University of California
Santa Barbara

State Estimation of Sampled-Data Systems including Applications to Vehicle Navigation and Tracking

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

Sharad Shankar

Committee in charge:

Professor João Hespanha, Chair
Professor Katie Byl
Professor James Rawlings
Professor Andrew Teel

June 2022

The Dissertation of Sharad Shankar is approved.

Professor Katie Byl

Professor James Rawlings

Professor Andrew Teel

Professor João Hespanha, Committee Chair

June 2022

State Estimation of Sampled-Data Systems including Applications to Vehicle Navigation and Tracking

by

Sharad Shankar

To my parents and siblings

Acknowledgements

Despite media depictions, the completion of a PhD is never a solitary affair, and I know that I would have had little success or joy throughout my degree without the people discussed hereafter. These include people in my life from before my PhD, who have continued to support me, as well as the vibrant and familial community I have had at UCSB.

Working under my advisor, Professor João Hespanha, exposed me to a wide range of powerful ideas and important applications. As a cursory glance through his publications page will reveal, Professor Hespanha is a brilliant and multifaceted researcher, and under his supervision I had the freedom to explore a number of difficult problems and useful tools that I will carry throughout my career. More importantly however, João was an advocate and resource in the most critical moments of my degree. João always ensured that I was over-prepared for any exam or presentation, and when I was submitting my first paper he stayed on campus well into the evening of the deadline to help perfect the submission. Professor Hespanha also supported my involvement in non-research related activities, including my completion of the Certificate in College and University Teaching (CCUT), as well as service to the department on various initiatives.

One of the biggest benefits of working for Professor Hespanha was the labmates I got to work with. Longtime fellow students Raphael Chinchilla, Murat Erdal, Henrique Ferraz, and Matthew Kirchner were always willing to hear each others ideas and work through difficult problems. Any time I spent with my labmates was a joy and we socialized outside of work whenever we had a chance. Postdoc Guosong Yang was also an incredible resource who, in addition to exploring the local food scene with me, often served as essentially a second advisor. Former labmates like David Copp, Jason Isaacs, and Justin Pearson were great mentors to me both at the beginning of my PhD as well as later on

during my job search. Newer labmates and visiting scholars like Sean Anderson, Matthew Blishke, Ramon Costa, Pedro Sequeira, Cuili Jin, Lucas Egidio, and Hikaru Akutsu were each wonderful additions to the troupe. I feel incredibly lucky to have worked with all these people who were equally brilliant and caring.

I could not have asked for a more perfect committee than Professors Katie Byl, Jim Rawlings, and Andy Teel. My committee provided invaluable guidance and feedback on my thesis, and I am thankful to each of them as my Professors as well. Access to experts like my committee members, as well as Professors Francesco Bullo, Bassam Bamieh, Jason Marden, Mengyang Gu, Diba Mirza, and Clint Schow was invaluable to both my teaching and research. I also TAed several times for Professor Teel, somehow learning foundations of control at a deeper level each time. Many of these professors helped organize the Center for Control, Dynamical-Systems, and Computation (CCDC), which through seminars and workshops, made me a more knowledgeable, resourceful, and thoughtful researcher.

In addition, I wanted to make sure to thank Professor Jason Speyer at UCLA who inspired my interest in filtering, estimation, and stochastic processes and advised me in my graduate school selection. Other professors at UCLA who deserve mention include Paolo Tabuada, Suhas Diggavi, Flavio Lorenzelli, Mike Briggs, and Yahya Rahmat-Samii. Of course, former postdoc Yasser Shoukry helped me get my start in research and encouraged me to attend UCSB for graduate school. I would be remiss without mentioning UCLA TAs Joshua Kovitz, Mostafa Majidpour, and Jerry Lien who taught me so much, gave me indispensable confidence, and who I have tried to emulate as a teacher.

I am so appreciative of UCSB staff who have been tireless in their support. Valerie de Veyra always made sure my paperwork was in order, and more importantly was a friend during my PhD. ECE staff like Beth English, Gayle Gonzalez, and Paul Gritt, along with Daryl Lee, Raul Ramirez, and Christopher Wimmel are always infinitely reliable. Maria

Napoli, Samantha Davis, and Ofelia Aguirre-Paden at CSEP gave me the opportunity to work with students and grow as a mentor. Kallie Hill and Stephanie Löfgren made working on the CCDC website and seminars a great experience. Lisa Berry supported me as a TA throughout my PhD and was a huge part of my being able to complete CCUT.

Of course, without friends and brilliant peers, no endeavor is worth completing. Close friends from UCLA who continued to support me during my PhD include Haris Akram, Michael Giordano, Mark Tvrz, Doria Deen, and Jose Hernandez. Without these people, I most certainly would not have made it out of undergrad much less graduate school. My amazing longtime housemates Katharine Dickson and Matthew Fritzler were always around to pick each other up and find distractions when necessary. Through the UCSB Greenhouse and Garden Project, my primary escape from being a graduate student, I made amazing friends like Seth Peterson, Kristen LaBonte, and garden buddies Angela Zhang and Evan Beiershmitt. Geoffrey Jewel and Christina Kerndal-Hansen shared their love of the local nature and natural history with me, for which I will forever be thankful. My best memories from graduate school were being outside, and I was fortunate to always have wonderful friends to go on hikes and adventures with like Elizabeth Huang, Daniel Lazar, Grace Kuo, Gabriel Daldegan, Jaakko Takkinen, and Vincent Luna, not to mention tennis partners Shararaghawardhan Balakrishnan and Keith Paarporn. Music is my favorite hobby, and my bandmates Amanda Kelley and Timnit Kefela could always be counted on for a jam session. Dear friends whose support cannot be overstated include Melanie Adams, Roman Aguilera, Matina Baradaran, Rachel Brewer, Hyunjin Kim, Shadi Mohagheghi, Camille Pivard, and Simone Stewart. Without a doubt, the peer group I had at UCSB and through CCDC was consistently supportive and inspirational. There are so many more I could mention here, and I hope that each of them knows how much I appreciate them.

Finally, and most vitally, I want to acknowledge my family members: my mother

Anuradha, father Shankar, brother Anirudh, and sister Sanjana. I was thankful for every moment we got to spend together over the past six years and hope I was able to make you all proud with this thesis.

Curriculum Vitæ

Sharad Shankar

Education

2022	Ph.D. in Electrical and Computer Engineering (Expected), University of California, Santa Barbara.
2019	M.S. in Electrical and Computer Engineering, University of California, Santa Barbara.
2016	B.S. in Electrical Engineering, University of California, Los Angeles.

Publications

S. Shankar, K. Ezal, and J. P. Hespanha. "Finite horizon maximum likelihood estimation for integrated navigation with RF beacon measurements." *Asian Journal of Control* 21.4 (2019): 1470-1482, DOI: https://doi.org/10.1002/asjc.2213

S. Shankar, G. Yang, and J. P. Hespanha "State Estimation for Asynchronously Switched Sampled-Data Systems." *Submitted to 2022 IEEE Conference on Decision and Control*

Abstract

State Estimation of Sampled-Data Systems including Applications to Vehicle Navigation and Tracking

by

Sharad Shankar

State estimation is a crucial part of navigation and control methods, however most well-known state estimation techniques assume some combination of linearity, Gaussian noise, as well as uniform sampling and synchrony of the process and measurements. This is a limitation for problems like integrated aircraft navigation, switched circuit monitoring, and maneuvering vehicle tracking where these simplifying assumptions may not hold. In this thesis, state estimation methods that accommodate these facets of real-world problems are explored. Methods discussed include finite-horizon nonlinear real-time optimization methods and Switched Kalman filtering for asynchronously switching systems. Some theoretical convergence results are presented along with results in simulations based on the aforementioned real-world systems.

Contents

Chapter 1

Introduction

Navigation and Control have played a significant role in human history, from the exploration and settling of the pacific islands [46], to the steam engine [52], to the moon missions [29], to the spectrum of important problems we find today. Exciting modern problems include navigation with local measurements [72], vision [59], and signals of opportunity [56], obstacle tracking and collision avoidance [5][93], and monitoring and regulation of power systems [87] and chemical processes [68]. These problems use cutting-edge tools from planning and control, which work by feeding back information from measurements of our system. Unfortunately, available measurements are often indirect to the system, related to only a subset of the states, or contain significant random noise [61]. Statistical state estimation methods are fundamental to achieving control, decision, and planning objectives by inferring reliable information about system states from measurements.

Two significant achievements relating to state estimation came about near the turn of the 19th century, in the context of marine navigation. One was the invention of precision timing devices that allowed for the fusion of information like speed relative to water from rotor measurements, as well as global positioning from astral measurements [27].

The other was the initial formulation and solution of the recursive least squares problem by Gauss and Legendre [77]. In the 1940's the linear state estimation problem was addressed by Wiener [86], and in 1959 Rudolf Kalman found an efficient recursive solution, the Kalman filter (KF) [42], inspired those advancements from the 1800's. The intervening 150-plus years had also encompassed significant developments in control theory, signal processing, statistics, and computing, each of which paved the way for designing optimal and robust state estimators [77]. Kalman filters are capable of efficiently and optimally fusing data from a variety of sensors while also incorporating prior knowledge about the system dynamics. However, as we will discuss in Section 2.1 and the beginning of Section 3, Kalman filters have limitations when dealing with combinations of nonlinearities, non-Gaussian noise, asynchrony, or discrete switching behavior that arise in many modern applications.

Further developments in technology during the past sixty years have given us more tools to use, including real-time optimization (RTO) approaches. RTO allows for estimation with minimal approximation in nonlinear or non-Gaussian problems, as well as with equality and inequality constraints [65]. These methods rely on efficient numerical optimization solvers, of which there are more and better each year. Moving Horizon Estimation (MHE) is a RTO approach that optimizes a cost function over a finite horizon to produce either a state or trajectory estimate [3]. MHE is closely related to Model Predictive Control (MPC), similarly to how Kalman Filtering is part of the Linear-Quadratic Gaussian control problem. Variants of MHE allow for simultaneous estimation of system parameters [45], estimation with mixed continuous and discrete states [23][31], as well as approximation of the full information problem [66][67]. RTO approaches may have benefits over even very general filtering methods, like Particle Filters (PFs), for certain nonlinear or hybrid problems [66]. The formulation of MHE estimation solutions to a variety of interesting practical problems is a vibrant area of study.

In this thesis, we consider the framework of sampled-data systems, which combines continuous-time processes and discrete-time measurements. Sampled-data systems can better describe real-world cyber-physical systems than continuous- or discrete-time state-space models on their own [74]. Most existing state estimation methods including Luenberger observers, Wiener-Kolmogorov Filters, KFs, PFs, and MHEs are nominally stated in either purely continuous or purely discrete time. Standard discretization of the continuous-time process is not always practical, since in many applications measurements and events occur at non-constant or multi-rate intervals. The complexity and multi-component nature of systems like vehicles and power networks requires careful adaptation of existing methods to minimize errors and ensure algorithm convergence. Our goal is to provide tools and analysis that will aid in the design of state estimators in these contexts.

This document is divided into two parts: Chapter 2 which deals with estimation for nonlinear sampled-data systems with potentially nonlinear and non-Gaussian measurements, and Chapter 3 which explores estimation with discrete switching events that may occur asynchronously to the measurement times.

In Chapter 2 we develop an algorithm we call Finite Horizon Maximum Likelihood Estimation (FHMLE) inspired by MHE. First in Sections 2.1 and 2.2 we review existing state estimation methods including filters and MHE, along with their strengths and drawbacks. FHMLE is a MHE method that explicitly formulates the problem as a maximum likelihood trajectory optimization assuming continuous-time dynamics. We apply this technique to an aircraft integrated navigation application, as covered in Section 2.3 and demonstrate its performance in simulations of an aircraft carrier landing problem with local radio-frequency (RF) measurements in Section 2.4. Available data includes RF measurements like pseudo–ranges, angles of transmission (AoT), and Doppler shift measurements, as well as accelerometer and gyroscope measurements. The navigation

estimates are obtained by solving a finite-dimensional nonlinear optimization using a primal-dual interior point algorithm (PDIP). This is an important problem to address as landing requires very high accuracy in position, velocity, and orientation estimates even in situations where GPS is compromised [18]. Our simulations show that the FHMLE provides some benefits over unscented Kalman filtering for this nonlinear problem when our RF sources produce poor Geometric Dilution of Precision (GDoP).

In Chapter 3 we explore state estimation of systems with asynchronous switching behavior. Asynchronously switched sampled-data systems can help model power systems and vehicles that evolve in continuous-time with switching behavior and discrete time measurements. In Section 3.3 we adapt existing methods for synchronous switched-system estimation, like the interacting multiple-model extended-viterbi algorithm [36], for the asynchronous case by adding estimation of inter-sample switch times. In Section 3.1 we provide theorems which ensure the statistical convergence of switched KF error dynamics under reasonable assumptions, and give intuition for the design of these methods. Simulations of both a circuit monitoring problem as well as a maneuvering vehicle tracking problem are performed in Sections 3.2 and 3.4 respectively. Switched power converters are an important component of direct current power systems and motor controllers [22]. Vehicle tracking is part of both air-traffic control and onboard collision avoidance systems [57][64]. These simulations show how poor resolution in knowledge of switch times can lead to significant error in state estimates, motivating our theory and methods.

Chapter 2

Finite Horizon Maximum Likelihood Estimation

In this section we discuss finite horizon approaches to state estimation of sampled data systems like

$$\dot{x}(t) = f(x(t), u(t), t) \tag{2.1}$$

$$y_k = h_k(x(t_k)) + v_k \tag{2.2}$$

where $x(t)$ is our state, $u(t)$ is our input, y_k is our sampled measurement, and v_k is a random noise. Some systems may also include an additive disturbance, although in continuous time care must be taken in formulating and discretizing stochastic dynamics [78][28].

Finite Horizon Estimation is an extension of ideas from Moving Horizon Estimation which we discuss in Section 2.2. When sensors like accelerometers and gyroscopes are available, we may be able to estimate our state without the dynamics model f, as we discuss in Section 2.3 in the context of Integrated Navigation. First, however, we will

examine popular filtering methods and how they perform in this setting.

2.1 Brief review of common state estimation methods and limitations

State estimation itself actually has three phases: Filtering, Smoothing, and Prediction. The goal of each is to use a history of measurements up to the current time k, $y_s, \ldots, y_{k-1}, y_k$ with $k > s$, to estimate states at various times, present, past, and future. In a filtering problem we use these measurements to estimate just the current state, x_k. In smoothing we may instead be estimating previous states like x_{k-i} with $i > 0$. Prediction problems try to estimate future states like x_{k+i} with $i > 0$. In this section we will discuss common filtering methods, though these methods depend on the ability to do prediction over one timestep, and can be extended to perform smoothing as well.

An important distinction with estimators is whether they provide a "point estimate" or an estimate of a distribution/interval. For example, if we are estimating a static parameter μ where measurements are received as $y_k \sim \mathcal{N}(\mu, \sigma^2)$ for some $\sigma > 0$, then we could just report the average value, $\text{avg}\{y_1, \ldots, y_k\}$, or we could report the average along with either a confidence interval or variance to represent our uncertainty in the mean estimate. The methods described in this section are not point estimators, as they all fit into the recursive Bayesian estimation framework in which we are estimating posterior distributions.

In filtering methods we are able to recursively solve the full-horizon (also called full-information) posterior distribution estimation (FIE) problem

$$p(x_k|y_1, y_2, \ldots, y_k) \tag{2.3}$$

using our solution to the problem

$$p(x_{k-1}|y_1, y_2, \ldots, y_{k-1}) \tag{2.4}$$

and our knowledge of $p(x_k|x_{k-1})$ and $p(y_k|x_k)$. In the case of linear systems with additive Gaussian noise,

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + d_{k-1}, \quad d_{k-1} \sim \mathcal{N}(0, Q_k), \quad Q_k = Q_k' > 0 \tag{2.5}$$

$$y_k = H_k x_k + n_k, \quad n_k \sim \mathcal{N}(0, R_k), \quad R_k = R_k' > 0 \tag{2.6}$$

a KF gives us the following closed form equations that solve this problem

$$\hat{x}_{k|k-1} = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} \tag{2.7}$$

$$K_k = (F_{k-1}P_{k-1}F_{k-1}' + Q_{k-1})H_k' \left(H_k(F_{k-1}P_{k-1}F_{k-1}' + Q_{k-1})H_k' + R_k\right)^{-1} \tag{2.8}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}) \tag{2.9}$$

$$P_k = (I - K_k H_k)(F_{k-1}P_{k-1}F_{k-1}' + Q_{k-1}) \tag{2.10}$$

where $\hat{x}_k$ is the mean, and P_k the variance of $p(x_k|y_1, \ldots, y_k)$ for each value of k. These equations do not require uniform sampling, nor consistent number of measurements. The properties necessary for convergence of the standard Kalman filter are given in [6], and are called uniform observability/detectability and uniform controllability/stabilizability. However, in the nonlinear or non-Gaussian case, these recursive methods require some level of approximation. For a nonlinear system, $x_{k+1} = f(x_k, u_k) + d_k$ and $y_k = h(x_k) + n_k$,

the standard extended Kalman filter uses linearization about the estimated mean [78],

$$\begin{aligned} F_k &= \frac{\partial}{\partial x} f_k(x,u)|_{\hat{x}_k,u_k} \\ H_k &= \frac{\partial}{\partial x} g_k(x)|_{\hat{x}_k} \end{aligned} \tag{2.11}$$

to compute the Kalman gain and covariance, where the prediction $\hat{x}_{k+1|k}$ and predicted measurement $h(\hat{x}_{k+1|k})$ can be computed from the nonlinear equations. If linearization leads to large errors in our estimates and estimated covariances for our sampling frequency, then this version of the EKF may lead to divergence of $\mathcal{N}(\hat{x}_k, P_k)$ from the true $p(x_k|y_1,\ldots,y_k)$ [71][85]. The unscented Kalman filter (UKF) suggested by Julier and Uhlmann [41] avoids linearization by instead propagating through our nonlinear equations a set of "sigma points" representing the one standard deviation ellipse of a Gaussian approximation of the relevant distributions. In the event that Gaussians do not describe the posterior distribution well, for example in the case of measurements with bimodal likelihood or heavy tails [54][37], this assumption may also be problematic [72]. We will see an example of this in Section 2.4.

