ROBUST APPROACHES TO NONLINEAR FILTERING WITH APPLICATIONS TO NAVIGATION

A Dissertation

by

GUNNER SMITH FRITSCH

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Kyle J. DeMars
Committee Members,	Manoranjan Majji
	Suman Chakravorty
	Srikanth Saripalli
Head of Department,	Ivett Leyva

August 2022

Major Subject: Aerospace Engineering

ABSTRACT

Linear estimators, like the extended Kalman filter (EKF), find continual use (especially in the field of navigation) mostly for their familiarity and computational efficiency. Often, these estimation must be safeguarded from the realistic elements of physical systems, such as nonlinearities, non-Gaussian noises, and unmodeled effects. To this end, existing linear estimators are frequently outfitted with procedure-first robustness techniques—*ad hoc* mechanisms designed specifically to prevent filter failure—such as measurement editing, gain underweighting, filter resets, and more. As an alternative, this dissertation elects a model-first ethos, proposing nonlinear Gaussian mixture (GM) filters that are derived from first principles to be robust. These inherently robust algorithms are split into two approaches—1) non-Bayesian filters and 2) fault-cognizant filters—the end result being a collection of filters that challenge the status quo of current practical estimation; instead of reusing preexisting filter frameworks for the sake of ease, customized filters can be designed specifically for the system at hand.

1) Bayes' rule, while the archetypal basis for measurement fusion, relies on a fundamental assumption; all specified models, such as prior distributions and measurement likelihoods, are presumed to exactly reflect reality. In practice, this is rarely the case, warranting an investigation into non-Bayesian alternatives to traditional measurement updates. Fortunately, generalized variational inference (GVI) provides an established foundation for such updates and is used in this work to prototype several robust non-Bayesian filters. As closed-form filters are usually preferred, an iterative confidence-based update is derived, which, through Monte Carlo analyses, is shown to be selectively conservative, such that a desired level of robustness can be user-appointed.

2) Whereas traditional filtering screens out undesirable, or faulty, measurements, fault-cognizant filtering attempts to directly model these erroneous measurements, yielding estimators inherently capable of processing returns that conflict with the conventional model of a sensor. As the nature of both valid and faulty measurements can differ significantly between systems, several different fault-cognizant updates (FCUs) are derived, each purposed for a specific application. Subsequent analyses illustrate the robustness of the FCU to faulty measurements, both known and unknown.

DEDICATION

To my wife, Sadie, who continues to be my biggest supporter.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Kyle J. DeMars, Manoranjan Majji, and Suman Chakravorty of the Department of Aerospace Engineering and Professor Srikanth Saripalli of the Department of Mechanical Engineering.

All work conducted for the dissertation was completed by the student independently under advisor Dr. Kyle J. DeMars.

Funding Sources

This research was supported by the Department of Defense through the National Defense Science & Engineering Graduate Fellowship Program under fellowship number F-8445592924.

NOMENCLATURE

IFR	intrinsic fault resistance
GNC	guidance, navigation, and control
NASA	National Aeronautics and Space Administration
IMU	inertial measurement unit
VI	variational inference
GVI	generalized variational inference
FISST	finite set statistics
MMSE	minimum mean square error
pdf	probability density function
pmf	probability mass function
IID	independent and identically distributed
GMM	Gaussian mixture model
GM	Gaussian mixture
KF	Kalman filter
EKF	extended Kalman filter
GSF	Gaussian sum filter
GMF	Gaussian mixture filter
FCMM	fault-cognizant measurement model
FCU	fault-cognizant update
FAR	false alarm rate
FAR	field-of-view
MC	Monte Carlo

KLD	Kullback-Leibler divergence
NLL	negative log loss
EM	expectation-maximization
a	italics, lowercase, unbolded characters $\rightarrow$ scalar variables
$\boldsymbol{a}$	italics, lowercase, bolded characters $\rightarrow$ vector variables
$\boldsymbol{A}$	italics, uppercase, bolded characters $\rightarrow$ matrix variables
$\boldsymbol{x}$	state vector
$\boldsymbol{z}$	measurement vector
$\boldsymbol{A}^T$	matrix transpose
$\boldsymbol{A}^{-1}$	matrix inverse
$\boldsymbol{I}_n$	identity matrix of size $n \times n$
$\lvert a \rvert$	absolute value
$\lvert \boldsymbol{A} \rvert$	determinant
$\lVert \boldsymbol{A} \rVert$ or $\lVert \boldsymbol{A} \rVert_S$	matrix spectral norm
$\lVert \boldsymbol{A} \rVert_{\mathrm{F}}$	matrix Frobenius norm
$\Pr(\cdot)$	probability
$p(\cdot)$	probability density function
σ	standard deviation
$\mathcal{D}[\cdot \lVert \cdot]$	divergence
$\mathcal{L}(\cdot, \cdot)$	loss function
$\mathbb{E}_{p(\boldsymbol{a})}\big[f(\boldsymbol{a})\big]$	expectation operator: $\int p(\boldsymbol{a}) f(\boldsymbol{a}) \mathrm{d}\boldsymbol{a}$
$p_g(\boldsymbol{a} \lvert \boldsymbol{m}, \boldsymbol{P})$	multivariate Gaussian distribution of $\boldsymbol{a}$ with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{P}$
$a \sim p(\cdot)$	random variable a is distributed to pdf $p(\cdot)$
$\mathcal{O}(\boldsymbol{a}^n)$	collection of all terms dependent on $\boldsymbol{a}$ of order n and higher
$\binom{n}{k}$	combination operator or binomial expansion: $\frac{n!}{k!(n-k)!}$

$n\mathrm{P}k$	permutation operator: $\frac{n!}{(n-k)!}$
$\boldsymbol{A} = \mathrm{diag}\{a_1,\, a_2,\, \ldots,\, a_n\}$	denotes diagonal matrix where a_i is i^{th} diagonal of $\boldsymbol{A}$
$\delta(a)$	Dirac delta function where $\delta(a) = \begin{cases} +\infty\,, & \text{if } a = 0 \\ 0\,, & \text{if } a \neq 0 \end{cases}$

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

The term filtering, as it pertains to the estimation of uncertain systems, generally refers to methodologies that fuse incoming data with preexisting knowledge of the system at hand in order to produce an estimate of some variables of interest. Since its rise to popularity in the mid-20th century, which many credit to Kalman and Bucy [2, 3], statistical filtering has been an essential component in solving many of today's modern engineering problems. Indeed, it was the Apollo missions of the 1960s that necessitated the creation of the extended Kalman filter (EKF) by Schmidt [4, 5], which not only remains a staple of spaceflight navigation to this day [6, 7], but has arguably become the most popular filter for a wide range of engineering fields.

Regardless of the specific application or estimator employed, one of the most common traits sought after in filtering is robust behavior; failure of the filtering algorithms is to be avoided whenever possible. Failures may be caused by various factors, such as unmodeled effects or computational errors, and can quickly lead to system-wide problems with large consequences. To avoid this, most filtering architectures install robust mechanisms into their filters. In practice, where the field of estimation is dominated by linear estimators like the EKF, this typically involves the addition of *ad hoc* procedures that screen incoming data or soften the rate of information gain. This dissertation is concerned with investigating alternative methods of robust filtering, comparing them against common practices from navigation applications, and obtaining nonlinear filtering realizations that surpass the limitations of linear estimators.

1.1 Filtering

First and foremost, it is worthwhile to describe filtering as interpreted by this work. Most generally, filters are the solution to the problem of estimating a set of latent variables contained in $\boldsymbol{x}$, referred to herein as the state vector, that are indirectly observed via some vector $\boldsymbol{z}$ of noisy measurements, which are a function of the state dictated by the real measurement model of the system. The variables in the state vector $\boldsymbol{x}$ are often assumed to vary over time according to the

system's real dynamics, which, in a discrete sense, describes the transformation of the state at time t_{k-1} to time t_k, and thus, elements are commonly indexed by k to denote the corresponding time step. As it is impossible to directly observe $\boldsymbol{x}_k$, a filter can be applied to the real system as shown in Fig. (1.1).

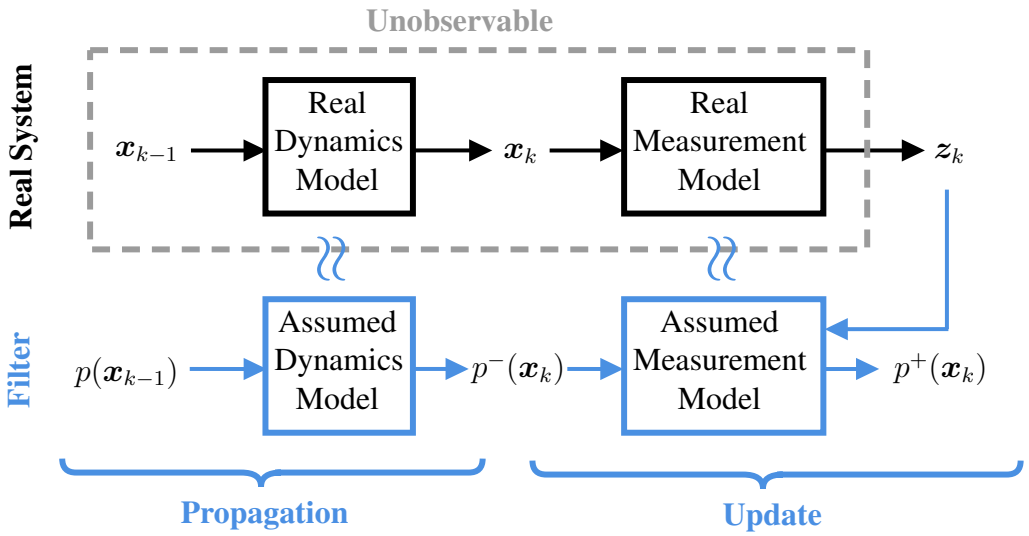


Figure 1.1 Example of a statistical filter being applied to a real system

Note that the filter models for the dynamics and measurements in Fig. (1.1) may not be equivalent to the models of the real system. This is reflective of reality; oftentimes, exact knowledge of the real system models is impossible and/or impractical, and thus, the filter assumes models that approximate the behavior of the physical system. Defining these assumed models is a key step in building a filter, as its derivation can be completely described by the assumptions it establishes.

While some filters may directly seek a state estimate $\hat{\boldsymbol{x}}_k$, such as least squares estimators [8], it is more informative to operate with respect to a probability density function (pdf) of which it is assumed $\boldsymbol{x}_k$ is a random variable, such that

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k) \, . \tag{1.1}$$

This concept is typically referred to as a probabilistic state-space model [9, 10]. By appropriately manipulating the pdf, the filter inherently retains a complete statistical description of the state from which quantities such as a state estimate or uncertainty measure can be extracted. This pdf manipulation is typically separated into two main stages—the propagation and the update. This work is primarily concerned with the update portion of filtering, but both are described here for reference.

1.1.1 Propagation

Assuming the filtering process of Fig. (1.1) can be initialized with the pdf $p(\boldsymbol{x}_{k-1})$, which contains preexisting knowledge of the state at the preceding time step, the filter propagation proceeds to transform $p(\boldsymbol{x}_{k-1})$ according to the assumed dynamical model, producing the propagated pdf of the state $p^{-}(\boldsymbol{x})$, which is referred to as the prior distribution. To propagate the distribution, the Chapman-Kolmogorov equations are typically paired with the classic Markovian assumption on the dynamical system [9, 11], producing

$$p^{-}(\boldsymbol{x}_k) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1})\mathrm{d}\boldsymbol{x}_{k-1}\,. \tag{1.2}$$

The term $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is known as the transition density that corresponds to the assumed dynamics model of Fig. (1.1) and is a pdf of $\boldsymbol{x}_k$ that quantifies its likelihood for a given $\boldsymbol{x}_{k-1}$. Practically speaking, $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is completely defined by the (assumed) dynamics of the state of Fig. (1.1), meaning dynamic models for $\boldsymbol{x}$ must be established before the transition density can be calculated.

1.1.1.1 Discrete Dynamics

In some cases, it is possible (and highly useful) to describe the dynamics of the state in a discrete sense, such that $\boldsymbol{x}_{k-1}$ is assumed to transform into $\boldsymbol{x}_k$ over a time interval from t_k to t_{k-1} according to the difference equation

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1}\,, \tag{1.3}$$

where $\boldsymbol{w}_{k-1}$ is a random variable used to model the process noise of a dynamical system.

Note that in the special case where: 1) $\boldsymbol{f}(\boldsymbol{x}_{k-1})$ is constrained to be a linear function of the state $\boldsymbol{F}\boldsymbol{x}_{k-1}$ and 2) $\boldsymbol{w}_{k-1}$ is taken to be zero-mean, Gaussian noise with covariance $\boldsymbol{Q}_{k-1}$, then the transition density can be expressed as

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = p_g(\boldsymbol{x}_k|\boldsymbol{F}\boldsymbol{x}_{k-1}, \boldsymbol{Q}_{k-1})\,.$$

If it is additionally assumed that the distribution $p(\boldsymbol{x}_{k-1})$ is a Gaussian, Eq. (1.2) will become equivalent to the propagation equations of the discrete Kalman filter.

