentially the space of possible orbit solutions for one observation $\boldsymbol{z}_1$, is restricted using the admissible region concept. Milani et al., (2004) and Tommei et al., (2007) allow only stable orbits around a central body. Hyperbolic and parabolic trajectories connecting the two observations are excluded by defining a region where the energy of all candidates is negative

$$\mathcal{C}_{\mathcal{E}}(\boldsymbol{x}) = \{\boldsymbol{x} : \mathcal{E}(\boldsymbol{x}) \leq 0\}, \tag{7.14}$$

with the energy per unit mass given by (2.15). Inserting (3.22) and (3.23) into (2.15) gives the energy (or alternatively the semi-major axis a) dependent on ρ and $\dot{\rho}$

$$2\mathcal{E} = \frac{\mu}{a} = \dot{\rho}^2 + 2\,\dot{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{u}\,\dot{\rho} + f(\rho) \tag{7.15}$$

where

$$f(\rho) = \dot{\boldsymbol{u}}^{\mathsf{T}}\dot{\boldsymbol{u}}\rho^2 + 2\,\dot{\boldsymbol{R}}^{\mathsf{T}}\dot{\boldsymbol{u}}\,\rho + \dot{\boldsymbol{R}}^{\mathsf{T}}\dot{\boldsymbol{R}} + \frac{2\,\mu}{\|\boldsymbol{R} + \rho\,\boldsymbol{u}\|}\,. \tag{7.16}$$

The energy equation is quadratic in the range-rate variable $\dot{\rho}$, with the symmetry axis located at the vertex

$$\dot{\rho}_{\mathrm{v}} = -\dot{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{u}\,. \tag{7.17}$$

If a value for the energy and a range ρ are provided, the equation is solved with

$$\dot{\rho}_{1,2} = \dot{\rho}_{\mathrm{v}} \pm \sqrt{\Delta_{\mathcal{E}}} \tag{7.18}$$

where the discriminant

$$\Delta_{\mathcal{E}} = \dot{\rho}_{\mathrm{v}}^2 - f(\rho) - 2\mathcal{E} \tag{7.19}$$

must be positive. An analytic expression for the energy boundary is therefore given, which is beneficial for a coordinate transformation in section 7.1.3. This property causes the following: for each fixed feasible range ρ, the solution with minimal energy is at the vertex $\dot{\rho}_{\mathrm{v}}$. Alternatively, for each energy level, the maximum range value ρ_{v} is also at $\dot{\rho}_{\mathrm{v}}$.

In addition to the energy constraint, Maruskin et al., (2009) force the perigee of the solutions to stay above a feasible height

$$\mathcal{C}_{\mathrm{peri}}(\boldsymbol{x}) = \{\boldsymbol{x} : \|\boldsymbol{r}_{\mathrm{peri}}(\boldsymbol{x})\| \leq r_{\mathrm{min}}\}, \tag{7.20}$$

in order to guarantee that the candidate objects do not de-orbit within the next revolutions. A common choice for a least allowed perigee height is 200 km above surface.

Energy and perigee bounds are natural constraints, i.e. no real objects are expected to lie outside the region. In most cases, however, a specific orbital domain is of interest, e.g. the geostationary ring. The solutions can be additionally bounded using constraints on the orbital elements (DeMars and Jah, 2013; DeMars et al., 2012). These could be the semi-major axis a

$$\mathcal{C}_a(\boldsymbol{x}) = \{\boldsymbol{x} : a_{\mathrm{min}} \leq a(\boldsymbol{x}) \leq a_{\mathrm{max}}\}. \tag{7.21}$$

or eccentricity e

$$\mathcal{C}_e(\boldsymbol{x}) = \{\boldsymbol{x} : e(\boldsymbol{x}) \leq e_{\mathrm{max}}\}. \tag{7.22}$$

The intersection of all constraints defines the final admissible region:

$$\mathcal{C}(\boldsymbol{x}) = \mathcal{C}_a \cap \mathcal{C}_e \cap \mathcal{C}_P. \tag{7.23}$$

The different constraints are illustrated in Figure 7.7 for the geostationary example object.

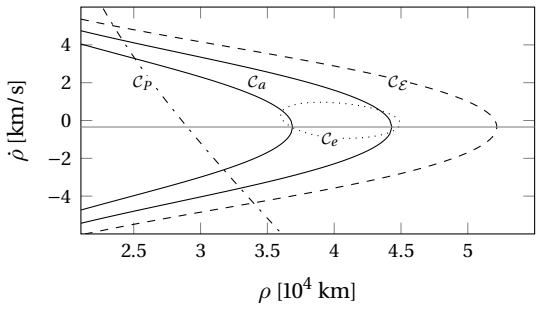


Figure 7.7: *Admissible region boundaries for a geostationary satellite. Details are provided in the text.*

Semi-major axis and eccentricity constraints cannot be easily applied to geostationary transfer orbits, as they vary considerably. The apogee height is more suitable, as objects can only be observed above a certain height. Objects in lower altitudes move to fast w.r.t. the observer and are typically not observed in survey scenarios. The constraint is

$$\mathcal{C}_{\mathrm{apo}}(\boldsymbol{x}) = \{\boldsymbol{x} : \|\boldsymbol{r}_{\mathrm{apo}}(\boldsymbol{x})\| \geq r_{\mathrm{obs}}\}, \tag{7.24}$$

where r_{obs} is the minimum height where objects can be observed. The admissible region constraints must

be carefully selected in order not to be too strict or to loose. Loose constraints increase the computational burden, while strict ones lead to a loss of associations.

As discussed in section 3.3.1, the inclination of near-geostationary objects typically varies between ±15°. As can be observed in Figure 3.4, the actual population of objects follows a certain pattern due to the long-term dynamics in the respective region. The orbital planes precess around the Laplace-plane over approximately 53 years. This prior knowledge can be incorporated in the admissible region definition. The feasible orbital planes are separated of the Laplace-plane normal $\boldsymbol{h}_L$ by an angle of about $\gamma = 7.5°$, thus, the boundary is defined by

$$\mathcal{C}_L(\boldsymbol{x}) = \{\boldsymbol{x} : \gamma - \Delta\gamma \leq \arccos \boldsymbol{h}(\boldsymbol{x}) \cdot \boldsymbol{h}_L \leq \gamma + \Delta\gamma\}, \tag{7.25}$$

where $\Delta\gamma$ is used to cover also objects drifting away from the perfect precession axis. The normal is computed from the initial position and velocity

$$\boldsymbol{h}(\boldsymbol{x}) = \frac{\boldsymbol{r}(\boldsymbol{x}) \times \dot{\boldsymbol{r}}(\boldsymbol{x})}{\|\boldsymbol{r}(\boldsymbol{x}) \times \dot{\boldsymbol{r}}(\boldsymbol{x})\|}. \tag{7.26}$$

The bounds covering the likely distribution of orbits are shown in the Figure 7.8.

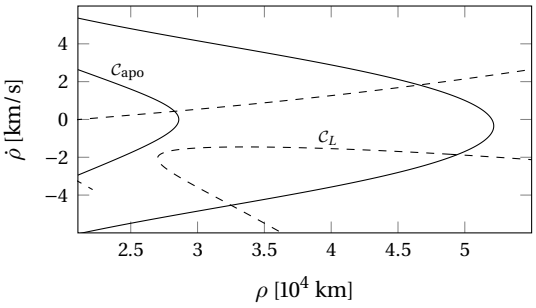


Figure 7.8: *Admissible region boundaries for a geostationary satellite considering the likely distribution of the geostationary region.*

Region segmentation

Besides separating feasible orbits from unfeasible ones, the admissible region itself is segmented using similar constraints. As described in Chapter 4 it is favorable for numerical optimization methods to minimize functions which contain only one local and thus global minimum. If this is not possible, the likelihood of catching every local minimum and therefore also the global one using heuristics increases with a lower number of local minima. As shown in the previous section, the valleys of the loss function are located along and within certain boundaries. Thus, as suggested in (Siminski et al., 2014a), it is beneficial to restrict the solution space to a fixed number of completed half orbital revolutions k. Additionally, the direction of the range-rate at the second epoch $\dot{\rho}_2$ can be used as a boundary definition. The number of revolutions in between measurements is unknown, but can be gated using the same requirements as above (e.g. allowing only stable orbits around Earth). However, with an increasing number of revolutions, more multi-revolution solutions become physically feasible. An extreme case is given if the object is observed two times at the same location. Then, there are multiple solutions with equal angular rates but different semi-major axis values. The number of completed orbital revolutions is given by

$$n_{\text{rev}}(\boldsymbol{x}) = \left\lfloor \frac{t_2 - t_1}{P} \right\rfloor + \frac{\nu_2 - \nu_1}{2\pi}, \tag{7.27}$$

where the orbital period P is computed from (2.16) using the semi-major axis (or energy). The brackets denote the floor function which reduces the number to the next lower integer value. The true anomaly difference, however, requires the knowledge about both geocentric positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. The restriction is thus applied after the propagation of the initial state to the second epoch. Alternatively, the number can be approximated for circular orbits with

$$n_{\mathrm{rev}}(\boldsymbol{x}) \approx \frac{t_2 - t_1}{P}\,, \tag{7.28}$$

which reduces the computation burden as no orbit must propagated. Similar to the parametrization in Lambert's problem, the k-variable is used to add information about completed half revolutions:

$$k = \begin{cases} 2\lfloor n_{\mathrm{rev}} \rfloor & \text{if } n_{\mathrm{rev}} - \lfloor n_{\mathrm{rev}} \rfloor < 0.5 \\ 2\lfloor n_{\mathrm{rev}} \rfloor + 1 & \text{if } n_{\mathrm{rev}} - \lfloor n_{\mathrm{rev}} \rfloor > 0.5\,. \end{cases} \tag{7.29}$$

The sign of the k-variable can be used to distinguish between positive and negative range-rate solutions. Thus, if $k < 0$, then $\dot{\rho}_2 < 0$ must be fulfilled and correspondingly if $k > 0$, then $\dot{\rho}_2 > 0$.

Instead of using a global admissible region, a set of n_k separate constrained domains are then created with

$$\mathcal{C}_{k,i}(\boldsymbol{x}) = \{\boldsymbol{x} : k(\boldsymbol{x}) = \|k_i\|,\ \operatorname{sgn}\dot{\rho}_2(\boldsymbol{x}) = \operatorname{sgn} k_i\} \quad \text{for } i = 1, \ldots, n_k\,. \tag{7.30}$$

with k_i going twice from the least feasible number $k_{\min}$ to the maximum possible $k_{\max}$ with once positive and once negative sign. The bounding values for the number of completed revolutions are calculated from the circular assumption (7.28) together with a feasible semi-major axis interval $a \in [a_{\min}, a_{\max}]$:

$$n_{R,\min} = \frac{t_2 - t_1}{P(a_{\max})} \quad \text{and} \quad n_{R,\max} = \frac{t_2 - t_1}{P(a_{\min})}\,. \tag{7.31}$$

The $k_{\min}$ and $k_{\max}$ are obtained from these values. The semi-major axis interval is not a tight bound as defined above with (7.21), but only limits the number of revolutions. However, the circular approximation might not cover the complete range depending on the geometry. Computing the actual values requires evaluating all possible points on the semi-major axis region boundary. Alternatively, another half revolution is added for safety to the range and tested even if it might not be required in some cases. As a consequence of forming distinct domains, a set of n_k individual optimization problems have to be solved instead of one, i.e.

$$\hat{\boldsymbol{x}}_k = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x},\, k)\,. \tag{7.32}$$

The global minimum corresponds to one number of completed full and half revolutions and one range-rate direction:

$$\hat{\boldsymbol{x}} = \arg\min_{k} L(\hat{\boldsymbol{x}}_k) \,. \tag{7.33}$$

The loss function together with the k-restricted sub-domains are shown in Figure 7.9 and Figure 7.10 (neglecting the $\dot{\rho}_2$ boundaries to improve visibility). The first shows the geostationary example from before, while the second illustrates the case of an object in some geostationary transfer orbit. The latter is included for completeness, the principle shape and method does not change. However, it can be observed that the number of possible multi-revolution solution increases for these observation geometries. The large number of possible k-values motivates the use of a restricted semi-major axis interval instead of using the global energy and perigee height constraints.

Region uncertainty

So far, the uncertainty of the initial observation is not included in the derivation of the admissible region boundaries. Worthy III and Holzinger, (2015) point out that this assumption is not always valid and illustrate it with an example object in a lower Earth orbit. Even though the effects of neglecting the uncertainty in the admissible region for near-geostationary objects is not as large as for lower altitude objects, they still need to be considered. However, if arbitrarily tight bounds on some elements are selected, the uncertain bounds become equally arbitrary. In order to cover a specific orbital region and reduce the computational burden, the size of the admissible region can be manually increased instead using some margins from experience or an approximation. The latter is derived here. For notational simplicity, the effect of the measurement on the energy bound is derived and later adapted to the semi-major axis bounds, i.e.

$$\sigma_{\mathcal{E}}^2(\boldsymbol{x}) \approx \frac{\partial \mathcal{E}}{\partial \boldsymbol{z}_1}^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{z}_1} \frac{\partial \mathcal{E}}{\partial \boldsymbol{z}_1}^{\mathsf{T}} \tag{7.34}$$

with the contribution from right ascension and its rate computed with

$$\frac{\partial \mathcal{E}}{\partial \alpha_1} = \frac{\partial \mathcal{E}}{\partial \boldsymbol{u}_1} \cdot \frac{\partial \boldsymbol{u}_1}{\partial \alpha_1} + \frac{\partial \mathcal{E}}{\partial \dot{\boldsymbol{u}}_1} \cdot \frac{\partial \dot{\boldsymbol{u}}_1}{\partial \alpha_1} \quad \text{and} \quad \frac{\partial \mathcal{E}}{\partial \dot{\alpha}_1} = \frac{\partial \mathcal{E}}{\partial \dot{\boldsymbol{u}}_1} \cdot \frac{\partial \dot{\boldsymbol{u}}_1}{\partial \dot{\alpha}_1} \tag{7.35}$$

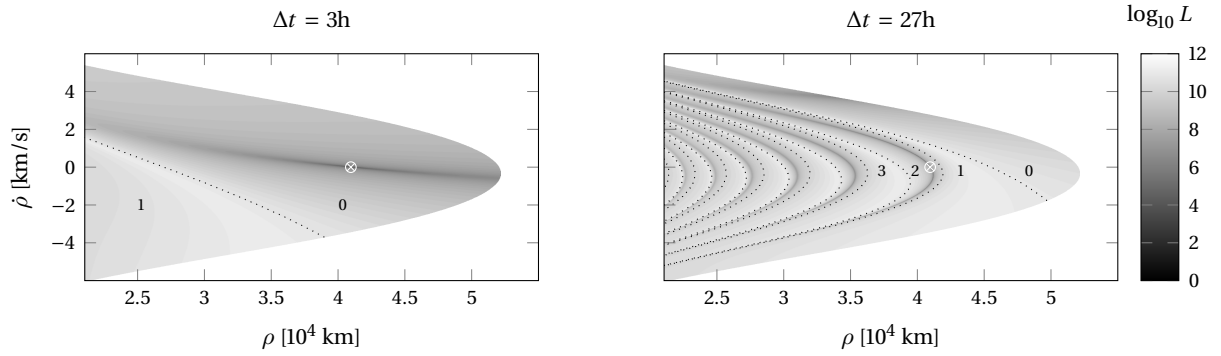


Figure 7.9: *Loss function for a geostationary object re-observed after 3 (left) and 27 hours (right). Additionally, the k-restricted domains are illustrated with the dashed lines.*

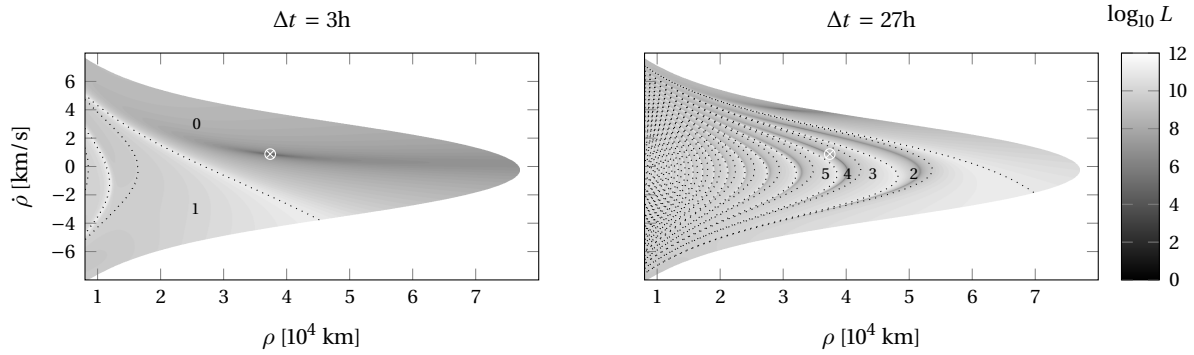


Figure 7.10: *Loss function for an object on the geostationary transfer orbit re-observed after 3 (left) and 27 hours (right). Additionally, the k-restricted domains are illustrated with the dashed lines.*

where

$$\frac{\partial \dot{\boldsymbol{u}}_1}{\partial \dot{\alpha}_1} = \frac{\partial \boldsymbol{u}_1}{\partial \alpha_1} \quad \text{and} \quad \frac{\partial \dot{\boldsymbol{u}}_1}{\partial \alpha_1} = \dot{\alpha}\frac{\partial^2 \boldsymbol{u}_1}{\partial \alpha_1^2} + \dot{\delta}\frac{\partial^2 \boldsymbol{u}_1}{\partial \delta \partial \alpha} \tag{7.36}$$

and the partials w.r.t. to line-of-sight and derivative are given by

$$\frac{\partial \mathcal{E}}{\partial \boldsymbol{u}_1} = \dot{\rho}_1 \dot{\boldsymbol{R}}_1 - \rho_1 \frac{\mu_\oplus}{(\boldsymbol{r}_1 \cdot \boldsymbol{r}_1)^{\frac{3}{2}}} \boldsymbol{R}_1 \quad \text{and} \quad \frac{\partial \mathcal{E}}{\partial \dot{\boldsymbol{u}}_1} = \rho_1^2 \dot{\boldsymbol{u}}_1 + \rho_1 \dot{\boldsymbol{R}}_1 \,. \tag{7.37}$$

The partials w.r.t. declination and rate are computed analogously. Of particular interest is the effect of the uncertainty at the boundaries of the region. The discussion is restricted to the semi-major axis bounds $a \in [a_{\min}, a_{\max}]$ corresponding to $k_{\min}$ and $k_{\max}$. The value $\sigma_a(\boldsymbol{x})$ can be directly obtained from $\sigma_{\mathcal{E}}$ as it is a linear transformation. Neglecting the impact of the angular uncertainty on the energy boundary (assuming small angular errors), the largest energy deviation is found for the maximum range ρ_{v} at the vertex. Thus, it is sufficient to compute σ_a once instead of testing the whole boundary to get an upper limit.

Fujimoto et al., (2014a) suggests to extend the admissible region by some σ-interval to include all feasible orbits. Worthy III and Holzinger, (2015) model the probability in the transition and propose to use it in the estimation process. In the special case of the k-restricted admissible region, only the outermost boundaries are interesting. The internal boundaries play no role, as they are covered by surrounding domains. Before computing k-intervals, the admissible region of an observation $\boldsymbol{z}_1$ is investigated by computing the $\rho_{\mathrm{v}}(a_{\min})$ and $\rho_{\mathrm{v}}(a_{\max})$ and the corresponding semi-major axis deviations $\sigma_{a,\max}$ and $\sigma_{a,\min}$. The respective semi-major axis bounds are then updated with a certain σ-interval (e.g. 3σ). An extended admissible region is shown in Figure 7.11. Afterward, the optimization proceeds as normal.