The most general form of a recursive Bayesian estimator is called the particle filter, or Sequential Monte Carlo (SMC) method. Instead of approximation with Gaussians, we now represent our distributions using a set of samples from the distribution, our "particles". Similarly to the UKF, our particles are simulated through the nonlinear dynamics and measurement equations. However, rather than using the unscented transform to be able to use the standard KF update equations, we instead update the weights on our particles based on the likelihood of measurements given each particle. As our number of particles approaches infinity, we should be able to accurately represent the posterior distribution [20].

In particle filters, we might want to save computational cost by using fewer samples.

However, with few samples the nature of our resampling (or lack thereof) could lead to either few informative particles, otherwise known as sample degeneracy, or poor coverage of our space, otherwise known as sample impoverishment. For particle filters, [17] and [92], for example, deal with the impoverishment and degeneracy issues and suggest resampling strategies to prevent impoverishment. Particle filters can easily become computationally expensive as we increase either the number of particles or complexity of our resampling scheme, and therefore may not be appropriate for real-time applications [66]. Despite this, particle filters do have good mean-squared convergence properties without the curse of dimensionality as shown in [20].

Particle filters further suffer from degeneracy issues when we wish to simultaneously estimate system parameters in addition to states [48]. For example a car might accelerate as $\ddot{x} = \frac{1}{m}u$ for input force u. If we don't know the exact value of the vehicle mass m, then to correctly filter and control the system we will need to estimate this value in addition to our position and velocity. In this type of situation, both Kalman and particle filters are limited in their ability to produce accurate estimates without significant memory and computation costs. Multiple-model [33] and adaptive Kalman filters [55][53], as well as Rao-Blackwellized particle filters [70][48] are examples of approaches to this type of problem.

In the context of sampled-data systems, some of these drawbacks can be quite significant. For example, each of these methods assume that the nonlinear discretization $x_{k+1} = f(x_k, u_k) + d_k$ is known from the continuous time equations obtained through physical modeling. However, as mentioned in the introduction of this Chapter, such discretizations may not be easily obtained, and errors can accumulate over time when using single-timestep prediction. Furthermore, sufficiently detailed representations of our posterior may be difficult to compute within single sample period, leading to potential latencies in receiving estimates for control.

2.2 Review of moving horizon estimation literature

We now explore some alternatives to filtering methods that, with some trade-offs, might avoid some of the problems encountered by filtering methods. Moving Horizon Estimation computes trajectory estimates which are suboptimal in the sense of filtering and smoothing at each timestep, but can be computed quickly even in nonlinear and non-Gaussian settings by leveraging the power of numerical optimization solvers.

In moving horizon estimation we generally have discrete-time systems of the form

$$\begin{aligned} x_{k+1} &\sim \mathcal{F}_k(x_k, u_k) \\ y_k &\sim \mathcal{H}_k(x_k) \end{aligned} \tag{2.12}$$

where u_k is known, and we essentially know $p(x_{k+1}|x_k, u_k)$ as well as $p(y_k|x_k)$. To estimate x_k, we optimize the maximum a posteriori cost over a finite window of N measurements

$$J_{MHE} = p(y_{k-N+1}, \ldots, y_k|x_{k-N+1}, \ldots, x_k)p(x_{k-N+1}, \ldots, x_k|x_{k-N})p(x_{k-N}) \tag{2.13}$$

To make this more recognizable, consider the system with additive Guassian noise,

$$\begin{aligned} x_{k+1} &= f_k(x_k, u_k) + d_k \quad d_k \sim \mathcal{N}(0, Q_k) \\ y_k &= h_k(x_k) + n_k \quad n_k \sim \mathcal{N}(0, R_k) \end{aligned} \tag{2.14}$$

and suppose we know that $x_{k-N} \sim \mathcal{N}(\bar{x}_{k-N}, P_{k-N})$. If we have independence assumptions $p(d_i, d_j) = p(d_i)p(d_j)$ and $p(n_i, n_j) = p(n_i)p(n_j)$ for $i \neq j$, and $p(n_i, d_j) = p(n_i)p(d_j)$ for

all i, j, then we can write our optimization using logarithm of cost as

$$\max_{x_{k-N},x_{k-N+1},\ldots,x_k} \sum_{i=k-N+1}^{k} \|y_i - h(x_i)\|^2_{R_i} + \sum_{i=k-N+1}^{k} \|x_i - f(x_{i-1}, u_{i-1})\|^2_{Q_{i-1}} + \|x_{k-N} - \bar{x}_{k-N}\|^2_{P_{k-N}} \tag{2.15}$$

Using some auxiliary variables d_i we can rewrite this as

$$\max_{x_{k-N},x_{k-N+1},\ldots,x_k,d_{k-N},\ldots,d_{k-1}} \sum_{i=k-N+1}^{k} \|y_i - h(x_i)\|^2_{R_i} + \sum_{i=k-N}^{k-1} \|d_i\|^2_{Q_{i-1}} + \|x_{k-N} - \bar{x}_{k-N}\|^2_{P_{k-N}}$$
$$\text{s.t. } d_i = x_{i+1} - f(x_i, u_i) \text{ for all } i \in \{k-N, \ldots, k-1\}, \tag{2.16}$$

which is a constrained nonlinear least-squares problem that we can then solve using a numerical optimization method. Examples of solvers optimized for MHE include CasADi [7], and Tenscalc [34, 35]. Some formulations will omit terms relating to x_{k-N} from the problem if we have no information about about its distribution [67]. It is noteworthy that an optimal filter or smoother of a state at a specific time i would marginalize over states at all other timesteps in the trajectory. This is generally expensive, and so we perform only the trajectory optimization and then use the obtained estimates as sub-optimally filtered and smoothed values for feedback control [63][67]. Another benefit of omitting marginalization is that it ensures that our estimate is representative of some likely feasible solution. For example if we marginalized over a variable with a bimodal distribution, we might end up with an optimal estimate that lies between the two peaks rather than at one or the other.

As discussed in Section 2.1, it is common to have static system parameters that we may need to simultaneously estimate along with our state. In a simultaneous state and

parameter MHE [45], we now have cost function

$$J_{MHE} = p(y_{k-N+1}, \ldots, y_k|\theta, x_{k-N+1}, \ldots, x_k)p(x_{k-N+1}, \ldots, x_k|x_{k-N}, \theta)p(x_{k-N}, \theta) \quad (2.17)$$

which we optimize for $x_{k-N}, \ldots, x_k$ and θ, again potentially omitting the step of marginalizing, although techniques like Expectation-Maximization can be used to marginalize out θ in this case [25].

In these formulations we no longer use measurements from before time $k-N+1$ that were part of the FIE problem. However, in order to achieve full-information estimation, we can replace $p(\theta, x_{k-N})$ with $p(\theta, x_{k-N}|y_1, \ldots, y_{k-N})$ or an approximation thereof. Our prior then becomes an "arrival cost", the computation/approximation and effectiveness of which is explored in [66], [45], [81], and [49], for example. One benefit of arrival cost computation in MHE versus posterior computation in filters is that through parallel processing more time can be allotted for the computation [81]. In many cases however, we pay only a small penalty for excluding information from before our window [67], and so arrival costs can often be omitted.

A significant benefit of MHE, that it shares with MPC, is the ability to include additional constraints in our problem. This is especially helpful in MPC when our inputs are inequality bounded [67], and this can also apply in MHE when we must simultaneously estimate inputs as in [62]. However, this can also be beneficial in handling states and measurements from non-Euclidean spaces as we will see in Section 2.3. For example, if our state evolves on the surface or interior of a sphere we can naturally encode this information using an equality or inequality constraint respectively. It is worth noting that there are KF variants which can operate on manifolds like [12] as well as optimization algorithms on manifolds without embedding [14]. We will build on the insights and tools used in MHE to develop our finite horizon method.

2.3 Finite horizon maximum likelihood aircraft integrated navigation

In this section we expand on the concept of FHMLE by considering the example of integrated navigation using common radio frequency (RF) measurements. First we review literature about integrated navigation, then we develop an FHMLE which we will compare to a UKF in simulation in Section 2.4.

2.3.1 Introduction to the integrated navigation problem

Integrated navigation here refers to the estimation of the position and orientation of a moving rigid body, based on a fusion of measurements collected from radio frequency (RF) beacons and an Inertial Measurement Unit (IMU). The RF beacons provide timing information from which we extract pseudo-range measurements (i.e., ranges up to an unknown constant due a clock offset), angle of transmission (AoT) measurements constructed by processing the signals transmitted/received by an antenna array, and velocity measurements obtained from the Doppler shift on the RF carrier frequency.

The fusion of pseudo-range with IMU measurements is a well-studied problem for Global Navigation Satellite Systems (GNSS) [11], in which the satellites play the role of the RF beacons. However, our setup presents a few key differences: First, we are interested in scenarios where the RF beacons are localized to a relatively small region of space, which would lead to very poor Geometric Dilution of Precision (GDoP) [11] in a GNSS setting. Second, we are also interested in scenarios where the number of beacons is small, in particular, fewer than the minimum of four beacons required for a system operating just with pseudo-range measurements and asynchronous clocks. Third, in addition to the usual pseudo-range measurements, we also have angle of transmission

measurements, which as we will show, can compensate for the lack of beacons and poor GDoP. The use of angle of transmission or angle of arrival measurements in integrated navigation is discussed in [39], while localization using Doppler measurements is discussed in [4]. This work is especially relevant to GPS-denied scenarios, where only a small number of ground-based beacons are available, for example when navigating using Signals of Opportunity [56].

In this thesis, we construct a maximum-likelihood estimator (MLE) for the position and orientation of a moving rigid body that fuses all available measurements (RF and IMU) over a finite window of time with constant length T, and ending at the current time t_0. Considering a finite window of time (and consequently a finite number of measurements) and finitely parameterizing the position/orientation over the time window $(t_0 - T, t_0]$, enables us to compute the maximum likelihood estimate numerically using a nonlinear programming solver. In this work, we use a primal-dual interior-point (PDIP) method that has been optimized for real-time computations [35].

A UKF approach for navigation in the GNSS setting appears in [83] and an attitude observer and EKF combination is considered in [16]. Our approach is in contrast to this type of Kalman filtering based approaches that recursively compute estimates and require approximations to nonlinear dynamics and non-Gaussian noise. Extended Kalman Filters (EKFs) locally linearize nonlinear dynamics and measurement equations, while the Unscented Kalman Filter (UKF) assumes that the prior and posterior distributions are adequately approximated by a finite set of "sigma points" [41]. A key benefit of the approach proposed here is that we need not assume that the posterior distribution of the state estimate is Gaussian, which is an important assumption in the derivations of both the EKF and UKF. Nonlinear measurements or dynamics often result in non-Gaussian posterior distributions, which can cause Kalman filter-based methods

to diverge or converge to sub-optimal solutions, depending on initialization and measurement history [24, 71]. However, in a FHMLE approach, we are able to re-initialize independently at each step, as well as compute solutions for multiple initializations. This prevents the filter from getting "stuck" on a suboptimal trajectory. Furthermore, EKFs and UKFs are designed for additive Gaussian noises, however such an assumption on the AoT measurements is problematic because an error in direction should be constrained to the unit-sphere. There are several well-studied unit-sphere distributions from the field of Directional Statistics, including the von Mises-Fisher and Kent distributions.

Finite horizon MLE techniques also differ from Kalman filtering-based approaches in that the former discards old measurements. While this appears to be a disadvantage at first, it also means that the effect of outliers or poor initializations is quickly attenuated. In the context of our problem, we will see that the use of a finite horizon helps the system rapidly recover from the effect of measurements taken over unfavorable geometries (i.e. all beacons localized to a small region of space). Our finite horizon MLE is heavily inspired by the literature on Moving Horizon Estimation (MHE) [76], which also considers a finite window of measurements, but typically also includes a penalty term to account to the "missing" measurements. A sliding window approach to GNSS/INS integration appeared in [30]. The effect of changing the length of the sliding window is considered in [95].

One final benefit of an approach that reduces state estimation to a numerical optimization lies in the ability to incorporate constraints into the design of the estimator. These constraints can be used, e.g., to restrict estimates to known bounds. For example, we can add constraints on bias terms for IMU sensors or on the position/orientation/velocity of the rigid body. Such constraints add additional information that can greatly improve the estimation accuracy. We note that equality constrained versions of both EKFs and UKFs also exist [76], but they rely on projections, and suffer from many of the challenges

mentioned above.

To illustrate the potential of our approach we present simulation results obtained from a scenario in which an aerial vehicle lands on an aircraft carrier and estimates its positions using an affordable IMU unit and measurements from RF beacon on board of the aircraft. The vehicle starts a few kilometers away from the aircraft, which presents a very unfavorable geometry [11] since all RF beacons are localized very close to each other (in comparison to the distance between aircraft and beacons). Our results show that the FHMLE is especially attractive to estimate the parameters mostly affected by this challenging geometry, consistently leading to smaller estimation error in regions with poor geometry.

This Section is organized as follows: In Subsection 2.3.2, we introduce the measurement models and likelihood functions to formulate the FHMLE optimization. The conversion of the maximum likelihood estimation to a numerical optimization is discussed in Section 2.3.3, together with a brief description of the nonlinear programming solver that we use. The next Section, Section 2.4 presents simulation results for an integrated aerial navigation system based on RF beacons aboard an aircraft carrier. We use these simulations to compare the performance of the FHMLE and a UKF approach to navigaion.

2.3.2 Problem formulation and measurement models

Consider a coordinate frame B attached to a rigid body that moves with respect to an inertial frame C and denote by $(p, R) \in \mathrm{SE}(3)$ the instantaneous position and orientation of the frame B with respect to C. Our goal is to estimate $p(t) \in \mathbb{R}^3$ and $R(t) \in \mathrm{SO}(3)$ over an interval $t \in (t_0 - T, t_0]$ based on a set of measurements $Y_{(t_0-T,t_0]}$ taken by sensors attached to B during the interval $(t_0 - T, t_0]$. Here, $SE(3)$ denotes the special Euclidean

group and $SO(3)$ the special orthogonal group in three dimensions [58].

We denote by $\ell\big(Y_{(t_0-T,t_0]} \mid p, R\big)$ the likelihood of the set of measurements $Y_{(t_0-T,t_0]}$ given the time evolutions p and R of the position and orientation of B with respect to C, respectively. This notation enables us to express the maximum-likelihood estimator (MLE) as

$$(\hat{p}, \hat{R}) \coloneqq \underset{p\in\mathcal{P}, R\in\mathcal{R}}{\arg\max}\, \ell\big(Y_{(t_0-T,t_0]} \mid p, R\big),$$

where $\mathcal{P}$ and $\mathcal{R}$ denote "admissible" sets for the functions $p(t)$ and $R(t)$, $t \in (t_0 - T, t_0]$ that define the position and orientation of B with respect to C, respectively. For p and R, to be "admissible", they must be compatible with the dynamics of the rigid body, with the understanding that in order to solve (2.25) numerically, the sets $\mathcal{P}, \mathcal{R}$ must be parameterized using a finite number of parameters. We defer further discussion on the parameterization of these sets to Section 2.3.3 and focus the remainder of this section on the construction of the likelihood function that appears in (2.25).

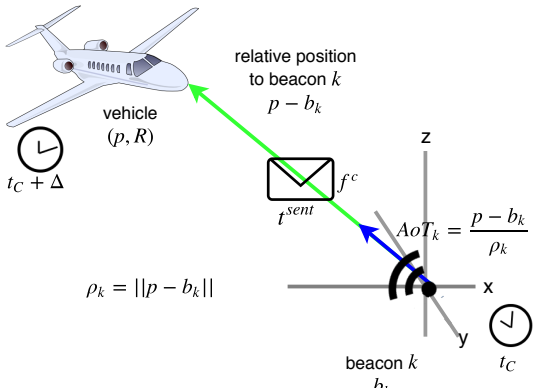


Figure 2.1: RF measurements: pseudorange, angle of transmission, Doppler shift.

Pseudorange Measurements

The pseudorange measurement $\rho_k(t)$ associated with the reception of an RF message from beacon k at time t is typically defined as

$$\rho_k(t) \coloneqq c(t - t^{\text{sent}}),$$

where t^{sent} denotes the time at which the message was sent and c is the RF propagation speed. Let b_k be the position in the coordinate frame C of the beacon that transmitted the message. Assuming that the receive antenna is at the origin of the coordinate frame B, under ideal conditions we would have

$$\rho_k(t) = \|p(t) - b_k\|,$$

where $\|\cdot\|$ denote the Euclidean norm of a vector (see Figure 2.1). However, the send time t^{sent} and the receive time t are measured with respect to two clocks that are not perfectly synchronized. In addition, the RF detection circuit introduces a stochastic error in the measurement of the receive time. Denoting by $\Delta(t)$ the offset between the transmitter and receiver clocks and assuming a zero-mean Normal distribution with variance σ_ρ^2 for the stochastic component of the error, the pseudorange measurement $\rho_k(t)$ is a Normal random variable with mean $\|p - b(t))\| + c\Delta(t)$ and variance σ_ρ^2, leading to the following likelihood function:

$$\ell\big(\rho_k(t)|p(t), \Delta(t)\big) = \frac{\exp\Big(-\frac{\big(\rho_k(t) - \|p(t) - b_k\| + c\Delta(t)\big)^2}{2\sigma_\rho^2}\Big)}{\sqrt{2\pi}\sigma_\rho}. \tag{2.18}$$

Angle of Transmission Measurements

The angle of transmission measurement $AoT_k(t)$ associated with the reception of an RF message from beacon k at time t is defined by the unit vector pointing from the beacon to the receive antenna. Under ideal conditions we would have

$$AoT_k(t) = \frac{p(t) - b_k}{\|p(t) - b_k\|} \tag{2.19}$$

(see Figure 2.1), but RF detection errors will also introduce a stochastic error in this measurement. We assume that the measurement follows a von Mises-Fisher unimodal isotropic distribution on the unit sphere in $\mathbb{R}^3$, with mean direction given by (2.19) and concentration parameter $\kappa > 0$ [1], which leads to the following likelihood function:

$$\ell\big(AoT_k(t)|p(t)\big) = \frac{\kappa \exp\left(\kappa \frac{(p(t)-b_k)' AoT_k(t)}{\|p(t)-b_k\|}\right)}{2\pi(e^{\kappa} - e^{-\kappa})}. \tag{2.20}$$

Note that a high value for the concentration parameter $\kappa > 0$ of a von Mises-Fisher distribution corresponds to a distribution tightly clustered around the mean, and consequently less noise on AoT_k, whereas the limiting case $\kappa \to 0^+$ corresponds to the uniform distribution. For non-isotropic distributions of the AoT error, one could base the likelihood on the Kent distribution instead [43].