1.1.1.2 Continuous Dynamics

While it is convenient to express dynamics in the discrete fashion of Eq. (1.3), there are many systems where the time-evolution of the state is naturally described by the differential equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t)\big) + \boldsymbol{w}(t)\,, \tag{1.4}$$

where $\boldsymbol{w}(t)$ is a white noise process that accounts for the stochasticity of the dynamical system. In systems with continuous dynamics, the transition density of Eq. (1.2) is not as straightforward to calculate. Instead of Eq. (1.4), the Fokker-Planck equation must be used, given by [9]

$$\frac{\partial p(\boldsymbol{x})}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\big\{p(\boldsymbol{x}) f_i(\boldsymbol{x})\big\} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\big\{p(\boldsymbol{x})[\boldsymbol{Q}^s]_{i,j}\big\}\,, \tag{1.5}$$

where $\mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}^T(\tau)] = \boldsymbol{Q}^s\delta(t-\tau)$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{Q}^s$ is the power spectral density of the process noise, and $f_i(\cdot)$ is the i^{th} element of $\boldsymbol{f}(\cdot)$. Therefore, given the distribution $p(\boldsymbol{x}_{k-1})$ as an initial condition, the prior distribution $p^-(\boldsymbol{x}_k)$ is produced via integration of Eq. (1.5) over the interval $[t_{k-1}, t_k]$. Unfortunately, few exact solutions to Eq. (1.5) exist, with the most popular solution requiring the same linear-Gaussian assumptions mentioned in Section 1.1.1.1.

1.1.2 Update

Again, referring to Fig. (1.1), at time t_k the filter update incorporates the observed information contained in measurement $\boldsymbol{z}_k$ to form the posterior distribution $p^+(\boldsymbol{x}_k)$. This is accomplished by fusing $\boldsymbol{z}_k$ according to the assumed measurement model with the prior information of the state in $p^-(\boldsymbol{x}_k)$. Most often, this process takes the form of Bayes' rule as

$$p^+(\boldsymbol{x}_k) \propto p^-(\boldsymbol{x}_k)\ell(\boldsymbol{z}_k|\boldsymbol{x}_k)\,, \tag{1.6}$$

where $\ell(\cdot|\cdot)$ is the measurement likelihood function, a detailed description of which is left for Section 4.1. However, while Bayes' rule is certainly prominent, the posterior is not required to take the form of Eq. (1.6); the primary motivation behind Chapter 3 is, in fact, non-Bayesian approaches to measurement updates. Following the update, the entire filtering procedure is usually iterated by recursively defining $p(\boldsymbol{x}_{k-1}) = p^+(\boldsymbol{x}_k)$, as filters are often designed to be sequential in practice.

1.1.3 Estimate Extraction

Often a part of a much larger architecture—i.e. guidance, navigation, and control (GNC)—each filter iteration typically includes the extraction of a point estimate $\hat{\boldsymbol{x}}_k$ from the posterior distribution, which can be passed to other subsystems. Point estimates can be generated many ways, but most methods involve minimizing the expectation of some risk function $\mathcal{R}(\cdot,\cdot)$ as [10]

$$\hat{\boldsymbol{x}}_k = \min_{\tilde{\boldsymbol{x}}} \left\{ \mathbb{E}_{p^+(\boldsymbol{x}_k)}\big[\mathcal{R}(\boldsymbol{x}_k, \tilde{\boldsymbol{x}})\big] \right\}. \tag{1.7}$$

The precise nature of the point estimate is directly determined by the definition of the risk function; for example, if

$$\mathcal{R}(\boldsymbol{x}_k, \tilde{\boldsymbol{x}}) = (\boldsymbol{x}_k - \tilde{\boldsymbol{x}})^T(\boldsymbol{x}_k - \tilde{\boldsymbol{x}})\,, \tag{1.8}$$

then the point estimate of Eq. (1.7) is the mean of $p^+(\boldsymbol{x}_k)$. Note that this is analogous to minimum mean square error (MMSE) estimation (or unbiased minimum variance estimation), and is known to have the optimal estimate of

$$\hat{\boldsymbol{x}}_k = \mathbb{E}_{p^+(\boldsymbol{x}_k)}\left[\boldsymbol{x}_k\right], \tag{1.9}$$

which is the mean of the posterior distribution of $\boldsymbol{x}_k$. A supplementary derivation of this is provided in Section B.1 for reference.

1.2 Robust Statistics

The term "robust" is often liberally used across various engineering fields to the point that it can take on different meanings even within the realm of estimation alone. In response, this section begins with a general description of robustness as it relates to statistics to better frame the objectives of the research. As mentioned in Section 1.1, filtering algorithms are completely described by their inherent assumptions, whether they align with the reality of the physical system or not. These assumptions include any foundational presuppositions made concerning the distribution of random variables, such as the prior state or observed measurements, and extend to assumptions made relevant by practical application, such as the size and nature of computational errors. In the vast majority of systems, it is impossible to form a practical filtering algorithm while guaranteeing the underlying assumptions will not be violated. Thus, the **robustness** of a filter becomes a critical element in determining successful estimation [12]—*how resilient is the filtering solution when subjected to deviations in various assumptions?* The less robust a filter is, the more prone it is to failure or corruption when implemented in real systems, and thus, highly robust filters are often sought after in many practical applications.

Since robustness is directly related to violations of the assumptions, an obvious question arises when trying to develop a robust filter: how does one rectify the filtering architecture to protect against these violations? The answer to this can differ significantly within the field of statistics. For instance, Box, who many credit with coining the term robust as understood by statisticians

[13], believes that robustness should be imparted to the underlying statistical models (**model-first**) such that resulting filtering algorithms can be strictly adhered to [14]. Conversely, others consider the best approach is to tailor the filtering procedures directly (**procedure-first**), as some fallacies in underlying model assumptions may be unavoidable [15]. Figure (1.2) is a concise visualization contrasting the model-first and procedure-first schools of thought. Robust filtering techniques derived from both philosophies exist, as well as techniques not belonging to either ideology, though most practical applications of robust filters involve the stability of a filter update to measurement outliers [16–18]. The research herein takes motivation from this previous work and seeks to develop and investigate model-first techniques that promote more robust filtering.

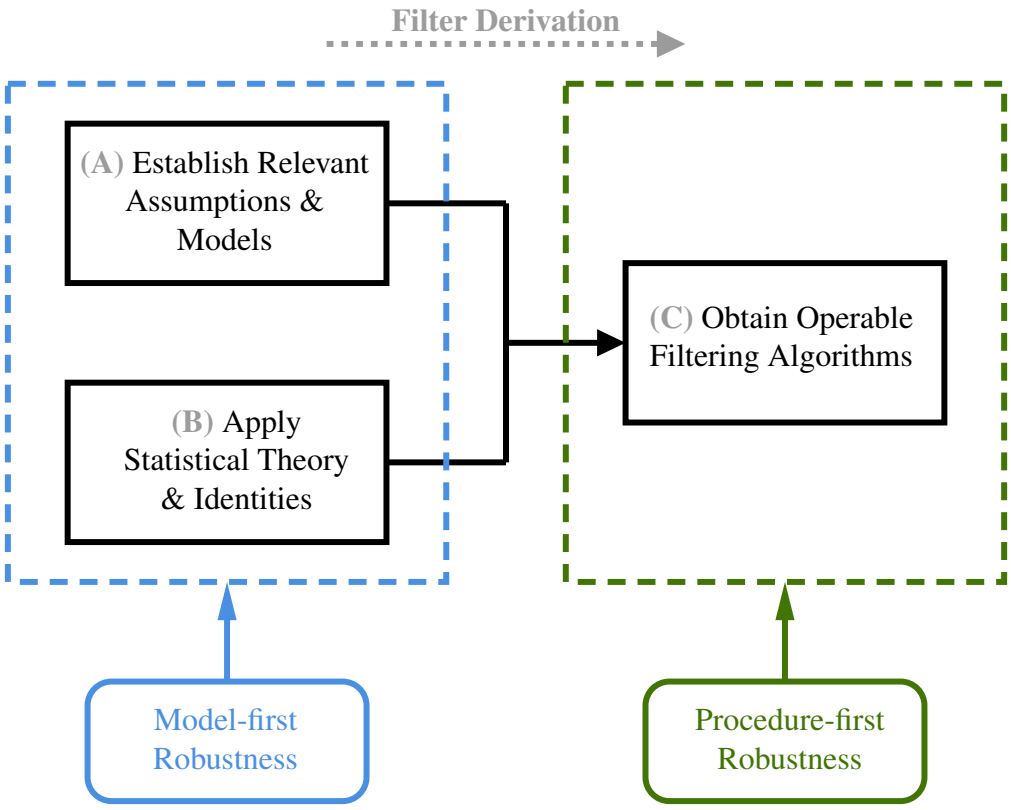


Figure 1.2 Procedure-first vs. model-first approaches to deriving robust filters

1.3 Current Filtering Practices

Following the birth of the Kalman filter, filtering quickly found popularity in the engineering community [19]. To many, the relatively straightforward application of statistics to state-space theory was attractive, in part, as it was readily adapted to many pre-existing systems. The prominence of filtering in engineering applications continues to this day, especially within the field of spacecraft navigation, resulting in decades of advancements being made in the field.

1.3.1 Spacecraft Navigation

Spacecraft navigation is a long-standing field and one of the first to push filtering theory into practical application. During the Apollo era, as NASA set its sights on landing a crewed-spacecraft on the moon, human-reliant navigation was no longer a sufficient means by which to navigate [20]. These missions required a reliable solution to real-time sensor fusion, as accurately estimating vehicle pose was necessary for a safe and successful lunar landing. This need pushed Schmidt and his team to develop the first version of the EKF, which many consider to be the birth of modern navigation.

Today, spacecraft navigation remains a thriving field, having been a core subsystem on many missions, such as the Space Shuttle and Orion [6, 7]. Though the precise configuration varies from mission to mission, most spacecraft navigation architectures contain some common elements [21]. For instance, the state vector $\boldsymbol{x}$ almost always consists of three-dimensional position and attitude, and frequently contains velocity as well. The propagation of the state estimate through time (i.e. the assumed dynamics model of Fig. (1.1)) is typically governed by an inertial measurement unit (IMU), which contains a gyroscope that receives measurements of the vehicle's rotational rate and an accelerometer that senses linear accelerations. The remaining sensor suite—which may include star cameras, rangefinders, GPS, optical cameras, LiDAR, magnetometers, and more—provides measurement data that is incorporated into the state estimate during the update stage of Fig. (1.1). The collection of algorithms responsible for propagating and updating the state estimate is referred to as the navigation filter.

Though many filtering methodologies exist, the vast majority of spacecraft navigation is realized through the EKF, or, more precisely, one of its variants. The multiplicative EKF (MEKF) is a filter frequently used in spaceflight as it designed specifically for the difficulties of attitude estimation [22]. Factorized forms of the EKF, such as square-root or UDU filters, are popular for their increased computational efficiency and numerical stability [23]. To prevent the navigation filter from failing, the two most common robustness procedures are residual editing (Section 4.7.1) —where measurement residuals are screened and any outliers are ignored by the filter—and underweighting (Section 3.4.1)—where information gain from measurements is artificially slowed to defend against linearization and modeling errors, both procedure-first approaches for designing robust filters. With these methodologies in mind, it is clear that linear, procedure-first filters are the dominant class of estimators in spacecraft navigation. Thus, throughout this work, the operation and limitation of these linear, procedure-first estimators is examined thoroughly, and improved estimation performance is sought by way of nonlinear, model-first approaches.

1.3.2 Additional Applications

While many current filtering practices are historically rooted in spaceflight, the application of such estimation algorithms extends far beyond navigation alone. Generally speaking, a filter is useful for any system where a state consisting of latent variables, often dynamically time-varying, is indirectly observed via noisy measurements. Such systems are frequently encountered and, in the modern era of computer-based technology, are typically well-suited to directly incorporate filtering architecture. Relevant examples include engineering applications like space domain awareness and autonomous vehicular navigation [24–26], along with systems found in more commonplace technologies such as phones [27, 28], virtual reality [29], and GPS [30]. Albeit a large motivation behind this research is spaceflight navigation and its continued reliance on the EKF, the principles can be applied to a wide variety of systems, and the work herein is presented in a generality that is readily adapted to many applications.

1.4 Dissertation Objectives

Considering the fact that current practice still almost universally relies on linear estimators with procedure-first robustness, this dissertation proposes alternative robust filtering methods with the following objectives:

1. Utilize methods of **nonlinear filtering** to allow for more accurate estimation of non-Gaussian systems and wider applicability.

2. Develop **model-first** approaches to robustness by addressing the statistical foundation and modeling assumptions of the filter.

3. Realize **closed-forms** to be compared with current filtering practices relevant to **navigation**.

The culmination of this dissertation is a set of novel filtering methodologies that seek to compete against and replace conventional filtering practices, with an emphasis on achieving robustness by directly addressing the inadequacy of modeling assumptions.

The remainder of the dissertation is segmented into three main chapters. Chapter 2 is a preliminary discussion on nonlinear filtering, which begins by contrasting the class of nonlinear estimators with that of linear estimators. The scope of Chapter 2 is then narrowed by focusing on nonlinear filters realized through Gaussian mixture (GM) approximations which includes the algorithms of the classic Gaussian sum filter as well as some common component management methods. Chapter 2 concludes with a Monte Carlo simulation that emphasizes the advantages of nonlinear filters relative to their linear counterparts.