Penalty function

The admissible region is introduced into the optimization problem by using a penalty function. The penalty function should be zero if a $\boldsymbol{x}$ hypothesis is tested in the allowed range of the loss function and should increase quadratically if it is not. This allowed range is defined by an interval of allowed half

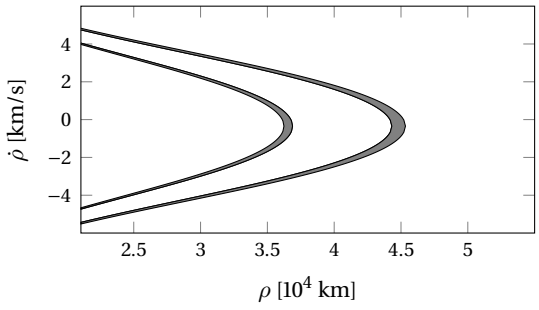


Figure 7.11: *Admissible region with 3σ upper and lower margin illustrated with gray area.*

revolutions. A possible penalty function is then given by

$$\phi_k(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \tilde{k}(\boldsymbol{x}) \in [k, k+1] \\ \lambda_k \left(\tilde{k}(\boldsymbol{x}) - (k+1)\right)^2 & \tilde{k}(\boldsymbol{x}) > k+1 \\ \lambda_k \left(\tilde{k}(\boldsymbol{x}) - k\right)^2 & \tilde{k}(\boldsymbol{x}) < k \end{cases} \tag{7.38}$$

where λ_k is a scaling parameter and $\tilde{k}(\boldsymbol{x})$ as computed from (7.29). The range-rate boundary is included in the same way, i.e.

$$\phi_{\dot{\rho}_2}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \operatorname{sgn} \tilde{\dot{\rho}}_2(\boldsymbol{x}) = \operatorname{sgn} k \\ \lambda_k \, \tilde{\dot{\rho}}_2(\boldsymbol{x})^2 & \text{else.} \end{cases} \tag{7.39}$$

If a probability density function for the admissible region is provided as done e.g. by Worthy III and Holzinger, (2015), then this can be directly included into the optimization as a penalty function. So far it is assumed that the prior $f(\boldsymbol{x}|\boldsymbol{z}_1)$ is constant and is therefore dropped in the estimation process. However, if it is somehow modeled, its natural logarithm can be added to the loss function. In this work, the above mentioned practical approach using an arbitrary penalty is selected with an appropriate scaling parameter.

7.1.3 Change of variables

Because of the curved shape of the different constrained regions for each k, it is favorable to change the coordinate system. Otherwise optimizers will run into difficulties as explained in Chapter 4. An obvious choice for a coordinate is the semi-major axis as it is highly correlated to the number of completed half revolutions k (in fact the relation is linear for circular orbits as shown above). Together with the range, an orbital solution is defined using (7.18). However, the range itself is not suitable as the second variable, as each range-semi-major axis combination allows two solutions. This can be solved by either splitting up the problem into two parts and minimizing each individually or by using an alternative grid. The relative range ρ' is therefore introduced with the respective vertex range as a reference, i.e. the maximum range for a given semi-major axis. The quadratic energy function from equation (7.15) is used to set up a root-finding problem for ρ_v

$$0 \stackrel{!}{=} \dot{\rho}_\mathrm{v}^2 + 2\,\dot{\boldsymbol{R}}^\mathsf{T}\boldsymbol{u}\,\dot{\rho}_\mathrm{v} + f(\rho_\mathrm{v}) - \frac{\mu}{a} \tag{7.40}$$

The new coordinates are then expressed by

$$\rho(a, \rho') = \rho_\mathrm{v}(a) - |\rho'| \tag{7.41}$$

and

$$\dot{\rho}(a,\rho') = \begin{cases} \dot{\rho}_\mathrm{v} - \sqrt{\Delta_\mathcal{E}}(a,\rho), & \text{if } \rho' < 0 \\ \dot{\rho}_\mathrm{v}, & \text{if } \rho' = 0 \\ \dot{\rho}_\mathrm{v} + \sqrt{\Delta_\mathcal{E}}(a,\rho), & \text{if } \rho' > 0 \end{cases} \tag{7.42}$$

The resulting loss function is shown in Figure 7.12. It shows how some boundaries are rectified. Transitions to even k-values denote a difference in completed full revolutions. The boundaries are thus equal to lines of constant semi-major axis. The other regions are improved as well in terms of beneficial shape but keep their curved boundaries.

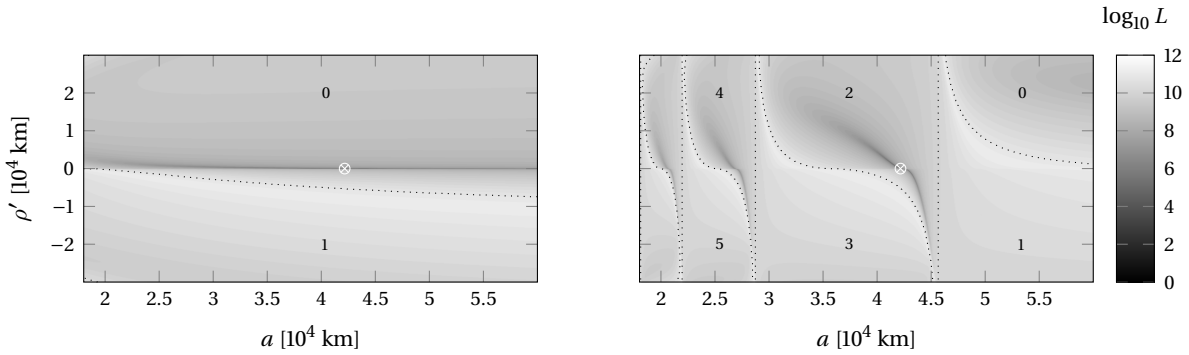


Figure 7.12: *Loss function for a geostationary object re-observed after 3 (left) and 27 hours (right) using the alternative coordinate system defined in the text. Additionally, the k-restricted domains are illustrated with the dashed lines.*

7.1.4 Angular-momentum conservation

The size of the solution space can be furthermore reduced by including the angular-momentum conservation from (2.11) (effectively an equality constraint)

$$\boldsymbol{r}_1 \times \dot{\boldsymbol{r}}_1 = \boldsymbol{r}_2 \times \dot{\boldsymbol{r}}_2 \,. \tag{7.43}$$

This constraint can be used to transform the two-dimensional into a one-dimensional problem, i.e. it defines a path through the ρ-$\dot{\rho}$ space. The global minimum does not have to lie on the line due to erroneous rate measurements. Still, the approach can be used to identify local valleys and then continue with good initial starting points in two dimensions.

The angular-momentum for each epoch depending on the unknown quantities ρ and $\dot{\rho}$ is computed with

$$\begin{aligned} \boldsymbol{r} \times \dot{\boldsymbol{r}} &= (\boldsymbol{R} + \rho\,\boldsymbol{u}) \times (\dot{\boldsymbol{R}} + \rho\,\dot{\boldsymbol{u}} + \dot{\rho}\,\boldsymbol{u}) \\ &= \dot{\rho}\,(\underbrace{\boldsymbol{R} \times \boldsymbol{u}}_{\boldsymbol{a}} + \underbrace{\rho\,\boldsymbol{u} \times \boldsymbol{u}}_{0}) + \rho^2\,\underbrace{(\boldsymbol{u} \times \dot{\boldsymbol{u}})}_{\boldsymbol{b}} \\ &\quad + \rho\,\underbrace{(\boldsymbol{R} \times \dot{\boldsymbol{u}} + \boldsymbol{u} \times \dot{\boldsymbol{R}})}_{\boldsymbol{c}} + \underbrace{(\boldsymbol{R} \times \dot{\boldsymbol{R}})}_{\boldsymbol{d}} \\ &= \dot{\rho}\,\boldsymbol{a} + \boldsymbol{f}(\rho) \end{aligned} \tag{7.44}$$

Insert (7.44) into (7.43) and cancel out range-rate terms $\dot{\rho}_1$ and $\dot{\rho}_2$ by scalar projection

$$\dot{\rho}_1 \boldsymbol{a}_1 - \dot{\rho}_2 \boldsymbol{a}_2 = \boldsymbol{f}_2(\rho_2) - \boldsymbol{f}_1(\rho_1) \qquad | \cdot (\boldsymbol{a}_1 \times \boldsymbol{a}_2)$$

gives a quadratic equation for the range ρ_2 at the second epoch

$$\begin{aligned} 0 &= (\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot (\boldsymbol{f}_2(\rho_2) - \boldsymbol{f}_1(\rho_1)) \\ &= (\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{f}_2(\rho_2) - (\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{f}_1(\rho_1) \\ &= \rho_2^2 \underbrace{(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{b}_2}_{q_1} + \rho_2 \underbrace{(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{c}_2}_{q_2} \\ &\quad + \underbrace{(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{d}_2 - (\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{f}_1(\rho_1)}_{q_3} . \end{aligned}$$

The equation is solved with

$$\rho_2 = \frac{-q_2 \pm \sqrt{q_2^2 - 4q_1q_3}}{2q_1} \tag{7.45}$$

where the positive root is selected. The range at the second epoch is required to compute the range-rate for the first epoch using

$$\dot{\rho}_1 (\boldsymbol{a}_1 \times \boldsymbol{a}_2) = (\boldsymbol{f}_2(\rho_2) - \boldsymbol{f}_1(\rho_1)) \times \boldsymbol{a}_2 ,$$

which gives two possible values

$$\dot{\rho}_1 = \pm \frac{\|(\boldsymbol{f}_2(\rho_2) - \boldsymbol{f}_1(\rho_1)) \times \boldsymbol{a}_2\|}{\|(\boldsymbol{a}_1 \times \boldsymbol{a}_2)\|} .$$

Thus, two paths are created that lead through the two-dimensional solution space: one for the positive and one for the negative value of $\dot{\rho}_1$. It is possible to compute the range-rate at the second epoch as well. However, the solution has no physical meaning, as it only fulfills the requirement of angular momentum conservation, not necessarily energy conservation. Taff and Hall, (1977) set up a system of equations using the angular momentum conservation and energy conservation equalities. Gronchi et al., (2011) build up on this approach and exchange the energy constraint with the equality of the Laplace-Lenz vector. Both methods rely on the solution of a high-order polynomial root-finding problem as outlined in section 1.3.2.

In this work, the solution for the range-rate depending on a free hypothesized range

$$\dot{\rho}_1(\rho_1, \boldsymbol{u}_1, \dot{\boldsymbol{u}}_1, \boldsymbol{u}_2, \dot{\boldsymbol{u}}_2) \tag{7.46}$$

is inserted along with the range into the 2-dimensional loss function. The principle is illustrated on the left side of Figure 7.13. The right side shows the loss values for one of the two possible paths (upper one). The best fitting orbit fulfilling the constraints is then

$$\hat{\rho}_1 = \arg\min_{\rho_1} L(\rho_1, \dot{\rho}_1(\rho_1)) . \tag{7.47}$$

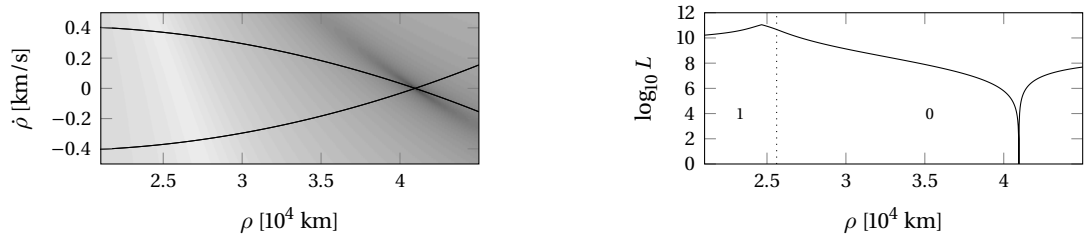


Figure 7.13: *Path with conserved angular-momentum on top of the two-dimensional loss function (left). On the right side is the one-dimensional loss function for the same object and observation geometry along with the k-intervals illustrated with dotted lines.*

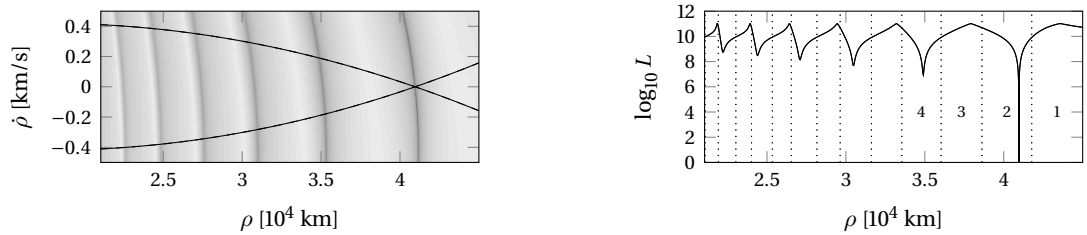


Figure 7.14: *Path with conserved angular-momentum on top of the two-dimensional loss function (left). On the right side is the one-dimensional loss function for the same object and observation geometry along with the k-intervals illustrated with dotted lines.*

As written before, the minimum is not necessarily located at the same position as in the two-dimensional case, i.e. $\hat{\boldsymbol{x}} \neq (\hat{\rho}_1, \dot{\rho}_1(\hat{\rho}_1))^\mathsf{T}$. However, it is used as a starter to the more difficult two-dimensional problem.

The angular-momentum conservation is furthermore used to restrict the solution space in ρ_1-$\dot{\rho}_1$-domain. The effect of the uncertainty in angles and angular rates on the range-rate at the first epoch is approximated with

$$\sigma_{\dot{\rho}_1} \approx \left(\frac{\partial \dot{\rho}_1}{\partial \boldsymbol{z}_1}\right) \boldsymbol{C}_{\boldsymbol{z}_1} \left(\frac{\partial \dot{\rho}_1}{\partial \boldsymbol{z}_1}\right)^\mathsf{T} \tag{7.48}$$

Analytical expressions for the partials can be derived or alternatively a finite difference scheme can be used. Considering the uncertainty, the solution space is reduced to a stripe surrounding the before mentioned path. The stripe width should be selected accounting for a feasible σ-interval (e.g. $3\,\sigma$).

7.2 Boundary-value approach

The association of two tracklets is also achieved by using the alternative boundary-value formulation of orbital states. The initial-value approach is asymmetric as the full information of one tracklet is used to hypothesize orbits, while the other one is used only for the discrimination. The boundary-value approach allows using the same information of both tracklets for the same purpose. As for the initial-value approach, the observed quantities are thus divided into state and discriminating variables. The angular observations

$$\boldsymbol{z} = (\alpha_1, \delta_1, \alpha_2, \delta_2)^\mathsf{T} \tag{7.49}$$

serve to define a state after augmenting them with range hypotheses at both epochs using

$$\tilde{\boldsymbol{y}}_{1:3}(t_1) = \boldsymbol{r}(\rho_1) \quad \text{and} \quad \tilde{\boldsymbol{y}}_{1:3}(t_2) = \boldsymbol{r}(\rho_2)\,. \tag{7.50}$$

The range values are combined for convenience into the free variable $\boldsymbol{x} = (\rho_1, \rho_2)^\mathsf{T}$. The tilde notation is used to denote the hypothesized orbit solution. It is obtained from the solution to the orbital boundary-value problem in section 2.2. The remaining angular rates

$$\dot{\boldsymbol{z}} = (\dot{\alpha}_1,\, \dot{\delta}_1, \dot{\alpha}_2,\, \dot{\delta}_2)^\mathsf{T} \tag{7.51}$$

are used as the discriminator. The orbit hypothesis $\tilde{\boldsymbol{y}}(k)$ is used to generate modeled observations

$$\tilde{\dot{\boldsymbol{z}}}(\boldsymbol{z}, \boldsymbol{x},\, k) = (\tilde{\dot{\alpha}}_1, \tilde{\dot{\delta}}_1, \tilde{\dot{\alpha}}_2, \tilde{\dot{\delta}}_2)^\mathsf{T} \tag{7.52}$$

which are compared to the actually measured ones. Note that the number of completed half revolutions k is included in the orbit generation.

Starting again with the relative probability of a hypothesis $\boldsymbol{x}$ and k according to Bayes' theorem

$$f(\boldsymbol{x}, k|\dot{\boldsymbol{z}}, \boldsymbol{z}) \propto f(\dot{\boldsymbol{z}}|\boldsymbol{x}, k,\, \boldsymbol{z}) f\,(\boldsymbol{x}, k|\boldsymbol{z})\,, \tag{7.53}$$

an optimization problem is created. The most likely $\boldsymbol{x}$ maximizes the probability as the MAP estimate. The prior is assumed to be constant for all feasible range-range combinations in an admissible region. Again, the likelihood is modeled as a multivariate normal distribution similar to (5.15) with

$$f(\dot{\boldsymbol{z}}|\boldsymbol{x}, k,\, \boldsymbol{z}) = \mathcal{N}\left(\tilde{\dot{\boldsymbol{z}}} - \dot{\boldsymbol{z}}; \mathbf{0},\, \boldsymbol{C}_{\tilde{\dot{\boldsymbol{z}}}} + \boldsymbol{C}_{\dot{\boldsymbol{z}}}\right) \tag{7.54}$$

where the covariance of the modeled angular rates is computed numerically using finite differences due to the complexity of the Lambert's problem solution. A step size of a few arc seconds is a feasible choice for the finite difference approximation. Alternatively, a set of sigma-points can be used as explained in section 5.4. Arora et al., (2015) recently derived partials of the solution to the Lambert's problem, which could be used in an linear transformation of the uncertainty. Schumacher et al., (2015) gives a detailed discussion on the uncertainty attached to the solution of the problem.

With typical sensor properties, the angular positions are well known up to an arc second. The effect on the orbit hypothesis is then negligible w.r.t. the uncertainty of the measured angular rates, i.e. $\sigma_{\tilde{\theta}} \ll \sigma_{\dot{\theta}}$. Figure 7.15 shows the evolution of the modeled angular rate uncertainty for a geostationary satellite. The modeled uncertainty stays an order below the measured one, but for geometries where the object completed a full or half revolution and the orientation of the orbital plane becomes indeterminable. For a geostationary object, the declination rate is influenced while the right ascension rate remains fairly determinable. However, as long as the objects are not re-observed at the same (or close to the same) position on their orbits, the above noted uncertainty assumption is valid.

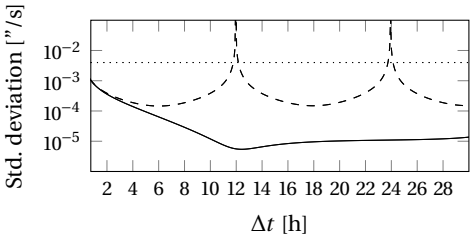


Figure 7.15: *Evolution of standard deviation estimate for modeled angular rates (diagonal components of $\boldsymbol{C}_{\tilde{\dot{\boldsymbol{z}}}}$. The solid line shows the declination rate standard deviation $\sigma_{\dot{\delta}}$ and dashed line the right ascension rate deviation $\sigma_{\dot{\alpha}}$. The dotted line illustrates the measurement uncertainty $\sigma_{\dot{\alpha},\dot{\delta}} = 0.004$ "/s.*

Instead of maximizing the likelihood, the negative natural logarithm is taken to simplify the equations of the normal distribution

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathcal{C}}{\arg\max}\ f(\dot{\boldsymbol{z}}|\boldsymbol{x}, k, \boldsymbol{z}) \tag{7.55}$$

$$= \underset{\boldsymbol{x}\in\mathcal{C}}{\arg\min}\ \left\{\tfrac{1}{2}\ln\det(2\pi(\boldsymbol{C}_{\tilde{\dot{z}}} + \boldsymbol{C}_{\dot{z}})) + (\dot{\boldsymbol{z}} - \tilde{\dot{\boldsymbol{z}}})^{\mathsf{T}}(\boldsymbol{C}_{\dot{z}} + \boldsymbol{C}_{\tilde{\dot{z}}})(\dot{\boldsymbol{z}} - \tilde{\dot{\boldsymbol{z}}})\right\}, \tag{7.56}$$

where $\mathcal{C}$ denotes the admissible region in which the prior is constant. The first term of (7.56) stays constant under the previously stated assumption of a negligible modeled uncertainty. Instead of maximizing the likelihood, the following loss function is minimized:

$$L(\boldsymbol{x}, k) = (\dot{\boldsymbol{z}} - \tilde{\dot{\boldsymbol{z}}})^{\mathsf{T}}(\boldsymbol{C}_{\dot{z}} + \boldsymbol{C}_{\tilde{\dot{z}}})(\dot{\boldsymbol{z}} - \hat{\dot{\boldsymbol{z}}})\,. \tag{7.57}$$

This comes with the advantage that the minimum

$$L(\hat{\boldsymbol{x}}) = d^2 \tag{7.58}$$

can be again gated using the chi-squared distribution.