Figure 2.2 shows the Weighted Geometric Dilution of Precision (WGDOP) for various sets of measurement types over the sample trajectory and beacon positions used in the simulations of Section **??**. WGDOP is an approximation of the position and clock offset estimation error that is obtained from the inverse of the Fisher Information matrix of the weighted least squares problem with linearized measurements [69]. Note that GDOP is infinite with only three pseudorange measurements and no AoT. Adding AoT measurements significantly improves WGDOP, with diminishing returns between two and three

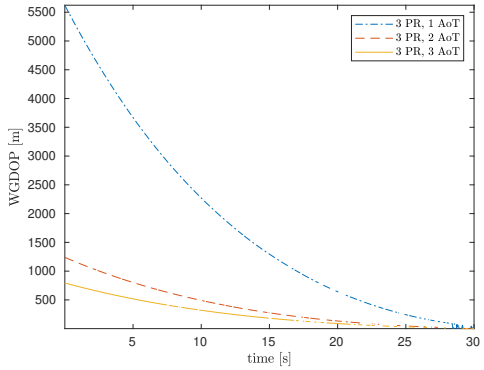


Figure 2.2: Weighted GDOP for various measurements

angle measurements.

Doppler Shift Measurements

The relative motion between transmitter and receiver antennas creates a Doppler shift

$$\Delta f_k(t) \coloneqq f_k^{rec}(t) - f_k^c$$

between the carrier frequency $f_k^{rec}(t)$, measured by the RF detector at time t, and the original carrier frequency f_k^c transmitted by beacon k. Under ideal conditions we would have

$$\Delta f_k(t) = -\frac{f_k^c}{c}\left(\frac{d}{dt}\|p(t) - b_k\|\right) = -\frac{f_k^c}{c}\frac{(p(t) - b_k)'\dot{p}(t)}{\|p(t) - b_k\|},$$

but due to measurement errors we assume that the perceived Doppler shift is a zero-mean Normal distribution with variance σ_f^2, leading to the following likelihood function:

$$\ell\big(\Delta f_k(t)|\dot{p}(t)\big) = \frac{\exp\Big(-\frac{\Big(\Delta f_k(t)+\frac{f_k^c}{c}\frac{(p(t)-b_k)'\dot{p}(t)}{\|p(t)-b_k\|}\Big)^2}{2\sigma_f^2}\Big)}{\sqrt{2\pi}\sigma_f}. \tag{2.21}$$

IMU Measurements

An accelerometer attached to the frame B produces a measurement $a_{\mathrm{m}}(t)$ of the proper acceleration expressed in B, which is given by

$$a_{\mathrm{m}}(t) \coloneqq R(t)(\ddot{p}(t) - g) + a_{\mathrm{b}}(t),$$

where g denotes the local gravity vector expressed in the inertial frame C, for example $g = \begin{bmatrix} 0 & 0 & 9.8065 \end{bmatrix}$ in the North-West-Up (NWU) inertial navigation frame. Additionally, $a_{\mathrm{b}}(t)$ is an additive bias introduced by the sensor. For simplicity, we assume that the accelerometer is calibrated to remove the effect of other deterministic errors such as scale factor and non-orthogonality errors [82], and that this measurement is further corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_a, so that we obtain the following likelihood function

$$\ell\big(a_{\mathrm{m}}(t)|\ddot{p}(t), R(t)\big) = \frac{\exp\left(-\frac{1}{2}\left\|a_{\mathrm{m}}(t) - R(t)(\ddot{p}(t)+g) - a_{\mathrm{b}}(t)\right\|^2_{\Sigma_a^{-1}}\right)}{\sqrt{2\pi^3 \det \Sigma_a}} \tag{2.22}$$

where $\|x\|^2_\Lambda$ denotes the quadratic form $x'\Lambda x$.

A rate-gyro attached to B produces a measurement $\omega_{\mathrm{m}}(t)$ of the angular velocity

expressed in B, given by

$$\omega_{\mathrm{m}}(t) \coloneqq R(t)\omega(t) + \omega_{\mathrm{b}}(t)$$

where the entries of the angular velocity vector $\omega = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix}$ are obtained from

$$\dot{R}(t)R'(t) = \begin{bmatrix} 0 & -\omega_z(t) & \omega_y(t) \\ \omega_z(t) & 0 & -\omega_x(t) \\ -\omega_y(t) & \omega_x(t) & 0 \end{bmatrix}$$

and $\omega_{\mathrm{b}}(t)$ denotes an additive bias introduced by the sensor. We similarly assume that the gyroscope is calibrated to remove other deterministic errors, and is corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_ω, and we obtain the following likelihood function:

$$\ell\big(\omega_{\mathrm{m}}(t) | R(t), \dot{R}(t)\big) = \frac{\exp\left(-\frac{1}{2}\left\|\omega_{\mathrm{m}}(t) - R(t)\omega(t) - \omega_{\mathrm{b}}(t)\right\|^2_{\Sigma_\omega^{-1}}\right)}{\sqrt{2\pi^3 \det \Sigma_\omega}}. \tag{2.23}$$

A magnetometer attached to B produces a measurement $m_{\mathrm{m}}(t)$ of the local magnetic field expressed in B, given by

$$m_{\mathrm{m}}(t) \coloneqq R(t)m$$

where m denotes the direction/intensity of the local magnetic north expressed in the inertial frame C. Assuming that this measurement is corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_m, we obtain the following likelihood function:

$$\ell\big(m_{\mathrm{m}}(t) | R(t)\big) = \frac{\exp\left(-\frac{1}{2}\left\|m_{\mathrm{m}}(t) - R(t)m\right\|^2_{\Sigma_m^{-1}}\right)}{\sqrt{2\pi^3 \det \Sigma_m}}. \tag{2.24}$$

For simplicity, we assume here that the magnetometer has been calibrated to remove the

effect of any soft and hard iron distortions.

Total Likelihood Function

While our main goal is to estimate the position $p(t)$ and orientation $R(t)$ of the rigid body attached to the coordinate frame B, the likelihood of the sensor measurements that we have introduced above depend on three other quantities that need to be estimated in addition to $p(t)$ and $R(t)$: the offset $\Delta(t) \in \mathbb{R}$ between the clock of the beacons and the clock of the RF receiver, the accelerometer bias $a_{\mathrm{b}}(t) \in \mathbb{R}^3$, and the rate-gyro bias $\omega_{\mathrm{b}}(t) \in \mathbb{R}^3$. In light of this, our maximum-likelihood estimator (MLE) is actually of the form

$$(\hat{p}, \hat{R}, \hat{\Delta}, \hat{a}_{\mathrm{b}}, \hat{\omega}_{\mathrm{b}}) \coloneqq \underset{\substack{p \in \mathcal{P}, R \in \mathcal{R}, \Delta \in \mathcal{D}, \\ a_{\mathrm{b}} \in \mathcal{A}_{\mathrm{b}}, \\ \omega_{\mathrm{b}} \in \Omega_{\mathrm{b}}}}{\arg\max} \ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big), \tag{2.25}$$

where $\mathcal{D}$, $\mathcal{A}_{\mathrm{b}}$, and Ω_{b} denote admissible sets for the functions $\Delta(t)$, $a_{\mathrm{b}}(t)$, and $\omega_{\mathrm{b}}(t)$.

Assuming that the random variables associated with the likelihood functions given by (2.18), (2.20), (2.21), (2.22), (2.23), and (2.24) are all conditionally independent, and given the trajectories $p \in \mathcal{P}$, $R \in \mathcal{R}$, $\Delta \in \mathcal{D}$, $a_{\mathrm{b}} \in \mathcal{A}_{\mathrm{b}}$, and $\omega_{\mathrm{b}} \in \Omega_{\mathrm{b}}$, the likelihood function $\ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big)$ in (2.25) is obtained as the product of the pdfs in (2.18), (2.20), (2.21), (2.22), (2.23), (2.24) for each of the measurements available in $Y_{(t_0-T,t_0]}$ taken by sensors in the interval $(t_0 - T, t_0]$.

Since each of the pdfs in (2.18), (2.20), (2.21), (2.22), (2.23), (2.24) is from the exponential family, rather than performing the maximization in (2.25), we instead solve

the following optimization

$$(\hat{p}, \hat{R}, \hat{\Delta}, \hat{a}_{\mathrm{b}}, \hat{\omega}_{\mathrm{b}}) \coloneqq \underset{\substack{p\in\mathcal{P}, R\in\mathcal{R}, \Delta\in\mathcal{D}, \\ a_{\mathrm{b}}\in\mathcal{A}_{\mathrm{b}}, \\ \omega_{\mathrm{b}}\in\Omega_{\mathrm{b}}}}{\arg\min} -\log \ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big), \tag{2.26}$$

which has the same set of minima $\{p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\}$.

2.3.3 Numerical Optimization

This section addresses two key aspects of involved in solving (2.26): first, the parameterization of the sets of admissible functions, and second, the numerical method used to solve the resulting optimization.

Trajectory Parameterization

To solve (2.26) using a numerical solver, we need to use finite parameterizations for the sets of admissible functions $\mathcal{P}, \mathcal{R}, \mathcal{D}, \mathcal{A}_{\mathrm{b}}, \Omega_{\mathrm{b}}$.

For temperature compensated IMU sensors, the accelerometer and rate-gyro biases drift very slowly so we can typically neglect changes in the biases over time intervals of length T on the order of tens of seconds [60]. This corresponds to sets $\mathcal{A}_{\mathrm{b}}, \Omega_{\mathrm{b}}$ consisting of signals in $\mathbb{R}^3$ that are constant over the whole interval $(t_0 - T, t_0]$ and can be parameterized, for example by

$$\mathcal{A}_{\mathrm{b}} \coloneqq \big\{a \in \mathbb{R}^3 : \|a\| \leqslant L_a\big\}, \qquad \Omega_{\mathrm{b}} \coloneqq \big\{\omega \in \mathbb{R}^3 : \|\omega\| \leqslant L_\omega\big\},$$

where L_a and L_ω denote maximum norms for the accelerometer and biases, typically obtained from the manufacturers datasheets.

Temperature-compensated clock circuits typically exhibit a very small drift over time

intervals of tens of seconds [51]. However, pseudoranges are extremely sensitive to such errors (recall that, at light speed, 1 nano second corresponds to about 0.3 meters), so we consider a richer class of admissible functions for the clock offset $\mathcal{D}$. Specifically, we allow the set $\mathcal{D}$ of admissible clock offsets to contain not only constant offsets but also constant clock drifts, denoted by β, over the time interval of interest. This corresponds to a set of linear functions of the form:

$$\mathcal{D} \coloneqq \big\{\beta(t - t_0) + \phi : |\beta - 1| \leqslant L_\Delta, \phi \in \mathbb{R}\big\},$$

where L_Δ denotes the maximum mismatch between the frequencies of the two clocks.

For a model-free parameterization of the sets $\mathcal{P}$ and $\mathcal{R}$, we draw inspiration from direct collocation methods for trajectory optimization [8], and parameterize admissible trajectories $p(t)$, $t \in (t_0 - T, t_0]$ with the values of the function $p(t)$ at a fixed set of N times

$$\tau_1 \coloneqq t_0 - T < \tau_2 < \cdots < \tau_{N-1} < \tau_N \coloneqq t_0, \tag{2.27}$$

which means that $\mathcal{P}$ is parameterized by

$$\big(p(\tau_1), p(\tau_2), \ldots, p(\tau_N)\big) \in \mathbb{R}^{3N}. \tag{2.28}$$

Following the collocation methods terminology, we refer to the τ_i as *knots*. While splines provide a very general and attractive option to interpolate the value of $p(t)$ and its derivatives between the knots, for numerical efficiency, the results presented here use a different and more computationally efficient approach: We take the knots, τ_i, to be the times at which we have measurements, meaning that we do not need to interpolate $p(t)$

between the knots, but we do need to compute the first and second derivatives of $p(t)$ for the likelihood functions of the Doppler and accelerometer measurements in (2.21) and (2.22), respectively. We evaluate these derivatives using the Lyness and Moler algorithm [50], which for second-order polynomials and equally spaced knots $\tau_i - \tau_{i-1} =: h$ leads to

$$\dot{p}(\tau_i) = \frac{p(\tau_{i+1}) - p(\tau_{i-1})}{2h}, \quad \ddot{p}(\tau_i) = \frac{p(\tau_{i+1}) - 2p(\tau_i) + p(\tau_{i-1})}{h^2}, \quad \forall k \in \{2, \ldots, N-1\}. \tag{2.29}$$

Formally, this means that $\mathcal{P}$ is the linear subspace of functions from $(t_0 - T, t_0]$ to $\mathbb{R}^3$ whose first and second derivatives at each knot τ_i are equal to those of the third-order Lagrange polynomial that interpolates the function at the three knots closest to τ_i. In view of (2.29), this particular set $\mathcal{P}$ has the desirable feature that the values of $p(\tau_i)$ and its derivatives $\dot{p}(\tau_i)$ and $\ddot{p}(\tau_i)$ are all linear combinations of at most three of the optimization parameters in (2.28).

To parameterize the admissible set of orientations $\mathcal{R}$, we follow a similar approach, except that we represent rotation matrices using unit quaternions. In particular, we parameterize an admissible trajectory $R(t)$, $t \in (t_0 - T, t_0]$ by a sequence of unit quaternions,

$$\big(q(\tau_1), q(\tau_2), \ldots, q(\tau_N)\big) \in SU(2)^N$$

corresponding to the orientations at the knots in (2.27). We then take $\mathcal{R}$ to be the linear subspace of functions from $(t_0 - T, t_0]$ to $SU(2)^N$ whose first and second derivatives at each knot τ_i are equal to those of the 3rd order Lagrange polynomial that interpolates the function at the 3 knots closest to τ_i, which means that we can again obtain simple formulas for $\dot{q}(\tau_i)$ and $\dot{R}(\tau_i)$ using the Lyness and Moler algorithm. These derivatives are needed for the likelihood function of the rate-gyro measurements in (2.23).

Remark 1 (Known dynamics). The above approach to constructing the sets $\mathcal{P}$ and $\mathcal{R}$ has the advantage that it does not require knowledge of the specific rigid body dynamics. However, when such a model is known incorporating these dynamics may potentially decrease the estimation errors. To understand how this could be accomplished within the framework of this thesis, suppose that the motion of the rigid body attached to the coordinate frame B follows dynamics of the following general form:

$$\dot{x} = f(x,t), \qquad p(t) = G\big(x(t)\big), \qquad R(t) = H\big(x(t)\big), \qquad x \in \mathbb{R}^n. \tag{2.30}$$

The sets $\mathcal{P}$ and $\mathcal{R}$ are naturally parameterized by the initial conditions to (2.30) as follows:

$$\mathcal{P} \coloneqq \{p_{x_0} : x_0 \in \mathbb{R}^n\}, \qquad \mathcal{R} \coloneqq \{R_{x_0} : x_0 \in \mathbb{R}^n\},$$

where

$$p_{x_0}(t) \coloneqq G(\varphi(t; x_0, t_0 - T)), \qquad R_{x_0}(t) \coloneqq H(\varphi(t; x_0, t_0 - T)), \tag{2.31}$$

and $\varphi(t; x_0, t_0 - T)$ denotes the solution to (2.30) with initial condition $x(t_0 - T) = x_0$. To follow this approach, we would not need to solve (2.30) explicitly as (2.31) seems to imply. Instead, and inspired again by direct collocation methods, we would parameterize the sets by the values of $x(\tau_i)$ at the knot points and regard (2.30) as optimization constraints at the knot points. □

Primal-Dual Interior-Point Solver

For the parameterizations of the admissible sets discussed in Section 2.3.3, the optimization in (2.26) is of the form

$$f(\hat{x}) = \min \left\{ f(x) : F(x) \geqslant 0, G(x) = 0, x \in \mathbb{R}^{n_x} \right\}, \tag{2.32}$$

where $\hat{x}$ denotes the optimizer of f. We solve this optimization numerically using a second-order Primal-Dual Interior Point (PDIP) method. While this class of methods guarantees convergence only when the function $f(x)$ is convex and the constraints $F(x) \geqslant 0, G(x) = 0$ define a convex set [15], in practice PDIP methods they perform well on a much wider class of problems. The particular solver that we use is generated by the MATLAB® toolbox `TensCalc` [34, 35]. This toolbox generates specialized C-code solvers for optimization problems like (2.32) and is especially fast for optimizations up to a few thousands of optimization variables and constraints. Short solve times are achieved by a combination of features that include automatically detecting and exploring the sparsity structure of the Hessian matrix of $f(x)$ and the Jacobian matrices of $F(x)$ and $G(x)$ to speed up each iteration of the PDIP iteration, reusing intermediate computations within and across iterations, and the use of code that improves the efficiency of micro-processor pipelining and caching [35].

Remark 2. Solver Complexity for Sparse Problems The key challenge with FHMLE is that, at each time step, we need to solve the optimization in (2.26). However, it turns that the specialized C-code solvers produced by the MATLAB® toolbox `TensCalc` [34, 35] are very efficient for the type of optimizations needed by FHMLE. Due to the independence of measurements across time and the numerical method used to handle dynamics in (2.29), we note that the Hessian matrix associated with the optimization in (2.32) is

sparse because

$$\frac{\partial^2 f}{\partial(p,R)(\tau_i)\partial(p,R)(\tau_j)} = 0,$$

for all indices i and j such that $|i-j| > 1$. This means that the number of nonzero entries of the Hessian increases linearly with the number of knots. Solve times for state-of-the-art methods of inverting sparse matrices scale essentially linear with the number of nonzero elements [26]. The computational complexity of a second-order PDIP method is determined by the complexity of inverting the Hessian matrix $\frac{\partial^2 f}{\partial x x'}$, therefore PDIP solve times increase linearly with the size of time window considered [15].