Next, Chapter 3 investigates the practicality of non-Bayesian filtering, which begins by discussing Bayes' rule and the issues inherent to its use. Referring to Fig. (1.2), model-first robustness is applied at **(B)** by replacing the traditional Bayesian update with generalized variational inference (GVI), an optimization-based data fusion paradigm capable of more conservative estimation. Four GVI updates are investigated, showcasing the varying behavior that GVI is capable of, before a linear, closed-form, confidence-based update is derived that is more suitable for practical implementation. This new update is then adapted to nonlinear systems and made nonlinear itself by

Gaussian mixture approximations, and is subsequently analyzed via several simulations. During these analyses, the non-Bayesian update is specifically tested against the procedure-first robustness technique of underweighting, emphasizing its applicability to navigation.

Whereas Chapter 3 achieves robustness by restating its statistical update, Chapter 4 proposes robustness by altering the models and assumptions necessary to filter derivations (which targets (A) of Fig. (1.2)). Instead of traditional measurement models that only account for valid measurements, four different fault-cognizant measurement models (FCMMs) are proposed, ranging from sensors with single measurement returns to scans of unlabeled measurements that are more common to image-based feature tracking applications. Possible spatial and temporal measurement distributions are explored, and various nonlinear fault-cognizant updates (FCUs) are derived by pairing an appropriate FCMM with user-selected measurement distributions. The performance of the FCU is evaluated across several different simulations, where its robustness is compared against residual editing.

The dissertation concludes with Chapter 5, which summarizes the key findings of the research and remarks on potential future work to be done. In general, it is found that higher levels of robustness can be achieved, both through non-Bayesian and fault-cognizant filtering. If, additionally, an update is realized via a nonlinear GM filter, it also becomes more robust to non-Gaussianity and linearization errors.

2. NONLINEAR FILTERING

The term "nonlinear," with respect to filtering, can take on different meanings, so it is useful to clarify the definition to avoid any confusion. In this work, nonlinear filtering is best defined by differentiating it from linear filtering. As such, this chapter begins by summarizing linear estimators to enhance the subsequent discussion on nonlinear estimation.

2.1 Linear Estimators

In many respects, the prominence of linear estimators can be attributed to minimum mean square error (MMSE) estimation. MMSE estimators accomplish Bayesian inference via the optimization described by Eqs. (1.7) and (1.8). Analogous to minimum variance estimators (MVE) under unbiased conditions, most MMSE estimates are difficult to calculate analytically. However, if it is assumed that measurements are linear with respect to the state, closed forms for the MMSE become more-readily available. This is the foundation for linear minimum mean square error (LMMSE) estimators, which are referred to as linear estimators in this work.

Linear estimators are the predominant class of estimators used in filtering and are attractive in part due to their computational efficiency and algorithmic simplicity relative to nonlinear filters. Operating only on the first two moments, mean and (co)variance, of a pdf [31], linear estimators are only approximately optimal unless the linearity of a system is guaranteed. The Kalman filter, for example, is the optimal sequential MMSE estimator for linear systems. As dynamical and observational linearity is rare in physical systems, these LMMSE algorithms are often specifically tailored to perform well in the presence of system nonlinearities, giving rise to estimators like the extended Kalman filter (EKF). Such estimators are oftentimes poorly dubbed as nonlinear filters; in truth, any filter that forms estimates via a linear combination of the measurements is a linear estimator, such that the EKF is best described as a *linear* estimator for *nonlinear* systems.

There exist many variants of such linear estimators, the key differentiation between them being the manner in which they resolve nonlinearities. The EKF, for example, is an estimator that relies

on first-order Taylor series approximations to linearize the dynamical and measurement models of the system about the estimate. The iterated EKF, discussed in [32], attempts to improve upon the EKF by recursively linearizing about the estimate until a desired iteration objective is reached. In situations where higher-order effects are deemed non-negligible, the first-order approximation of the EKF may not be sufficient. In such cases, filters that employ second-order Taylor series linearizations, such as the Gaussian and truncated second-order filters [9, 33], are sometimes selected, although there is a tendency to avoid these algorithms as they require calculating significantly more derivatives than their first-order counterparts. To avoid calculating all derivatives inherent to Taylor series-based filters, Gaussian filters are considered preferable alternatives [34]. These filters are so named for the assumption of Gaussianity that permits point-based statistical linearization of the system models, with the most notable methods being the unscented, quadrature, and cubature Kalman filters [35–37].

2.1.1 A Traditional Linear Estimator: The EKF

As mentioned in Section 1.3.1, the EKF came about as a tool for spacecraft navigation in the 1960s as an adaptation of Kalman's linear filtering theory to nonlinear systems [20]. As it is such a prominent filter, it is used frequently in this work as a "baseline" to which other filters are compared, and thus, this section gives a brief summary of the core algorithms, which can be segmented into discrete stages similar to Fig. (1.1). In this work, the Bayesian-derived EKF is considered, where the system dynamics are built upon the models presented in Section 1.1.1. For the discrete dynamics of Eq. (1.3), the process noise $\boldsymbol{w}_{k-1}$ is assumed to be Gaussian distributed such that

$$\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1} \quad \text{where} \quad \boldsymbol{w}_{k-1} \sim p_g(\boldsymbol{w}_{k-1}|\boldsymbol{0}, \boldsymbol{Q}_{k-1})\,, \tag{2.1a}$$

where $\boldsymbol{Q}_{k-1}$ is the discrete process noise covariance. If the dynamics of the system are continuous, as in Eq. (1.4), $\boldsymbol{w}(t)$ is assumed to be a Brownian motion process (and thus Gaussian [9]), such

that

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t)\big) + \boldsymbol{w}(t) \quad \text{where} \quad \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}^T(\tau)] = \boldsymbol{Q}^s\delta(t-\tau)\,, \tag{2.1b}$$

where $\boldsymbol{Q}^s$ is the power spectral density of the process noise. Much like the dynamics models, the measurement model is also taken to be a nonlinear function of the state with additive, Gaussian noise as

$$\boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{x}_k) + \boldsymbol{v}_k \quad \text{where} \quad \boldsymbol{v}_k \sim p_g(\boldsymbol{v}_k|\boldsymbol{0}, \boldsymbol{R}_k)\,. \tag{2.1c}$$

Again, it should be noted that a key assumption of the models in Eqs. (2.1) is that the noises are zero-mean, Gaussian, and additive.

2.1.2 EKF Initialization

The EKF is initialized at time t_0 with an initial state estimate of the form

$$\boldsymbol{x}_0 \sim p_g(\boldsymbol{x}_0|\boldsymbol{m}_0, \boldsymbol{P}_0)\,, \tag{2.2}$$

such that it is assumed that the initial state $\boldsymbol{x}_0$ is randomly drawn from a Gaussian distribution of mean $\boldsymbol{m}_0$ and covariance $\boldsymbol{P}_0$. To initialize the recursions of the filter, note that $\boldsymbol{m}_0^+ = \boldsymbol{m}_0$ and $\boldsymbol{P}_0^+ = \boldsymbol{P}_0$, which is a notational change made purely for clarity.

2.1.3 EKF Propagation

The discrete propagation stage of the EKF is responsible for transitioning the state estimate through time as

$$\boldsymbol{m}_k^- = \boldsymbol{f}_k(\boldsymbol{m}_{k-1}^+) \tag{2.3a}$$

$$\boldsymbol{P}_k^- = \boldsymbol{F}_k(\boldsymbol{m}_{k-1}^+)\boldsymbol{P}_{k-1}^+\boldsymbol{F}_k^T(\boldsymbol{m}_{k-1}^+) + \boldsymbol{Q}_{k-1}\,, \tag{2.3b}$$

where $\boldsymbol{F}_k(\boldsymbol{m})$ is the Jacobian of the dynamics function $\boldsymbol{f}_k(\cdot)$ evaluated at $\boldsymbol{x} = \boldsymbol{m}$. Thus, the propagated state pdf is given as

$$p^-(\boldsymbol{x}_k) = p_g(\boldsymbol{x}_k|\boldsymbol{m}_k^-, \boldsymbol{P}_k^-)\,. \tag{2.3c}$$

If the dynamical model is continuous, as in Eq. (2.1b), the mean and covariance of Eq. (2.3c) must be calculated by integrating

$$\dot{\boldsymbol{m}}(t) = \boldsymbol{f}\big(\boldsymbol{m}(t)\big) \tag{2.3d}$$

$$\dot{\boldsymbol{P}}(t) = \boldsymbol{F}\big(\boldsymbol{m}(t)\big)\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{F}^T\big(\boldsymbol{m}(t)\big) + \boldsymbol{Q}^s\,, \tag{2.3e}$$

where $\boldsymbol{F}(\cdot)$ is the Jacobian of $\boldsymbol{f}(\cdot)$.

2.1.4 EKF Update

The EKF's update stage fuses the information contained in measurement $\boldsymbol{z}_k$ of Eq. (2.1c) into the propagated state estimate from Eq. (2.3c) to form the posterior state estimate

$$p^+(\boldsymbol{x}_k) = p_g(\boldsymbol{x}_k|\boldsymbol{m}_k^+, \boldsymbol{P}_k^+)\,, \tag{2.4a}$$

where

$$\boldsymbol{m}_k^+ = \boldsymbol{m}_k^- + \boldsymbol{K}\Delta\boldsymbol{z}_k \tag{2.4b}$$

$$\boldsymbol{P}_k^+ = \boldsymbol{P}_k^- - \boldsymbol{K}\boldsymbol{H}_k(\boldsymbol{m}_k^-)\boldsymbol{P}_k^- \tag{2.4c}$$

$$\boldsymbol{K} = \boldsymbol{P}_k^-\boldsymbol{H}_k^T(\boldsymbol{m}_k^-)\boldsymbol{W}^{-1} \tag{2.4d}$$

$$\boldsymbol{W} = \boldsymbol{H}_k(\boldsymbol{m}_k^-)\boldsymbol{P}_k^-\boldsymbol{H}_k^T(\boldsymbol{m}_k^-) + \boldsymbol{R}_k \tag{2.4e}$$

$$\Delta\boldsymbol{z}_k = \boldsymbol{z}_k - \boldsymbol{h}_k(\boldsymbol{m}_k^-)\,, \tag{2.4f}$$

where $\boldsymbol{H}_k(\boldsymbol{m})$ is the Jacobian of the measurement function $\boldsymbol{h}_k(\cdot)$ evaluated at $\boldsymbol{x} = \boldsymbol{m}$.

Traditional EKF Vs. Bayesian EKF

Many times, there is little to no differentiation between the Bayesian EKF of Eqs. (2.3) and (2.4) and the traditional, MMSE-derived EKF, as they are algorithmically identical. The distinction between the two lies in the assumptions—most prevalently that the Bayesian EKF requires an assumption of Gaussianity that the traditional EKF does not. As a consequence, the means and covariances of the traditional EKF are often mistakenly prescribed to Gaussian distributions, which is not necessarily true.

2.2 Nonlinear Estimators

While describing linear estimators in Section 2.1 is rather straightforward, it can be more difficult to precisely explain what a nonlinear estimator is, as they can be quite varied in appearance. In fact, it is perhaps easiest to classify a nonlinear estimator as any filter that does not strictly fall within the definition of a linear estimator. But even with this differentiation between linear and nonlinear filters, this does not explain why nonlinear filters are needed, or what advantages they have over their linear counterparts. As such, this section helps clarify these points.

While linear estimators are extremely useful in their own right, there will always exist certain drawbacks when subjected to the nonlinear reality of physical systems. For example, many of these filters rely on assumptions of Gaussianity, such that failure is likely to occur in the presence of non-Gaussian noises. Furthermore, errors from repeated linearizations can accumulate to the point of debasing the filtering solution, resulting in less robust, less accurate estimation. It is for these reasons that nonlinear filtering is such a prominent field of research and a focal point of this work.

2.2.1 Particle Filtering

There exist numerous approaches to nonlinear filtering [38–40], but particle filtering and Gaussian mixture (GM) filtering are considered the most prevalent. In fact, many consider the particle filter, and its many variants, to be the truest form of nonlinear filtering, as it involves generating a large number of "particles" by directly sampling the prior distribution, performing filtering

operations on each individual particle, and forming a posterior distribution constructed from the propagated/updated particles. One advantage to particle filtering is that a closed-form generally requires less restrictive assumptions, such as not needing identical input and output distributions as required by many filters. However, particle filtering performance improves as the number of particles increases, such that some applications require a vast number of particles to ensure accurate estimation, which in turn escalates the computational burden. Furthermore, as the filter iteratively operates on the particles, the state distribution often morphs such that some particles may no longer effectively describe the statistical nature of the system, which warrants a resampling procedure. As such, particle filtering is generally viewed as too computationally demanding for use in onboard, real-time spacecraft navigation, but is used frequently in robotics applications.