7.2.1 Loss function topography

The loss function of the boundary-value formulation is shown for feasible $\boldsymbol{x}$ values in Figure 7.16 and Figure 7.17. The figures are generated for observations from the same example geostationary object as for the initial-value method. The satellite is located at longitude $\lambda = 100°$, while the sensor is at $\lambda = 30°$ and $\phi = 40°$. The time interval between the tracklets is 3 hours in the first and 27 hours in the second figure. Measurement uncertainty is again modeled with $\sigma_{\alpha,\delta} = 2\,''$ and $\sigma_{\dot{\alpha},\dot{\delta}} = 0.004\,''/\text{s}$. The loss functions are evaluated for a reasonable choice of k-values, i.e. close to the true solution. It can be observed that for both time intervals, only one k shows a distinct loss function minimum in the range of 10^0. The function shows a beneficial topography, i.e. only one local and thus global minimum inside the feasible domain. The shape resembles the one of a quadratic function and shows no curved local valleys as in the initial value formulation. The first point is advantageous when using quasi-Newton optimization themes, which approximate the function with a quadratic one.

However, the function can become more challenging to minimize for special cases. For the analysis,

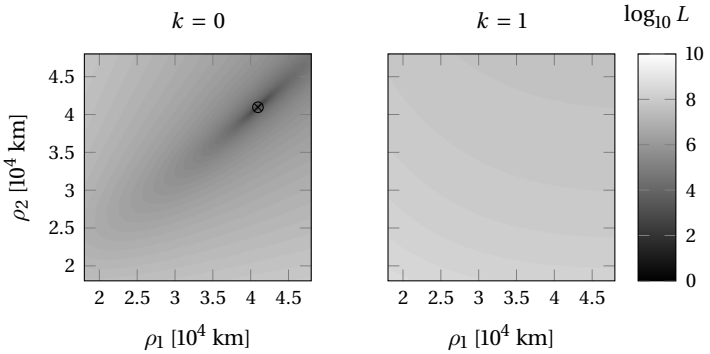


Figure 7.16: *Loss function topography for example geostationary object (explained in text) re-observed after 3 hours. The two plots depict different loss functions for specified k-values.*

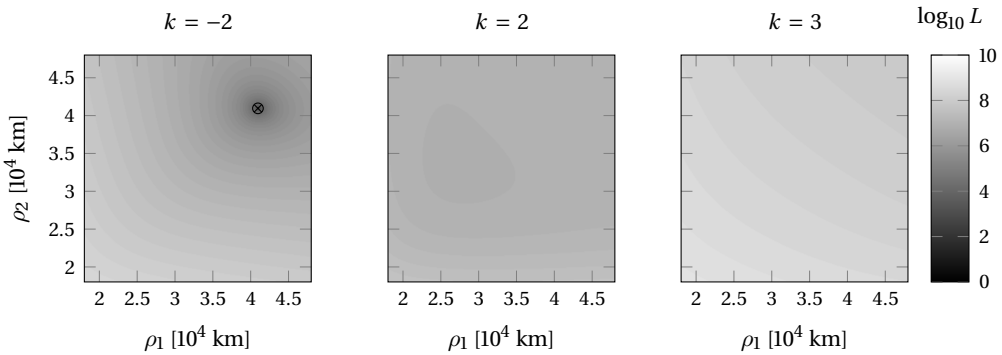


Figure 7.17: *Loss function topography for example geostationary object re-observed after 27 hours. The three plots depict different loss functions for specified k-values.*

the loss function is divided into two components: the angular rate differences at the first and at the second epoch. The decomposition is shown in Figure 7.18 for the 27 hour time interval case. Additionally, the line with no topocentric range velocity is shown for both epochs. For geostationary objects or objects close to the perigee, the orbital solution lies on (or close to) this line. For others, this path effectively splits the domain into two feasible areas: one with a positive and one with a negative range-rate. The

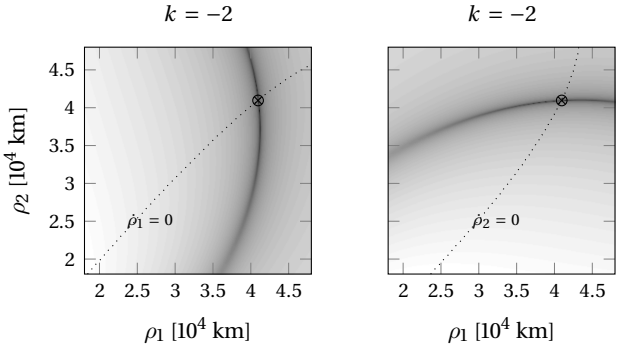


Figure 7.18: *Admissible region decomposition: each side shows the loss on one side of the problem, i.e. the scaled angular rate differences at t_1 and t_2.*

decomposition illustrates two difficulties that can arise.

Firstly, when the angular rates are better determined at one epoch than at the other, one side will gain weight and contribute more to the shape of the function. Reasons for better determined angular rates at one epoch are e.g. when the object moves out of the field of view at the other epoch and less measurements per tracklet are collected. If one side dominates, the loss function becomes a curved, narrow, and possibly flat valley. Optimization schemes as the quasi-Newton ones might fail to approximate the gradient properly after reaching the valley. Thus, the mostly slower but safer choice is to select the line search as presented in section 4.2.

Secondly, the loss function can become more challenging for geometries where the two valleys of the decomposition overlay. This geometry is e.g. obtained if the object is observed two times at approximately the same location on its orbit. In case of a geostationary orbit, this unfavorable but likely case happens when observing an object after one sidereal day. Figure 7.19 illustrates the problem for a re-observation time of 24 hours (which is slightly more than a sidereal day). As discussed, the shape is again unfavorable for optimization schemes. Even though, the two valleys can be divided using the range-rate criterion, again a narrow and flat valley is obtained.

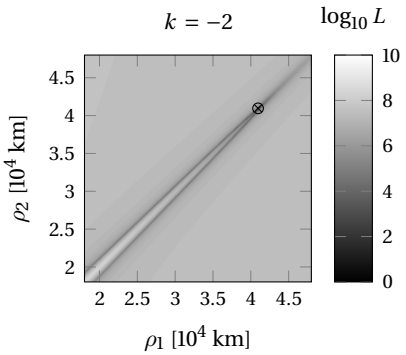


Figure 7.19: *Challenging loss function topography for example geostationary object re-observed after 24 hours.*

As a consequence, these geometries should be avoided during observation planning itself. It should be noted, that these geometries are unfavorable in terms of orbit improvement as well, as less information can be yielded from the observations (Cordelli, 2015). A more thorough discussion of the issue is provided in Chapter 8.

7.2.2 Admissible region

An advantage of the boundary-value formulation over the initial-value formulation is that different multi-revolution solutions are automatically separated by the use of k in the orbit computation. Unlike in the initial-value formulation, a separation of the region is not required anymore. Each loss function corresponding to one k is expected to contain one minimum. However, global boundaries which restrict the solution space can be likewise restricted. They can help to find good starting points or exclude solutions that diverge into unfeasible areas.

In principle, there are two ways of restricting the orbit solutions. One before actually computing the orbit $\tilde{\boldsymbol{y}}$ using Lambert's theorem, and one after computing it. Roscoe et al., (2013) state criteria for the elimination of orbits in the first case. They introduce perigee and apogee bounds to restrict the range at each epoch before combining the observations

$$a_{\min}(1 - e_{\max}) \leq \|\boldsymbol{r}\| \leq a_{\max}(1 + e_{\max})\,. \tag{7.59}$$

Inserting (3.22) gives the bounds

$$\rho_{\max} = -\boldsymbol{R}\cdot\boldsymbol{u} + \sqrt{(\boldsymbol{R}\cdot\boldsymbol{u})^2 + a_{\max}(1 + e_{\max}) - \boldsymbol{R}\cdot\boldsymbol{R}} \tag{7.60}$$

and

$$\rho_{\min} = -\boldsymbol{R}\cdot\boldsymbol{u} + \sqrt{(\boldsymbol{R}\cdot\boldsymbol{u})^2 + a_{\min}(1 - e_{\max}) - \boldsymbol{R}\cdot\boldsymbol{R}}\,. \tag{7.61}$$

They can be used to get a rectangular feasible domain. Additionally, combinations are excluded where the least possible semi-major axis

$$a_{\mathrm{L,min}} = \frac{\boldsymbol{r}_1(\rho_1) + \boldsymbol{r}_2(\rho_2) + c(\rho_1, \rho_2)}{2} \tag{7.62}$$

is above the maximum allowed limit $a_{\max}$. Likewise, the minimum eccentricity solution in Lambert's

theorem

$$e_{\mathrm{L,min}} = \frac{\boldsymbol{r}_1(\rho_1) - \boldsymbol{r}_2(\rho_2)}{c(\rho_1, \rho_2)} \tag{7.63}$$

is used to remove hypotheses with $e_{\mathrm{L,min}} > e_{\max}$. They also introduce angular rates and their uncertainties into the boundary definition.

Again, constraints can be used to incorporate a-priori knowledge of the orbital distribution of the objects caused by the long-term dynamics, i.e.

$$\mathcal{C}_L(\boldsymbol{x}) = \{\boldsymbol{x} : \gamma - \Delta\gamma \leq \arccos\left(\boldsymbol{h}(\boldsymbol{x}) \cdot \boldsymbol{h}_L\right) \leq \gamma + \Delta\gamma\}, \tag{7.64}$$

where the normal vector is computed with

$$\boldsymbol{h}(\boldsymbol{x}) = \frac{\boldsymbol{r}_1(\rho_1) \times \boldsymbol{r}_2(\rho_2)}{\|\boldsymbol{r}_1(\rho_1) \times \boldsymbol{r}_2(\rho_2)\|}\,. \tag{7.65}$$

The constraints, which can be applied before solving Lambert's problem iteratively, are illustrated in Figure 7.20. Another approach is to bound the orbits after computing the Lambert's solution, which consequently increases the computational burden.Figure 7.21 shows the region boundaries as defined in section 7.1.2 for semi-major axis $\mathcal{C}_a$, eccentricity $\mathcal{C}_e$, and perigee height $\mathcal{C}_{\mathrm{p}}$ bounds. The semi-major axis boundary can be used to obtain the interval of admissible k-values.

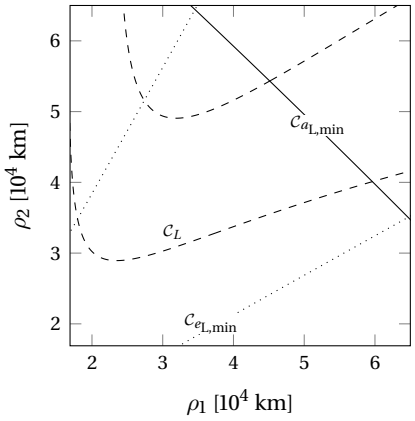


Figure 7.20: *Admissible region: lines denote the boundaries which are computed before the actual Lambert's problem solution is computed.*

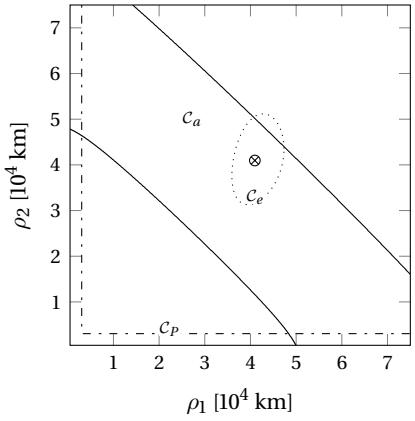


Figure 7.21: *Admissible region: lines show boundaries which are obtained after computing the Lambert's problem solution.*

7.2.3 Angular-momentum conservation

Taking all observed quantities into account, the range at the second observation is directly expressed in terms of the first epoch range from (7.45)

$$\rho_2(\rho_1, \boldsymbol{u}_1, \dot{\boldsymbol{u}}_1, \boldsymbol{u}_2, \dot{\boldsymbol{u}}_2)$$

using the angular-momentum conservation

$$\boldsymbol{r}_1 \times \dot{\boldsymbol{r}}_1 = \boldsymbol{r}_2 \times \dot{\boldsymbol{r}}_2 \tag{7.66}$$

Consequently, if the angular rates are considered accurate enough, the problem reduces to a one-dimensional minimization with

$$\hat{\rho}_1 = \min_{\rho_1} L(\rho_1, \rho_2(\rho_1)) \,. \tag{7.67}$$

Even when they are not accurate, the value can still be used as a starter for the two-dimensional problem. Furthermore, if the loss is already below a reasonable threshold, another improvement step is not required anymore to decide whether two observations belong to a common object. The benefit of using the one-dimensional loss function is safer and faster convergence.

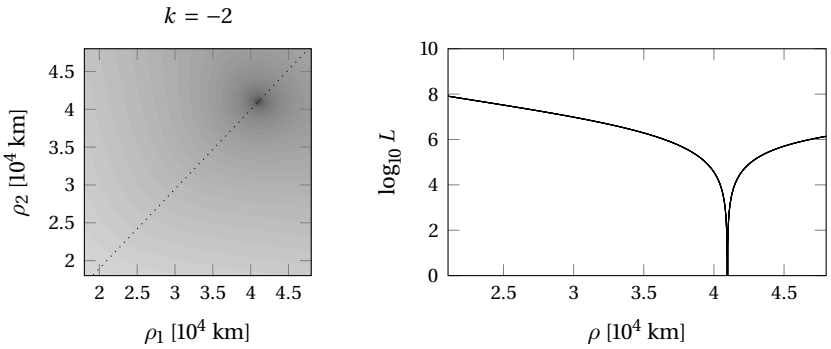


Figure 7.22: *Path with conserved angular-momentum on top of the two-dimensional loss function (left). On the right side is the one-dimensional loss function for the same object and observation geometry.*

An additional use of the angular-momentum conservation is the reduction of the solution space. When approximating the uncertainty of the second range using the linear transformation

$$\sigma_{\rho_2} \approx \left(\frac{\partial \rho_2}{\partial(\boldsymbol{z}, \dot{\boldsymbol{z}})}\right) \boldsymbol{C}_{z,\dot{z}} \left(\frac{\partial \rho_2}{\partial(\boldsymbol{z}, \dot{\boldsymbol{z}})}\right)^{\mathsf{T}} , \tag{7.68}$$

a feasible stripe around the previously defined path is computed allowing for a specific σ-margin. The partials are computed either analytically or using a finite difference scheme.

7.3 Comparison

Although both methods use the same principle, namely a Bayesian' association based on the maximum-likelihood estimate, they differ significantly in their performance. The boundary-value formulation allows the easy separation of different multi-revolution solutions and the resulting loss function is typically easier to solve for optimization algorithms. The simplicity of the problem assures some robustness of the approach. A minimization of the initial-value loss function can be made equally robust using the k-restricted domains and a coordinate conversion, however, this comes with considerable additional implementation efforts and a thus a loss in transparency.

As the initial-value problem from section 2.1 allows the direct propagation of the satellite orbit using an arbitrary complex force model, the same can be straightforward implemented in the association and initial orbit determination approach. The same force model can be, in principle, used for the boundary-value formulation, but requires more implementation efforts such as a multiple-shooting or numerical continuation solver (cf. section 2.2).

The computational costs of using both formulations are compared in (Siminski et al., 2014a). Around 200 tracklets collected in two nights are tested against each other (all-vs-all), resulting in a total of $\sim 16,000$ association tests. The initial-value approach is implemented using k-restricted domains, the alternative grid, and a differential evolution minimization scheme (as explained in section 4.3) in order to increase the chances of catching every local minimum. The boundary-value approach uses a quasi-Newton solver (as in section 4.1) due to the simplicity of the loss function. The association based on the boundary-value formulation outperforms the initial-value one by a factor of 50 for the used data set (i.e. 2% of the computational time is required for the boundary-value method compared to the initial-value one). The speed difference is caused by various reasons: the different optimization methods used, the selection of starting values, and the loss function topography. However, a fair comparison should use exactly the same settings for both formulations.

When incorporating a high-fidelity force model into the orbit propagation, the initial-value formulation can become more efficient than the boundary-value one. Each evaluation of the boundary-value loss function will require a solution of a perturbed-dynamics Lambert's problem. Depending on the efficiency of the used solver, this requires several orbit propagations for one function evaluation. The accuracy results in section 8.2,however, indicate that a high-fidelity model is not always necessary for geostationary objects (of course neglecting special cases such as high-area-to-mass ratio objects). Furthermore, as long as the two-body orbit solution obtained by the loss function minimization is good enough it can be also improved using a least squares adjustment and a complex force model.

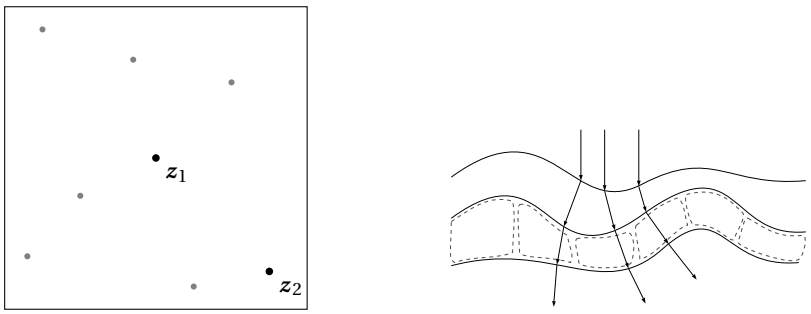
The association and initial orbit determination for a large set of measurements (e.g. from one or several nights) can be trivially parallelized as each test is independent on each other. As long as enough computational resources are available, both formulations can be made sufficiently fast for regular observation processing and catalog maintenance. However, another performance criterion becomes more important than the computational burden and complexity of the implementation, namely the sensitivity to errors in the measurements. It will be calculated in the next chapter together with the mentioned accuracy assessment.

8. Sensitivity analysis

Both formulations, the boundary as well as the initial-value one, exhibit certain advantages and disadvantages over each other. The methods use the same amount of information of the two tracklets, namely the angles and angular rates including their uncertainties. However, the robustness w.r.t. variations in the noise and transformation errors differ between the two formulations.

The uncertainty estimate is based on a noise model. The noise model describes the distribution of errors in the measurements. Good knowledge on the error distribution allows a more robust and more unambiguous association of observations (see also the discussion in section 5.5). Overestimating the errors leads to larger variances and covariances in each respective loss function of the two formulations. This effectively decreases the statistical distance d^2 between the observations and hence leads to more pairs passing the association test. The noise model and accordingly an association threshold can be calibrated using tracklets of objects with well known orbits.

However, as discussed in section 3.2, the noise and bias changes, amongst others, with the weather, viewing direction or aging of the sensor. The errors even differ for objects appearing on the same image but within a different area, e.g. one which is less well represented by the plate model. The latter effect is illustrated on the left side of Figure 8.1. The observation $\boldsymbol{z}_1$ is surrounded by reference stars, which allows a good plate representation. However, the second observation $\boldsymbol{z}_2$ is less well represented as it lies close to the image edge and is locally surrounded by less stars.