Implementation for FHMLE

One limitation of second-order PDIP methods is that the optimization criterion $f(x)$ and the constraint function $F(x)$, $G(x)$ must be twice-differentiable. This is not the case in our problem, because the likelihood functions associated with the angle of transmission in (2.20) and the Doppler shift in (2.21) contain a division by $\|p(t) - b_k\|$ that is not differentiable at the point $p(t) = b_k$. However, one can confidently exclude such points from the set of admissible $p(t)$, because this would imply that the rigid body was collocated with the beacon. Nevertheless, we have found that the convergence of the solver was improved by introducing the additional slack optimization variables

$$r_k(t) \coloneqq \frac{(p(t) - b_k)' AoT_k(t)}{\|p(t) - b_k\|}, \tag{2.33}$$

which enabled us to rewrite (2.20) as

$$f\big(AoT_k(t)|p(t)\big) = \frac{\kappa \exp(\kappa r_k(t))}{2\pi(e^{\kappa} - e^{-\kappa})} \tag{2.34}$$

subject to the following constraint that enforces (2.33):

$$\|p(t) - b_k\| \, r_k(t) = (p(t) - b_k)' AoT_k(t). \tag{2.35}$$

Note that both (2.34) and (2.35) are smooth everywhere, as opposed to the original likelihood in (2.20). A similar procedure can be applied to smooth the likelihood of the Doppler measurements in (2.21). While the use of slack variables increases the size of the optimization, the added numerical stability generally decreases the number of iterations for the solver, and overall results in smaller solve times.

Remark 3 (Regularization and a-priori information). Solver convergence is sometimes improved by introducing L_2-regularization terms on some variables, e.g., in the IMU biases a_b and ω_b. From a Bayesian perspective, this is equivalent to placing a Gaussian prior on those variables and simply amounts to including additional quadratic additive terms in (2.26). Similarly, box inequality constraints placed on optimization variables, which may also be viewed as priors, can improve solver convergence. □

2.4 Simulations of aircraft integrated navigation

The FHMLE and a tightly integrated UKF were compared in a simulated scenario where an aerial vehicle approaches an aircraft carrier from the south in a North-West-Up (NWU) inertial frame. Three RF beacons were placed on board the aircraft carrier: two at the start of the landing strip (one on each side separated by 40 meters) and one on the control tower (44 meters north of the western forward beacon and 35 meters above the deck). The noise parameters for the sensors are summarized in Table 2.1. The values corresponding to the RF measurements (pseudorange, angle of transmission, and Doppler shift) were obtained through detailed Monte Carlo simulations of the RF

signal processing that leads to these measurements and the IMU sensor parameters were obtained from InvenSense's MPU-6050 MEMS IMU datasheets [38].

Parameter	Symbol	Value
RF propagation speed	c	299792458 m/s
RF carrier frequency	f_k^c	5.4 GHz
pseudorange noise variance	σ_ρ	1 m
angle of transmission noise concentration	κ	$2 \cdot 10^4$
Doppler shift noise variance	σ_f	.1 Hz
gravity field	g	$\begin{bmatrix} 0 \\ 0 \\ 9.8065 \end{bmatrix}$ m/s^2
accelarometer noise covariance matrix	Σ_a	$.15\, I_{3\times 3}$ [m/s^2]
rate-gyro noise covariance matrix	Σ_ω	$.05\, I_{3\times 3}$ [rad/s]
local magnetic field	m	$\begin{bmatrix} 40000 \\ 0 \\ 0 \end{bmatrix}$ [nT]
magnetometer noise covariance matrix	Σ_m	$400\, I_{3\times 3}$ [nT]
measurements sampling frequency	h	20Hz
FHMLE horizon length	T	1 sec

Table 2.1: Simulation parameters.

A FHMLE horizon length of one second with a measurement frequency of 20 Hz leads to an optimization over 20 knots. The accelerometer bias, gyroscope bias, and clock drift are assumed constant over each window, since each changes relatively slowly to the scale of a second. This leads to an optimization in 268 variables with 140 equality constraints, corresponding to the simulation displayed in Figure 2.3. Despite the Hessian being 408× 408, at most 9000 elements are nonzero, or around 5.4% of the matrix entries. Since the number of nonzero matrix entries scales linearly, while the size of the matrix scales quadratically, we expect the percentage of nonzero elements to decrease with horizon length. No regularization was used for the FHMLE, however box constraints were placed on the position and bias terms to improve convergence.

Unscented Kalman Filter Estimator

To test the effectiveness of our approach against the prior state of the art, we used an Unscented Kalman Filter (UKF) to solve the same estimation problem. In essence, the

UKF uses the unscented transform to extend Kalman Filtering to scenarios with nonlinear dynamics or measurements. The unscented transform uses a set of "sigma points" to represent the first two moments of a given distribution. In a UKF, the unscented transform takes the estimated mean $\hat{x}(t)$ and variance $\hat{P}_x(t)$ at time t, and computes sigma points $s^1(t), s^2(t), ..., s^{n+1}(t)$ each in $\mathbb{R}^n$, for n the dimension of the state x. It can be shown that a correct choice of the $n+1$ sigma points fully describes the first two moments of a distribution [41]. These sigma points are simulated through the noisy process and measurements to obtain sigma points of the predicted state and measurement at time at a subsequent time. The UKF then performs the same update as the usual Kalman filter through the inverse unscented transform. The key benefit of an UKF is that the it is generally less sensitive to filter divergence [24] than Kalman filters extensions that rely on local linearization like the EKF [84]. The UKF in this simulation was designed with the same modeling assumptions as the FHMLE, except that the AoT measurements were modeled as independent and identical Gaussian noises in the azimuth and elevation angles of the origin of B with respect to C. The resulting simulated distribution is similar to the von Mises-Fisher distribution, especially for large values of κ, and preserves the notion of angle as belonging to the unit sphere.

Filter Initialization

The FHMLE and the UKF estimators both depend on how they are initialized: At every time, the FHMLE nonlinear solver needs to be initialized with a state estimate that is the starting point for the PDIP iterations. The state of the UKF filter needs to be initialized at the start of an experiment (and whenever the filter diverges).

Each pair (AoT_i, AoT_j) of angle of transmission measurements, corresponding to beacons i and j, provides a position estimate $\bar{p}_{ij}$ at the point p that is closest to the corre-

sponding AoT rays:

$$\bar{p}_{ij} \coloneqq \min_{p,\alpha_i,\alpha_j} \|p - b_i + \alpha_i AoT_i\|^2 + \|p - b_j + \alpha_j AoT_j\|^2.$$

We initialized the FHMLE estimator using the straight line trajectory that better fits these $\bar{p}_{ij}$ and that is compatible with the Doppler velocity measurements (which is a simple least squares problem). While this initial estimate may not be particularly good at fitting the pseudo-range measurements and the IMU measurements, the PDIP was able to very reliably converge from this initialization to the maximum likelihood estimate in a small number of iterations. We note that with three angle of transmission measurements, we can get three estimates $\bar{p}_{ij}$ by selecting (i, j) equal to $(1, 2)$, $(2, 3)$, or $(2, 3)$. Initializing with these 3 options, we selected the one which converged to the largest value for the likelihood function. The same algorithm is then used to initialize the UKF at the start of each trial.

Results and Discussion

Figure 2.3 compares the estimation errors and 1-σ bounds for the UKF and FHMLE over 1000 Monte Carlo trials of measurements generated as in Section 2.4. These plots show that the UKF has a significant initial position bias and is noisier towards the beginning of the trajectory. Figure 2.4 shows the empirical distributions of the norms of UKF and FHMLE position errors over two segments of trajectory: the the first 10 seconds of the trajectory where geometry is poor, as well as the final 20 seconds where geometry is improved. This figure confirms that the UKF is heavier tailed in the first 10 seconds, while the performance of the two methods is comparable in the final 20. This is likely due to a combination of poor geometry and the fact that the UKF uses an approximation of the von Mises-Fisher distribution, while the FHMLE is able to handle

the exact equations. Figure 2.5 demonstrates that by increasing κ to $2.5 \cdot 10^4$ and using a higher cost IMU (In this example a Honeywell HG-1700 IMU [60] with $\Sigma_a = 0.03 I_{3\times 3}$ m/s^2), we enter a regime where the UKF is actually more effective than the FHMLE at estimating position, especially in the final 20 seconds. Figure 2.3 also shows that FHMLE estimates of attitude, velocity, and bias terms are all either comparable or clearly better than those of the UKF.

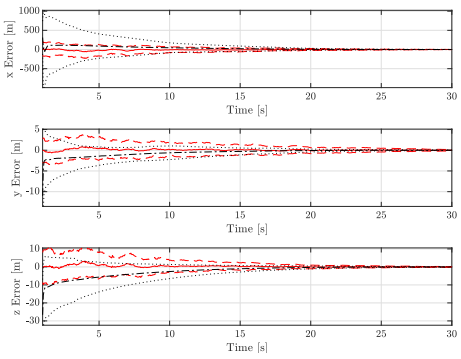
Figure 2.6 shows a sample MC trial for position and attitude errors, demonstrating that the UKF position estimates are smoother than the FHMLE estimates. However this comes with the drawback that the effects of noisy measurements taken in segments with poor geometry on estimates persist for longer than in the FHMLE, which can recover more quickly from unexpectedly poor quality measurements. This is the primary tradeoff between the two methods. Both FHMLE and UKF attitude estimates are nonsmooth, but errors are quite small.

Figure 2.7 shows the errors in position and velocity of the FHMLE without angle measurements. Because the initialization described above is not available without angle measurements, the filter is initialized with truth at the start of the trials and then warm starts are used throughout the rest of the trajectory. Regularization as in Remark 3 was observed to be necessary for convergence of the FHMLE without AoT measurements. Comparing Figures 2.3 and 2.7 demonstrates that the FHMLE performs significantly better in position and velocity estimates with AoT measurements than without. This matches the intuition given by Weighted GDOP in Figure 2.2.

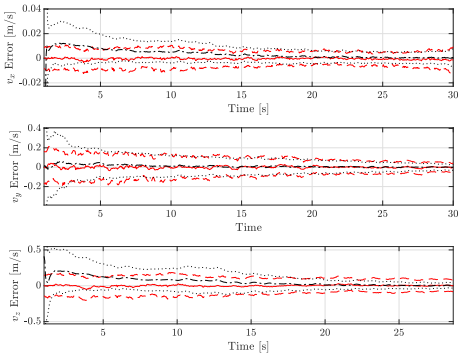
To validate the usefulness of the proposed FHMLE algorithm for more general maneuvers, we consider the case where the aircraft completes a banked circle of diameter 600 meters, centered 1500 meters from the deck of the carrier. Such a maneuver is common in the lead up to the carrier approach, and banked turns and descent combined capture

many possible aircraft motions. The Monte Carlo error results in histogram form over 100 Monte Carlo trials of 30 second flights are shown in Figure 2.8. We see the effect of bias and multimodal behavior in the UKF estimates. The UKF estimates appear to have lower variance, albeit with significant bias, than the FHMLE estimates in most cases. The FHMLE estimates, while less smooth, appear to be unbiased for every variable.

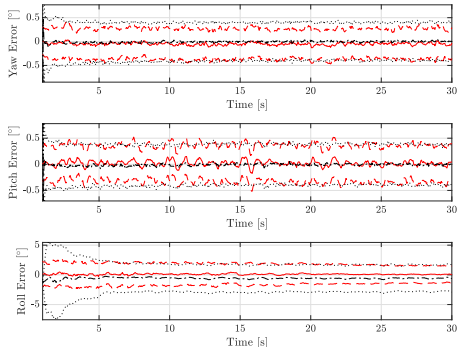
Figure 2.9 shows the effect of horizon length on runtime and root mean-squared error (RMSE). The FHMLE runtime for the one second horizon used for Figures 2.3-2.7 is around 11ms, while the UKF runtime in Matlab is around 5ms, and both are well below our sampling time for RF measurements (50ms). This plot shows that we can use a window up to around 2 seconds in length without average runtimes exceeding 30 ms. The computation times also grow approximately linearly with the window length, as suggested by the discussion in Section 2 The RMSE also decreases with horizon length, however these gains are offset by the latency introduced in estimation, especially for a rapidly moving vehicle. The determination of optimal window lengths is explored further in [95].



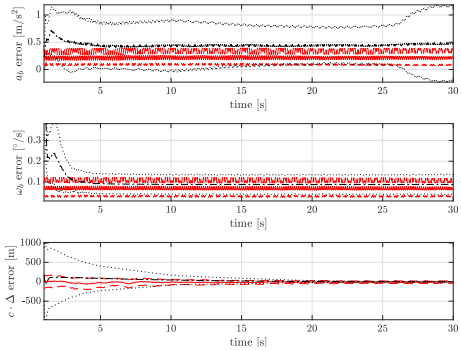
(a) position estimation errors



(b) velocity estimation errors

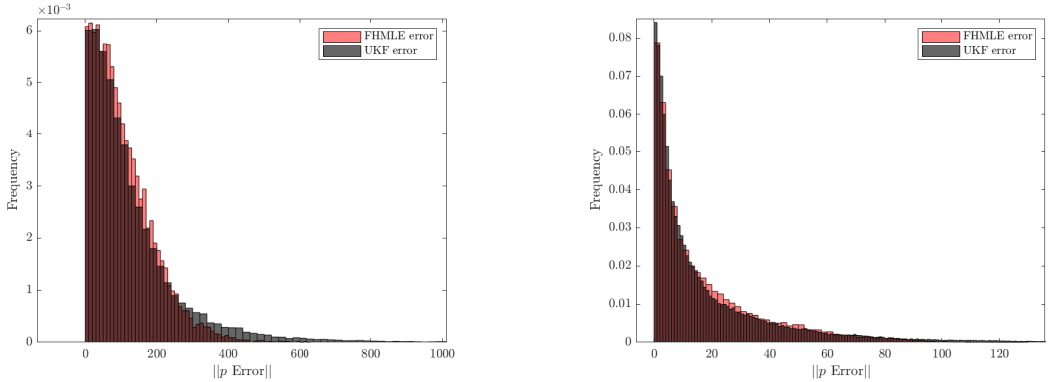


(c) attitude estimation errors



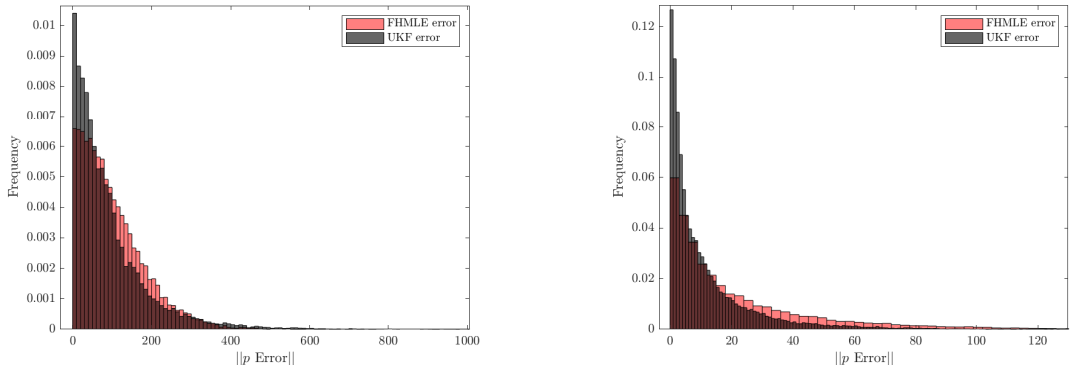
(d) bias estimation errors

Figure 2.3: Monte Carlo positions, velocities, orientation, and bias average errors with sigma bounds (solid red: FHMLE average error, dashed red: FHMLE 1-σ error, dot-dashed black: UKF average error, dotted black: UKF 1-σ error).



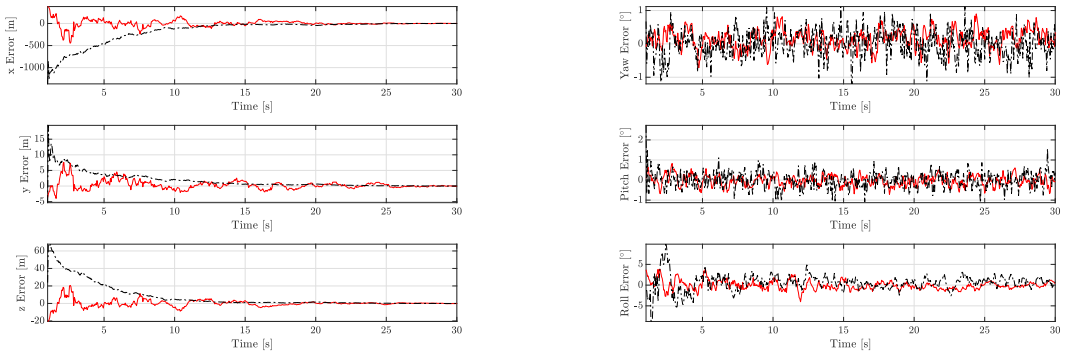
(a) first 10 seconds (b) final 20 seconds

Figure 2.4: Empirical distribution of $\|\hat{p} - p\|$ in the first 10 and final 20 seconds respectively. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap.



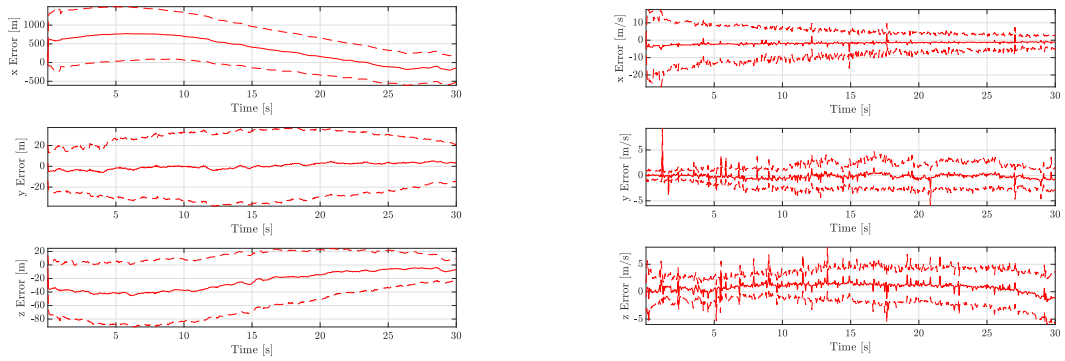
(a) first 10 seconds (b) final 20 seconds

Figure 2.5: Empirical distribution of $\|\hat{p} - p\|$ in the first 10 and final 20 seconds respectively, with $\kappa = 2.5 \cdot 10^4$ and $\Sigma_a = 0.03\, I_{3\times 3}\ m/s^2$. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap.