2.2.2 Gaussian Mixture Filtering

GM filtering is, in many ways, quite similar to particle filtering. Where particle filtering approximates pdfs with a collection of sampled particles, a GM filter uses a summation of weighted Gaussian distributions to represent the pdf. A useful, though overly simplified, comparison of the two methods is that a GM can be interpreted as a collection of particles, with the exception that each GM "particle"—usually referred to as a GM component—possesses a probabilistic "volume", attributed to the Gaussian of each component, which a standard particle does not have. Generally, this implies that a GM can more efficiently describe a distribution, as it typically requires a fewer number of "particles" to accurately represent a distribution than a particle filter does, albeit each GM component contains multiple parameters. Additionally, the nature of the Gaussian distribution is such that it possesses many useful and advantageous properties, which can significantly aid the derivation and implementation of filtering equations. For instance, the Gaussian distribution is well-known for its use in reaching a closed-form solution to the Bayesian-derived Kalman filter [10]. Furthermore, while extracting estimates from particle filters can be complicated, many times involving complex clustering operations, GM representations have efficient and straightforward techniques for estimate extraction [41]. Therefore, this work elects GM filters as the primary method for nonlinear filtering.

2.2.2.1 The Gaussian Distribution

The Gaussian, or normal, distribution is perhaps the most well-known and widely-used distribution in all of statistics. With its discovery dating as far back as the early 1700s [42], the popularity of the Gaussian is not by chance; it frequently appears in natural systems and has many appealing mathematical characteristics [43]. Given in (multivariate) pdf form by

$$p_g(\boldsymbol{a}|\boldsymbol{b},\boldsymbol{C}) = \left|2\pi\boldsymbol{C}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\boldsymbol{a}-\boldsymbol{b}]^T\boldsymbol{C}^{-1}[\boldsymbol{a}-\boldsymbol{b}]\right\}, \tag{2.5}$$

which is a Gaussian distribution of random vector $\boldsymbol{a}$ having mean $\boldsymbol{b}$ and covariance $\boldsymbol{C}$, the Gaussian distribution is unimodal and symmetric about its mean. Coincidentally, the Gaussian distribution is completely (and succinctly) described by its first and second moments: mean and covariance. By the central limit theorem and law of large numbers, many sequences of random variables converge to (or can be approximated by) the Gaussian distribution, thus making it a common choice when modeling many random variables, such as sensor noise. This common appearance of Gaussian noise is only one reason for its wide use in estimation, another reason being the ease with which Gaussians are mathematically manipulated. For example, Gaussian random variables are known to remain Gaussian under linear transformations, and the product of multiple Gaussians is known to be a scaled Gaussian itself, an identity noted in Eq. (A.6). Traits such as these make closed-form estimators more attainable, as Gaussians tend to remain Gaussian under select transformations. As such, this work frequently relies on Gaussians—and the properties they possess—to derive various filtering architectures. However, as to not relegate these filters to the assumption of Gaussian distributed priors and noise, Gaussian mixture models (GMMs) are used.

2.2.2.2 Gaussian Mixture Models

Gaussian mixture models are widely used in estimation to approximate different types of pdfs. The concept is based on describing a generic pdf, say $p(\boldsymbol{x})$, as a summation of weighted Gaussian

pdfs of the form [44]

$$p(\boldsymbol{x}) = \sum_{\xi=1}^{L} w_\xi p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi)\,, \tag{2.6}$$

where w_ξ, $\boldsymbol{m}_\xi$, and $\boldsymbol{P}_\xi$ correspond to the weight, mean and covariance of the ξ^{th} GM component. Note that, to be considered a valid pdf, Eq. (2.6) must remain positive and integrate to unity over the support of $\boldsymbol{x}$, which is easily enforced via

$$w_\xi \geq 0 \,\forall\, \xi \quad \text{and} \quad \sum_{\xi=1}^{L} w_\xi = 1\,.$$

Though there exist numerous configurations, a closed-form GM filter is effectively any filtering framework where the prior and posterior distributions are both proper GMs in the form of Eq. (2.6). While this restricts the manner by which statistical information is stored and operated on, it allows for a significantly more liberal description of the estimated distribution. For instance, even though the pdfs $p^-(\boldsymbol{x})$ and $p^+(\boldsymbol{x})$ must both be GMs, the weights, means, and covariances can change such that $p^-(\boldsymbol{x})$ may appear nothing like $p^+(\boldsymbol{x})$. In this way, a GM filter is able to operate on complex, multi-modal, asymmetrical distributions while still making use of the handy properties of Gaussians mentioned in Section 2.2.2.1.

2.2.2.3 Generating GMMs using the EM Algorithm

The GMM's ability to flexibly represent a wide variety of pdfs makes it a powerful statistical tool. However, how the GMM takes on the appropriate form is less clear. In some cases, such as when a GMM posterior is formed via Bayes' rule, the GMM posterior is naturally shaped by the Bayesian fusion of the prior and measurement likelihood. However, this implies that a correctly formed GMM posterior is contingent upon both the prior and measurement likelihood being properly shaped as well. Thus, the question still remains as to how to generate a GMM that appropriately represents the statistics it is intended to model.

The solution to this can vary and depends on the system at hand. For instance, if the accuracy

with which the GMM approximates the true statistics is relatively unimportant, then invoking a generic, yet plausible, GMM with large amounts of uncertainty may be sufficient. For example, when initializing a filter, the pdf of the state $p(\boldsymbol{x}_0)$ can take on various configurations without, in general, causing filter failures, as long as the uncertainty is sufficiently large. However, this lack of rigor is inappropriate in other cases, such as when defining measurement distributions. If the measurement distribution is malformed (i.e., the filter measurement model does not reflect reality), even slightly so, filter performance will degrade significantly after enough measurements are ingested. In this regard, some approach to accurately shape a GMM is needed. Luckily, the expectation-maximization (EM) algorithm is well-suited for this purpose.

The EM algorithm is a general approach to finding a maximum-likelihood estimate for a set of latent variables, which is adapted in [45] specifically for GMMs. Given an initial GM of the form of Eq. (2.6) (with time indices removed), the EM algorithm iteratively modifies the GM components using N samples of the target distribution $\boldsymbol{x}_n \sim p(\boldsymbol{x})$, where $n = 1, 2, \ldots, N$. The exact process is given by

$$
\begin{aligned}
\gamma_{n,\xi} &= \frac{w_\xi p_g(\boldsymbol{x}_n|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi)}{\sum_{i=1}^{L} w_i p_g(\boldsymbol{x}_n|\boldsymbol{m}_i, \boldsymbol{P}_i)} \\
N_\xi &= \sum_{n=1}^{N} \gamma_{n,\xi} \\
w_\xi^{\text{new}} &= \frac{N_\xi}{N} \\
\boldsymbol{m}_\xi^{\text{new}} &= \frac{1}{N_\xi} \sum_{n=1}^{N} \gamma_{n,\xi} \boldsymbol{x}_n \\
\boldsymbol{P}_\xi^{\text{new}} &= \frac{1}{N_\xi} \sum_{n=1}^{N} \gamma_{n,\xi} (\boldsymbol{x}_n - \boldsymbol{m}_\xi^{\text{new}})(\boldsymbol{x}_n - \boldsymbol{m}_\xi^{\text{new}})^T ,
\end{aligned}
$$

which is iterated until

$$
\epsilon_{\text{EM}} = \sum_{n=1}^{N} \ln \left\{ \sum_{\xi=1}^{L} w_\xi p_g(\boldsymbol{x}_n|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi) \right\}
$$

has converged to a steady-state value. Note that the GM size L is user-specified, and should be set according to the desired computational complexity of the model. An example of the EM's capabilities is shown in Section 4.8.2.2.2, where an exponential distribution is approximated by a five component GM.

2.2.2.4 Classic example: The Gaussian Sum Filter

First attributed to Alspach and Sorenson in [44], the Gaussian sum filter (GSF) is considered by most to be the first Gaussian mixture filter. The GSF here is specifically outfitted for the same nonlinear Gaussian systems as the EKF, which are given in Eqs. (2.1).

2.2.2.4.1 GSF Propagation

Consider $p^+(\boldsymbol{x}_{k-1})$, which is the posterior pdf of state $\boldsymbol{x}$ at time t_{k-1}. The GSF assumes that the distribution is represented by a GM of L^+_{k-1} components as

$$p^+(\boldsymbol{x}_{k-1}) = \sum_{\xi=1}^{L^+_{k-1}} w^+_{k-1,\xi} p_g(\boldsymbol{x}_{k-1}|\boldsymbol{m}^+_{k-1,\xi}\,,\ \boldsymbol{P}^+_{k-1,\xi})\,, \tag{2.7}$$

where $w^+_{k-1,\xi}$, $\boldsymbol{m}^+_{k-1,\xi}$, and $\boldsymbol{P}^+_{k-1,\xi}$ are the weight, mean, and covariance of the ξ^{th} GM component, respectively. Therefore, if the dynamical system of Eq. (2.1a) is taken, the GSF prediction step is given by [44]

$$L^-_k = L^+_{k-1} \tag{2.8a}$$

$$w^-_{k,\xi} = w^+_{k-1,\xi} \tag{2.8b}$$

$$\boldsymbol{m}^-_{k,\xi} = \boldsymbol{f}_k(\boldsymbol{m}^+_{k-1,\xi}) \tag{2.8c}$$

$$\boldsymbol{P}^-_{k,\xi} = \boldsymbol{F}_k(\boldsymbol{m}^+_{k-1,\xi})\boldsymbol{P}^+_{k-1,\xi}\boldsymbol{F}^T_k(\boldsymbol{m}^+_{k-1,\xi}) + \boldsymbol{Q}_{k-1}\,, \tag{2.8d}$$

such that the predicted pdf of $\boldsymbol{x}$ at time t_k becomes the GM

$$p^-(\boldsymbol{x}_k) = \sum_{\xi=1}^{L_k^-} w_{k,\xi}^- p_g(\boldsymbol{x}_k | \boldsymbol{m}_{k,\xi}^- \,,\, \boldsymbol{P}_{k,\xi}^-) \,, \tag{2.9}$$

where $\boldsymbol{F}_k(\boldsymbol{m})$ is the Jacobian of the nonlinear dynamics function $\boldsymbol{f}_k(\cdot)$ of Eq. (2.1a) evaluated at $\boldsymbol{x} = \boldsymbol{m}$. In the case that the dynamics of the state are in the continuous form of Eq. (2.1b), then the means and covariances of each component in Eq. (2.7) must be integrated to t_k according to Eqs. (2.3d) and (2.3e). Note that, according to Eqs. (2.8a) and (2.8b), the component number and weights remain unchanged during propagation.

2.2.2.4.2 GSF Update

Assuming that the GM predicted pdf $p^-(\boldsymbol{x}_k)$ of Eq. (2.9) is available, the posterior pdf produced by the GSF update is

$$p^+(\boldsymbol{x}_k) = \sum_{\xi=1}^{L_k^+} w_{k,\xi}^+ p_g(\boldsymbol{x}_k | \boldsymbol{m}_{k,\xi}^+ \,,\, \boldsymbol{P}_{k,\xi}^+) \,, \tag{2.10}$$

where

$$L_k^+ = L_k^- \tag{2.11a}$$

$$w_{k,\xi}^+ = \frac{w_{k,\xi}^- \kappa_\xi}{\eta} \tag{2.11b}$$

$$\boldsymbol{m}_{k,\xi}^+ = \boldsymbol{m}_{k,\xi}^- + \boldsymbol{K}_\xi [\boldsymbol{z}_k - \boldsymbol{h}_k(\boldsymbol{m}_{k,\xi}^-)] \tag{2.11c}$$

$$\boldsymbol{P}_{k,\xi}^+ = \boldsymbol{P}_{k,\xi}^- - \boldsymbol{K}_\xi \boldsymbol{H}_k(\boldsymbol{m}_{k,\xi}^-) \boldsymbol{P}_{k,\xi}^- \,, \tag{2.11d}$$

and where

$$\eta = \sum_{\xi=1}^{L_k^+} w_{k,\xi}^- \kappa_\xi \tag{2.12a}$$

$$\kappa_\xi = p_g(\boldsymbol{z}_k | \boldsymbol{h}_k(\boldsymbol{m}_{k,\xi}^-), \boldsymbol{W}_\xi) \tag{2.12b}$$

$$\boldsymbol{K}_\xi = \boldsymbol{P}_{k,\xi}^- \boldsymbol{H}_k^T(\boldsymbol{m}_{k,\xi}^-) \boldsymbol{W}_\xi^{-1} \tag{2.12c}$$

$$\boldsymbol{W}_\xi = \boldsymbol{H}_k(\boldsymbol{m}_{k,\xi}^-) \boldsymbol{P}_{k,\xi}^- \boldsymbol{H}_k^T(\boldsymbol{m}_{k,\xi}^-) + \boldsymbol{R}_k \,. \tag{2.12d}$$

The matrix $\boldsymbol{H}_k(\boldsymbol{m})$ is the Jacobian of the nonlinear observation function $\boldsymbol{h}_k(\cdot)$ of Eq. (2.1a) evaluated at $\boldsymbol{x} = \boldsymbol{m}$, and since all elements of Eq. (2.12b) are known, κ_ξ is an evaluated likelihood, not a pdf. Note that Eqs. (2.8a) and (2.11a) make it clear that the number of GM components remains unchanged during each filter iteration, such that any increase or decrease to the number of components must be done by some additional filtering mechanism.