(a) *Plate model* **(b)** *Seeing*

Figure 8.1: *The error sources for astrometric observations which are difficult to eliminate through calibration. The left side (a) shows the effect of the reference star distribution on the plate model. Satellite images near the center of the image are less affected by distortions and can thus be measured more accurately than objects near the edges. The right side (b) shows the influence of seeing: unpredictable turbulences in atmosphere affect individual observations. Both can be theoretically estimated from the residuals of the plate model solution.*

Thus, a complete and proper modeling of the exact noise model using calibrated values becomes challenging. Long-term changes can be accounted by recalibrating the noise model. Short-term changes, e.g. caused by the seeing (see right side of Figure 8.1), could be, in principle, estimated using the residuals of the plate model fit. In practical applications, a fixed large enough noise estimate is assumed to suc-

cessfully link most observation pairs, also under bad viewing conditions. The effect of false noise model assumptions on the association performance is studied in this chapter.

8.1 Association performance

The association performance is quantified by two different measures: the success rate of the associations and the rate of false associations. The first describes the number of observation pairs generated by the same object which are successfully matched. Ideally, a number of 100 % is wanted, as the information of the tracklets is otherwise lost (or at least lost until it can be matched to an object in the catalog or another measurement). When subtracting the success-rate from 100 %, the number of missed correlations can be assessed.

The rate of false associations, or false positives, causes no loss in information but an increase in computational burden. Every object candidate which is created by two tracklets needs to be assessed and further processed. A low number is favorable for this rate. False positive rates must be carefully interpreted. A few percent might give an impression of a negligible problem. However, the absolute number of unphysical objects is important as it directly scales the computational efforts required. If 1000 observations are collected which could not be correlated to cataloged objects, around half a million pairs must be tested. 1 % of false positives then creates around 5000 object candidates (where observations are used multiple times for different candidates). The problem increases with the numbers of uncorrelated tracklets used for the association tests. Another difficulty which arises when discussing false positive rates, is the proper sample selection. When testing measurement pairs from a schedule which observes objects on completely different orbits (e.g. different orbital plane, or large angular separation), the false-positive rate will consequently be low. In contrast, when using a measurement set which contains observations of co-located closely spaced objects, mistagging becomes highly likely and the rate will increase. Co-located satellites typically perform regular maneuvers to stay in their relative position. Unless the maneuvers are identified in the measurement to catalog object correlation process, the orbits must be newly determined with the presented methods. The difficulty of misassociation is thus a regularly appearing problem when observing object on the geostationary ring. Another scenario which requires fast and unambiguous initial orbits, is right after a break-up of a satellite. The fragments will most likely be distributed close to the original object, before they are drifted apart due to the dynamics.

Both methods rely on approximate transformations of uncertainties and use the compressed information of the tracklet, i.e. angles and angular rates. Thus, a successive least squares orbit improvement is always advised in order to remove unfeasible candidates. A final decision (confirmation) after the linking based on e.g. 3σ bounds for the least squares residuals is a common technique and also proposed by Milani et al., (2011) and Fujimoto et al., (2014b). This improvement is also advised in order to obtain the most likely orbital solution using the complete information of the two tracks. Typical false positive rates are discussed in (Siminski et al., 2014a). A systematic approach to test the false association probability is used in this thesis and presented in the Chapter 9.

8.1.1 Test measurement set

Tracklets collected by GSOC's surveillance telescope (as explained in section 3.1) are used to test the performance. The telescope setup is still in development and testing phase while writing this thesis. Consequently, the performance parameters differ depending on the used camera or other setup changes. Observations are selected from 13 nights in December 2014 and January 2015, where the setup and weather was fairly stable. The location of the telescope (for testing) is the Zimmerwald observatory in Switzerland. The nights of the observations are summarized in Figure 8.2.

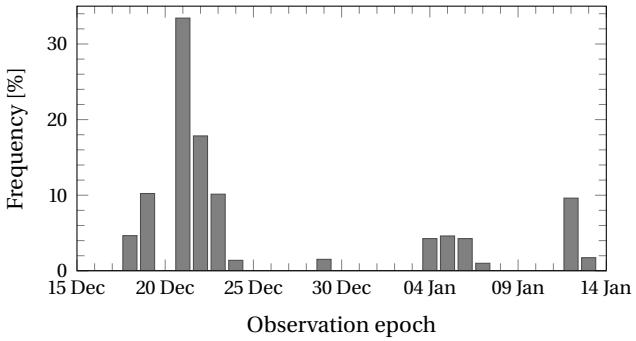


Figure 8.2: *Distribution of the collected measurements in tracking mode for selected time frame between 15 December 2014 and 15 January 2015.*

Instead of using tracklets of unknown origin, a special set was selected with known association to catalog objects. These truth associations serve as a reference to test the proposed methods. The measurements are collected in tracking-mode, i.e. the objects of interest are deliberately observed multiple times a night and the field of view is centered around the object location. The object observed in each tracklet is well known and accordingly also the associated tracklets. Additionally, the noise differences resulting from uneven distribution of reference stars is reduced as the objects are located in the center. Other noise varying effects are still present but the overall model variations are reduced when compared to a survey-mode or larger sets of measurements.

The re-observations times and geometries play an important role in the orbit determination process. If an object is re-observed at the same location within its orbit, less information gain is obtained than at another orbit position. In section 3.5 the typical re-observation conditions of a survey are discussed. However, as the measurement set here is obtained through deliberate tracking, a different distribution is expected. The resulting distribution of re-observation times is shown in Figure 8.3.

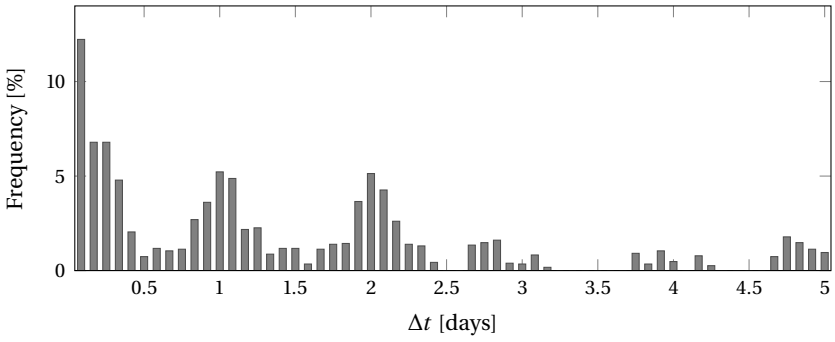


Figure 8.3: *Distribution of re-observation times for the selected sample. Time differences above 5 days are not shown. The number of re-observation times is computed from all subsequent observations and accordingly scaled with the same.*

The re-observation times are similar to a survey scenario and consistent with the distribution in Figure 3.12. Most objects are separated by time spans of 1 to 3 days. The distribution is consistent with the dates in Figure 8.2. Rare re-observation times above 5 days are neglected. The re-observation geometry in Figure 8.4, however, differs from the distribution in the survey simulation in Figure 3.12. Restricting the observations to low solar phase angles during the scheduling process forces the objects to be observed repeatedly at the same position of their orbit and results in a peak of the distribution at small angular separations $\Delta\nu$. Here the peak is at 30° angular separation between the observations, which corresponds to a re-observation time of 2 hours in addition to an integer number of days. The 60° peak is also visible in

the simulated distribution. However, the overall distribution is more even and favorable than the one obtained from the survey. Nevertheless, smaller separations are still present and can cause difficulties.

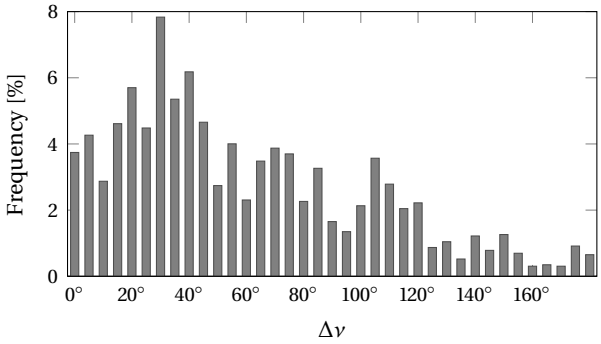


Figure 8.4: *Distribution of re-observation geometries. The geometry is illustrated with the angle between the observations covered of the orbit. The total number is computed from all subsequent observations.*

8.1.2 Threshold dependency

The success rate of each method is not only dependent on the loss function formulation and sensitivity but also on the minimization process. The loss function of the initial-value approach is more challenging to minimize (see the topography discussion is the previous chapter). Both formulations allow multiple-revolution solutions. As a consequence, observations can be matched to each other assuming a false number of completed half-revolutions in between. However, the orbital solution of the result is largely off the target and does not allow a subsequent orbit improvement. Thus, such a match is not considered a successful association. The same force model is used for assessing the initial and boundary value formulation, namely two-body dynamics.

Each loss function minimum d^2 is compared against a predefined threshold d_{u}^2. Whenever the loss value falls below the threshold, the test is counted as a success. The overall success rate is computed by dividing the number of passed d^2 values by the total number of performed association tests. The threshold is varied in order to allow a calibration. The number of correctly matched observation pairs are shown in Figure 8.5 for different noise model assumptions. The plots can also be interpreted as the cumulative distribution of the loss values. The cumulative distribution is properly calibrated when the values resemble the theoretical ones from Table 5.1.

The influence of the bias terms is negligible as it can be seen in the figures one the right hand side. Even an exaggerated bias model of 5 arc minutes changes the output only marginally. This behavior is caused by different factors in each formulation. The bias insensitivity of the two approaches is discussed in section 7.1 and section 7.2.

When scaling up the σ_{N} value, the association performance in terms of successful association is increased. However, the noise model then does not represent the physical properties of the sensor and the viewing conditions. The next section discusses the theoretical and real distribution of the loss function minima used for the gating.

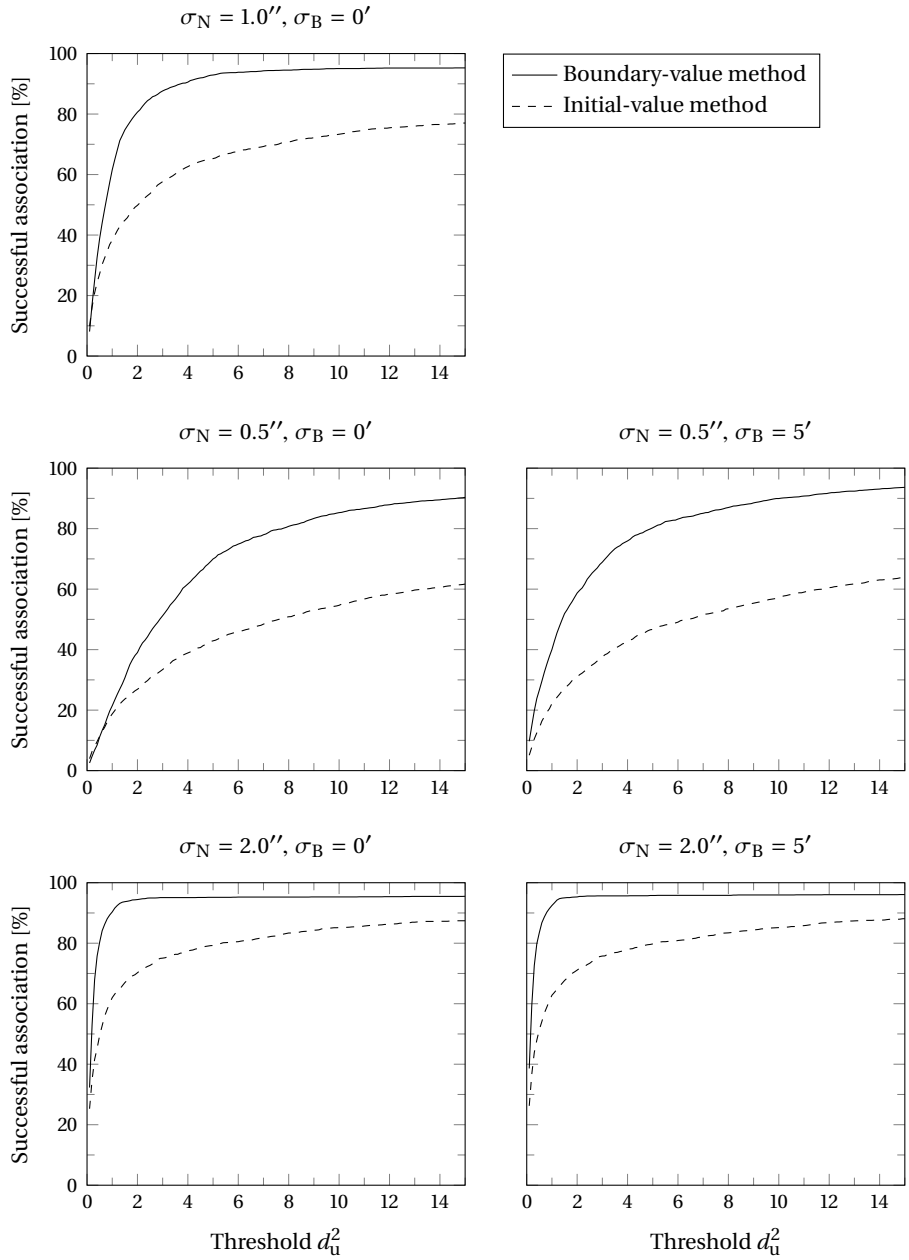


Figure 8.5: *Successful association rate for different noise models. Dashed lines show rates for initial-value formulation while solid lines are used for the boundary-value one.*

8.1.3 Calibration using the theoretical distribution

As described in section 7.1 and section 7.2, a loss function minimum d^2 is, in principle, distributed according to the chi-squared distribution. However, due to erroneous state propagation and uncertainty transformations, the actual distribution can differ from the theoretical one (as discussed in section 5.4). Figure 8.6 shows the actual distribution of the minima of all associated tracklet pairs against the theoretical chi-squared distribution. The different ranges for the different noise models along the vertical axis should be noted.

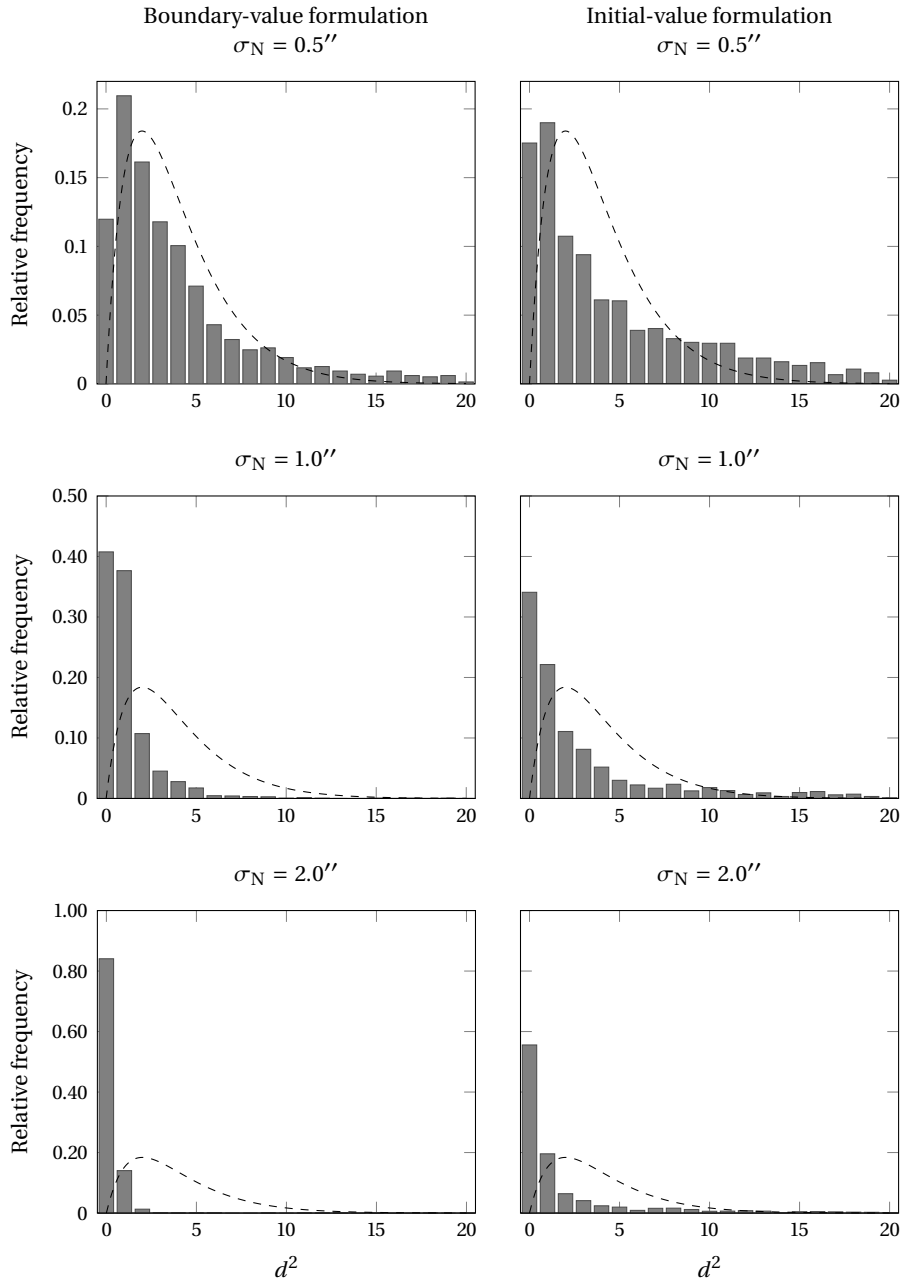


Figure 8.6: *Comparison between theoretical chi-squared distribution and sample histogram for different noise model assumptions. The left side shows distribution for boundary-value and the right side shows it for initial-value formulation.*

A deviation from the theoretical distribution can be observed for all cases. By adjusting the noise estimate, the distribution can be fitted to the theoretical function illustrated with the dashed lines. The fitting works better for the boundary-value formulation, where an assumed standard deviation σ_N around 0.5" is an appropriate value. Both formulations exhibit thicker tails, i.e. more larger loss values than predicted from theory. The difference is more pronounced for the initial-value formulation.

This discrepancy can be caused by various reasons. In order to exclude false noise assumptions, the correlation tests are repeated but the measurements are simulated instead with random errors added according to a known noise model. The two-line elements from the database are propagated to the respective measurement epoch in order to calculate the astrometric positions. The linear motion model is fitted to these in order to obtain the angles and angular rates. The resulting distribution of loss values is shown in Figure 8.7.

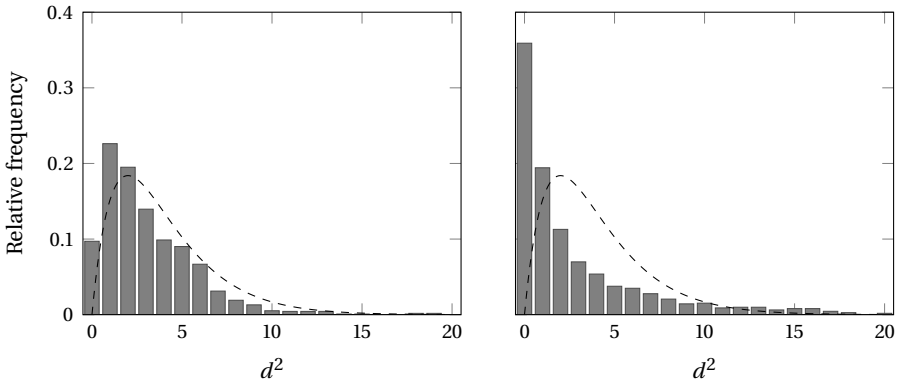


Figure 8.7: *Comparison between theoretical chi-squared distribution and sample histogram for known error distribution $\sigma_N = 0.5''$. Left side shows distribution for boundary-value and right side shows for initial-value formulation.*

In case of the boundary-value formulation, the number of large loss values is reduced and the overall function almost resembles the theoretical one. The initial-value formulation shows the same different distribution as in the previous figures with real measurements. The tail is thicker, and most values are located close to zero and not around the predicted distribution. The differences between both formulations point out another cause for the invalidity of the assumptions, namely the unrealistic transformation of uncertainties. As discussed in section 5.5, the assumption of having a normally distributed uncertainty in the modeled observations is not always valid. The actual distributions can be skewed to one side or even have an other arbitrary shape (as illustrated in Figure 5.15).