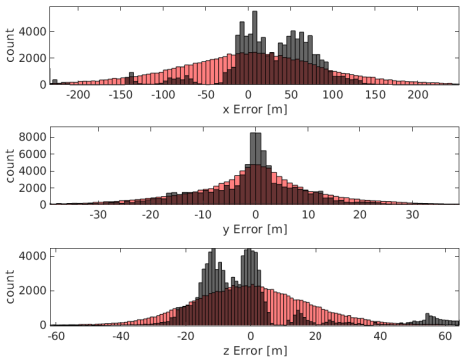
(a) position estimation errors (b) attitude estimation errors

Figure 2.6: Position and attitude estimation errors over a singe MC trial (solid red: FHMLE error, dot-dashed black: UKF error).

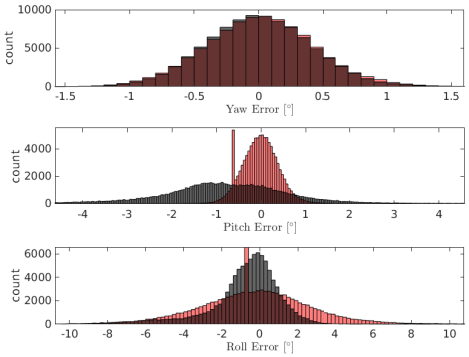


(a) position error without AoT (b) velocity error without AoT

Figure 2.7: Position and velocity Monte Carlo estimation errors without angle of transmission measurements (solid red: FHMLE average error, dashed red: FHMLE 1-σ error).

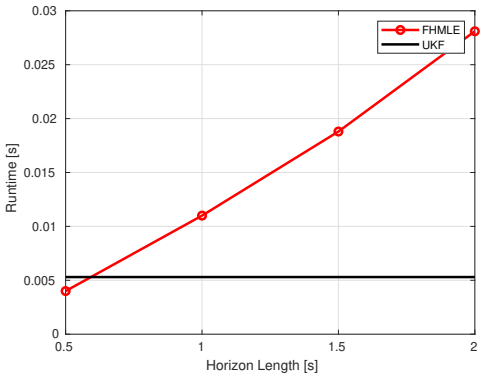


(a) position estimation errors

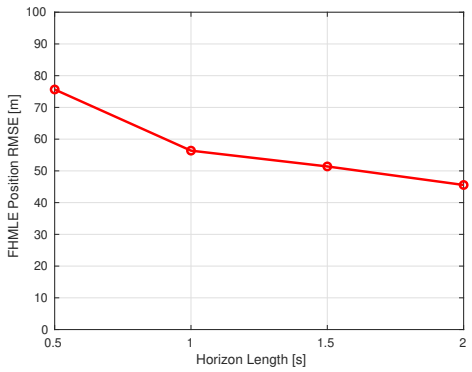


(b) attitude estimation errors

Figure 2.8: Position and attitude estimation errors over 100 MC trials of a 600m diameter banked circle. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap..



(a) Average runtimes for various horizon lengths



(b) RMSE for various horizon lengths

Figure 2.9: Average Runtimes and RMSE for various horizon lengths.

2.4.1 Conclusions

Navigation using RF beacons is useful in GPS-denied environments or when relying on Signals of Opportunity. We derived a Finite Horizon Maximum Likelihood Estimator for integrated navigation using realistic RF measurements. Benefits of the finite horizon approach include avoiding linearization, naturally accommodating constraints, and allowing for a richer class of noise models. This thesis shows that, by using appropriate numerical methods, FHMLE in integrated navigation is effective and compares favorably to a UKF in when using low-cost IMUs and in situations with poor geometry, especially when making use of AoT measurements.

Our simulation results show that the FHMLE tends to produce estimates that, while having a smaller estimation error can be less smooth than those obtained from a UKF. In moving horizon estimation (MHE), this is often resolved by adding to the optimization cost terms that penalize the distance between the current estimate and the one obtained with the previous window of measurements [30, 65]. Experimenting with this option and understanding its implications in the context of maximum likelihood estimation is an important topic for future research. Other topics for future research include the development of FHMLE approaches to the "loose integration" of RF and IMU measurements (as in [11]) and the simultaneous estimation of position/orientation and the parameters of the noise distributions (covariance matrices for Gaussian distributions and concentration parameters for the von Mises-Fisher distribution).

Chapter 3

State Estimation with Asynchronous Switching Behavior

Real-world systems are often best modeled in continuous time, for example using equations of motion, but with measurements taken at discrete instants [75]. Many systems also vary their behavior between discrete modes, either by their construction or to simplify control [90]. Examples include vehicles with a gearbox transmission, power systems using switched circuits or sources, or an aircraft with several trim conditions including cruising and banked turning. In real-world systems we must also consider noise in our measurements, usually represented by random additive noise. A practical formulation for such systems is a stochastic sampled-data switched system [73], given by

$$\dot{x}(t) = f(\sigma(t), x(t), u(t)) + w(t)$$
$$y(t_k) = h(\sigma(t_k), x(t_k)) + v(t_k),$$

where $x(t)$ is the state, $u(t)$ is an input, $w(t)$ is a disturbance, $y(t_k)$ is a measured output subject to random noise $v(t_k)$, $\sigma(t)$ is a "switching signal" taking values in a finite set

that tells us the active mode at time t, and t_k are discrete times indexed by k. The control of such systems is addressed, for example, in [47].

State estimation of discrete-time switched systems has attracted considerable attention, including works by Alessandri et al [2]. In these papers, the unknown switching signal is estimated using a Maximum-Likelihood method combined with either Kalman filtering or Moving Horizon Estimation of the continuous states. In contrast, Interacting Multiple-Model (IMM) approaches to hybrid system state estimation have been suggested in [10] and [40]. Ho [36] augmented these methods using Viterbi algorithm concepts to obtain pseudo Maximum-A-Posteriori (MAP) solutions to the windowed estimation problem. In [32], a review of estimation methods for switched systems is provided.

In these prior works, it is always assumed that switches occur only at times that measurements are obtained, in other words the sampling times. There are papers that consider estimation of continuous-time switched systems like [88], [44], and [79]. In [89], the authors consider switches that occur at a constant offset from the measurement times. However we could not find prior works that consider the problem of fully asynchronous switches with sampled measurements.

In this thesis we address state estimation when switches can occur at any time between measurement samples. In Section 3.1 we provide results on the convergence of Kalman filtering methods in the setting where the switching signal is known at sampling times, but exact switching times are unknown. We build upon analysis first done by Anderson and Moore [6], and more recently extended by Zhang [94]. Our results provide bounds on the mean error and mean-squared (MSE) of the estimates, which can be useful in the context of control [19].

In Section 3.3 we provide a method for simultaneously estimating the state $x(t)$ and switching signal $\sigma(t)$. This method is inspired by the IMM extended-Viterbi (IMM-EV1) approach [36]. In Sections 3.2 and 3.4, we show simulations that demonstrate our

theoretical results and validate the performance of our modified algorithms.

3.0.1 Problem description

We consider a linear sampled-data output-error switched system,

$$\dot{x}(t) = A(\sigma(t))x(t) + B(\sigma(t))u(t) \tag{3.1}$$

$$y(t_k) = Hx(t_k) + v_k, \tag{3.2}$$

for $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^\ell$, $v_k, y_k \in \mathbb{R}^m$, and $A(\sigma(t)) \in \{A(1), \ldots, A(L)\}$ a $n \times n$ matrix, $B(\sigma(t)) \in \{B(1), \ldots, B(L)\}$ a $n \times \ell$ matrix, with switching signal $\sigma(t) \in \{1, \ldots, L\}$. Our goal is to jointly estimate the switching signal and state at discrete periodically sampled timesteps $t_k = kT$, where T is the sampling period. We denote the state, input, and active mode at the discrete timesteps as $x_k = x(t_k)$, $u_k = u(t_k)$ and $\sigma_k = \sigma(t_k)$ respectively, as well as the active system matrices $A_k = A(\sigma(t_k))$ and $B_k = B(\sigma(t_k))$. We impose a dwell time $\tau_d > T$ so that switches occur at least τ_d apart from each other and at most once per sample. We can then parametrize the signal σ_t by the sequences $\{\sigma_k\}$ and $\{\bar{t}_k\}$, where the latter specifies the exact time at which a switch occurs within the interval $[t_k, t_{k+1})$.

We assume a zero order hold (ZOH) for the input, so that we have an exact discrete time update equation

$$x_{k+1} = F_k x_k + G_k u_k, \tag{3.3}$$

where

$$F_k = e^{A_{k+1}(T-\bar{t}_k)} e^{A_k \bar{t}_k}. \tag{3.4}$$

and

$$G_k = F_k \int_0^{\bar{t}_k} e^{-A_k \tau} B_k d\tau + \int_{\bar{t}_k}^{T} e^{A_{k+1}(T-\tau)} B_{k+1} d\tau. \tag{3.5}$$

3.1 Error bounds for state estimation

3.1.1 Assumptions

We consider the following definition of observability for a time-varying discrete-time linear system [94].

Definition 1 (Uniform Observability)**.** The sequence (F_k, H) is uniformly observable, i.e. there exist constants $h \in \mathbb{Z}_{>0}$ and $\rho_1 \in \mathbb{R}_{>0}$ such that for all $x \in \mathbb{R}^n$

$$\rho_1 \|x\|^2 \leqslant x' \left(\sum_{i=k}^{k+h} \Phi'_{k+h,i} H' R^{-1} H \Phi_{k+h,i} \right) x$$

where $\Phi_{i,k} \coloneqq F_{i-1} \cdots F_{k+1} F_k$.

In many cases, uniform observability of time-varying systems like switched systems is difficult to verify for all possible switching signals [9]. By imposing a dwell time, uniform observability of each mode can generate uniform observability of the switched system.

Assumption 1 (Each mode observable)**.** Each unswitched pair $(e^{A(i)T}, H)$ represents a uniformly observable system with constants $h_1, \ldots, h_L$, $\rho_1^1, \ldots, \rho_1^L$.

Lemma 1. *Suppose that we have Assumption 1 and $\tau_d > \overline{h}T$, where $\overline{h} \coloneqq \max\{h_1, \ldots, h_L\}$. Then the switched system in (3.3), (3.2) is uniformly observable for every admissible switching sequence with constants $h = 2\overline{h} - 1$ and $\rho_1 = \min\{\rho_1^1, \ldots, \rho_1^L\}$, that do not depend on the sequence.*

Proof: Given that $\tau_d > \bar{h}T$, the system must spend greater than h_j timesteps in any mode j. In order to guarantee that the time window $[t_k, t_{k+h})$ contains at least h_j samples uninterrupted in a single mode j, then our window must be at least $h = 2\bar{h} - 1$

samples long. In any window of this length we must have,

$$\rho \|x\|^2 \leqslant x' \left(\sum_{i=k}^{k+2\bar{h}-1} \Phi'_{i,k} H' R^{-1} H \Phi_{i,k} \right) x$$

where $\rho = \min\{\rho_1^1, \ldots, \rho_1^L\}$. □

Assumption 2 (Some Modes Observable)**.** Suppose that some of our modes are uniformly observable and we place a restriction on the ratio of time spent in unobservable modes, i.e. exists $r_t \in [0, 1)$ and $T_0 \in \mathbb{R}$ such that the time spend in unobservable modes, T_u, in a given interval (t_1, t_2) satisfies

$$T_u(t_1, t_2) \leqslant \tau_d + r_t(t_2 - t_1) \qquad \forall t_2 > t_1 \geqslant 0 \tag{3.6}$$

Lemma 2. *For some modes observable, let the dwell time be $\tau_d > \overline{h}T$, the maximum h_i among observable modes. Then the system is uniformly observable with constant*

$$h = \left\lceil \frac{2\overline{h} - 1 + \frac{\tau_d}{T}}{1 - r_t} \right\rceil \tag{3.7}$$

and ρ_1 the minimum ρ_1^j associated with observable modes.

Proof: Similarly to Lemma 1, any window of length h is guaranteed to spend at least $\overline{h}$ timesteps uninterrupted in an observable mode. The rest of the proof is identical to that of Lemma 1. □

3.1.2 Errors in system matrices

A Kalman filter is the MAP state estimator of a discrete-time system. Kalman filters compute state estimates $\hat{x}_k$ and their associated covariance matrices P_k at sample k.

We assume that our initial conidition is a random variable $x(0) \sim \mathcal{N}(\hat{x}_0, P_0)$ and that $v_k \sim$ iid $\mathcal{N}(0, R)$ for R a $m \times m$ symmetric positive-definite matrix. We compute the estimate at sample $k+1$ by combining y_{k+1} with a prediction $\hat{x}_{k+1|k}$ based on the previous estimat $\hat{x}_k$. These sources of information are combined through the Kalman gain matrix K_k, which depends on the system matrices as well as the measurement noise variance R. When we do not know system matrices F_k and G_k exactly, due to uncertainty in switching times, but have estimates $\hat{F}_k$ and $\hat{G}_k$, then our output-error Kalman filter update equations are of the form,

$$\hat{x}_{k+1|k} = \hat{F}_k \hat{x}_k + \hat{G}_k u \tag{3.8}$$

$$P_{k+1|k} = \hat{F}_k P_k \hat{F}_k' \tag{3.9}$$

$$K_k = (\hat{F}_k P_k \hat{F}_k') H' (H(\hat{F}_k P_k \hat{F}_k') H' + R)^{-1} \tag{3.10}$$

$$\hat{x}_{k+1} = (I - K_k H)\hat{x}_{k+1|k} + K_k y_{k+1} \tag{3.11}$$

$$P_{k+1} = (I - K_k H) P_{k+1|k}. \tag{3.12}$$

First we provide error bounds for our estimated system matrices assuming that we know the correct sequence $\{\sigma_k\}$ but not the exact switching times $\{\bar{t}_k\}$, instead using estimates $\{\hat{t}_k\}$ plugged into (3.4), (3.5) to compute $\hat{F}_k$ and $\hat{G}_k$. In this scenario we will bound the error of our state estimates using bounds on the error of the estimated state transition matrices due to switching time uncertainty.

Lemma 3 (Error in Estimation of System Matrices)**.** *For a transition between modes i and j, let the error in switching time estimation be denoted $\tilde{t} \coloneqq \hat{t} - \bar{t}$, then the estimation*

error, $\widetilde{F} \coloneqq \hat{F} - F$ *is bounded in norm as*

$$\begin{aligned} \left\|\widetilde{F}\right\| &\leqslant |\widetilde{t}|\,\|A(j) - A(i)\|\, e^{(\|A(j)-A(i)\|+3\|A(i)\|+\|A(j)\|)T} \\ &\leqslant T\,\|A(j) - A(i)\|\, e^{(\|A(j)-A(i)\|+3\|A(i)\|+\|A(j)\|)T} \end{aligned} \tag{3.13}$$

and the estimation error, $\widetilde{G} \coloneqq \hat{G} - G$ *is bounded in norm as*

$$\left\|\widetilde{G}\right\| \leqslant \left\|\widetilde{F}\right\| e^{\|A(i)\|T}\,\|B(i)\| + |\widetilde{t}|e^{\|A(j)\|T} \cdot \left(e^{\|A(i)\|T}\,\|B(i)\| + e^{\|A(j)\|T}\,\|B(j)\|\right). \tag{3.14}$$

Proof: Call $F_j = e^{A(j)(T-\bar{t})}$, $F_i = e^{A(i)\bar{t}}$, $E_j \coloneqq e^{-A(j)\tilde{t}}$ and $E_i \coloneqq e^{A(i)\tilde{t}}$. We then have that $\hat{F} = \hat{F}_j\hat{F}_i = F_jE_jE_iF_i$, so $\widetilde{F} = F_j(E_jE_i - I)F_i$, then

$$\begin{aligned} \left\|\widetilde{F}\right\| &\leqslant \|E_jE_i - I\|\,\|F_j\|\,\|F_i\| \\ &\leqslant \left\|(E_j - E_i^{-1})E_i\right\| e^{(\|A(j)\|+\|A(i)\|)T} \\ &\leqslant \left\|(e^{-A(j)\tilde{t}} - e^{-A(i)\tilde{t}})e^{A(i)\tilde{t}}\right\| e^{(\|A(j)\|+\|A(i)\|)T}. \end{aligned}$$

Then using the fact that $\left\|e^{X+Y} - e^X\right\| \leqslant \|Y\|\, e^{\|X\|+\|Y\|}$ [91] where $Y = -A(j)\widetilde{t} - A(j)\widetilde{t}$ and $X = -A(i)\widetilde{t}$, we obtain (3.13).

Our error in G, after some manipulation, can be written as

$$\widetilde{F}\int_0^{\hat{t}} e^{-A(i)\tau}B(i)d\tau + e^{A(j)(T-\bar{t})}\int_0^{\tilde{t}} e^{-A(i)\tau}B(i)d\tau - e^{A(j)(T-\bar{t})}\int_0^{\tilde{t}} e^{A(j)\tau}B(j)d\tau$$

from which we obtain

$$\left\|\widetilde{G}\right\| \leqslant \left\|\widetilde{F}\right\|\int_0^{\hat{t}} e^{\|A(i)\|\tau}\,\|B(i)\|\,d\tau + e^{\|A(j)\|T}\int_0^{\left\|\widetilde{t}\right\|} \left(e^{\|A(i)\|\tau}\,\|B(i)\| + e^{\|A(j)\|\tau}\,\|B(j)\|\right)d\tau$$

which gives us (3.14) after computing integrals. □

The bounds in (3.13) and (3.14) guarantee that the errors in $\hat{F}$ and $\hat{G}$ go to zero as the error in $\hat{t}$ goes to zero, which happens when our sampling period goes to zero. These bounds also improve as the $A(i)$'s become more similar to each other.