GSF vs. GMEKF

While originally named the Gaussian sum filter (GSF) by Sorenson and Alspach [46], these algorithms are also frequently referred to as the Gaussian mixture extended Kalman filter (GMEKF) as they closely resemble a collection of EKFs due to both using first-order Taylor series linearizations about the means. However, the GSF/GMEKF of Section 2.2.2.4 is derived directly from Bayes' rule under Gaussian assumptions, not from the EKF, which makes "GMEKF" a bit of a misnomer. A key difference between the GSF/GMEKF and a collection of EKFs is that the GSF/GMEKF inherently performs a type of "hypothesis testing" when calculating the posterior weights of Eq. (2.11b), something that the EKF cannot account for on its own.

2.2.3 GMM Component Management Schemes

Generally, a higher number of GM components leads to improved filtering accuracy at the cost of computational complexity. This improvement in accuracy is a result of not only increasing the

number of linearization points in the pdf, but from the fact that it becomes more appropriate to neglect higher order moments/effects for symmetric, concentrated (i.e. small variance) distributions [9]. As a GMM acquires more components, the variances of each component will generally decrease, leading to less linearization error throughout filtering procedures [47]. It is therefore important to monitor the number of components, seeking a compromise between filtering accuracy and computational efficiency—a task accomplished by component management schemes such as pruning, merging, and splitting.

In practice, some GM filters, such as the GM probability hypothesis density (GMPHD) filter of [41], naturally produce additional GM components during propagation/update, such that the number of components increases with each iteration. As such, it is often good practice to introduce component management schemes, such as pruning and merging, to limit the number of GM components generated. Additionally, depending on the system at hand, it may be necessary to implement GM splitting, where a single Gaussian component is separated into multiple Gaussian components, which is often used to reduce linearization errors.

2.2.3.1 Pruning

The process of pruning a GM is typically straightforward and can be accomplished in several ways, but all methods involve eliminating the components of Eq. (2.6) belonging to the lowest weights [48]. In this work, it is desired to omit any components with negligible associated probabilities, and thus a weight threshold is defined as $0 < w_{\text{thresh}} \ll 1$, such that any ξ^{th} GM component where $w_\xi < w_{\text{thresh}}$ is eliminated from the GM of Eq. (2.6). Therefore, noting that any ξ^{th} component having weight $w_\xi \geq w_{\text{thresh}}$ belongs to the set of non-pruned components $\mathcal{D}$, the pruned posterior GM is expressed as

$$p^p(\boldsymbol{x}) = \sum_{\xi \in \mathcal{D}} w_\xi^p p_g(\boldsymbol{x}|\boldsymbol{m}_\xi,\ \boldsymbol{P}_\xi)\,, \qquad \text{with} \qquad w_\xi^p = \frac{w_\xi}{\sum\limits_{\xi \in \mathcal{D}} w_\xi} \quad \forall\, \xi \in \mathcal{D}\,, \tag{2.13}$$

where the weights w_ξ^p are normalized to ensure Eq. (2.13) remains a valid GM.

2.2.3.2 *Merging*

The precise manner in which GM components are merged is open to considerable interpretation, leading to many different possible merging frameworks [49, 50]. The unifying focus behind these different methods is the compression of a GM with L components into a GM of size L^m—where $L^m < L$—by combining Gaussian components. Typically, these processes attempt to merge the components deemed statistically similar. The method used in this work is from [48] and is included for its conceptual simplicity relative to other, more involved, methods. The entire merging process can be explained in two stages: a similarity search and the compression of the GM.

First, a selection criteria is established that segregates GM components into groups to be merged. This is accomplished via the squared Mahalanobis distance between pairs of GM components from Eq. (2.6), which is calculated as

$$d_{M_{ij}} = (\boldsymbol{m}_i - \boldsymbol{m}_j)^T (\boldsymbol{P}_i + \boldsymbol{P}_j)^{-1} (\boldsymbol{m}_i - \boldsymbol{m}_j)\,, \tag{2.14}$$

and a similarity threshold d_{thresh} that is appointed such that any two GM components i and j having $d_{M_{ij}} < d_{\text{thresh}}$ are designated "similar enough" to be merged. After this similarity search is performed over the entire GM of Eq. (2.6), components deemed similar enough to merge are assigned to the same set $\mathcal{C}_\ell$, where $\ell = 1, \ldots, L^m$ and $L^m \leq L$.

The compression of the GM involves the direct calculation of the merged GM components. Through the similarity search, it is determined that the GM of Eq. (2.6) can be expressed as

$$p^m(\boldsymbol{x}) = \sum_{\ell=1}^{L^m} w_\ell^m p_g(\boldsymbol{x} | \boldsymbol{m}_\ell^m,\, \boldsymbol{P}_\ell^m)\,, \tag{2.15}$$

where each ℓ^{th} GM component of Eq. (2.15) is the merged product of the components belonging

to set $\mathcal{C}_\ell$. The components of Eq. (2.15) are thus calculated as

$$
\begin{aligned}
w_\ell^m &= \sum_{i \in \mathcal{C}_\ell} w_i \\
\boldsymbol{m}_\ell^m &= \sum_{i \in \mathcal{C}_\ell} \frac{w_i}{w_\ell^m} \boldsymbol{m}_i \\
\boldsymbol{P}_\ell^m &= \sum_{i \in \mathcal{C}_\ell} \frac{w_i}{w_\ell^m} \Big(\boldsymbol{P}_i + (\boldsymbol{m}_i - \boldsymbol{m}_\ell^m)(\boldsymbol{m}_i - \boldsymbol{m}_\ell^m)^T \Big) .
\end{aligned} \tag{2.16}
$$

Note that it is advised to execute any pruning procedures before merging the GM, as this avoids merging computations that are made irrelevant when the components are pruned [48].

2.2.3.3 Splitting

Much in the way resampling procedures are common practice for particle filtering methods, GM filters often benefit from the ability to generate additional GM components in order to improve filtering operations. Increasing the number of components not only increases the number of linearization points, but also decreases the overall linearization error accrued throughout the filtering procedures [47]. Many methods exist for splitting Gaussians; some operate through precomputed splitting libraries [51], while others utilize optimization methods [50]. Selecting an appropriate splitting method can be difficult, as it typically depends on the system at hand, as well as a desired balance between estimation accuracy and computational efficiency. A best practice is to define a automated splitting process, such as in [52], which seeks to split the GM along directions of the highest relevant nonlinearity.

The GM splitting procedure elected by this work utilizes the aforementioned splitting libraries of [53], where each library contains precomputed weights $\tilde{w}_\ell$, means $\tilde{m}_\ell$, and variances $\tilde{\sigma}^2$ that are

formed via the optimization

$$\sum_{\ell=1}^{L^s_{\text{dim}}} \tilde{w}_\ell \tilde{p}_g(x|\tilde{m}_\ell, \tilde{\sigma}^2) = \min_{\tilde{p}_g} \left\{ \mathcal{D}_{\beta=2}\big[p_g(x|0,1) \big\| \sum_{\ell=1}^{L^s_{\text{dim}}} \tilde{w}_\ell \tilde{p}_g(x|\tilde{m}_\ell, \tilde{\sigma}^2)\big] + \lambda \tilde{\sigma}^2 \right\} \tag{2.17}$$

$$\text{subject to } \sum_{\ell=1}^{L^s_{\text{dim}}} \tilde{w}_\ell = 1\,,$$

where $\mathcal{D}_{\beta=2}[\cdot\|\cdot]$ is the β divergence with parameter $\beta = 2$. The weighting term λ effectively adjusts the importance that the variance has within the cost function; increasing λ will decrease the size of variances $\tilde{\sigma}^2$. Note that the optimization of Eq. (2.17) corresponds to the splitting of a univariate Gaussian, but is used to split multivariate Gaussians as well, one dimension at a time. Splitting along the k^{th} dimension of a n dimensional multivariate Gaussian produces

$$w_\xi p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi) \approx \sum_{\ell=1}^{L^s_{\text{dim}}} w^s_{\xi,\ell} p_g(\boldsymbol{x}|\boldsymbol{m}^s_{\xi,\ell},\, \boldsymbol{P}^s_{\xi,\ell})\,, \tag{2.18}$$

where

$$w^s_{\xi,\ell} = \tilde{w}_\ell w_\xi \qquad\qquad \boldsymbol{m}^s_{\xi,\ell} = \tilde{m}_\ell \boldsymbol{m}_\xi \qquad\qquad \boldsymbol{P}^s_{\xi,\ell} = \tilde{\boldsymbol{S}}_\xi \tilde{\boldsymbol{S}}^T_\xi\,, \tag{2.19}$$

and where $\tilde{\boldsymbol{S}}_\xi$ is the square-root factor formed from the ξ^{th} covariance as

$$\tilde{\boldsymbol{S}}_\xi = \begin{bmatrix} \boldsymbol{s}^1_\xi & \boldsymbol{s}^2_\xi & \cdots & \tilde{\sigma}\boldsymbol{s}^k_\xi & \cdots & \boldsymbol{s}^n_\xi \end{bmatrix}.$$

where $\boldsymbol{s}^k_\xi$ corresponds to the k^{th} row of the square-root factor $\boldsymbol{P}_\xi = \boldsymbol{S}_\xi \boldsymbol{S}^T_\xi$. To split the GM component of Eq. (2.18) across multiple dimensions, the procedure of Eqs. (2.18) and (2.19) can be applied consecutively.

A straightforward, albeit naive, approach to splitting is taken in this work. To ensure a sufficiently large number of components exists within the nonlinear filter, splitting is enforced any time the number of components becomes too few according to a user specified threshold L^s_{thresh}. For ex-

ample, consider the case where the GM of Eq. (2.6) is selected to be split across the k^{th} dimension because $L < L^s_{\text{thresh}}$. Each Gaussian component $w_\xi p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi)$ would be split into a GM of L^s_{dim} components such that

$$\sum_{\xi=1}^{L} w_\xi p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi) \approx \sum_{\xi=1}^{L} \sum_{\ell=1}^{L^s_{\text{dim}}} w^s_{\xi,\ell} p_g(\boldsymbol{x}|\boldsymbol{m}^s_{\xi,\ell},\ \boldsymbol{P}^s_{\xi,\ell})\,, \tag{2.20}$$

according to Eq. (2.19). In this work, it is usually elected to apply splitting across every dimension of the state $\boldsymbol{x} \in \mathbb{R}^n$, meaning that a newly split GM pdf $p^s(\boldsymbol{x})$ would be of the form

$$p^s(\boldsymbol{x}) = \sum_{\xi=1}^{L^s} w^s_\xi p_g(\boldsymbol{x}|\boldsymbol{m}^s_\xi,\ \boldsymbol{P}^s_\xi)\,, \tag{2.21}$$

where $L^s = L \times (L^s_{\text{dim}})^n$. More nuanced approaches to splitting GMs certainly exist [51, 52], and are advisable for practical applications, but as much of this work is not focused on the development and testing of splitting procedures, this naive approach is deemed appropriate.

2.2.3.4 Estimate Extraction

Recalling the discussion of Section 1.1.3, while it is most fitting to perform estimation over the entire pdf of the state, other subsystems (i.e. guidance and control) often require a single point estimate to be reported. Again, as illustrated by Eq. (1.7), there are various ways to extract this estimate depending on how one defines the risk function $\mathcal{R}(\cdot,\cdot)$. Unfortunately, few risk functions produce both useful and computationally practical point estimates. If $\mathcal{R}(\cdot,\cdot)$ takes on the form of Eq. (1.8), however, it is shown in Section B.1 that the point estimate of the state is simply the mean, which for the GM pdf of Eq. (2.6) is calculated as

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] \\
&= \int \boldsymbol{x} \sum_{\xi=1}^{L} w_\xi p_g(\boldsymbol{x}; \boldsymbol{m}_\xi, \boldsymbol{P}_\xi) \mathrm{d}\boldsymbol{x} \\
&= \sum_{\xi=1}^{L} w_\xi \int \boldsymbol{x} p_g(\boldsymbol{x}; \boldsymbol{m}_\xi, \boldsymbol{P}_\xi) \mathrm{d}\boldsymbol{x}
\end{aligned}$$

$$= \sum_{\xi=1}^{L} w_\xi \boldsymbol{m}_\xi \,, \tag{2.22}$$

which is not only statistically meaningful (all puns aside), but also has an efficient closed-form solution.