The boundary-value formulation requires the transformation of normally distributed angular errors into the modeled angular rates using the solution to the Lambert's problem. As the uncertainty of the modeled angular rates contributes only marginally to the loss scaling in most geometries (see Figure 7.15), it can be mostly neglected and does not cause such a deformation of the loss distribution. The transformation of uncertainty in case of the initial-value formulation is performed using the same approximation for fair comparison.

Alternative methods to express the uncertainty region of states and measurements are outlined in Chapter 5 and can help to overcome this problem with the initial-value formulation. This difficulty, however, highlights a major advantage of the boundary-value formulation. The formulation does not require a complex propagation of uncertainties for most geometries.

Another reason for the deviation of the theoretical distribution from actual one can be a bad re-observation geometry. Bad geometries can affect the solution of both formulations. However, the individual contributions are difficult to separate in case of the initial-value formulation as it is already strongly affected by the transformation. Hence, only the boundary-value formulation is studied in the following. Figure 8.8 shows the distribution of d^2-values, once for very large angular separation between the tracklets and once for very short separations.

The angular separation is computed as the difference in true anomaly between the two epochs. While the left case with good orbital coverage of at least 90° produces an almost perfect chi-squared-distribution, the low coverage case ($\Delta\nu \leq 5°$) does not.

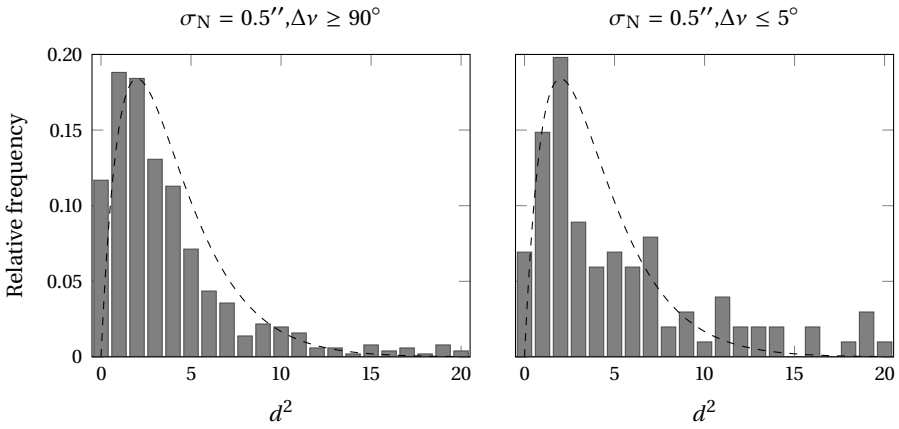


Figure 8.8: *Comparison between theoretical chi-squared distribution and sample histogram for known error distribution $\sigma_N = 0.5''$ and different angular separations $\Delta\nu$.*

The difficulty which arises at these full-revolution re-observation times, is furthermore illustrated in Figure 8.9 by fixing a threshold and counting the number of successful gate passes for different angular separations. A low noise assumption with $\sigma_N = 0.5''$ along with the $P = 0.95\%$ gate according to Table 5.1 is selected to increase the effect. A larger noise value would result in an almost flat bar graph.

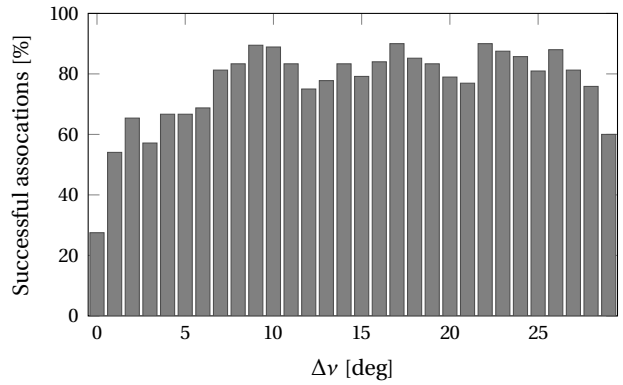


Figure 8.9: *Association performance w.r.t. central angle $\Delta\nu$. Detailed description in text.*

8.1.4 Noise model sensitivity

The required level of detail in the noise modeling is now studied. As discussed in the introduction of this chapter, the actual noise differs due to various reasons. Practically, a fixed value of the expected noise is used for all measurements of one sensor during correlation. However, a more detailed representation could be implemented if it shows a benefit. For this purpose, a reasonable threshold of $P = 0.95\%$ is selected and the assumed noise value σ_N is varied. The number of associations due to the variation then shows the impact of the mismodeled noise.

The function increases up to approximately 1'', where it converges towards a stable level. The final level covers around 95%. Consequently, the figure can be also used to calibrate the noise model. The

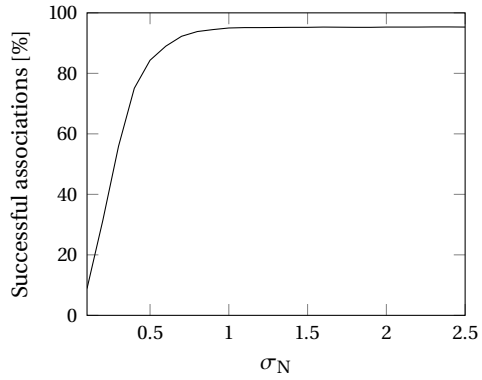


Figure 8.10: *Association performance for fixed $P = 95\%$ gate w.r.t. changing noise assumptions σ_N.*

feasibly noise value for the measurements is in the range of $0.5 - 1$ arc second. If a fixed value above $1''$ is selected for the association, a small variation in the sub arc second range does not play a role. However, if a lower value is selected, small changes in the sensor performance or noise variations can lead to missed associations. The benefit of using a lower value is the better exclusion of false positives.

8.2 Accuracy of solutions

Aside from correlating measurements with each other, the association methods generate first orbital solutions. The success of a correlation is thus also quantified by the accuracy of the orbit it computes. As most preliminary orbit determination methods, the presented approaches rely on reduced information. That is, assuming only two-body dynamics and using the angles and angular-rates instead of the series of angles. Consequently, the result will in most cases be less accurate than when performing a least squares orbit improvement using a high-fidelity force model. The latter improvement is, however, only possible if a starting value close enough to the wanted solution is provided. Otherwise, the iterations may diverge. An absolute value for the required closeness cannot be provided, as the convergence always depends on the individual re-observation geometry, orbit dynamics and the used nonlinear least squares solver. While a simple differential correction process could diverge into a completely different region, damped or trust-region approaches are able to find the best fitting orbit. Observations of high-altitude orbits around the geostationary domain typically allow for courser initial guesses than lower orbits. The approximate accuracy required for stable convergence in the near-geostationary domain is assumed to be around 100 km in position and 0.03 km/s in velocity. Initial states with errors in this order showed still good convergence in orbit improvement tests using a trust-region approach.

This work is motivated by the challenges which arise when observing satellites in a survey scenario. The obtained accuracies of the initial orbits do not represent the actual accuracies of the cataloged states. Most of the observations in a regular survey will be generated by known objects and hence not require an initial orbit. In order to assess the general accuracy of catalog states, detailed observability studies are required accounting for consider parameters (cf. Montenbruck and Gill, (2000)). Assessing the accuracy of orbits using real observations requires precise knowledge at best from a different source. As this information is not available, a simulated data set with known positions is used instead. Consequently, this section discusses the accuracy of initial orbits based on the simulated optical sensor network from Chapter 3. Due to the transformation errors in the initial-value formulation of the association test, the assessment from here on focuses on the boundary-value formulation. The accuracy assessment follows the work presented in (Siminski et al., 2014b).

8.2.1 Simulation framework

In accordance with the plans of GSOC to build up a telescope network with first stations being Sutherland (South Africa) and Zimmerwald (Switzerland), the same network is simulated. Telescopes are operated in survey-mode as described in section 3.4, i.e. no follow-up observations or special tracking campaigns are planned. The programmed survey does not represent an optimal solution but a base-line scenario. Future work can use this program to compare its performance with results of more advances schedulers.

After collecting all observations, a random error with the normal distribution representing the measurement noise is added to each component of the simulated tracklet. Additionally, a constant random normally distributed error representing the bias is added to each element of a tracklet. The details about the observation scenario and errors are summarized in Table 8.1.

Table 8.1: *Simulation setup*

	Value
Noise σ_N	$1''$
Bias σ_B	$5''$
Time spacing	10 s
Observations per tracklet	5

Once the tracklet is formed, it is stored in a list together with the object identification numbers. This allows the later identification and accuracy assessment. The whole year 2015 is simulated in order to capture and visualize temporal variations. A set of around 80,000 tracklets is created in this way.

8.2.2 Results

Tracklets belonging to the same object are extracted and tested with each other. In order to statistically capture the likely geometries, one observation arc is always tested with the tracklet of the next occurrence of the object. If the minimum of the combined loss function d^2 is below the predefined threshold ($P = 0.95\%$), the pair is counted as a match. The orbital solution of this match is then compared with the actual state from the simulation. The differences between preliminary orbits and simulated truth in position and velocity are illustrated with Figure 8.11. Additionally, the errors in orbital elements is shown in Figure 8.12.

The figures summarize the overall performance of the method, but do not provide detailed information on geometry or time dependency. However, the performance of the method is promising as around 90% of the orbits lie in the feasible range of solutions. This performance is particularly promising as the base-line observation scenario still allows an improvement e.g. in terms of re-observation geometry.

A peak can be observed at around 20 km and 0.002 km/s, which is caused by the differences between two-body dynamics and the SGP4 model in case of the typical re-observation time-span of a few days. The median errors in the Radial-Tangential-Normal frame (RTN) of all results are given in Table 8.2. To suppress the influence of the large outliers, the median of the errors is used instead of the mean.

Table 8.2: *Median errors in RTN frame*

	Position			**Velocity**		
Direction	Radial	Tangential	Normal	Radial	Tangential	Normal
Median error	20 km	2 km	2 km	10^{-3} km/s	10^{-3} km/s	$5 \cdot 10^{-4}$ km/s

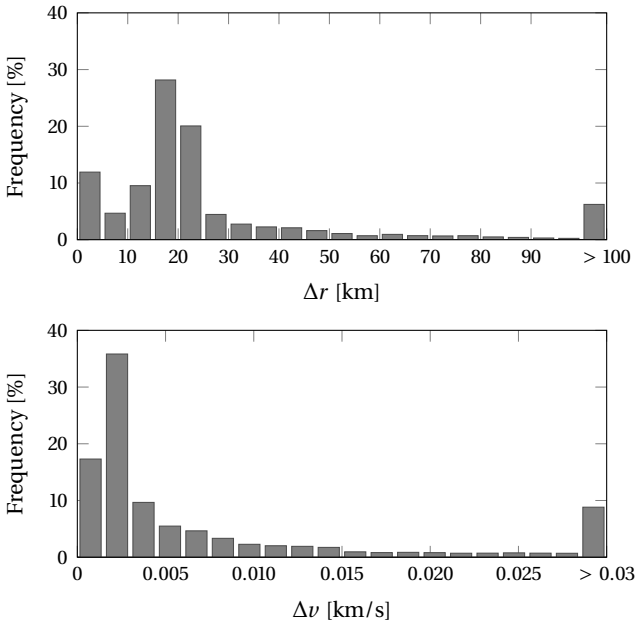


Figure 8.11: *Distribution of position (upper part) and velocity (lower part) errors for complete dataset. Note that the last bin shows all errors above the respective value. The number of pairs denotes the percentage of association test results with an error defined in the bin.*

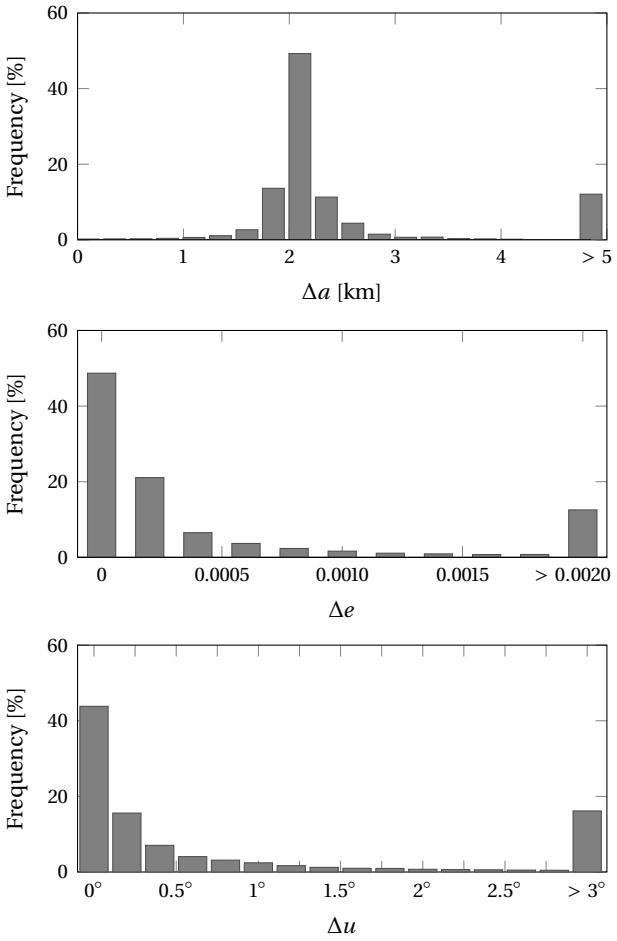


Figure 8.12: *Distribution of errors in semi-major axis Δa, eccentricity Δe, and argument of latitude Δu ($u = \omega + \nu$) for complete dataset. Note that the last bin shows all errors above the respective value. The number of pairs denotes the percentage of association test results with an error defined in the bin.*

Even if the overall performance is promising, a significant amount of solutions still exhibit larger errors. These object candidates would be effectively lost when the final criterion fails, i.e. an orbit improvement and a residuals test. It is important to understand the reasons for this failures in order to either avoid them beforehand while planning observations or by introducing ways to overcome this challenges. Hence, the different causes of failure are analyzed in the following paragraphs and illustrated by showing the dependency on certain variables.

The first analyzed parameter is the re-observation time. The position and velocity errors w.r.t. the re-observation time for all tracklet pairs are shown in Figure 8.13. The performance does not degrade with time. It should be noted that the density of samples decreases with time (a typical re-observation occurs within the first few days). The force model error remains fairly constant for the observed time frame in the near-geostationary domain. This property can be exploited, when the reduction of computational burden is necessary. However, a decrease in performance, i.e. many candidate solutions with larger errors, can be observed for time gap with an integer number of days between the observation epochs. One day corresponds approximately to one orbital revolution of an object on the geostationary orbit. Thus, the angle covered between the tracklet locations on the orbit will be small. The difference here denotes the geometric angular separation and not the physical angular separation between the two states. The dependency on Δv is visualized in Figure 8.14.

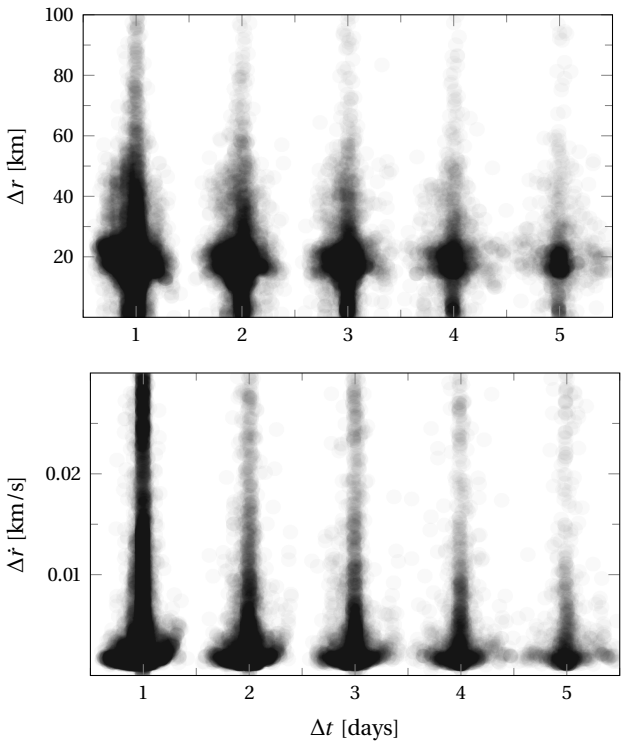


Figure 8.13: *Distribution of position (upper part) and velocity (lower part) errors w.r.t. the re-observation time.*

The scatter plots should be interpreted carefully. While they provide good visualization of the error dependencies, they do not give a good representation of the actual density. The density is roughly visualized by using transparent circles for each tracklet pair. For precise density values (percentage of successful pairs), the above used histograms are more meaningful (Figure 8.11 and Figure 8.12). A relation of the arc covered to the errors of the obtained orbits can be observed. The dependency has two main reasons. The first being that more measurements are collected for certain angular separations, consequently more matching tests can fail. This is particularly visible at 0° and 30° angular separation, where for the latter in

principle a feasible arc is covered. But, due to the observation strategy as outlined in section 3.4, more observations are collected. The same is true for the peak at 60°.

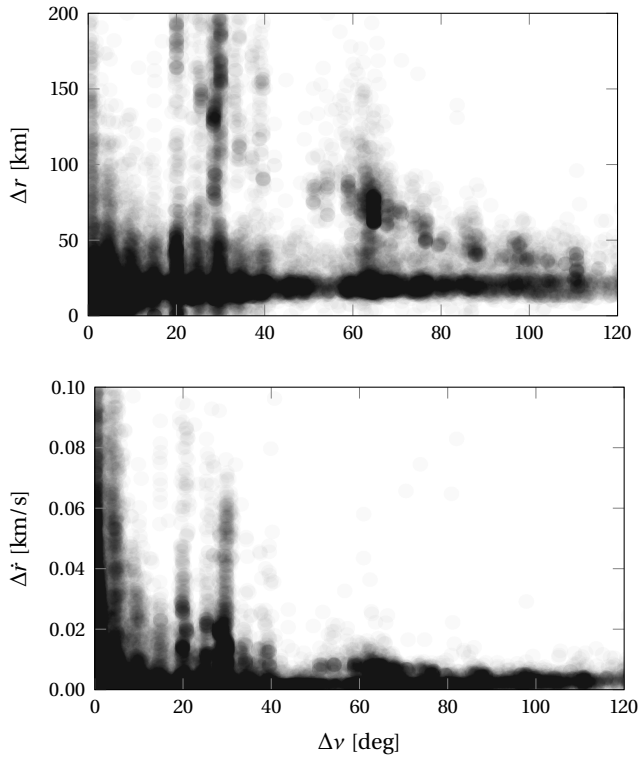


Figure 8.14: *Distribution of position (upper part) and velocity (lower part) errors w.r.t. the orbital arc $\Delta\nu$ covered by the two tracklets.*

The second reason is the reduced orbital information obtained by observing an object at approximately the same geometry and the same location of their orbit. In the case of a perfect full revolution between the two tracklet epochs, the information gain is the smallest possible. Besides being able to draw conclusions on the orbital period (or semi-major axis), several orbits with that period can be fitted to the two arcs. This can be illustrated with the lines of constant semi-major axis in Figure 7.7 and Figure 7.21 for the boundary and initial-value formulation.

A least squares improvement, which uses possibly more information than angles and angular rates, will likely have its difficulties with these geometries as the equation system $\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H}$ becomes nearly singular. Cordelli, (2015) studies the effect of geometry on accuracies using least squares covariance computation and finds the same conclusions.

As pointed out before, the scheduler can be improved in order to reduce the amount of unfavorable geometries. This improvement, however, is limited by the visibility constraints. Other measures have to be taken in order to use the information of such tracklets. A simple approach to overcome the lack of information in the orbit determination process, is to restrict the orbits to certain types, e.g. assuming circular orbits or apogee observations as described in section 6.3.