3.1.3 Estimation error bounds

To bound the estimation error of our filter, we denote the filter error by $e_k \coloneqq \hat{x} - x$, and the prediction error be $z_{k+1} \coloneqq x_{k+1} - \hat{F}\hat{x}_k - \hat{G}_k u_k$. With switching time uncertainty, the estimation error propagates as

$$e_{k+1} = (I - K_k H) z_{k+1} + K_k v_{k+1} \tag{3.15}$$

where in a sampling period in which no switch occurs,

$$z_{k+1} = \hat{F}_k e_k \tag{3.16}$$

and in a period where a switch occurs,

$$z_{k+1} = \hat{F}_k e_k + \widetilde{F}_k x_k + \widetilde{G}_k u_k. \tag{3.17}$$

We define the mean squared errors $\Sigma_k \coloneqq \mathbb{E}[e_k e_k']$ and $\Omega_k \coloneqq \mathbb{E}[z_k z_k']$. These update as

$$\Sigma_{k+1} = (I - K_k H)\Omega_{k+1}(I - K_k H)' + K_k R K_k'. \tag{3.18}$$

where when no switch occurs,

$$\Omega_{k+1} = \hat{F}_k \Sigma_k \hat{F}_k' \tag{3.19}$$

and when a switch occurs,

$$\begin{aligned}\Omega_{k+1} = &\hat{F}_k\Sigma_k\hat{F}_k' + \hat{F}_k\mathbb{E}[e_kx_k']\widetilde{F}_k' + \widetilde{F}_k\mathbb{E}[x_ke_k']\hat{F}_k \\ &+ \widetilde{F}_k\mathbb{E}\left[x_kx_k'\right]\widetilde{F}_k' + \hat{F}_k\mathbb{E}[e_k]u_k'\widetilde{G}_k \\ &+ \widetilde{F}_k\mathbb{E}\left[x_k\right]u_k'\widetilde{G}_k' + \widetilde{G}_ku_k\mathbb{E}\left[x_k\right]'\widetilde{F}_k' \\ &+ \widetilde{G}_ku_k\mathbb{E}[e_k]'\hat{F}_k' + \widetilde{G}_ku_ku_k'\widetilde{G}_k'.\end{aligned} \tag{3.20}$$

We will need the following:

Assumption 3. $P_{k|k-1}$ is positive definite and bounded above for all k. The upper bound is shown in [94]. Let $\overline{\lambda}$ denote the maximum, and $\underline{\lambda}$ the minimum eigenvalue that $P_{k|k-1}^{-1}$ can have.

Fact 1 (Observability of error dynamics)**.** *In [94] it is shown that if the sequence $(\hat{F}_k, H)$ uniformly observable then the sequence $(\hat{F}_k(I_n - K_{k-1}H), H)$ is also uniformly observable, i.e. there exists $\rho_3 \in \mathbb{R}_{>0}$ such that for the same h as in Definition 1,*

$$\rho_3 \left\|e\right\|^2 \leqslant e'\left(\sum_{i=k}^{k+h}\bar{\Phi}_{i,k}'H'R^{-1}H\bar{\Phi}_{i,k}\right)e$$

for all e, where $\bar{\Phi}_{i,k} \coloneqq F_{i-1}(I_n - K_{i-2}H)\cdots F_k(I_n - K_{k-1}H)$.

We now present a theorem bounding the expected prediction error and mean-squared prediction error.

Theorem 1 (Bounds on prediction error)**.** *Given Assumptions 1 and 3, and suppose $\mathbb{E}\left[x_k'x_k\right] < \gamma^2$, and $\left\|u_k\right\| < \delta$ for all k, let*

$$d \coloneqq \frac{\alpha_3}{\rho_3}\overline{\lambda}(\gamma^2\left\|\widetilde{F}_k\right\|^2 + 2\gamma\delta\left\|\widetilde{F}_k\right\|\left\|\widetilde{G}_k\right\| + \delta^2\left\|\widetilde{G}_k\right\|^2)$$

where $\alpha_3 = 1 + \alpha_1/\alpha_2$, $\alpha_1 > 0$ the largest possible eigenvalue of $H'P_{k|k-1}H$ for all k, and

$\alpha_2 > 0$ the smallest eigenvalue of R. Then there exist constants $\beta > 0$ and $\xi > 0$ such that if

$$c(a) \coloneqq \frac{\overline{\lambda}}{\underline{\lambda}}\left(a + \beta\sqrt{a} + \xi\right)$$

for $a \in \mathbb{R}_{>0}$, then for any $i \in \mathbb{Z}_{>0}$

$$\left\|\mathbb{E}\left[z_{k+i}\right]\right\|^2 \leqslant \max\left\{c\left(\left\|\mathbb{E}\left[z_k\right]\right\|^2\right), c\left(c(d)\right)\right\}. \tag{3.21}$$

Furthermore, there exist constants $\omega_h > 0$ and $\omega_{h-1} > 0$, such that the prediction MSE is bounded for all times $k + j$, $j \in \mathbb{Z}_{>0}$. as

$$tr(\Omega_{k+j}) \leqslant \max\left\{\frac{\overline{\lambda}}{\underline{\lambda}}\left(tr(\Omega_k) + 2\omega_{h-1}\right), \frac{\overline{\lambda}}{\underline{\lambda}}\left(\frac{2\overline{\lambda}\sigma_3}{\rho_3}\omega_h + 2\omega_h\right)\right\}. \tag{3.22}$$

A proof is provided just below in Section 3.1.4.

Remark 4 (Estimation error bounds). Given the bounds in Theorem 1, we can also bound $\mathbb{E}\left[e_k\right]$ and Σ_k for arbitrary k using

$$\left\|\mathbb{E}\left[e_k\right]\right\| \leqslant \left\|I - K_{k-1}H\right\| \left\|\mathbb{E}\left[z_k\right]\right\| \tag{3.23}$$

and

$$\operatorname{tr}(\Sigma_k) \leqslant \left\|I - K_{k-1}H\right\|^2 \operatorname{tr}(\Omega_k) + \left\|K_{k-1}\right\|^2 \operatorname{tr}(R) \tag{3.24}$$

which follow from (3.15) and (3.18) respectively.

Notably, the assumption $\mathbb{E}\left[x_k' x_k\right] < \gamma^2$ also serves as a bound on $\left\|\mathbb{E}\left[x_k\right]\right\|^2$ and variance of x_k. This theorem and remark state that the dwell time condition ensures that intermittent model uncertainties due to switching do not lead to unbounded growth

in our state estimation errors.

3.1.4 Proof of Theorem 1

From (3.15)-(3.17) we get that

$$\mathbb{E}\left[z_{k+1}\right] = \hat{F}_k(I - K_{k-1}H)\mathbb{E}\left[z_k\right]$$

when no switch occurs between samples k and $k+1$, and

$$\mathbb{E}\left[z_{k+1}\right] = \hat{F}_k(I - K_{k-1}H)\mathbb{E}\left[z_k\right] + \widetilde{F}_k\mathbb{E}\left[x_k\right] + \widetilde{G}_k u_k$$

when a switch occurs. We use the Lyapunov function

$$V_k \coloneqq \mathbb{E}\left[z_k\right]' P_{k|k-1}^{-1}\mathbb{E}\left[z_k\right]$$

which from Assumption 3 is positive definite and upper bounded. We have that

$$V_{k+1} - V_k = -\mathbb{E}\left[z_k\right]' H' S_k^{-1} H\mathbb{E}\left[z_k\right] \tag{3.25}$$

when no switch occurs, where $S_k \coloneqq HP_{k+1|k}H' + R$. When a switch occurs,

$$V_{k+1} - V_k = -\mathbb{E}\left[z_k\right]' H' S_k^{-1} H\mathbb{E}\left[z_k\right] + 2q' P_{k+1|k}^{-1}\Lambda_k\mathbb{E}\left[z_k\right] + q' P_{k|k-1}^{-1} q. \tag{3.26}$$

where $\Lambda_k \coloneqq \hat{F}_k(I - K_k H)$ and $q \coloneqq \widetilde{F}_k\mathbb{E}\left[x_k\right] + \widetilde{G}_k u_k$. It is derived in [94] that,

$$\mathbb{E}\left[z_k\right]' H' S_k^{-1} H\mathbb{E}\left[z_k\right] \leqslant -\frac{1}{\alpha_3}\mathbb{E}\left[z_k\right]' H' R^{-1} H\mathbb{E}\left[z_k\right] \tag{3.27}$$

for $\alpha_3 > 0$ defined in our theorem. For a switch occurring between times k and $k+1$ but no switches in the interval $\{k+1, \ldots, k+h\}$ we then know that $V_{k+h} - V_k$ is bounded above by

$$-\frac{1}{\alpha_3}\mathbb{E}\left[z_k\right]'\left(\sum_{i=k}^{k+h}\bar{\Phi}_{i,k}'H'R^{-1}H\bar{\Phi}_{i,k}\right)\mathbb{E}\left[z_k\right] + 2q'P_{k+1|k}^{-1}\Lambda_k\mathbb{E}\left[z_k\right] + q'P_{k|k-1}^{-1}q.$$

From uniform observability and Fact 1 we have

$$V_{k+h} - V_k \leqslant -\frac{\rho_3}{\alpha_3}\left\|\mathbb{E}\left[z_k\right]\right\|^2 + 2q'P_{k+1|k}^{-1}\Lambda_k\mathbb{E}\left[z_k\right] + q'P_{k|k-1}^{-1}q. \tag{3.28}$$

In other words, we now know that

$$\bar{\Phi}_{k+h,k}'P_{k+h|k+h-1}^{-1}\bar{\Phi}_{k+h,k} - P_{k|k-1}^{-1} \leqslant -\frac{\rho_3}{\alpha_3}I. \tag{3.29}$$

We want to show that for a switch occurring between samples k and $k+1$, the expected prediction error at sample $k+h$ satisfies,

$$\begin{cases} V_{k+h} < V_k & \text{if } \left\|\mathbb{E}\left[z(k)\right]\right\|^2 > d \\ \left\|\mathbb{E}\left[z(k+h)\right]\right\|^2 < c(d) & \text{if } \left\|\mathbb{E}\left[z(k)\right]\right\|^2 \leqslant d \end{cases}$$

for some constant $d > 0$, and positive continuous function $c(\cdot)$. We proceed by considering the two cases:

1. Suppose $\left\|\mathbb{E}\left[z_k\right]\right\|^2 > d$. We want to show that

$$\left\|\mathbb{E}\left[z_k\right]\right\|^2 > d \Rightarrow V_{k+h} - V_k < 0 \tag{3.30}$$

Applying S-procedure [80] to (3.28), we know that (3.30) is true if and only if (3.31)

is true.

$$\exists \pi > 0 \text{ s.t. } \pi \begin{bmatrix} -I & 0 \\ 0 & d \end{bmatrix} - \begin{bmatrix} -\frac{\rho_3}{\alpha_3} I & \Lambda_k' (P_{k|k-1}^{-1})' q \\ q' P_{k|k-1}^{-1} \Lambda_k & q' P_{k|k-1}^{-1} (\widetilde{F}_k x_k + \widetilde{G}_k u_k) \end{bmatrix} > 0 \tag{3.31}$$

By Schur complement, this is equivalent to (3.32).

$$\exists \pi > 0 \text{ s.t. } \frac{\rho_3}{\alpha_3} - \pi > 0 \text{ and } \pi d - q' P_{k|k-1}^{-1} q - (\frac{\rho_3}{\alpha_3} - \pi) q' P_{k|k-1}^{-1} \Lambda_k \Lambda_k' (P_{k|k-1}^{-1})' q > 0 \tag{3.32}$$

If we choose $\pi = \frac{\rho_3}{\alpha_3} - \varepsilon$ for some small enough $\varepsilon > 0$, such that if

$$d > \frac{\alpha_3}{\rho_3} \overline{\lambda} (\gamma^2 \left\| \widetilde{F}_k \right\|^2 + \gamma\delta \left\| \widetilde{F}_k \right\| \left\| \widetilde{G}_k \right\| + \delta^2 \left\| \widetilde{G}_k \right\|^2)$$

then we satisfy the conditions in (3.32) and therefore show that the Lyapunov function decreases before the next switch occurs. Then the maximum value attained by $\|\mathbb{E}[z_{k+i}]\|^2$ for $i > 0$ satisfies

$$\|\mathbb{E}[z_{k+i}]\|^2 \leqslant \frac{\overline{\lambda}}{\underline{\lambda}} \big(\|\mathbb{E}[z_k]\|^2 + \beta \|\mathbb{E}[z_k]\| + \xi \big) \tag{3.33}$$

where

$$\beta \coloneqq 2 \|\Lambda\| (\gamma \left\| \widetilde{F}_k \right\| + \delta \left\| \widetilde{G}_k \right\|) \tag{3.34}$$

$$\xi \coloneqq \gamma^2 \left\| \widetilde{F}_k \right\|^2 + \gamma\delta \left\| \widetilde{F}_k \right\| \left\| \widetilde{G}_k \right\| + \delta^2 \left\| \widetilde{G}_k \right\|^2 \tag{3.35}$$

2. Suppose $\|\mathbb{E}[z_k]\|^2 \leqslant d$. Then by substituting d into (3.26) we get

$$V_{k+1} \leqslant (\overline{\lambda} - \lambda_{\min}(\left\| H' S_k^{-1} H \right\|)) d + \overline{\lambda} \beta \sqrt{d} + \overline{\lambda} \xi. \tag{3.36}$$

Since we showed that the Lyapunov function is non-increasing over a timestep with

no switch and must decrease over h or more timesteps with no switch, then (3.36) gives us an upper bound on V_i for $k < i < k + j$ where $k + j$ is the sample where the next switch occurs. Then

$$\|\mathbb{E}\left[z_{k+j}\right]\|^2 \leqslant c(d) \coloneqq \frac{\overline{\lambda}}{\underline{\lambda}}\left(d + \beta\sqrt{d} + \xi\right) \tag{3.37}$$

which is a bound greater than d. If $\|\mathbb{E}\left[z_{k+j}\right]\|^2 > d$ then applying (3.33) to (3.37) tells us the maximum value attained by $\|\mathbb{E}\left[z_{k+i}\right]\|^2$ for $i > 0$ must satisfy

$$\|\mathbb{E}\left[z_{k+i}\right]\|^2 \leqslant c(c(d)) \tag{3.38}$$

(3.33) and (3.38) produce (3.21).

To prove (3.22) we will use the following Lyapunov function,

$$W_k \coloneqq \operatorname{tr}(P_{k|k-1}^{-1} \cdot \Omega_k)$$

and proceed by similar analysis as with the expected error. We will use the following Ruhe trace inequality [13, Fact 5.12.4, p. 333]:

Fact 2. *For positive semi-definite Hermitian matrices A and B with eigenvalues ordered largest to smallest, $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n \geqslant 0$ and $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n \geqslant 0$ respectively, the following holds*

$$\sum_{i=1}^{n} a_{n-i+1} b_i \leqslant tr(AB) \leqslant \sum_{i=1}^{n} a_i b_i \tag{3.39}$$

which gives us

$$\underline{\lambda}\operatorname{tr}(\Omega_k) \leqslant \operatorname{tr}(P_{k|k-1}^{-1}\Omega_k) \leqslant \overline{\lambda}\operatorname{tr}(\Omega_k) \tag{3.40}$$

where $\overline{\lambda}$ and $\underline{\lambda}$ are the maximum and minimum eigenvalues respectively attainable by

$P_{k|k-1}^{-1}$ which are given by Assumption 3. We can bound the update of our Lyapunov function, $W_{k+1} - W_k$, over the step after a switch using (3.18)-(3.20), (3.39), (3.40), and the fact that $2\mathbb{E}[e'x] \leqslant \varepsilon\mathbb{E}[e'e] + \frac{1}{\varepsilon}\mathbb{E}[x'x]$ for arbitrary $\varepsilon > 0$ as

$$W_{k+1} - W_k \leqslant \operatorname{tr}\left(\left(\bar{\Phi}'_{k+1,k} P_{k+1|k}^{-1} \bar{\Phi}_{k+1,k} - P_{k|k-1}^{-1}\right)\Omega_k\right) + \overline{\lambda}\varepsilon \left\|\hat{F}_k\right\| \left\|\widetilde{F}_k\right\| \operatorname{tr}(\Sigma_k) + \overline{\lambda} T_k \tag{3.41}$$

where

$$\begin{aligned} T_{k+1} = {} & (\hat{F}_k \mathbb{E}[e_k] + \widetilde{F}_k \mathbb{E}[x_k])' \widetilde{G}_k u_k \\ & + u_k' \widetilde{G}_k' (\hat{F}_k \mathbb{E}[e_k] + \widetilde{F}_k \mathbb{E}[x_k]) \\ & + \frac{1}{\varepsilon} \widetilde{F}_k^2 \mathbb{E}[x_k' x_k] + \widetilde{G}_k^2 \|u_k\|^2 \\ & + \hat{F}_k K_{k-1} R K_{k-1}' \hat{F}_k' \end{aligned}$$

with a switch. The Lyapunov function change over h steps, $W_{k+h} - W_k$, is then bounded by

$$\begin{aligned} & \operatorname{tr}\left(\left(\bar{\Phi}'_{k+h,k} P_{k+h|k+h-1}^{-1} \bar{\Phi}_{k+h,k} - P_{k|k-1}^{-1} + \varepsilon\eta I\right)\Omega_k\right) \\ & + \operatorname{tr}\left(P_{k+h|k+h-1}^{-1} \sum_{i=0}^{h-1} \bar{\Phi}'_{k+i|k} T_{k+h-1-i} \bar{\Phi}_{k+i|k}\right) \end{aligned}$$

where $\eta \coloneqq \overline{\lambda}\left\|\hat{F}_k\right\|\left\|\widetilde{F}_k\right\|\|I - K_{k-1}H\|$ and $T_i = \left\|\hat{F}_i K_{i-1} R K_{i-1}' \hat{F}_i'\right\|$ when $i \neq k$. From (3.40) and (3.29), we get

$$W_{k+h} - W_k \leqslant \left(\frac{-\rho_3}{\alpha_3} + \varepsilon\eta\right)\operatorname{tr}(\Omega_k) + \overline{\lambda}\operatorname{tr}\left(\sum_{i=0}^{h-1} \bar{\Phi}'_{k+h|k+i} T_{k+i} \bar{\Phi}_{k+h|k+i}\right) \tag{3.42}$$