Though the point estimate of Eq. (2.22) is useful, there are times when a concise measure of uncertainty is required as well. For example, linear estimators generally produce the mean and covariance of the estimated state naturally, making both quantities familiar and regularly utilized by those in the field of estimation. These quantities are not readily available to GM filters, however, as evidenced by Eq. (2.22) where the overall mean of the GM pdf requires extra calculation. Similarly, the overall covariance, while convenient for compact uncertainty quantification, must be calculated as well. Therefore, noting that the definition of covariance is

$$\boldsymbol{P} = \mathbb{E}_{p(\boldsymbol{x})}\left[(\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}])^T\right], \tag{2.23a}$$

the definition of the mean is recalled from Eq. (2.22) such that Eq. (2.23a) becomes

$$\begin{aligned}
\hat{\boldsymbol{P}} &= \mathbb{E}_{p(\boldsymbol{x})}\left[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^T\right] \\
&= \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}\boldsymbol{x}^T] - 2\hat{\boldsymbol{x}}\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}^T] + \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T \\
&= \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}\boldsymbol{x}^T] - \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T \,,
\end{aligned} \tag{2.23b}$$

where the notation $\hat{\boldsymbol{P}}$ signifies that Eq. (2.23b) is the estimated covariance. Expanding the expec-

tation of Eq. (2.23b) with the GM pdf of Eq. (2.6) yields

$$\begin{aligned}\hat{\boldsymbol{P}} &= \int \boldsymbol{x}\boldsymbol{x}^T \sum_{\xi=1}^{L} w_\xi p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi)\mathrm{d}\boldsymbol{x} - \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T \\ &= \sum_{\xi=1}^{L} \left\{ w_\xi \int \boldsymbol{x}\boldsymbol{x}^T p_g(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi)\mathrm{d}\boldsymbol{x} \right\} - \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T \\ &= \sum_{\xi=1}^{L} \left\{ w_\xi \mathbb{E}_{p_{g,\xi}(\boldsymbol{x})}[\boldsymbol{x}\boldsymbol{x}^T] \right\} - \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T , \end{aligned} \tag{2.23c}$$

where $\mathbb{E}_{p_{g,\xi}(\boldsymbol{x})}[\cdot]$ is the expectation with respect to the ξ^{th} Gaussian pdf. Noting that, similar to Eq. (2.23b), the covariance of the ξ^{th} GM component is given by

$$\boldsymbol{P}_\xi = \mathbb{E}_{p_{g,\xi}(\boldsymbol{x})}[\boldsymbol{x}\boldsymbol{x}^T] - \boldsymbol{m}_\xi \boldsymbol{m}_\xi^T ,$$

Eq. (2.23c) can be shown to be

$$\hat{\boldsymbol{P}} = \sum_{\xi=1}^{L} \left\{ w_\xi \big(\boldsymbol{P}_\xi + \boldsymbol{m}_\xi \boldsymbol{m}_\xi^T \big) \right\} - \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T . \tag{2.23d}$$

Fortunately, much like the mean of Eq. (2.22), Eq. (2.23d) demonstrates that the covariance of a GM pdf has a closed-form calculation. After straightforward manipulation, Eq. (2.23d) can be equivalently expressed as

$$\hat{\boldsymbol{P}} = \sum_{\xi=1}^{L} w_\xi \big(\boldsymbol{P}_\xi + (\boldsymbol{m}_\xi - \hat{\boldsymbol{x}})(\boldsymbol{m}_\xi - \hat{\boldsymbol{x}})^T \big) , \tag{2.24}$$

which is particularly useful for its computational efficiency.

2.3 Linear vs. Nonlinear: GSF and EKF

The differences between linear and nonlinear estimators are detailed in Sections 2.1 and 2.2, with nonlinear filters boasting several advantages over their linear counterparts. This section further illustrates these advantages by directly comparing the GSF to its linear equivalent, the EKF, via

two Monte Carlo simulations, the architecture of which is described in Section 2.3.1 and showcased via the first simulation—a simple two dimensional Keplerian system. The second simulation is one revisited throughout the dissertation, wherein a falling body experiencing unidimensional motion is tracked and is used to showcase the GSF's superior treatment of nonlinearities. Additionally, the procedures of pruning, merging, splitting, and estimate extraction described in Section 2.2.3 are also implemented such that they can be verified simultaneously.

2.3.1 Monte Carlo Analysis

The Monte Carlo (MC) method is widely regarded as one of the best probabilistic analysis techniques, dating back to the 1940s [54, 55]. It is used to empirically quantify the uncertainties in a system by way of simulating a multitude of Monte Carlo trials, where each trial is the outcome of a single system simulation that is unique from all other trials by way of sampling the random variables inherent to the system models. As it is so well-known and useful, there are several sections in this work that use the MC technique to analyze filtering performance, the exact manner by which is described here.

To initialize the i^{th} MC trial, an initial state $\boldsymbol{x}_0^i$ is drawn from the initial distribution $p(\boldsymbol{x}_0)$, which is then propagated through the duration of the trial according to some dynamic model (see Section 1.1.1), generating noisy measurements from the simulated truth $\boldsymbol{x}_k^i$ at each time step t_k according to some measurement model (see Section 4.1). At each of these time steps, the average filter error standard deviations $\sigma_{\text{filt},\iota}$ and Monte Carlo error standard deviations $\sigma_{\text{MC},\iota}$ of the ι^{th} state of $\boldsymbol{x}$ are calculated as

$$\sigma_{\text{filt},\iota} = \sqrt{\frac{1}{N_{\text{MC}}}\sum_{i=1}^{N_{\text{MC}}} \hat{\boldsymbol{P}}_k^i(\iota,\iota)} \qquad \text{and} \qquad \sigma_{\text{MC},\iota} = \sqrt{\frac{1}{N_{\text{MC}}}\sum_{i=1}^{N_{\text{MC}}} \left(e_\iota^i - \overline{e}_\iota\right)^2}, \tag{2.25}$$

where N_{MC} is the number of MC trials, and where e_ι^i is the ι^{th} element of the estimation error vector for the i^{th} MC trial defined as

$$\boldsymbol{e}_k^i = \boldsymbol{x}_k^i - \hat{\boldsymbol{x}}_k^i\,,$$

with the average filter error at time t_k calculated as

$$\overline{e}_\iota = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} e_\iota^i \,. \tag{2.26}$$

Note that $\hat{\boldsymbol{x}}_k^i$ and $\hat{\boldsymbol{P}}_k^i$ are the extracted state and covariance estimate, respectively, of the i^{th} trial at t_k, which for GM filters is calculated using Eqs. (2.22) and (2.24), and that $\hat{\boldsymbol{P}}_k^i(\iota, \iota)$ denotes the ι^{th} diagonal of $\hat{\boldsymbol{P}}_k^i$.

Generally, a well-behaved filter yields $\sigma_{\text{filt},\iota} \approx \sigma_{\text{MC},\iota}$, where $\sigma_{\text{filt},\iota} > \sigma_{\text{MC},\iota}$ indicates underconfident (or conservative) estimation, while $\sigma_{\text{filt},\iota} < \sigma_{\text{MC},\iota}$ signifies overconfident (or smug) filter performance. In practice, a conservative filter is customarily more desirable than one that is overconfident, as overconfidence often leads to issues such as filter divergence [56].

In addition to the metrics given in Eqs. (2.25) and (2.26), the behavior of propagated states $\boldsymbol{x}_k^i$ can describe much about the nature of a system. For example, consider a system governed by point-mass gravitational dynamics, given by

$$\dot{\boldsymbol{r}} = \boldsymbol{v} \tag{2.27a}$$

$$\dot{\boldsymbol{v}} = -\frac{\mu}{\|\boldsymbol{r}\|^3} \boldsymbol{r} \,, \tag{2.27b}$$

where $\boldsymbol{r} = [r_x \; r_y]^T$, $\boldsymbol{v} = [v_x \; v_y]^T$, and μ is the gravitational parameter of Earth. The state vector $\boldsymbol{x} = [\boldsymbol{r} \; \boldsymbol{v}]^T$ is taken to have the initial Gaussian distribution

$$\boldsymbol{x}_0 \sim p_g(\boldsymbol{x}_0 | \boldsymbol{m}_0, \, \boldsymbol{P}_0) \,, \tag{2.28}$$

where

$$\boldsymbol{m}_0 = \begin{bmatrix} 14600\,\text{km} \\ -3000\,\text{km} \\ 3\,\text{km/s} \\ 1\,\text{km/s} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P}_0 = \begin{bmatrix} 250000\,\text{km}^2 & 0 & 0 & 0 \\ 0 & 250000\,\text{km}^2 & 0 & 0 \\ 0 & 0 & 0.01\,(\text{km/s})^2 & 0 \\ 0 & 0 & 0 & 0.01\,(\text{km/s})^2 \end{bmatrix}.$$

500 MC samples are then drawn from Eq. (2.28), the positions of which are plotted are in Fig. (2.1) along with the 1-, 2-, and 3-σ intervals of $p(\boldsymbol{x}_0)$.

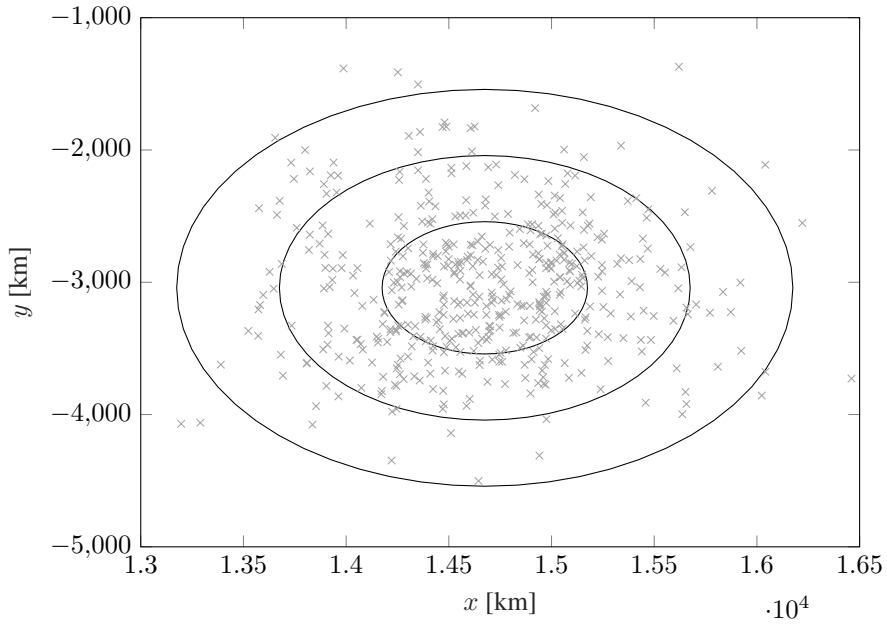


Figure 2.1 500 Monte Carlo samples drawn from $p(\boldsymbol{x}_0)$

The MC samples are then propagated via Eq. (2.27) for 2 hours, the collection of which are used to compare the performance of a GSF to that of an EKF. To do this, the initial distribution of Eq. (2.28) is propagated according to Eqs. (2.3) for an EKF, the contours of which are plotted over the propagated MC samples in Fig. (2.2).

Due to the obvious non-Gaussian nature of the propagated MC samples, the mean and covariance of the EKF is unable to properly describe the actual distribution of the propagated sam-

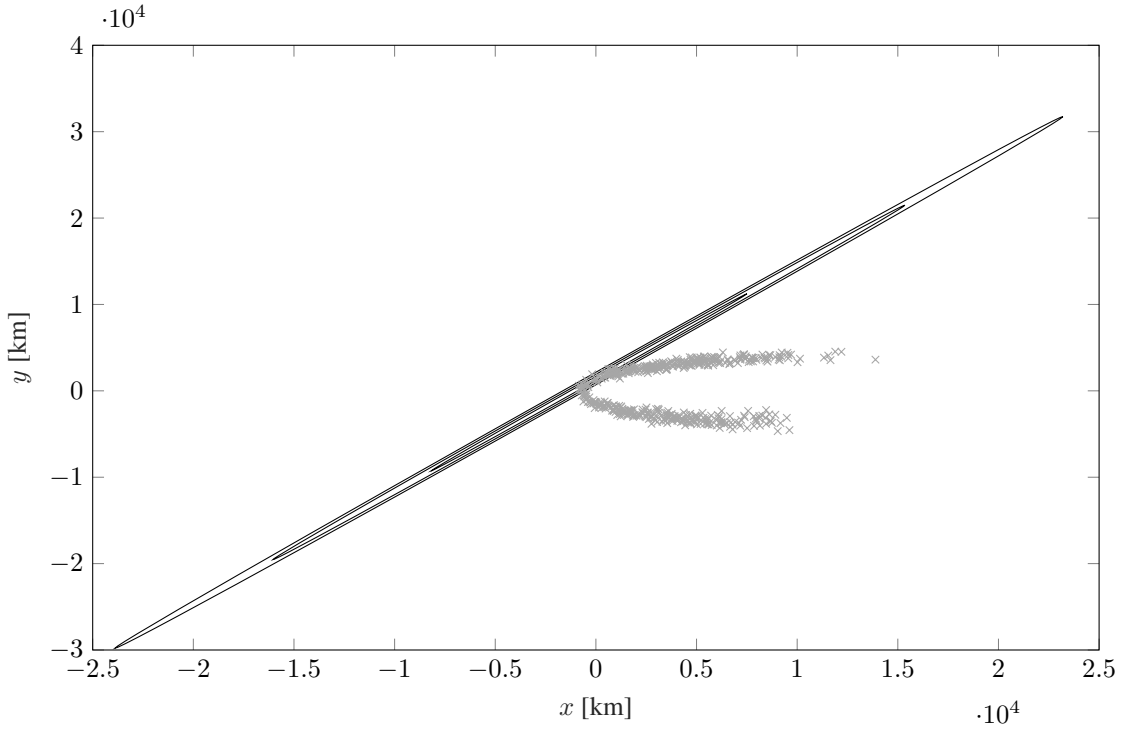


Figure 2.2 Propagated Monte Carlo samples against contours of EKF estimate

ples. On the other hand, if a GSF is used—which propagates the initial distribution pursuant to Eqs. (2.8) and (2.9) after splitting it into 625 components via a five component splitting library across all four dimensions of the state according to Section 2.2.3.3—the resulting estimated distribution matches the propagated MC samples much more appropriately, which is seen in Fig. (2.3). This is just one example of the benefits to using nonlinear filters.