The tools developed for this analysis can be used for other tasks than testing the accuracy performance, namely sensor or network design and assessment. To demonstrate the capabilities of the proposed sensor network, as described in section 3.5, the number of acceptable object candidates is assessed. This gives an estimate on the completeness of the derived catalog. An acceptable catalog candidate is obtained if at least one tracklet pair creates a feasible solution. Figure 8.15 shows the ratio of acceptable candidates per month of the year in order to visualize the monthly variations. Of course, objects which cannot be observed due to bad seasonal visibility, can also not be correlated. Thus, only observed objects are taken as a reference. Around 70-80 % of the observed objects can be correlated per month. Correlation does not necessarily mean that the orbital solution is sufficiently good. Hence, the rate of acceptable candidates is shown as well, i.e. orbits with a position error better than 100 km and velocity error below 0.03 km/s. The orbits with sufficient accuracy are furthermore separated into two different classes in Figure 8.16 accounting only for the positional error to better visualize the seasonal variations.

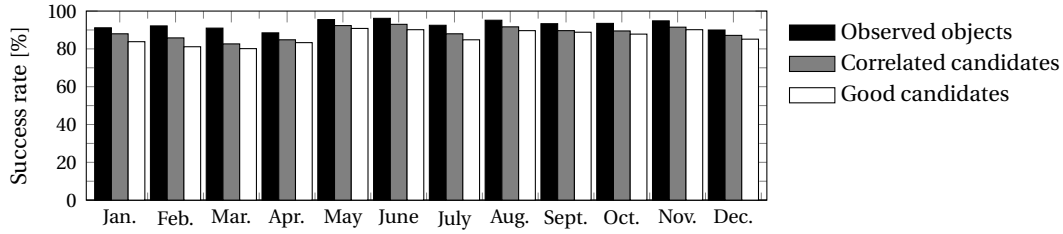


Figure 8.15: *Success rate per month in terms of observed and correlated objects. Additionally, the rate of good candidate solutions is shown.*

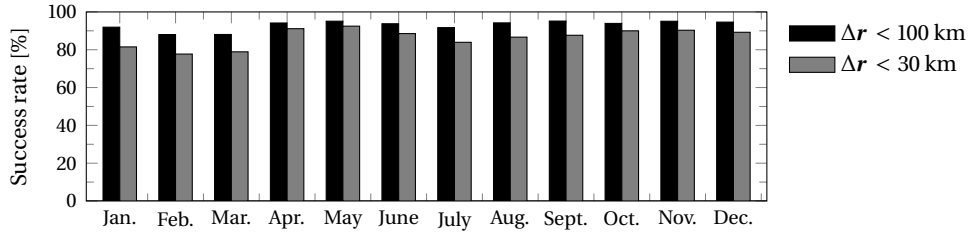


Figure 8.16: *Percentage of correlated objects with a positional error below 100 km and 30 km respectively.*

When observing the different months, it can be noted that some months (e.g. January) show worse results than later months (e.g. May) for the selected year. The effect is caused by the visibility constraints, e.g. due to the adverse influence of the Moon and Milky way. Thus, objects can only be re-observed at unfavorable geometries. The reason for the performance difference between the two different months mentioned before is shown in the following Figure 8.17. The figure shows the distribution of re-observation geometries w.r.t. the central angle once for May and for January. In January most objects are re-observed after an integer number of full revolutions, while May allows also different geometries.

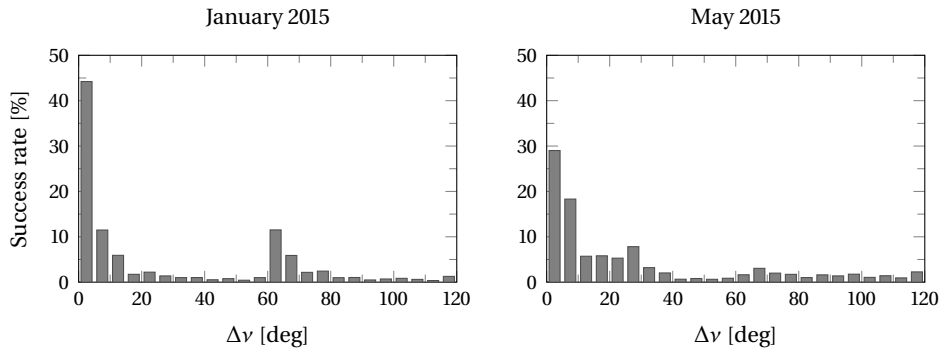


Figure 8.17: *Success rate w.r.t. the re-observation geometry for January (left side, bad performance) and May (right side, good performance).*

Figure 8.18 shows the rate of acceptable candidates (as defined above) w.r.t. the re-observation time in order to see if the performance is degrading with time. Besides the decrease in performance for full revolutions in between the measurements, the accuracy remains in feasible bounds over the selected time span of three days.

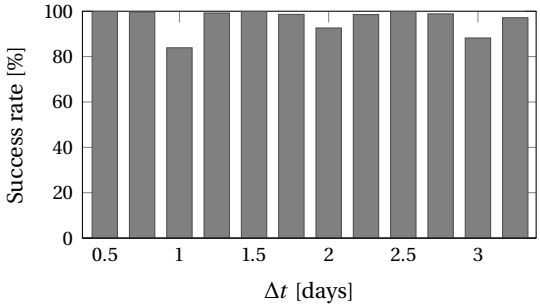


Figure 8.18: *Percentage of acceptable candidates ($\Delta \boldsymbol{r} < 100$ km and $\Delta \dot{\boldsymbol{r}} < 0.03$ km/s w.r.t. re-observation time.)*

This short analysis serves also as an example of using the tools for a network assessment. Other analyses are likewise possible, such as testing different observation strategies or sensor locations and setups. Additional parameters, e.g. final orbit accuracy using follow up observations, should be also considered for a final system analysis but is not included here.

9. Observability

Motivated by the previous chapter, the influence of the observation geometry on the association performance is studied. Capabilities of sensors are limited by the current state-of-the-art technology, location and atmospheric conditions. The observation duration and geometry, however, can be altered in order to optimally use the available observation time. This work studies the determination of first orbits in a survey scenario. If a-priori knowledge on the orbital states is accessible, sensors can be tasked to optimally increase the information of each object. In a survey scenario, the tracklet length and the survey schedule can be adjusted to increase the performance while still covering the complete population.

The observability conditions for a geostationary object are discussed in the first section of this chapter. The common separation of objects along different longitudes around the Earth motivates a systematic study of re-observation geometries. The second part discusses the implications on the observation planning and gives recommendations to unambiguously associate newly detected objects. An early version of this discussion was first presented in (Siminski and Fiedler, 2014).

9.1 Geostationary case study

The geostationary orbit is densely filled with commercial satellites, often co-located in clusters for redundancy and better coverage. False association of observations to other objects within the same cluster are very likely to occur, even if a catalog solution is already available. False associations are either revealed by a subsequent orbit improvement or by adding information from observations of the following nights. In critical situations, such as an approaching intercepting object, ambiguities must be avoided.

Instead of testing the false positive rate with a random set of observations, a systematic approach is selected. A false association scenario is simulated. The scenario is illustrated in Figure 9.1. An object on a perfect geostationary orbit is observed at t_1 from a telescope located at latitude $\phi = 47°$. The object is observed in south direction in order to avoid asymmetries in the results due to the observation geometry. The position is propagated to a second epoch t_2. The position of the satellite is illustrated with the black spheres in the figure. Instead of re-observing the object at the true location, observations of objects separated in longitude by $\Delta\lambda$ are associated. This artificial problem allows studying the principle effects, but it exaggerates the real situation, as the simulated objects lie exactly on the same orbit but with different longitude. Under real dynamics and with applied collocation strategies, the objects would lie on slightly inclined and eccentric orbits.

The resulting orbits for a longitudinal separation of $\Delta\lambda = 0.5°, 1.0°, 1.5°$ are shown in Figure 9.2. The orbits are transformed into the local topocentric coordinate system, i.e. declination δ and hour angle τ. The local hour angle is computed from the right ascension by

$$\tau = \Theta_{\mathrm{LST}} - \alpha\,, \tag{9.1}$$

where Θ_{LST} denotes the local mean sidereal time at the observing station. The position of an object on the geostationary orbit stays constant in this coordinate system and is illustrated with a circle in the plots.

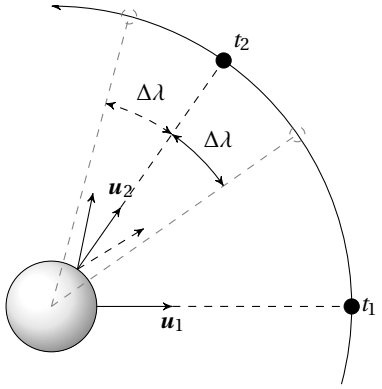


Figure 9.1: *Geometry for the geostationary case study. First observation illustrated with $\boldsymbol{u}_1$. The three vectors named with $\boldsymbol{u}_2$ point towards true and false objects. False objects are located on same orbit but shifted by $\Delta\lambda$.*

The circle also denotes the starting position at t_1 of the false association orbits. Multiple different cases are shown: orbits of a false association after 0.25, 0.75, and 1.75 revolutions and orbits after almost one and three completed revolutions. The arrows point towards the direction of flight on the path and consequently to the location of the second epoch t_2. The gray continuation of the orbits is shown to emphasize the location along the orbit at the observation epoch. The plots are centered around the starting declination δ_1.

The shift in longitude (and hour angle) is accomplished by adapting the semi-major axis. A smaller semi-major axis leads to a drift in prograde direction while a larger one leads to a retrograde shift. In all cases, extrema in declination of the paths are observed between the starting and ending point. The extrema appear at the apsides of the orbits. This symmetric behavior is caused by the formulation of the boundary-value approach which minimizes the angular-rate differences at both observation epochs equally large assuming an equal uncertainty for scaling.

The orbits connecting the observation of the positively shifted ($\Delta\lambda > 0$) false candidates with the initial observation (denoted with prograde solutions) are observed after the object passed the perigee at the minimum. In case of an association with an object with a negative separation (retrograde solution), the object passes the apogee at the maximum. Hence, solutions with the same longitudinal difference but opposite sign will appear with different apparent velocities on the sky. Consequently, the angular rates will be closer or farther apart from the reference ones depending on the observation geometry. The association problem is thus not symmetric for retrograde and prograde solutions, i.e. one direction is easier to associate than the other depending on the re-observation geometry.

The maximum of the declination depends on the re-observation time (see different scaling of the axis). Larger time spans allow for smaller semi-major axis differences required to shift the object to the observed hour angle. The angular-rate differences are likewise reduced as the combined orbit is closer to the reference one. Consequently, the association gets more ambiguous with an increasing number of days in between the observation epochs.

In case of a re-observations after almost full revolutions (second and third plot in Figure 9.2), the observation epochs are placed approximately at the apsides of the combined solution. This assures the smallest possible angular-rates differences. The semi-major axis of the solution is well defined by the re-observation time. In case of an exact integer number of full revolutions, the semi-major axis would be exactly known. The eccentricity can not be solved for and remains unobserved. Hence, a large range of eccentricity values allows a combined solution with an acceptable loss, i.e. small angular-rate differences.

In Figure 9.3, the differences in angular rates are plotted for different re-observation times and a fixed angular separation $\Delta\lambda = 1^\circ$. For exactly full revolutions, the solution of Lambert's problem becomes

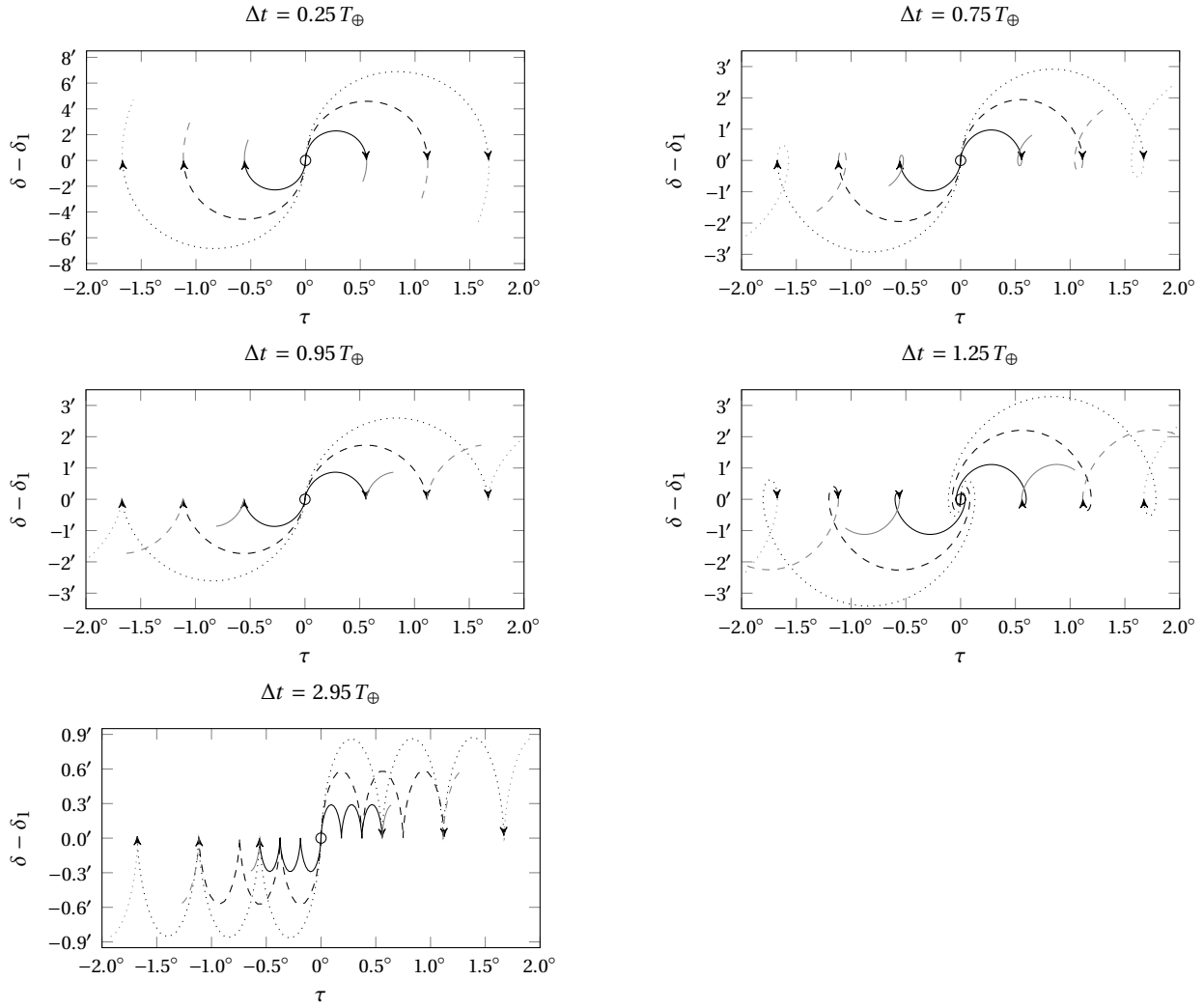


Figure 9.2: *Orbit solutions corresponding to false associations with objects shifted by* $\Delta\lambda = 0.5^\circ, 1.0^\circ, 1.5^\circ$ *shown in local hour angle and declination. The solutions are computed for different re-observation times.*

undefined as the orbital plane cannot be fixed. When fixing the orbital plane, the angular-rates differences will approach zero as described above. However, in order not to use different methods for different geometries, the sections are cut out of the plot. Re-observations at completed half revolutions also cause difficulties in the Lambert's solver, due to the plane uncertainty. However, as these re-observation geometries are anyway unlikely, they are excluded from the discussion. Re-observations of the same object after half a revolution are possibly occurring in a sensor network or in a space-based observation scenario. However, in these cases the geometries will be different as the sensors are located at a different position relative to the object.

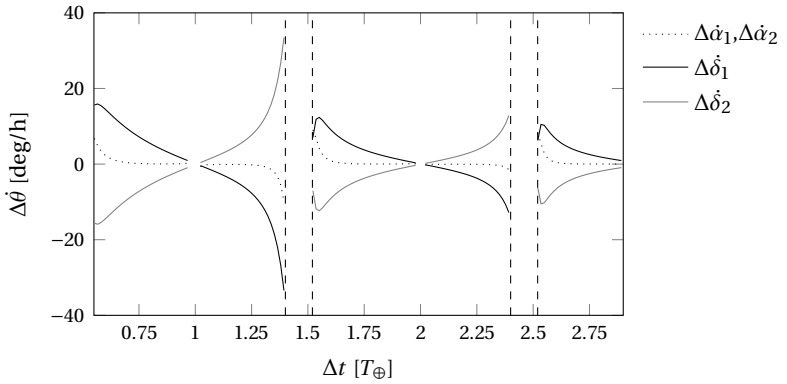


Figure 9.3: *Angular-rate differences between nominal solution* $\Delta\lambda = 0.0°$ *and corrupted one* $\Delta\lambda = 1.0°$.

Re-observation times below half a revolution are not shown in the figure as the differences are considerably larger and do not fit in the same scaling. Linking observations of the same night allows the most unambiguous decisions with the drawback of covering a smaller orbital arc. Declination rate differences $\Delta\dot{\delta}_{1,2}$ are larger than the right ascension rate differences $\Delta\dot{\alpha}_{1,2}$. Both rates decrease to zero when approaching the full-revolution geometries. Due to the symmetry (as discussed above) $\Delta\dot{\delta}_1$ and $\Delta\dot{\delta}_2$ are opposite in sign, while the right ascension rate differences are equal at both epochs.

The angular-rate differences are scaled by the uncertainty to compute the loss. The uncertainty in the modeled rates is dependent on the angular noise σ_N and the number of observations (see section 3.6). Figure 9.4 and Figure 9.5 show the loss d^2 at the minimum for different re-observation times and different retrograde and prograde solutions ($\Delta\lambda \in [-4°, 4°]$). The loss values are once shown for an assumed standard deviation $\sigma_N = 1''$ (Figure 9.4) and $\sigma_N = 0.5''$ (Figure 9.5).

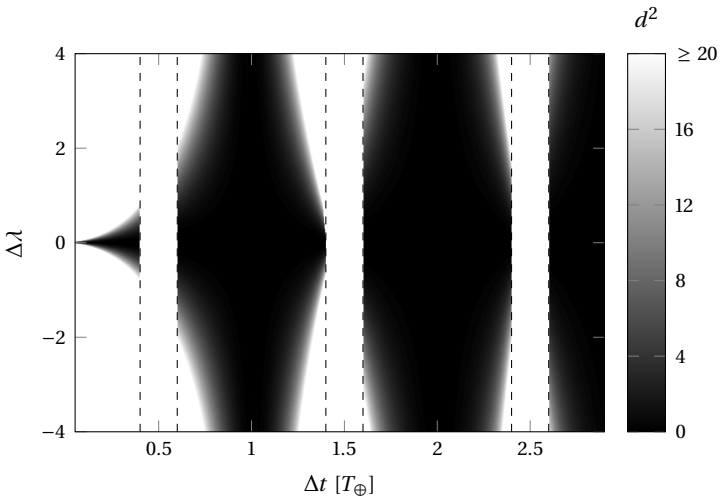


Figure 9.4: *Loss function minima* d^2 *of false association solutions for* $\Delta\lambda$ *range and different re-observation times for the noise model* $\sigma_N = 1''$.