We choose $\varepsilon = \rho_3/(2\alpha_3\eta)$, which also affects the value of T_k. Therefore we see that over any h steps, if the MSE at time k satisfies

$$\operatorname{tr}(\Omega_k) > \frac{2\overline{\lambda}\alpha_3}{\rho_3}\operatorname{tr}\left(\sum_{i=0}^{h-1} \bar{\Phi}'_{k+h|k+i} T_{k+i} \bar{\Phi}_{k+h|k+i}\right) \tag{3.43}$$

then $W_{k+h} - W_k \leqslant 0$. We note that it might be possible to achieve a better bound with different choice of ε. We must then consider the fact that unlike for the expected error in (3.27), the Lyapunov function W_i can now increase even in non-switch intervals due to the T_k terms. We will again deal with this by splitting into two cases. First let ω_j be defined as the upper bound derived from our upper bounds on F_k, K_k, etc., as well as bounds on $\widetilde{F}$ and $\widetilde{G}$ from Lemma 3, and bound on $\|\mathbb{E}[z_k]\|$ in (3.21), of the quantity

$$\operatorname{tr}\left(\sum_{i=0}^{j-1} \bar{\Phi}'_{k+j|k+i} T_{k+i} \bar{\Phi}_{k+j|k+i}\right) \leqslant \omega_j,$$

for any k. We know that for any $j > 0$,

$$W_{k+j} - W_k \leqslant \overline{\lambda}\omega_j$$

Let us consider the two cases:

1. Suppose $\operatorname{tr}(\Omega_k) > \frac{2\overline{\lambda}\alpha_3}{\rho_3}\omega_h$. Then (3.41) and (3.42) tell us that $W_{k+h} < W_k$ and the maximum value between k and $k+h$ is bounded as

$$\operatorname{tr}(\Omega_{k+j}) \leqslant \frac{\overline{\lambda}}{\underline{\lambda}}\left(\operatorname{tr}(\Omega_k) + \omega_{h-1}\right) \text{ for } j \in \{k, \ldots, k+h\}, \tag{3.44}$$

   which is also the maximum value attained until some $\operatorname{tr}(\Omega_{k+j}) \leqslant \frac{\overline{\lambda}\alpha_3}{\rho_3}\omega_h$, since the value cannot increase over h steps otherwise. This brings us to our next case:

2. Suppose $\operatorname{tr}(\Omega_k) \leqslant \frac{2\overline{\lambda}\alpha_3}{\rho_3}\omega_h$. Now the maximum value that $\operatorname{tr}(\Omega_{k+1})$ could attain is

$$\operatorname{tr}(\Omega_{k+1}) \leqslant \frac{\overline{\lambda}}{\underline{\lambda}}\left(\frac{2\overline{\lambda}\alpha_3}{\rho_3}\omega_h + \omega_1\right)$$

   If we achieved the maximum then $\operatorname{tr}(\Omega_{k+1}) > \frac{2\overline{\lambda}\alpha_3}{\rho_3}\omega_h$, so $W_{k+h+1} \leqslant W_{k+1}$ and

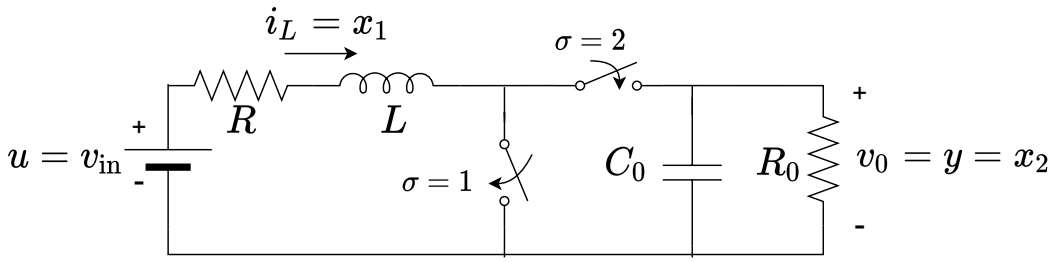


Figure 3.1: Boost Converter Circuit

therefore the maximum value of $\text{tr}(\Omega_{k+j})$ for all $j > 0$ is bounded as

$$\text{tr}(\Omega_{k+j}) \leqslant \frac{\overline{\lambda}}{\underline{\lambda}} \left(\frac{2\overline{\lambda}\alpha_3}{\rho_3}\omega_h + \omega_h \right) \quad j \in \mathbb{Z}_{>0} \tag{3.45}$$

with (3.44) and (3.45) combine to prove (3.22), with an additional ω_{h-1} or ω_h added to each to account for the case of starting in non-switch timestep. □

3.2 Case study: switching power conversion

A boost converter is a popular switching power converter for stepping up a DC voltage without transformers or amplifiers. This is necessary when a high-power source is not available to perform amplification. A model for a realistic boost converter is provided in [21]. We have dynamics as given in (3.1) where

$$A(1) = \begin{bmatrix} -R_1/L_1 & 0 \\ 0 & -1/R_0C_0 \end{bmatrix},$$
$$A(2) = \begin{bmatrix} -R_1/L_1 & -1/L_1 \\ 1/C_0 & -1/R_0C_0 \end{bmatrix},$$
$$B(1) = B(2) = [1/L_1 \;\; 0]'.$$

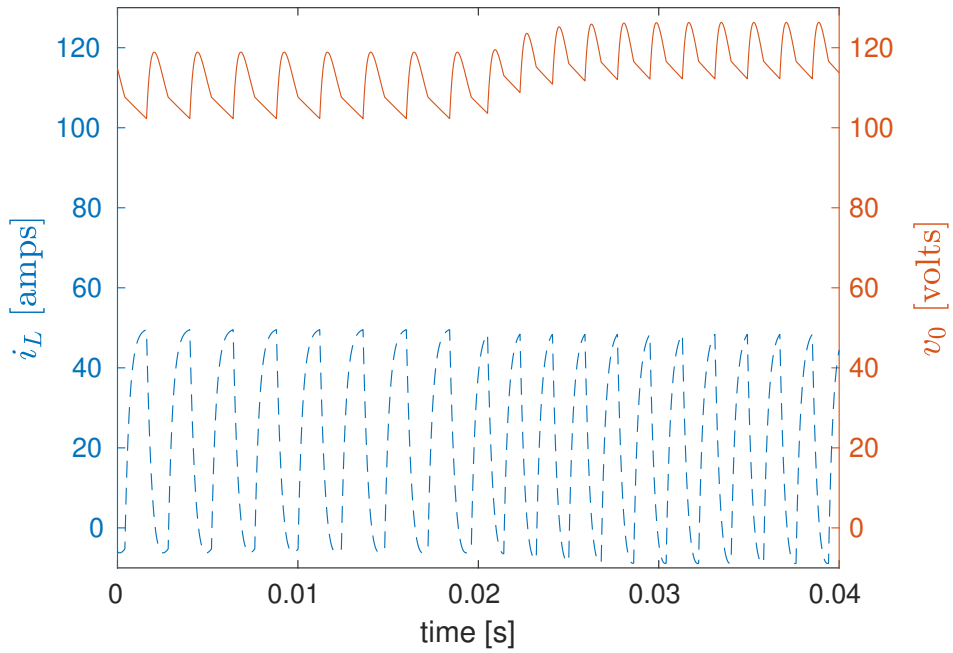


Figure 3.2: Example state evolution for Boost Converter starting with switches every 1.2 ms then increasing to every 0.9 ms at 0.02 seconds, blue dashed line indicates i_L and red solid line indicates v_0.

Here, $x = [i_L\ v_0]'$ and $u = v_{\text{in}}$. We additionally choose $y = v_0$, or in other words $H = [0\ 1]$. We use the values $R_1 = 2$ ohms, $L_1 = 500$ microhenrys, $R_0 = 50$ ohms, $C_0 = 470$ microfarads, and $v_{in} = 100$ volts from [21]. Figure 3.2 shows the result of simulating this system.

We simulated 10 seconds of operation with switch frequencies ranging from 1.2 to 0.9 ms/switch, and output voltages ranging between 100 and 120 volts, with measurement noise corresponding to $R = 5$ volts^2.

Table 3.1 shows the effect of sample period T on the Kalman filter estimation error when exact switching times are unknown but the active modes are known at switching times. Table 3.2 shows how effective the gridded estimation in (3.46) is when sampling at 0.5 ms for different values of g over 100 trials. As expected more precision in the switching time interval leads to more accuracy in the Kalman filter estimates, with diminishing

T (ms)	0.75	0.5	0.35	0.15	0.1
RMSE(i_L) (amps)	19.2	13.5	8.74	6.60	4.21
RMSE(v_0) (volts)	3.46	2.74	2.13	1.63	1.24

Table 3.1: Effect of Decreasing Sampling Period on Kalman filter RMSE

T/g (ms)	0.5	0.25	0.167	0.125	0.1
RMSE(i_L) (amps)	13.5	10.7	9.40	9.17	9.09
RMSE(v_0) (volts)	2.74	2.15	2.00	1.96	1.89

Table 3.2: Effect of increasingly precise gridded switching time estimation on Kalman filter RMSE, 100 trials

returns. This effect is greater with faster sampling than switching time optimization since faster sampling also gives us more measurements. However, in Table 3.1 the most improvement we would expect from having more measurements, in the case where our state was held constant, is a factor of $\sqrt{5} \approx 2.24$ between $T = 0.5$ and 0.1 ms. However, our accuracy improves by a factor of around 3, showing the effect of improved precision in switching time.

3.3 Joint estimation of state and switching

In Theorem 1, the error in the state estimates is driven by the switching time errors appearing in Lemma 3. We will augment the IMM Extended-Viterbi Kalman filter (IMM-EV1 KF) [36] with the maximum likelihood problem of estimating switching time within a single sample interval given by

$$J_k(\tau) \coloneqq p\big(y_{k+1} | x_{k+1} = \hat{x}_{k+1|k,\bar{t}=\tau}\big)$$

Where $\hat{x}_{k+1|k,\bar{t}=\tau}$ is found, for example, by computing (3.4), (3.5) and plugging into (3.3). We can search for the optimum of this cost by gridding the sample period $[t_k, t_{k+1})$ with

g points $\{\tau_i\}_1^g$ where $\tau_i \coloneqq \frac{iT}{g} - \frac{T}{2g}$. We can then compute

$$\hat{t}_k = \arg\max_{\tau_i} J_k(\tau_i). \tag{3.46}$$

We can now state Algorithm 1, a heuristic method which builds on the IMM-EV1 Kalman filter by including our gridded switching time estimation. To ensure that $\overline{\lambda}$ exists in Assumption 3, we add ϵI to each $P_{k|k-1}$ for some small $\epsilon > 0$.

Algorithm 1. *filter bank* $\{(\hat{x}_k^1, P_k^1), \ldots, (\hat{x}_k^L, P_k^L)\}$

mode probabilities $a_k^1, \ldots, a_k^L$

for *i from 1 to L* ***do***

for *j from 1 to L* ***do***

compute $\hat{t}_k^{ij}$ *for switch from i to j using* (3.46)

let $b_{ij} = J_k(\hat{t}_k^{ij})$

end for

$\hat{j} = \max_j b_{ij}$

compute $\hat{x}_{k+1}^i$, *and* P_{k+1}^i *from* $\hat{x}_k^{\hat{j}}$, $P_k^{\hat{j}}$, *and* $\hat{t}_k^{i\hat{j}}$,

$a_{k+1}^i = b_{i\hat{j}} a_k^{\hat{j}}$

end for

normalize a_{k+1}^i*'s*

3.4 Case study: maneuvering vehicle tracking

A model of a continuous-time switched system representing a vehicle moving in two dimensions with $x = [x_1 \ \dot{x}_1 \ x_2 \ \dot{x}_2]'$ is given by,

$$A(1) = A(2) = A(3) = I_2 \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$B(1) = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}'$$
$$B(2) = \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix}'$$
$$B(3) = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}'$$

where $\otimes$ denotes the Kronecker product, with $\tau_d > 1$ second and measurements sampled every 0.5 seconds. This double-integrator system corresponds to the discrete time switched systems used in [36], among others. Its discretization with ZOH over timestep T is given by

$$F(1) = F(2) = F(3) = I_2 \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \tag{3.47}$$

$$G(1) = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}'$$
$$G(2) = \begin{bmatrix} -\frac{T^2}{2} & -T & \frac{T^2}{2} & T \end{bmatrix}'$$
$$G(3) = \begin{bmatrix} \frac{T^2}{2} & T & -\frac{T^2}{2} & -T \end{bmatrix}'.$$

We consider a single trajectory over 10 seconds, with $u(t) = 1$, starting in mode 1, swiching to mode 2 at 1.65 seconds, to mode 3 at 2.75 seconds, and back to mode 1 at 3.9 seconds. The resulting trajectory is shown in Figure 3.3 along with a single trial of

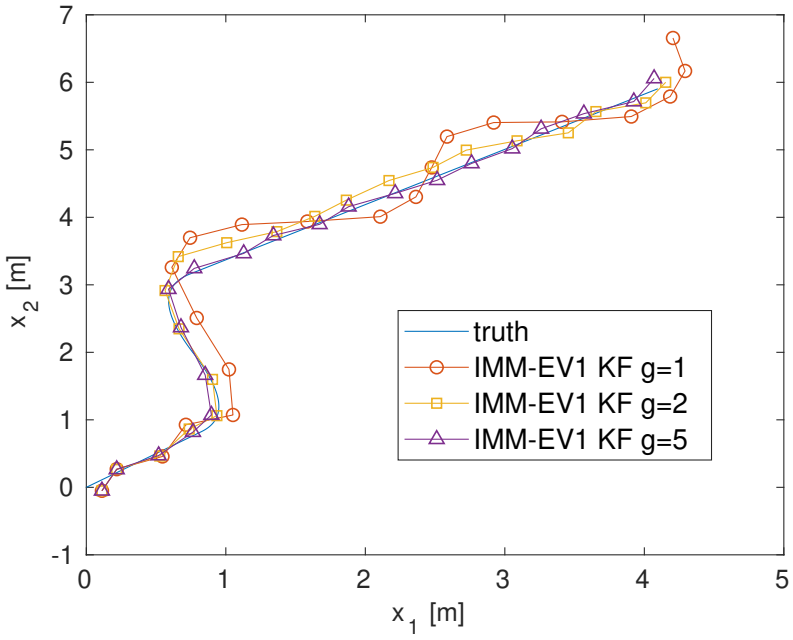


Figure 3.3: Single trial of vehicle true and estimated trajectories using Algorithm 1 for $g = 1, 2, 5$

estimate trajectories computed using Algorithm 1 where

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}.$$

We compute the RMSE over 500 monte carlo trials, and the results for varying divisions, g, of our sampling time are show in Figure 3.4.

3.4.1 Conclusion

We showed stability under dwell-time constraints of Switched System Kalman filtering errors with intermittent uncertainty in system dynamics due to unknown switching times. The bounds developed, while conservative, give us guarantees and intuition about filter implementations like the IMM-EV1 KF. Simulations of a boost converter and ma-

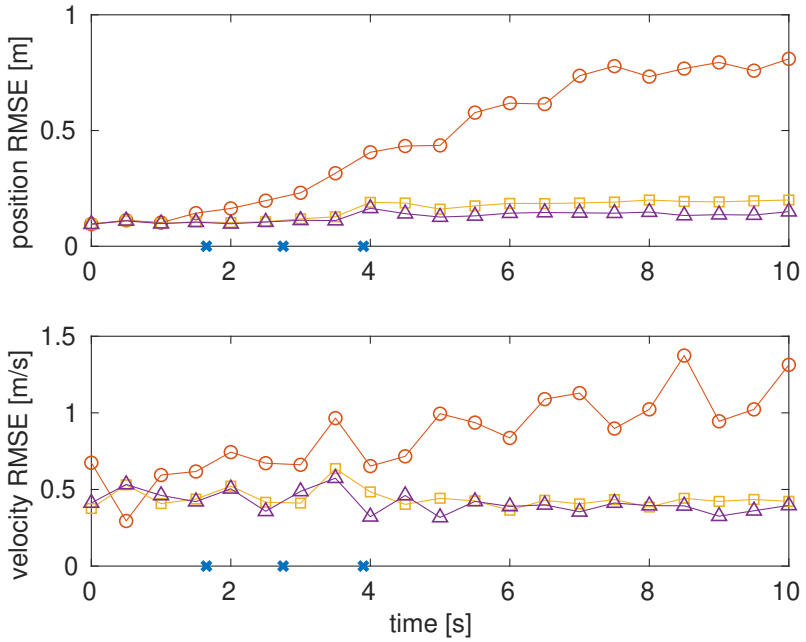


Figure 3.4: Vehicle position and velocity estimation RMSE over 500 MC trials, Algorithm 1 with $g = 2$ shown with red circles, $g = 5$ with yellow squares, and $g = 10$ with purple triangles. Blue x's mark the switching times.

neuvering vehicle showed improvement in the accuracy of filtering algorithms when we improved the precision of switching time estimates.

An immediate extension is to consider requirements on control algorithms to satisfy the assumptions in Theorem 1. It would be interesting to extend these results to nonlinear problems, for which our analysis could be applied to linearized error dynamics.