2.3.2 Falling Body Simulation

In this simulation, a system similar to that found in [57] is used wherein a stationary, ground-based observer obtains nonlinear range measurements of an object falling freely through the atmosphere, a diagram of which is provided in Fig. (2.4). This system is selected for several reasons, the first being that simulation and analysis is relatively straightforward for low dimensional systems. Higher dimensional systems, while more complex, can make the interpretation of results more tedious to convey, diluting the apparent importance of any results. Secondly, since this dissertation covers a wide variety of filters, it is useful to maintain a familiar simulation throughout,

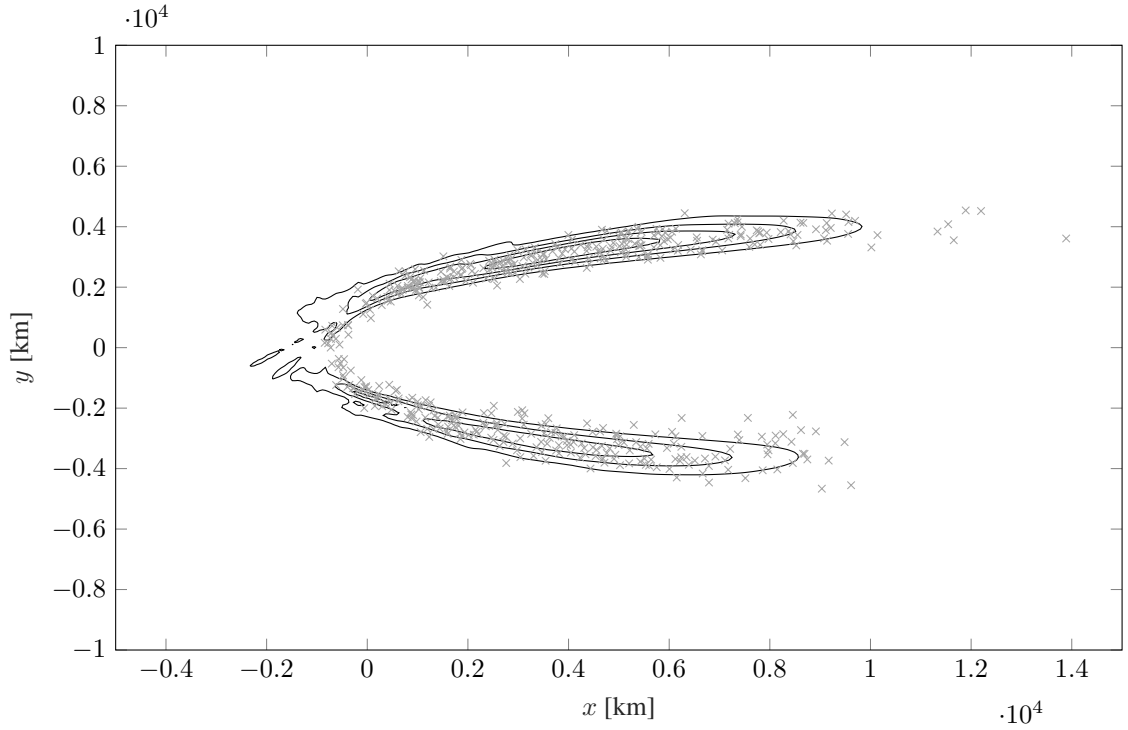


Figure 2.3 Propagated Monte Carlo samples against contours of GSF estimate

as this helps highlight the differences between the filters within a common context. As such, the system of Fig. (2.4) is revisited throughout this dissertation, with minor changes being made when appropriate.

True to Fig. (2.4), it is assumed that the object is falling straight down, such that its motion is one-dimensional, and the position (y) and velocity ($\dot{y}$) of the object relative to the ground are estimated, as well as the object's constant ballistic coefficient (β_c), such that the state vector takes the form

$$\boldsymbol{x}_k = \begin{bmatrix} y & \dot{y} & \beta_c \end{bmatrix}^T . \tag{2.29}$$

It is further assumed that the dynamic model is not affected by process noise, and the effects of drag are taken into account using a simple model given by

$$\ddot{y} = \frac{\rho_0 \mathrm{e}^{-\frac{y}{k_\rho}} \dot{y}^2}{2\beta_c} - g \, , \tag{2.30}$$

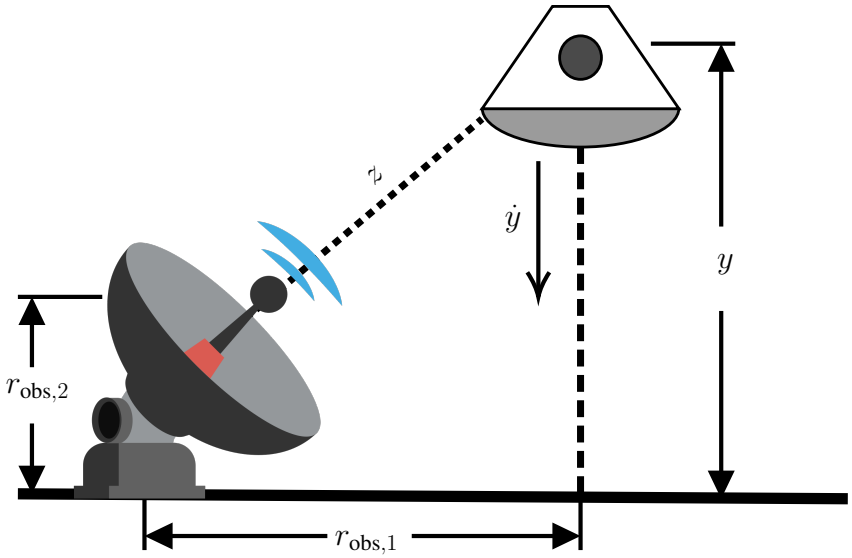


Figure 2.4 Falling body tracked via one-dimensional, nonlinear range measurement

where $g = 32.2^{\text{ft}/\text{s}^2}$ is the acceleration due to gravity, $\rho_0 = 3.106 \times 10^{-3\,\text{lb}/\text{ft}^3}$ is the atmospheric density, and $k_\rho = 22,000$ ft is a decay constant. The initial state is drawn from a multivariate Gaussian distribution as

$$\boldsymbol{x}_0 \sim p_g(\boldsymbol{x}_0 | \boldsymbol{m}_0, \boldsymbol{P}_0) \,, \tag{2.31}$$

where

$$\boldsymbol{m}_0 = \begin{bmatrix} 100000\,\text{ft} \\ -6000\,^{\text{ft}}/_{\text{s}} \\ 2000\,^{\text{lb}}/_{\text{ft}^2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P}_0 = \begin{bmatrix} 500\,\text{ft}^2 & 0 & 0 \\ 0 & 20000\,(^{\text{ft}}/_{\text{s}})^2 & 0 \\ 0 & 0 & 250000\,(^{\text{lb}}/_{\text{ft}^2})^2 \end{bmatrix} . \tag{2.32}$$

Measurements are generated every second using the model given by Eq. (2.1), where $\boldsymbol{h}_k(\cdot)$ is given by

$$h_k(\boldsymbol{x}_k) = \sqrt{r_{\text{obs},1}^2 + (y - r_{\text{obs},2})^2} \,, \tag{2.33}$$

where $r_{\text{obs},1} = 500$ ft and $r_{\text{obs},2} = 250$ ft describe the static position of the observer. The measure-

ment noise covariance is taken to be $R_k = 100$ ft^2. The duration of the simulation is 30 seconds, and the Monte Carlo analysis consists of 500 trials.

2.3.2.1 Filter Configuration

Two different filters, an EKF and a GSF, are both implemented within this simulation to estimate the state vector of Eq. (2.29). The EKF is assembled in its standard Bayesian form given in Eqs. (2.2)–(2.4). The GSF takes on the form described in Section 2.2.2.4, but with the component management methods of pruning, merging, and splitting applied after the update of Eqs. (2.10)–(2.12) for each iteration of the filter. Of the three procedures, pruning is implemented first according to Section 2.2.3.1 with a threshold value of $w_{\text{thresh}} = 1 \times 10^{-6}$, followed by the merging of Section 2.2.3.2 using a threshold value of $d_{\text{thresh}} = 0.25$. Splitting is applied last—according to Section 2.2.3.3—across all three dimensions of $\boldsymbol{x}$ using a splitting library of three components ($L^s_{\text{dim}} = 3$) with a threshold value of $L^s_{\text{thresh}} = 20$. Note that this method of splitting is also used upon the initial prior of Eq. (2.31) for the GSF, with the exception that a five component splitting library is used, such that it is initialized with a GM of 125 components. At each time step, the estimate of the EKF is simply taken as the posterior mean and covariance from Eq. (2.4), while the GSF estimate is extracted from the posterior GM according to Eqs. (2.22) and (2.24).

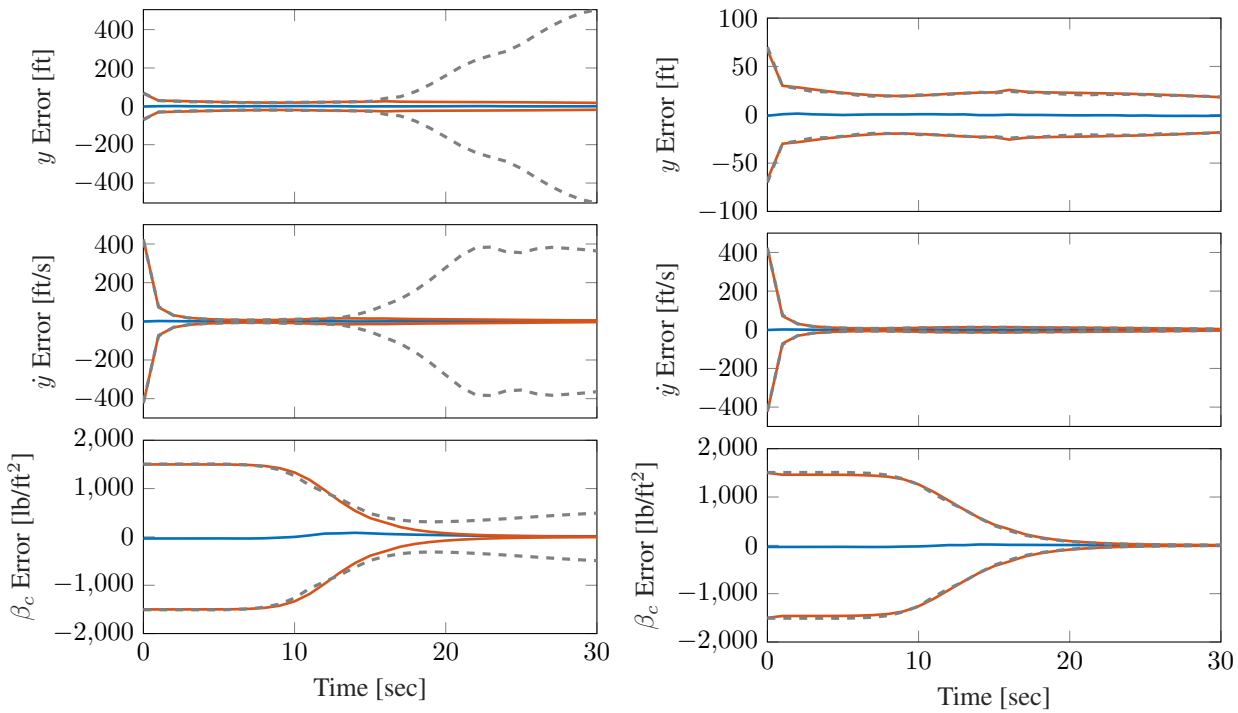
Many times, especially with high dimensional states common to spacecraft navigation, numerical propagation of the filter estimate becomes a computationally expensive process. Since real-time state estimation is typically necessary for GNC, integration fidelity of the dynamic state may need to be lowered to achieve sufficient computational speed. Such a scenario is mimicked here, where all filter propagation is performed using the Matlab function `ode45`, and the relative and absolute integration tolerances are modified to increase propagation speed at the cost of precision. The intended effect is to degrade the filtering propagation to the point that correct estimation becomes difficult, especially for the EKF.