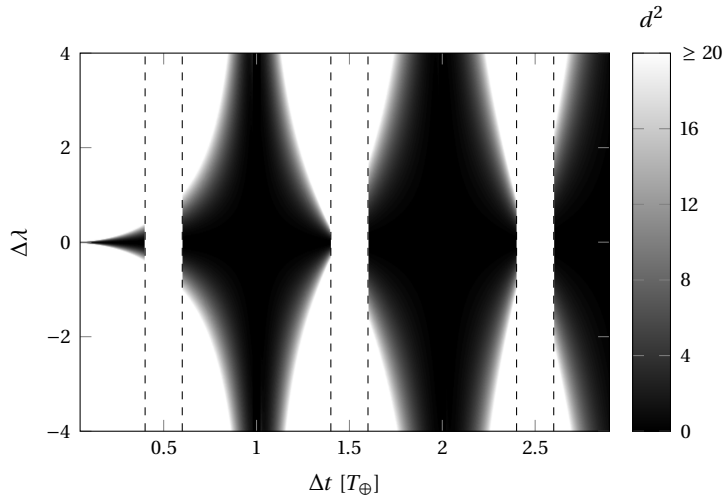


Figure 9.5: *Loss function minima d^2 of false association solutions for $\Delta\lambda$ range and different re-observation times for the noise model $\sigma_N = 0.5''$.*

Loss function values below the $P = 0.95\%$ gate ($d_u^2 \sim 9.86$), result in a false association. The dark areas show the false associations while the gray and white space expose the proper candidate rejections. The figures show three important properties. First, a false association of tracklets obtained from different nights is very likely. A proper selection of the re-observation geometry can improve the association performance in terms of false associations. Lastly, less uncertain angular-rates allow more unambiguous associations. The last two points motivate the next discussion, i.e. at which geometries and for how long should objects be optimally re-observed.

9.2 Implications for observation strategies

As the noise in the measured angles is fixed by sensor and observation conditions, only the duration of the tracklet can be altered to improve the angular-rate accuracy. Figure 9.6 shows the dependency between angular rate uncertainty (illustrated with $\sigma_{\dot{\theta}}$) and the number of individual observations in one tracklet. The time spacing between observations (integration and read-out) is fixed with 20 seconds and the assumed angular noise standard deviation is $\sigma_\theta = 1''$. A longer duration of a tracklet shortens the available observation time for each object in the night. A general recommendation for a long optimal tracklet duration cannot be given as only a few objects can be observed with the limited resources. Nevertheless, the following discussion can be used in the optimization process of finding a good balance between observation time and cataloging performance.

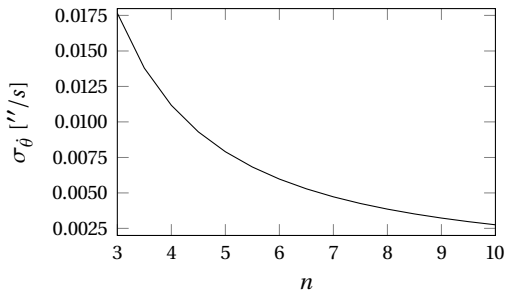


Figure 9.6: *Angular-rate uncertainty depending on number of observations in tracklet.*

Figure 9.7 shows the minimum number of observations within one tracklet needed to unambiguously associate tracklets with a longitudinal separation of $\Delta\lambda = 1°$. The plot is generated by altering the number of observations until the loss function minimum of the false association passes exactly the $P = 0.95\%$ gate. This is a root-finding problem, which is solved numerically using a bisection method.

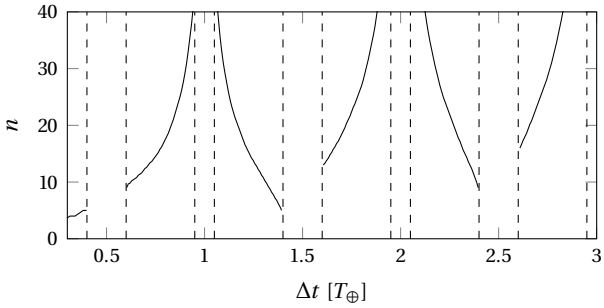


Figure 9.7: *Number of observations required per tracklet for an unambigous association ($\Delta\lambda < 1°$).*

If objects are re-observed within the same night, typical tracklet durations of around 5 measurements guarantee good performance. The following nights show the same pattern as discussed in the previous loss function plots, i.e. the shorter the arc covered by the two tracklets, the more ambiguous the association and the more observation would be needed. 40 observations amount up to a tracklet duration of $\sim$ 13.3 minutes in this setup, which would dramatically worsen the global coverage of the scheduler. While five observations per tracklet allows observing 36 slots in an hour, tracklets with $\sim$ 13.3 minutes duration would only provide 4.5 observed slots.

When observed at the right location, five to ten observations within the tracklet are sufficient to perform an unambiguous association of observations from two subsequent nights. If the objects are re-observed only after two nights, the minimum required tracklet duration is increased. Nights are longer in winter, which allows observing larger arcs than in summer time (up to almost half an orbit). While $n = 5 - 10$ is a feasibly number during winter time, it is not sufficient in summer time, where the arc of the collected tracklets can only cover one third of the orbit. Additionally, not all objects can be observed once in the beginning of the night and once at the end. Figure 3.12 shows typical geometries in a survey scenario. Due to the survey strategy, the arcs covered are typically around 15° and 30°. In order to guarantee unambiguous associations, n must be around 20 - 30. In order not to waste observation time, these long tracklet durations should not be used for all measurements. Instead, only newly detected objects could be verified in this way.

Planning re-observations

One tracklet does not provide enough information to schedule re-observations as the approximate position in the future is not well known. The admissible region provides a way to describe all possible orbits and consequently all possible future observation positions. Holzinger et al., (2014) describe a method to use the information to determine a time-interval left for re-observation. As the amount of possible re-observation locations increases with time, a scheduling is only feasible for short durations after the first sighting. The admissible region is filled with hypotheses, where each one is confirmed or rejected by scheduling an observation at an optimal re-observation epoch. This epoch is defined here in terms of unambiguous association, but could be likewise implemented in terms of information gain. Such an improved schedule is possible for longer time-gaps between the tracklets when restricting the solution space. Typically, when observing objects close to the near-geostationary domain, the orbital period is approximately known. This does not apply to objects on decaying transfer orbits, but to most around the geostationary ring.

Consequently, a re-observation can be scheduled by scanning a complete declination stripe after a specific time interval (as explained in section 3.5).

The strategy is altered in the following way. When detecting a new object, i.e. an observation which cannot be correlated to any object in the catalog, a re-observation at an optimal epoch can be planned. The first tracklet is collected with a fixed $n = 5$. The tracklet duration at the second epoch is then increased until the association performance is achieved (assumptions as above: $\Delta\lambda = 1^\circ$ and $P = 0.95\%$). The resulting recommendation of observation epochs and corresponding tracklet duration is shown in Figure 9.8.

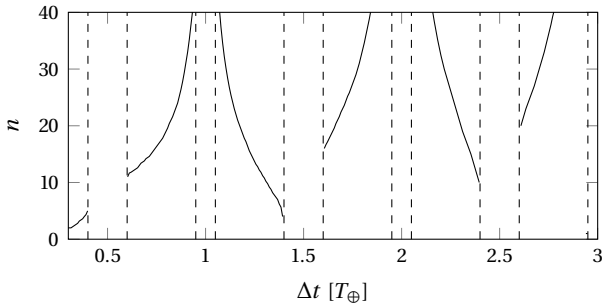


Figure 9.8: *Number of observations required for second tracklet for an unambiguous association ($\Delta\lambda < 1^\circ$).*

The figure mostly resembles Figure 9.7 although slightly longer tracklets are required for each geometry. The difference is small, which indicates that one high accuracy tracklet is enough to perform successful associations. If one tracklet at t_1 yields a more precise angular-rate estimate than the other tracklet at t_2, the loss function will be dominated by the rates at t_1. As an orbital solution is defined by six independent parameters, the angular-rates at one epoch are enough to find a unique solution together with the two angle pairs. A tracklet duration with $n = 20 - 30$ causes a minor loss of observation time when applied only for newly detected objects.

It should be noted that smaller angular separations, as apparent in satellite formations, increase the requirements to an unfeasible range. The consequences of false associations must also be assessed. Primarily, a false association causes an additional computational burden. The ambiguities are either resolved by an orbit improvement, or by subsequent measurements of the objects. Orbit improvement methods experience the same difficulties as the presented methods, i.e. they find feasible orbits connecting two measurement arcs of two different objects. Hence, the final association decision is shifted until the next observation of the objects is obtained. During normal operations, this delayed decision is acceptable and the observation strategy should focus on the maintenance of the complete catalog. In certain situations, however, rapid decisions are required, e.g. directly after a break-up event or if a newly detected object is in close proximity of another satellite. Then, an improved re-observation scheme as illustrated here is applicable.

10. Conclusions

The main goal of this work was to provide measurement correlation methods required for the build-up of a space object catalog. Major difficulties and current challenges when correlating observations have been outlined. Common approaches are either computationally intensive or oversimplify the problem. A general framework is presented in which the best fitting orbit is found by numerically minimizing a loss function. The thesis is structured in a way that formulations repeatedly show up in a different context. This purposely guides the reader through the individual steps of the process and highlights the consistency of the approach with other orbit estimation methods. The framework is applied to two different formulations, each coming with advantages and drawbacks. The individual methods are analyzed in depth for fair comparison and to give elaborate recommendations. The performance is studied theoretically using simulated measurements and practically with real measurements. The overall performance is shown to be very promising.

This final chapter is structured as follows: firstly, each chapter is summarized, where the main conclusion of each part is emphasized. Then, a recommendation for a practical implementation is given. Lastly, future research possibilities are outlined.

10.1 Summary

The thesis starts by recapitulating basic astrodynamics. The main notable aspect from the chapter is that a satellite orbit can be represented by either position and velocity at one epoch or two position vectors at two epochs. Both representation are in theory equivalent. In a practical numerical treatment of differential equations, the initial-value problem is easier to solve as it just involves numerical integration. The boundary-value formulation requires more advanced methods, but nevertheless can be solved.

Telescope setup, observation strategies and scheduler have a direct impact on the performance of the methods and are consequently in detail discussed. Using the long-term dynamics of the geostationary orbit, a base-line schedule is presented. The schedule serves to compute a typical distribution of observations, i.e. objects are mostly re-observed after 1 to 5 days. The objects are often re-observed at almost the same location of the orbit or shifted by 1 - 4 hours. The first is caused by requiring low solar phase angles, while the latter re-observation geometries are caused by the selected strategy.

Numerical optimization is a huge field of research in mathematics and consequently a lot of advanced algorithms are published. The thesis does not go in detail of each algorithm, but summarizes advantages and possible weaknesses. The latter part is important, as numerical optimization can only work properly if the problem is well-defined. Code which runs in a operational service, must be reliable and robust. Consequently, the chapter explains the benefits of appropriate coordinate systems and constraints.

A major difficulty during measurement correlation is the proper representation of uncertainties attached to object states and measurements. The uncertainty regions are required to decide if the association probability is above a certain threshold or not. Alternative representation are presented, which can help to overcome the issues. Additionally, methods to quantify and visualize erroneous uncertainty transformations are provided and later used in the tracklet-to-tracklet association performance assessment.

The chapter about orbit determination provides the statistical background for the later derived association methods. The chapter also provides a compendium of preliminary orbit determination methods under constraining assumptions, e.g. assuming circular orbits. Incorporating assumptions, based on e.g. statistical knowledge about the object population, is often a valid choice and the algorithms can serve to build up a set of tools.

The main research contribution is the chapter about the actual tracklet association and initial orbit determination. The association is performed by computing the minimum of the association loss. If the loss is below a predefined threshold, the observations are correlated. Two different loss functions, following from the two different orbital representations, are presented. Benefits of each representation are discussed.

The performance of the methods in presence of noise is assessed using real observations and simulated ones. Real tracklets are taken from a measurement set with good knowledge of the originating objects. Consequently, the algorithms are tested for the rate of successful associations. The overall performance is shown to be very promising. However, the boundary-value formulation is easier to calibrate and more robust to errors in the noise model.

The simulated observation set is used to test the achievable accuracy of the method. Even with a simplified force model, i.e. two-body dynamics, the methods have shown to deliver sufficiently accurate preliminary orbits. Both studies depict a major challenge by showing a drop in performance at full-revolution re-observation geometries.

The last chapter analyzes the geometry dependency of the performance using a case study approach. The study is used to illustrate the difficulties but also reveals a possibility to overcome them. The performance of the methods can be increased by selecting beneficial re-observation epochs and observing objects for longer durations. The outcome of the chapter can be used in a scenario where observations can be planned for newly detected objects.

Comparison of the two formulations

The two used formulations of the association problem show a different performance. The implemented boundary-value method requires considerably less computational resources, however, as discussed in section 7.3 this advantage is not always guaranteed for other force-models than two-body dynamics and depends also on the complexity of the implementation.

The initial-value formulation requires a special coordinate transformation and is, in general, more difficult to minimize. Main reason for this difficulty is the multi-modal shape of the loss function, i.e. multiple local minima can contain the searched-for orbital solution. Robustness is improved by segmenting the domain of feasible solutions into subspaces, each one containing at most one minimum. A one dimensional search helps to find starting values for each subspace minimization. The boundary-value formulation offers a more favorable loss function topography. Instead of one optimization problem, it provides multiple loss functions for different numbers of completed half revolutions. This automatic segmentation leads to a simple loss functions for each sub-problem. Chapter 8 shows the clear advantage of the symmetric method in terms less transformation errors and easier calibration. The initial-value method requires advanced methods for the transformation of the initial-state uncertainty to the observation space of the second tracklet. Due to the large uncertainties in the angular-rates, the uncertainty cannot be modeled anymore with a multivariate normal density function after propagation. The boundary-value method, on the other hand, generates the orbit hypothesis using the angles-only information of the tracklet. The modeled angular rate uncertainty is mostly negligible when compared to the observation noise, which can be well represented by a normal distribution. Thus, the distribution of the association results can be compared to a theoretical chi-squared distribution and accordingly calibrated.

The initial-value method can be modified using e.g. a Gaussian mixture representation of the uncertainties. If the transformation is properly modeled, the association performance can become similarly good as for the boundary-value approach. Good means that it can be calibrated to achieve a certain association success without creating to many false positives. However, the high-order representation of uncertainties increases not only the complexity of the method but also the computational burden.

To summarize the discussion: the boundary-value formulation offers a robust and fast association and initial orbit determination method without the need for a complicated implementation. The simplicity of the approach makes it transparent and less prone to errors. The used uncertainties are approximately distributed according to the normal distribution and thus allow a proper calibration without much efforts. The accuracy results in section 8.2 additionally show that the implementation assuming just two-body dynamics suffices for most objects. The computationally more costly perturbed Lambert's problem solver is thus only applied for the observations which cannot be linked with the fast implementation. More details about the recommended processing pipeline are given in the following paragraph.

Recommendation

When it comes to the association of tracklets, there is no one-size-fits-all solution. The selection of a proper method largely depends on the quality of the data and consequently on the observation strategy. The developed methods perform very well on tracklets with a duration of a couple of minutes. This observation scenario is a trade-off solution between good coverage of the geostationary domain and association efforts. Shorter tracklets, which do not provide accurate angular-rates, require a different approach and more measurements while longer observation arcs might not require any special treatment at all and can be solved using traditional approaches.

Any of the methods might fail in bad geometries. However, this failure is an intrinsic problem of the data. These geometries must be identified and handled differently. Various strategies can be selected to overcome the issues, i.e. incorporating knowledge about the population or a brute-force sampling approach which will create multiple state hypotheses.

In the end, a process chain should be implemented. Preliminary orbit determination assuming a circular orbit (as described in section 6.3) eliminates the sensitivity to bad geometry. The remaining data can be processed by the developed approaches. If observations still remain uncorrelated, they must be correlated with measurements to come or the solution space must be sampled in a brute-force approach if an estimate is absolutely necessary.

It is furthermore recommended to couple the association, orbit determination and observation planning. A scheduler should use the uncertainty information of cataloged data to decide which object to observe at which epoch. The data used for decision should contain verified but also unconfirmed object states (e.g. from a brute-force admissible region sampling approach).

10.2 Outlook

Instead of searching for feasible combinations of two tracklets, the search-space can be increased to three and more observation combinations. As pointed out in section 3.4, this will increase the combinational burden but also reduces the ambiguities and difficulties. In the mentioned multi-step association process it could be seen as an additional last step. Zittersteijn, (2015) developed a loss function approach based on the work in this thesis, which accounts for an arbitrary number of tracklets. Due to the large number of required tests, an all-vs-all association approach as used in this thesis becomes unfeasible. Consequently, they apply search heuristics to find matching triples.

This work is assuming normally distributed uncertainty regions. This is helpful in terms of transparency and simplicity but does not represent the reality. Future research should carefully balance out where higher order representations are beneficial and where simplifications lead to more robustness. This applies not only to the initial-value and object-to-measurement association but also to the boundary-value approach for certain geometries.

The presented association methods provide necessity tests for measurements, i.e. all measurements which actually originate from a common object fulfill this test. But, they can also fulfill the requirement if originating from different but closely-spaced objects. Unfavorable geometries increase this effect. In these cases, sufficiency can only be guaranteed if more observations are considered. Given a set of observations, a global optimization problem in terms of the combined association likelihood can be defined. Results of a prototype implementation using an expectation-maximization optimization approach (Dempster et al., 1977) are shown in Figure 10.1. Measurements of three closely-spaced co-located satellites are simulated and associated using 11 expectation maximization steps.

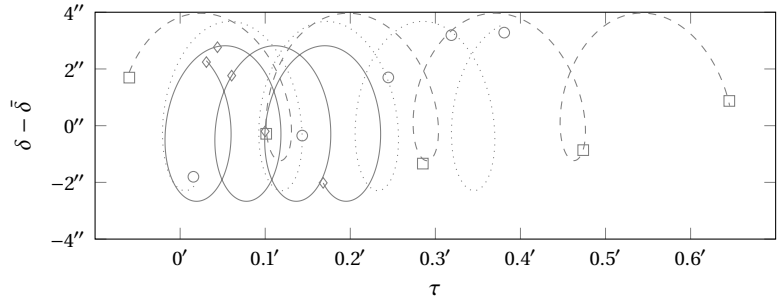


Figure 10.1: *Association of measurements to three closely-spaced objects with a semi-major axis separation of 1 km and eccentricity difference of* $0.0001°$. *Plot centered around* $-6.83°$. *The different markers denote the observations and the different lines the associated orbital solutions.*

Another difficulty arises when associating measurements where the object maneuvered in-between the measurement epochs. With no prior knowledge, the association becomes a completely ambiguous operation in a dense environment, as every observation can be linked to another. But, if prior knowledge is incorporated, e.g. a probability for a maneuver at a certain time, the association becomes less arbitrary. Maneuvers can be learned by classification methods either for each object individually, for each bus type, or for the whole domain (e.g. North-South maneuvers and East-West maneuvers for geostationary objects). If the normal association without a maneuver does not provide a match, the association is repeated allowing for probable maneuvers and considering for their probability. If a likely maneuver is found, the association is a success.

As can be seen in this section, the subject of measurement association and space debris orbit determination still offers a wide range of research opportunities. Even if initial orbit determination has been studied for asteroids already for centuries, new sensors and the increasing risk in the near-Earth space environment open up all new questions. As sensors become better, smaller debris particles will be observed which leads to an even larger amount of observations and consequently larger catalogs. The hazardous debris in our valuable space will eventually call for an international collaboration and hopefully to a measurement exchange. Merging the information adds another task to the list.

Bibliography

Aida, S., Kirschner, M., and Meissner, F. (2014), "Collision risk assessment and mitigation strategy for the GSOC GEO Satellites," *24th International Symposium on Space Flight Dynamics*, Laurel, Maryland, USA.

Ansalone, L. and Curti, F. (2013), "A genetic algorithm for initial orbit determination from a too short arc optical observation," *Advances in Space Research*, **52** (3), pp. 477–489.

Arora, N., Russell, R. P., Strange, N., and Ottesen, D. (2015), "Partial derivatives of the solution to the Lambert boundary value problem," *Journal of Guidance, Control, and Dynamics*, **38** (9), pp. 1563–1572.