Bibliography

[1] A. Abdi, J. A. Barger, and M. Kaveh. A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station. *IEEE Transactions on Vehicular Technology*, 51(3):425–434, May 2002. (cited in p. 19)

[2] A. Alessandri, M. Baglietto, and G. Battistelli. A maximum-likelihood Kalman filter for switching discrete-time linear systems. *Automatica*, 46(11):1870–1876, 2010. (cited in p. 42)

[3] F. Allgöwer, T. A. Badgwell, J. S. Qin, J. B. Rawlings, and S. J. Wright. Nonlinear predictive control and moving horizon estimation — an introductory overview. In P. M. Frank, editor, *Advances in Control*, pages 391–449, London, 1999. Springer London. ISBN 978-1-4471-0853-5. (cited in p. 2)

[4] A. Amar and A. J. Weiss. Localization of narrowband radio emitters based on Doppler frequency shifts. *IEEE Transactions on Signal Processing*, 56(11):5500–5508, Nov 2008. (cited in p. 14)

[5] A. Amditis, A. Polychronopoulos, I. Karaseitanidis, G. Katsoulis, and E. Bekiaris. Multiple sensor collision avoidance system for automotive applications using an IMM approach for obstacle tracking. In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997)*, volume 2, pages 812–817. IEEE, 2002. (cited in p. 1)

[6] B. D. O. Anderson and J. B. Moore. Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM J. Control Optim.*, 19(1):20–32, 1981. (cited in p. 7, 42)

[7] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1):1–36, 2019. (cited in p. 11)

[8] U. M. Ascher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1st edition, 1998. (cited in p. 25)

[9] M. Babaali and M. Egerstedt. Pathwise observability and controllability are decidable. In *42nd IEEE Conf. Decis. Control*, volume 6, pages 5771–5776, 2003. (cited in p. 44)

[10] Y. Bar-Shalom, K.-C. Chang, and H. A. P. Blom. Tracking a maneuvering target using input estimation versus the interacting multiple model algorithm. *IEEE Trans. Aerosp. Electron. Syst.*, 25(2):296–300, 1989. (cited in p. 42)

[11] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Jan. 2001. (cited in p. 13, 16, 40)

[12] A. Barrau and S. Bonnabel. Invariant Kalman filtering. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:237–257, 2018. (cited in p. 12)

[13] D. S. Bernstein. *Matrix Mathematics*. Princeton University Press, 2009. (cited in p. 54)

[14] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1): 1455–1459, 2014. (cited in p. 12)

[15] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. (cited in p. 28, 29)

[16] T. H. Bryne, J. M. Hansen, R. H. Rogne, N. Sokolova, T. I. Fossen, and T. A. Johansen. Nonlinear observers for integrated INS/GNSS navigation: Implementation aspects. *IEEE Control Systems Magazine*, 37(3):59–86, June 2017. (cited in p. 14)

[17] J. V. Candy. *Bayesian signal processing: classical, modern, and particle filtering methods*, volume 54. John Wiley & Sons, 2016. (cited in p. 9)

[18] S. M. Chaves, R. W. Wolcott, and R. M. Eustice. NEEC research: Toward GPS-denied landing of unmanned aerial vehicles on ships at sea. *Naval Engineers Journal*, 127(1):23–35, 2015. (cited in p. 4)

[19] R. Chinchilla and J. P. Hespanha. Optimization-based estimation of expected values with application to stochastic programming. In *58th IEEE Conf. Decis. Control*, pages 6356–6361, 2019. (cited in p. 42)

[20] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on signal processing*, 50(3):736–746, 2002. (cited in p. 8, 9)

[21] G. S. Deaecto, J. C. Geromel, F. S. Garcia, and J. A. Pomilio. Switched affine systems control design with application to DC–DC converters. *IET Control Theory & Appl.*, 4(7):1201–1210, 2010. (cited in p. 57, 58)

[22] L. N. Egídio. *Contributions to Switched Affine Systems Control Theory with Applications in Power Electronics*. PhD thesis, Ph. D. thesis, University of Campinas, 2020. (cited in p. 4)

[23] G. Ferrari-Trecate, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems. *IEEE Transactions on Automatic Control*, 47(10):1663–1676, 2002. doi: 10.1109/TAC.2002.802772. (cited in p. 2)

[24] R. Fitzgerald. Divergence of the Kalman filter. *IEEE Transactions on Automatic Control*, 16(6):736–747, 1971. (cited in p. 15, 32)

[25] M. Ge and E. C. Kerrigan. Noise covariance identification for nonlinear systems using expectation maximization and moving horizon estimation. *Automatica*, 77: 336–343, 2017. (cited in p. 12)

[26] J. Gilbert, C. Moler, and R. Schreiber. Sparse matrices in Matlab: Design and implementation. *SIAM Journal on Matrix Analysis and Applications*, 13(1):333–356, 1992. doi: 10.1137/0613024. URL https://doi.org/10.1137/0613024. (cited in p. 29)

[27] R. T. Gould. The marine chronometer: its history and development. *Woodbridge*, 1989. (cited in p. 1)

[28] C. Graham and D. Talay. *Discretization of Stochastic Differential Equations*, pages 155–195. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-39363-1. doi: 10.1007/978-3-642-39363-1_7. URL https://doi.org/10.1007/978-3-642-39363-1_7. (cited in p. 5)

[29] M. S. Grewal and A. P. Andrews. Applications of Kalman filtering in aerospace 1960 to the present [historical perspectives]. *IEEE Control Systems Magazine*, 30 (3):69–78, 2010. (cited in p. 1)

[30] H. U. Gul, Y. D. Kai, and M. Khan. Multi-sensor integrated filtering for highly dynamic system using recursive moving horizon estimation technique. In *2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pages 351–356, Jan 2017. (cited in p. 15, 40)

[31] Y. Guo and B. Huang. Moving horizon estimation for switching nonlinear systems.

Automatica, 49(11):3270 – 3281, 2013. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2013.08.028. URL http://www.sciencedirect.com/science/article/pii/S0005109813004299. (cited in p. 2)

[32] Y. Guo and B. Huang. Moving horizon estimation for switching nonlinear systems. *Automatica*, 49(11):3270–3281, 2013. (cited in p. 42)

[33] P. D. Hanlon and P. S. Maybeck. Multiple-model adaptive estimation using a residual correlation Kalman filter bank. *IEEE Transactions on Aerospace and Electronic Systems*, 36(2):393–406, 2000. (cited in p. 9)

[34] J. P. Hespanha. `tenscalc` — a Matlab toolbox for nonlinear optimization using symbolic tensor calculus. https://github.com/hespanha/tenscalc, June 2017. (cited in p. 11, 28)

[35] J. P. Hespanha. TensCalc — A toolbox to generate fast code to solve nonlinear constrained minimizations and compute Nash equilibria. Technical report, University of California, Santa Barbara, Santa Barbara, June 2017. Available at http://www.ece.ucsb.edu/~hespanha/techrep.html. (cited in p. 11, 14, 28)

[36] T.-J. Ho and B.-S. Chen. Novel extended Viterbi-based multiple-model algorithms for state estimation of discrete-time systems with Markov jump parameters. *IEEE Trans. Signal Process.*, 54(2):393–404, 2006. (cited in p. 4, 42, 59, 61)

[37] M. Idan and J. L. Speyer. Cauchy estimation for linear scalar systems. *IEEE transactions on automatic control*, 55(6):1329–1342, 2010. (cited in p. 8)

[38] InvenSense. MPU-6050 MEMs IMU. https://www.invensense.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf, 2013. (cited in p. 31)

[39] J. T. Isaacs, K. O. Ezal, and J. P. Hespanha. Local carrier-based precision approach and landing system. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 6284–6290, Dec 2016. (cited in p. 14)

[40] L. A. Johnston and V. Krishnamurthy. An improvement to the interacting multiple model (IMM) algorithm. *IEEE Trans. Signal Process.*, 49(12):2909–2923, 2001. (cited in p. 42)

[41] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, Mar 2004. (cited in p. 8, 14, 32)

[42] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552. (cited in p. 2)

[43] J. T. Kent. The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):71–80, 1982. (cited in p. 19)

[44] G. K. Kolotelo, L. N. Egidio, and G. S. Deaecto. H_2 and H_∞ filtering for continuous–time switched affine systems. In *9th IFAC Symp. Robust Control Des.*, volume 51, pages 184–189, 2018. (cited in p. 42)

[45] P. Kuhl, M. Diehl, T. Kraus, J. P. Schloder, and H. G. Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers & Chemical Engineering*, 35(1):71–83, 2011. ISSN 0098-1354. doi: https://doi.org/10.1016/j.compchemeng.2010.07.012. URL https://www.sciencedirect.com/science/article/pii/S0098135410002607. (cited in p. 2, 12)

[46] D. Lewis. Polynesian and micronesian navigation techniques. *The Journal of Navigation*, 23(4):432–447, 1970. (cited in p. 1)

[47] D. Liberzon. *Switching in Systems and Control*. Birkhäuser, 2003. (cited in p. 42)

[48] F. Lindsten, T. B. Schön, and L. Svensson. A non-degenerate Rao-Blackwellised particle filter for estimating static parameters in dynamical models*. *IFAC Proceedings Volumes*, 45(16):1149–1154, 2012. ISSN 1474-6670. doi: https://doi.org/10.3182/20120711-3-BE-2027.00184. URL https://www.sciencedirect.com/science/article/pii/S1474667015381088. 16th IFAC Symposium on System Identification. (cited in p. 9)

[49] R. Lopez-Negrete, S. C. Patwardhan, and L. T. Biegler. Constrained particle filter approach to approximate the arrival cost in moving horizon estimation. *Journal of Process Control*, 21(6):909–919, 2011. ISSN 0959-1524. doi: https://doi.org/10.1016/j.jprocont.2011.03.004. URL https://www.sciencedirect.com/science/article/pii/S0959152411000497. (cited in p. 12)

[50] J. N. Lyness and C. B. Moler. Van der Monde systems and numerical differentiation. *Numerische Mathematik*, 8(5):458–464, Aug. 1966. (cited in p. 26)

[51] R. J. Matthys. *Crystal oscillator circuits*. Wiley-Interscience, New York, 1983. (cited in p. 25)

[52] J. C. Maxwell. On governors. *Proceedings of the Royal Society of London*, (16): 270–283, 1868. (cited in p. 1)

[53] R. Mehra. Digital simulation of multidimensional Gauss-Markov random processes. *IEEE Transactions on Automatic Control*, 14(1):112–113, 1969. (cited in p. 9)

[54] R. J. Meinhold and N. D. Singpurwalla. Robustification of Kalman filter models. *Journal of the American Statistical Association*, 84(406):479–486, 1989. (cited in p. 8)

[55] A. Mohamed and K. Schwarz. Adaptive Kalman filtering for INS/GPS. *Journal of geodesy*, 73(4):193–203, 1999. (cited in p. 9)

[56] J. J. Morales, P. F. Roysdon, and Z. M. Kassas. Signals of opportunity aided inertial navigation. In *Proceedings of the 29th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2016)*, pages 1492–1501, 2016. (cited in p. 1, 14)

[57] A. Mukhtar, L. Xia, and T. B. Tang. Vehicle detection techniques for collision avoidance systems: A review. *IEEE transactions on intelligent transportation systems*, 16(5):2318–2338, 2015. (cited in p. 4)

[58] R. Murray, Z. Li, and S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. Taylor & Francis, 1994. URL https://books.google.com/books?id=D_PqGKRo7oIC. (cited in p. 17)

[59] N. J. Nilsson et al. Shakey the robot. 1984. (cited in p. 1)

[60] Novatel. Hg1700 span. https://www.novatel.com/assets/Documents/Papers/HG1700_SPAN62.pdf, 2008. (cited in p. 24, 34)

[61] S. C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, and S. L. Shah. Nonlinear bayesian state estimation: A review of recent developments. *Control Engineering Practice*, 20(10):933–953, 2012. ISSN 0967-0661. doi: https://doi.org/10.1016/j.conengprac.2012.04.003. URL https://www.sciencedirect.com/science/article/pii/S0967066112000871. 4th Symposium on Advanced Control of Industrial Processes (ADCONIP). (cited in p. 1)

[62] L. Pina and M. A. Botto. Simultaneous state and input estimation of hybrid systems with unknown inputs. *Automatica*, 42(5):755–762, 2006. (cited in p. 12)

[63] S. A. P. Quintero, D. A. Copp, and J. P. Hespanha. Robust UAV coordination for target tracking using output-feedback model predictive control with moving horizon estimation. In *2015 American Control Conference (ACC)*, pages 3758–3764, July 2015. (cited in p. 11)

[64] K. Ramachandra. *Kalman filtering techniques for radar tracking*. CRC Press, 2018. (cited in p. 4)

[65] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, Feb 2003. (cited in p. 2, 40)

[66] J. B. Rawlings and B. R. Bakshi. Particle filtering and moving horizon estimation. *Computers & Chemical Engineering*, 30(10):1529–1541, 2006. ISSN 0098-1354. doi: https://doi.org/10.1016/j.compchemeng.2006.05.031. URL https://www.sciencedirect.com/science/article/pii/S0098135406001566. Papers form Chemical Process Control VII. (cited in p. 2, 9, 12)

[67] J. B. Rawlings, D. Q. Mayne, and M. Diehl. *Model predictive control: theory, computation, and design*, volume 2. Nob Hill Publishing Madison, 2017. (cited in p. 2, 11, 12)

[68] N. L. Ricker. Model predictive control with state estimation. *Industrial & engineering chemistry research*, 29(3):374–382, 1990. (cited in p. 1)

[69] H. Sairo, D. Akopian, and J. Takala. Weighted dilution of precision as quality measure in satellite positioning. *IEE Proceedings - Radar, Sonar and Navigation*, 150(6):430–436, Dec 2003. ISSN 1350-2395. doi: 10.1049/ip-rsn:20031008. (cited in p. 19)

[70] S. Särkkä, A. Vehtari, and J. Lampinen. Rao-Blackwellized particle filter for multiple target tracking. *Information Fusion*, 8(1):2–15, 2007. (cited in p. 9)

[71] F. H. Schlee, C. J. Standish, and N. F. Toda. Divergence in the Kalman filter. *AIAA Journal*, 5:1114–1120, June 1967. (cited in p. 8, 15)

[72] S. Shankar, K. Ezal, and J. P. Hespanha. Finite horizon maximum likelihood estimation for integrated navigation with RF beacon measurements. *Asian Journal of Control*, 21(4):1470–1482, 2019. (cited in p. 1, 8)

[73] B. Shen, Z. Wang, and X. Liu. A stochastic sampled-data approach to distributed H_∞ filtering in sensor networks. *IEEE Trans. Circuits Syst. I Regul. Pap.*, 58(9): 2237–2246, 2011. (cited in p. 41)

[74] P. Shi. Filtering on sampled-data systems with parametric uncertainty. *IEEE Transactions on Automatic Control*, 43(7):1022–1027, 1998. (cited in p. 3)

[75] P. Shi. Robust Kalman filtering for continuous-time systems with discrete-time measurements. *IMA J. Math. Control Inf.*, 16(3):221–232, 1999. (cited in p. 41)

[76] D. Simon. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Applications*, 4(8):1303–1318, August 2010. (cited in p. 15)

[77] H. W. Sorenson. Least-squares estimation: from Gauss to Kalman. *IEEE spectrum*, 7(7):63–68, 1970. (cited in p. 2)

[78] J. L. Speyer and W. H. Chung. *Stochastic processes, estimation, and control*. SIAM, 2008. (cited in p. 5, 8)

[79] A. Tanwani, H. Shim, and D. Liberzon. Observability for switched linear systems: Characterization and observer design. *IEEE Trans. Automat. Contr.*, 58(4):891–904, 2013. (cited in p. 42)

[80] F. Uhlig. A recurring theorem about pairs of quadratic forms and extensions: A survey. *Linear Algebra Appl.*, 25:219–237, 1979. (cited in p. 52)

[81] S. Ungarala. Computing arrival cost parameters in moving horizon estimation using sampling based filters. *Journal of Process Control*, 19(9):1576–1588, 2009. ISSN 0959-1524. doi: https://doi.org/10.1016/j.jprocont.2009.08.002. URL https://www.sciencedirect.com/science/article/pii/S0959152409001383. (cited in p. 12)

[82] D. Unsal and K. Demirbas. Estimation of deterministic and stochastic IMU error parameters. In *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pages 862–868, April 2012. doi: 10.1109/PLANS.2012.6236828. (cited in p. 21)

[83] R. Van Der Merwe, E. Wan, and S. I Julier. Sigma-point Kalman filters for nonlinear estimation and sensor-fusion-applications to integrated navigation. In *Proceedings of the AIAA Guidance, Navigation & Control Conference*, volume 3, 08 2004. (cited in p. 14)

[84] E. A. Wan and R. V. D. Merwe. The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 153–158, 2000. (cited in p. 32)

[85] E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing,*

Communications, and Control Symposium (Cat. No. 00EX373), pages 153–158. Ieee, 2000. (cited in p. 8)

[86] N. Wiener, N. Wiener, C. Mathematician, N. Wiener, N. Wiener, and C. Mathématicien. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, volume 113. MIT press Cambridge, MA, 1949. (cited in p. 2)

[87] F. F. Wu. Power system state estimation: a survey. *International Journal of Electrical Power & Energy Systems*, 12(2):80–87, 1990. (cited in p. 1)

[88] W. Xiang, M. Che, C. Xiao, and Z. Xiang. Observer design and analysis for switched systems with mismatching switching signal. In *2008 Int. Conf. Intell. Comput. Technol. Autom.*, pages 650–654, 2008. (cited in p. 42)

[89] W. Xiang, J. Xiao, and M. N. Iqbal. Robust observer design for nonlinear uncertain switched systems under asynchronous switching. *Nonlinear Anal. Hybrid Syst.*, 6 (1):754–773, 2012. (cited in p. 42)

[90] G. Yang and D. Liberzon. Stabilizing a switched linear system with disturbance by sampled-data quantized feedback. In *2015 Am. Control Conf.*, pages 2193–2198, 2015. (cited in p. 41)

[91] G. Yang and D. Liberzon. Feedback stabilization of switched linear systems with unknown disturbances under data-rate constraints. *IEEE Transactions on Automatic Control*, 63(7):2107–2122, July 2018. (cited in p. 47)

[92] T. Yang, P. G. Mehta, and S. P. Meyn. Feedback particle filter. *IEEE transactions on Automatic control*, 58(10):2465–2480, 2013. (cited in p. 9)

[93] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13–es, 2006. (cited in p. 1)

[94] Q. Zhang. On stability of the Kalman filter for discrete time output error systems. *Syst. & Control Lett.*, 107:84–91, 2017. (cited in p. 42, 44, 49, 51)

[95] Z. Zhou and B. Li. GNSS windowing navigation with adaptively constructed dynamic model. *GPS Solut.*, 19(1):37–48, Jan. 2015. ISSN 1080-5370. doi: 10.1007/s10291-014-0363-y. URL http://dx.doi.org/10.1007/s10291-014-0363-y. (cited in p. 15, 35)