Bar-Shalom, Y., Li, X. R., and Thiagalingam, K. (2004), *Estimation with applications to tracking and navigation: theory algorithms and software*, New York: John Wiley & Sons, Inc.

Battin, R. H. (1999), *An introduction to the mathematics and methods of astrodynamics*, rev. ed., Reston, Virginia: American Institute of Aeronautics and Astronautics.

Bauschinger, J. (1906), *Die Bahnbestimmung der Himmelskörper*, 1st ed., Leipzig: W. Engelmann.

Beutler, G. (2004), *Methods of celestial mechanics: volume I: physical, mathematical, and numerical principles*, vol. 1, Springer-Verlag Berlin Heidelberg.

Brent, R. P. (1973), *Algorithms for minimization without derivatives*, Englewood Cliffs, New Jersey: Prentice-Hall.

Broucke, R. and Cefola, P. J. (1972), "On the equinoctial orbit elements," *Celestial Mechanics*, **5** (3), pp. 303–310.

Brouwer, D. (1959), "Solution of the problem of artificial satellite theory without drag," *The Astronomical Journal*, **64**, p. 378.

Clarke, A. C. (1945), "Extra-terrestrial relays," *Wireless world*, **51** (10), pp. 305–308.

Cordelli, E. (2015), "A covariance analysis to optimize the optical follow-up strategies," *Proceedings of the 66th International Astronautical Congress*, Jerusalem, Israel.

DeMars, K. J. (2010), "Nonlinear orbit uncertainty prediction and rectification for space situational awareness," PhD thesis, The University of Texas at Austin.

DeMars, K. J., Hussein, I., Früh, C., Jah, M. K., and Erwin, R. S. (2015), "Multiple-object space surveillance tracking using finite-set statistics," *Journal of Guidance, Control, and Dynamics*, **38** (9), pp. 1741–1756.

DeMars, K. J. and Jah, M. K. (2011), "Evaluation of the information content of observations with application to sensor management for orbit determination," *Proceedings of AIAA/AAS Astrodynamics Specialist Conference*, Girdwood, Alaska, USA.

DeMars, K. J. and Jah, M. K. (2013), "Probabilistic initial orbit determination using Gaussian mixture models," *Journal of Guidance, Control and Dynamics*, **36** (5), pp. 1324–1335.

DeMars, K. J., Jah, M. K., and Schumacher Jr., P. W. (2012), "Initial orbit determination using short-arc angle and angle rate data," *IEEE Transactions on Aerospace and Electronic Systems*, **48** (3), pp. 2628–2637.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, **39** (1), pp. 1–38.

Dorsey, W. W., Neyret, P., Lo, G. J. P., and Ozkul, A. (1986), "Colocation of geostationary communication satellites," *AIAA 11th Communication Satellite Systems Conference*, San Diego, California, USA.

Dubyago, A. D. (1961), *The determination of orbits*, 1st ed., New York: The Macmillan Company.

Escobal, P. R. (1965), *Methods of orbit determination*, New York: John Wiley & Sons, Inc.

Everitt, B. S. and Skrondal, A. (2010), *The Cambridge dictionary of statistics*, 4th ed., New York: Cambridge University Press.

Faubel, F., McDonough, J., and Klakow, D. (2009), "The split and merge unscented Gaussian mixture filter," *IEEE Signal Processing Letters*, **16** (9), pp. 786–789.

Fiedler, H., Herzog, J., Weigel, M., Montenbruck, O., and Schildknecht, T. (2014), "SMARTnet: a sensor for monitoring the geostationary ring," *24th International Symposium on Space Flight Dynamics*, Laurel, Maryland, USA.

Fiedler, H., Weigel, M., Herzog, J., Schildknecht, T., Prohaska, M., Ploner, M., and Montenbruck, O. (2015), "SMARTnet: first experience of setting up a telescope system to survey the geostationary ring," *25th International Symposium on Space Flight Dynamics*, Munich, Germany.

Fletcher, R. (1987), *Practical methods of optimization*, New York: John Wiley & Sons, Inc.

Flohrer, C. (2008), "Mutual validation of satellite-geodetic techniques and its impact on GNSS orbit modeling," Zürich, Switzerland: Schweizerische Geodätische Kommission, Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule Zürich.

Flohrer, T. (2012), "Optical survey strategies and their application to space surveillance," Zürich, Switzerland: Schweizerische Geodätische Kommission, Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule Zürich.

Flohrer, T., Krag, H., Klinkrad, H., and Schildknecht, T. (2011), "Feasibility of performing space surveillance tasks with a proposed space-based optical architecture," *Advances in Space Research*, **47** (6), pp. 1029–1042.

Flohrer, T., Schildknecht, T., Musci, R., and Stöveken, E. (2005), "Performance estimation for GEO space surveillance," *Advances in Space Research*, **35** (7), pp. 1226–1235.

Flury, W. (1991), "Collision probability and spacecraft disposition in the geostationary orbit," *Advances in Space Research*, **11** (12), pp. 67–79.

Früh, C. and Schildknecht, T. (2009), "Catalogue correlation of space debris objects," *5th European Conference on Space Debris*, Darmstadt, Germany.

Fujimoto, K., Alfriend, K. T., and Schildknecht, T. (2014a), "A boundary value problem approach to too-short arc optical track association," *24th AAS/AIAA Space Flight Mechanics Meeting*, Santa Fe, New Mexico, USA.

Fujimoto, K., Maruskin, J. M., and Scheeres, D. J. (2010), "Circular and zero-inclination solutions for optical observations of earth-orbiting objects," *Celestial Mechanics and Dynamical Astronomy*, **106** (2), pp. 157–182.

Fujimoto, K., Scheeres, D. J., and Alfriend, K. T. (2012), "Analytical nonlinear propagation of uncertainty in the two-body problem," *Journal of Guidance, Control, and Dynamics*, **35** (2), pp. 497–509.

Fujimoto, K., Scheeres, D. J., Herzog, J., and Schildknecht, T. (2014b), "Association of optical tracklets from a geosynchronous belt survey via the direct Bayesian admissible region approach," *Advances in Space Research*, **53** (2), pp. 295–308.

Gadaleta, S. M., Horwood, J. T., and Poore, A. B. (2012), "Short arc gating in multiple hypothesis tracking for space surveillance," *Proceedings of SPIE*, **8385**, International Society for Optics and Photonics.

Gibbs, J. W. (1889), "On the determination of elliptical orbits from three complete observations," *Memoires National Academy of Science*, **4** (2), pp. 79–104.

Gooding, R. H. (1990), "A procedure for the solution of Lambert's orbital boundary-value problem," *Celestial Mechanics and Dynamical Astronomy*, **48** (2), pp. 145–165.

– (1996), "A new procedure for the solution of the classical problem of minimal orbit determination from three lines of sight," *Celestial Mechanics and Dynamical Astronomy*, **66** (4), pp. 387–423.

Gronchi, G. F. (2004), "Classical and modern orbit determination for asteroids," *IAU Colloqium No. 196: Transits of Venus: New Views of the Solar System and Galaxy*, ed. by D. W. Kurtz, Preston, UK: Cambridge University Press, pp. 293–303.

Gronchi, G. F., Dimare, L., and Milani, A. (2010), "Orbit determination with the two-body integrals," *Celestial Mechanics and Dynamical Astronomy*, **107** (3), pp. 299–318.

Gronchi, G. F., Farnocchia, D., and Dimare, L. (2011), "Orbit determination with the two-body integrals. II," *Celestial Mechanics and Dynamical Astronomy*, **110** (3), pp. 257–270.

Herget, P. (1965), "Computation of preliminary orbits," *Astronomical Journal*, **70** (1), pp. 1–3.

Herrick, S. (1971), *Astrodynamics: orbit determination, space navigation, celestial mechanics*, London: Van Nostrand Reinhold Co.

Herzog, J., Früh, C., and Schildknecht, T. (2010), "Build-up and maintenance of a catalogue of GEO objects with ZimSMART and ZimSMART 2," *Proceedings of the 61st International Astronautical Congress*, Prague, Czech Republic.

Herzog, J., Schildknecht, T., Hinze, A., Ploner, M., and Vananti, A. (2013), "Space surveillance observations at the AIUB Zimmerwald observatory," *6th European Conference on Space Debris*, Darmstadt, Germany.

Hill, K., Alfriend, K. T., and Sabol, C. (2008), "Covariance-based uncorrelated track association," *Proceedings of AIAA/AAS Astrodynamics Specialist Conference*, Honolulu, Hawaii, USA.

Hill, K., Sabol, C., and Alfriend, K. T. (2012), "Comparison of covariance based track association approaches using simulated radar data," *The Journal of the Astronautical Sciences*, **59** (1-2), pp. 281–300.

Hirzinger, G., Landzettel, K., Brunner, B., Fischer, M., Preusche, C., Reintsema, D., Albu-Schäffer, A., Schreiber, G., and Steinmetz, B. (2004), "DLR's robotics technologies for on-orbit servicing," *Advanced Robotics*, **18** (2), pp. 139–174.

Holzinger, M. J., Luu, K. K., Sabol, C., and Hill, K. (2014), "Probabilistic tracklet characterization and prioritization using admissible regions," *Proceedings of the 2014 Advanced Maui Optical and Space Surveillance Technologies Conference*, Maui, Hawaii, USA.

Horwood, J. T. and Poore, A. B. (2014), "Gauss von Mises distribution for improved uncertainty realism in space situational awareness," *SIAM/ASA Journal on Uncertainty Quantification*, **2** (1), pp. 276–304.

Jones, B. A., Doostan, A., and Born, G. H. (2013), "Nonlinear propagation of orbit uncertainty using non-intrusive polynomial chaos," *Journal of Guidance, Control, and Dynamics*, **36** (2), pp. 430–444.

Julier, S. J. (2002), "The scaled unscented transformation," *Proceedings of the American Control Conference*, **6**, 4555–4559 vol.6.

Julier, S. J. and Uhlmann, J. K. (1997), "New extension of the Kalman filter to nonlinear systems," *Proceedings of SPIE*, **3068**, International Society for Optics and Photonics, pp. 182–193.

Kee, C., Parkinson, B. W., and Axelrad, P. (1991), "Wide area differential GPS," *Navigation*, **38** (2), pp. 123–145.

Kessler, D. J. (1994), "The current and future environment: an overall assessment," *Preservation of near-earth Space for future generations*, ed. by J. A. Simpson, Cambridge University Press, pp. 19–36.

Kirchner, G., Koidl, F., Friederich, F., Buske, I., Völker, U., and Riede, W. (2013), "Laser measurements to space debris from Graz SLR station," *Advances in Space Research*, **51** (1), pp. 21–24.

Klinkrad, H. (2006), *Space debris: models and risk analysis*, Springer-Verlag Berlin Heidelberg.

Kozai, Y. (1959), "The motion of a close earth satellite," *The Astronomical Journal*, **64**, p. 367.

Lancaster, E. R. and Blanchard, R. C. (1969), *A unified form of Lambert's theorem*, tech. rep.

Lenz, S. M., Bock, H. G., Schlöder, J. P., Kostina, E. A., Gienger, G., and Ziegler, G. (2010), "Multiple shooting method for initial satellite orbit determination," *Journal of guidance, control, and dynamics*, **33** (5), pp. 1334–1346.

Mahalanobis, P. C. (1936), "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences*, **2** (1), pp. 49–55.

Mardia, K. V. (1970), "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, **57** (3), pp. 519–530.

Maruskin, J. M., Scheeres, D. J., and Alfriend, K. T. (2009), "Correlation of optical observations of objects in earth orbit," *Journal of Guidance, Control and Dynamics*, **32** (1), pp. 194–209.

Milani, A., Gronchi, G., Vitturi, M. D. M., and Knezevic, Z. (2004), "Orbit determination with very short arcs. I admissible regions," *Celestial Mechanics and Dynamical Astronomy*, **90** (1), pp. 57–85.

Milani, A., Tommei, G., Farnocchia, D., Rossi, A., Schildknecht, T., and Jehn, R. (2011), "Correlation and orbit determination of space objects based on sparse optical data," *Monthly Notices of the Royal Astronomical Society*, **417** (3), pp. 2094–2103.

Molotov, I. et al. (2008), "International scientific optical network for space debris research," *Advances in Space Research*, **41** (7), pp. 1022–1028.

Montenbruck, O. and Gill, E. (2000), *Satellite orbits: models, methods and applications*, Springer-Verlag Berlin Heidelberg.

Musci, R., Schildknecht, T., Flohrer, T., and Beutler, G. (2007), "Evolution of the orbital elements for objects with high area-to-mass ratios in geostationary transfer orbits," *Advances in Space Research*, **41**, pp. 1071–1076.

Noordung, H. (1929), *Das Problem der Befahrung des Weltraums: der Raketen-Motor*, Berlin: Richard Carl Schmidt & Co.

Norris, P. (2015), "Satellite programs in the United States," *Handbook of space security: policies, applications and programs*, ed. by K.-U. Schrogl, P. L. Hays, J. Robinson, D. Moura, and C. Giannopapa, Springer-Verlag New York, pp. 743–774.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007), *Numerical recipes: the art of scientific computing*, 3rd ed., New York: Cambridge University Press.

Price, K. V., Storn, R. M., and Lampinen, J. A. (2005), *Differential evolution: a practical approach to global optimization*, Springer-Verlag Berlin Heidelberg.

Roscoe, C. W. T., Schumacher Jr., P. W., and Wilkins, M. P. (2013), "Parallel track initiation for optical space surveillance using range and range-rate bounds," *Proceedings of AIAA/AAS Astrodynamics Specialist Conference*, Hilton Head, South Carolina, USA.

Schildknecht, T. (1994), "Optical astrometry of fast moving objects using CCD detectors," Zürich, Switzerland: Schweizerische Geodätische Kommission, Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule Zürich.

– (2007), "Optical surveys for space debris," *The Astronomy and Astrophysics Review*, **14** (1), p. 41.

Schildknecht, T., Hugentobler, U., and Ploner, M. (1999), "Optical surveys of space debris in GEO," *Advances in Space Research*, **23** (1), pp. 45–54.

Schneider, M. D. (2012), "Bayesian linking of geosynchronous orbital debris tracks as seen by the Large Synoptic Survey Telescope," *Advances in Space Research*, **49** (4), pp. 655–666.

Schumacher Jr., P. W., Sabol, C., Higginson, C. C., and Alfriend, K. T. (2015), "Uncertain Lambert problem," *Journal of Guidance, Control, and Dynamics*, **38** (9), pp. 1573–1584.

Schumacher Jr., P. W., Wilkins, M. P., and Roscoe, C. W. T. (2013), "Parallel algorithm for track initiation for optical space surveillance," *6th European Conference on Space Debris*, Darmstadt, Germany.

Shampine, L. F. and Gordon, M. K. (1975), *Computer solution of ordinary differential equations*, New York: W. H. Freeman and Company.

Silha, J., Schildknecht, T., and Flohrer, T. (2014), "Improved space object orbit determination using CMOS detectors," *Proceedings of the 2014 Advanced Maui Optical and Space Surveillance Technologies Conference*, Maui, Hawaii, USA.

Siminski, J. A. and Fiedler, H. (2014), "Observability analysis for tracklet association and initial orbit determination," *40th COSPAR Scientific Assembly*, Moscow, Russia.

Siminski, J. A., Fiedler, H., and Schildknecht, T. (2013a), "Track association performance of the best hypotheses search method," *6th European Conference on Space Debris*, Darmstadt, Germany.

Siminski, J. A., Montenbruck, O., Fiedler, H., and Schildknecht, T. (2014a), "Short-arc tracklet association for geostationary objects," *Advances in Space Research*, **53** (8), pp. 1184–1194.

Siminski, J. A., Montenbruck, O., Fiedler, H., and Weigel, M. (2013b), "Best hypotheses search on iso-energy-grid for initial orbit determination and track association," *23rd AAS/AIAA Spaceflight Mechanics Meeting*, Kauai, Hawaii, USA.

Siminski, J. A., Weigel, M., and Fiedler, H. (2014b), "Catalog build-up for geostationary orbit using simulated short-arc tracklets," *Proceedings of the 2014 Advanced Maui Optical and Space Surveillance Technologies Conference*, Maui, Hawaii, USA.

Sivia, D. S. (2006), *Data analysis: a Bayesian tutorial*, 2nd ed., New York: Oxford University Press.

Smith, R. L. and Huang, C. (1986), "Study of a homotopy continuation method for early orbit determination with the Tracking and Data Relay Satellite System (TDRSS)," *NASA STI/Recon Technical Report N*, **86**.

Soop, E. M. (1994), *Handbook of geostationary orbits*, Dordrecht: Kluwer Academic Publishers.

Taff, L. G. and Hall, D. L. (1977), "The use of angles and angular rates," *Celestial mechanics*, **16** (4), pp. 481–488.

Takahama, T. and Sakai, S. (2006), "Constrained optimization by the ε constrained differential evolution with gradient-based mutation and feasible elites," *IEEE Congress on Evolutionary Computation*, IEEE, pp. 1–8.

Tommei, G., Milani, A., and Rossi, A. (2007), "Orbit determination of space debris: admissible regions," *Celestial Mechanics and Dynamical Astronomy*, **97** (4), pp. 289–304.

Väisälä, Y. (1939), "Eine einfache Methode der Bahnbestimmung," *Mitteilungen der Sternwarte der Universität Turku*, **1**, pp. 1–32.

Vallado, D. A. and Alfando, S. (2014), "Curvilinear coordinate transformations for relative motion," *Celestial Mechanics and Dynamical Astronomy*, **118** (3), pp. 253–271.

Virtanen, J., Muinonen, K., and Bowell, E. (2001), "Statistical ranging of asteroid orbits," *Icarus*, **154** (2), pp. 412–431.

Weeden, B. (2015), "SSA concepts worldwide," *Handbook of space security: policies, applications and programs*, ed. by K.-U. Schrogl, P. L. Hays, J. Robinson, D. Moura, and C. Giannopapa, Springer-Verlag New York, pp. 985–998.

Worthy III, J. L. and Holzinger, M. J. (2015), "Incorporating uncertainty in admissible regions for uncorrelated detections," *Journal of Guidance, Control, and Dynamics*, **38** (9), pp. 1–17.

Zittersteijn, M. (2015), "Associating optical measurements and estimating orbits of geocentric objects through population-based meta-heuristic methods," *Proceedings of the 66th International Astronautical Congress*, Jerusalem, Israel.

Acknowledgements

First of all, I would like to express my gratitude to Dr. Oliver Montenbruck. It was impressive to observe how clear and precise he could understand my research ideas, comment on them, and identify possible weaknesses and/or advantages instantaneously. His constructive criticism and suggestions have greatly improved my research and the quality of this thesis.

I would like to thank Prof. Dr. Felix Huber for being my doctoral advisor at the Bundeswehr University and for his helpful comments during seminars. Furthermore, I want to thank Dr. Hauke Fiedler, who gave me the opportunity to pursue my studies in GSOC's space situational awareness group, also provided valuable input for the thesis, and helped with various organizational issues. I also want to thank my other former colleagues in the work group, who contributed to my research through many discussions: Martin Weigel, Dr. Johannes Herzog, Andreas Hinze, and most notably Srinivas Setty.

During my research visit at the University of Bern, I could test and assess my methods with measurements taken by their observatory. I would like to thank Prof. Dr. Thomas Schildknecht for this opportunity and the valuable insight into observational data. The vivid exchange of ideas with Dr. Alessandro Vananti, Emiliano Cordelli, and Michiel Zittersteijn was very beneficial and helped to strengthen the robustness of my methods. I am also grateful for the insight into operational experience provided by my new colleagues from ESOC's space debris office, especially Dr. Tim Flohrer and Dr. Holger Krag.

The research for this thesis was funded by Munich Aerospace and I would like to thank everyone involved in the financing process.

Finally, I would like to thank my family for their encouragement and patience through the years. The support of my wonderful wife (and mother of our lovely daughter) was essential for completing my studies.