2.3.2.2 *Monte Carlo Results*

The main results of the falling body Monte Carlo analysis are contained in Fig. (2.5), with a zoomed version provided by Fig. (2.6) for an increased level of detail, where the plotted quantities are calculated from Eqs. (2.25) and (2.26). While the two filters perform remarkably similar for the first 12 seconds, it is immediately clear that the GSF outperforms the EKF towards the end of the simulation, which, to be clear, is an expected outcome. Around the 12 second mark, the MC uncertainties begin to grow larger than those of the average filter ($\sigma_{\text{MC},\iota} > \sigma_{\text{filt},\iota}$) for the EKF, a trend that is especially noticeable in the velocity channel of Fig. (2.6a). This indicates that the EKF is producing overconfident estimates, which is caused by a number of factors. As previously mentioned, the integration tolerances are specifically selected to decrease the accuracy of state estimates. This is coupled with the higher linearization errors of the EKF and the relatively low observability of the ballistic coefficient—a characteristic visible in Fig. (2.5), where the uncertainty of the β_c estimate doesn't decrease until about 10 seconds. Figure (2.6a) shows a relatively large average error in the EKF's ballistic coefficient, which causes the propagation of Eq. (2.30) to further degrade and causes the EKF to diverge.

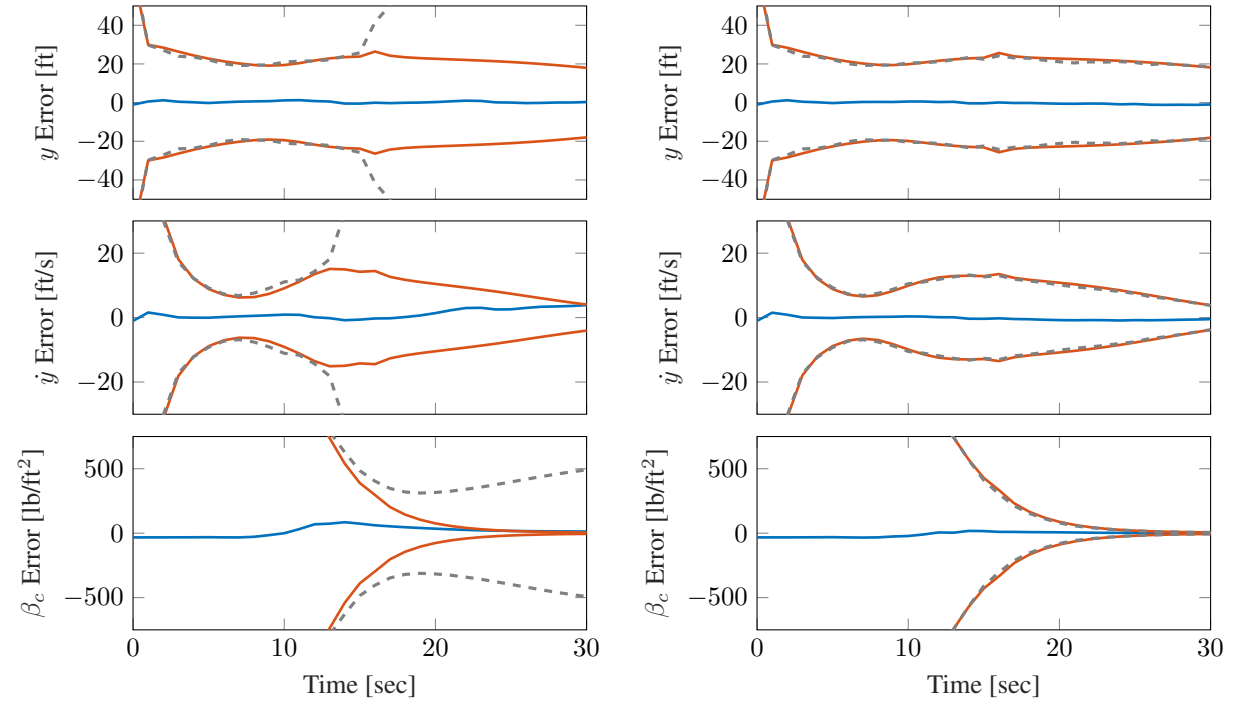
Meanwhile, the GSF results of Figs. (2.5b) and (2.6b) show much healthier performance, with $\sigma_{\text{MC},\iota} \approx \sigma_{\text{filt},\iota}$ throughout the entire simulation–a benefit of using a nonlinear filter. Figure (2.6b) shows how the GSF is more capable of estimating the ballistic coefficient, completely avoiding filter divergence altogether.

Beyond the average errors and uncertainties, the GM component behavior of the GSF should be discussed. In an unadorned GSF (in which noises are assumed to be single Gaussians), where no component management techniques are employed, the number of components remains unchanged, as previously mentioned in Section 2.2.2.4.2. However, as this simulation has induced pruning, merging, and splitting, the number of GM components changes significantly throughout the trials. Figure (2.7) contains the average number of GM components at each time step, as well as the average number of components removed by pruning/merging and created by splitting. Upon the acquisition of the first measurement ($t_1 = 1\text{sec}$), it is seen that a large number of components



a) Extended Kalman Filter **b) Gaussian Sum Filter**

Figure 2.5 Monte Carlo results between standard EKF and GSF, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)



a) Extended Kalman Filter **b) Gaussian Sum Filter**

Figure 2.6 Zoomed view of Monte Carlo results between standard EKF and GSF, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

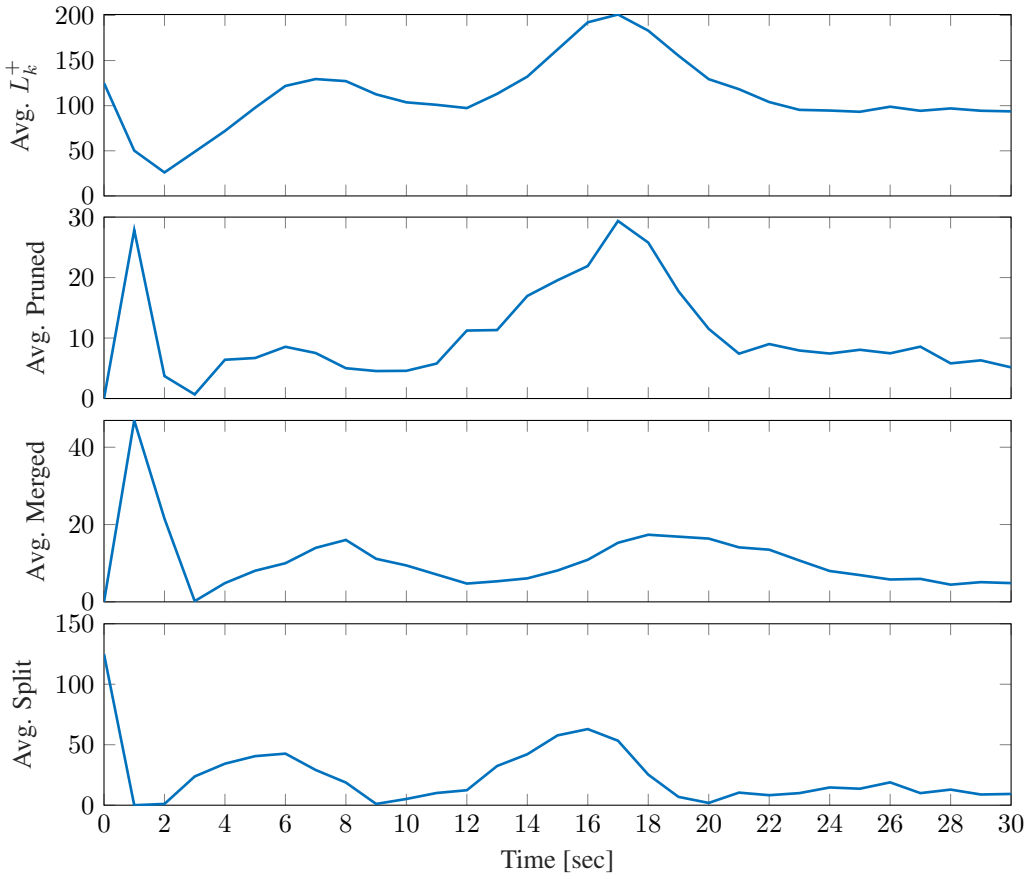


Figure 2.7 Average component behavior across all Monte Carlo trials

are merged and pruned. The cause behind this is clear: a large amount of information is gained from the first measurement relative to the amount of uncertainty present in the initial estimate. As such, the GSF update quickly down-weights GM components that exist in highly-unlikely regimes, which are removed via pruning. At the same time, the GSF also mean-shifts the GM components to regions of higher-probability (according to the measurement), which tends to decrease the relative squared Mahalanobis distance between the components; thus, many components become candidates for merging. In general, this trend of increased pruning and merging activity can be expected during periods of large information gain. Once enough components are removed via pruning and merging such that L_k^+ falls below L_{thresh}^s, the existing components are split to improve the statistical spread of the GM components and decrease linearization errors.

To further illustrate the capabilities of the GSF relative to the standard EKF, 500 samples are

drawn from the initial distributions of Eq. (2.31) and are propagated for 75 seconds according to the nonlinear dynamics of Eq. (2.30). The initial distribution $p(\boldsymbol{x}_0)$ is then propagated to 75 seconds via both filters, where splitting has been applied only at the initial time for the GSF. The contours of the propagated distributions are plotted in Fig. (2.8) for the EKF and Fig. (2.9) for the GSF. It is clear that after 75 seconds the sampled points have lost their Gaussianity, especially between the y and β_c states, such that the mean and covariance of the EKF is no longer sufficient in describing the system. The GSF, however, produces an overall pdf that fits the samples nicely, as the contours of Fig. (2.9) denote a good level of adaptation to the nonlinear effects of the dynamic model.

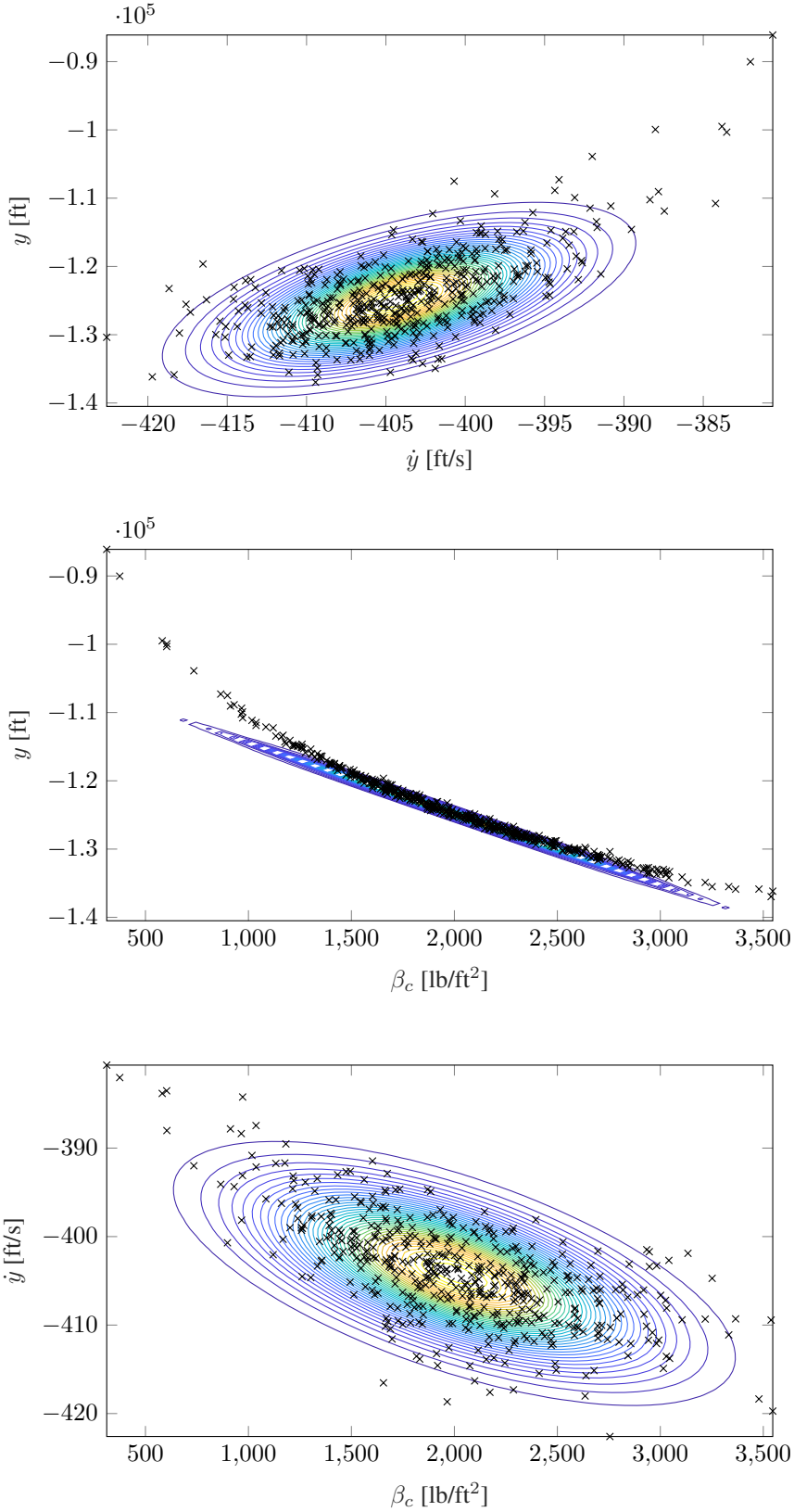


Figure 2.8 Contours of propagated EKF vs. 500 propagated MC samples drawn from $p(\boldsymbol{x}_0)$

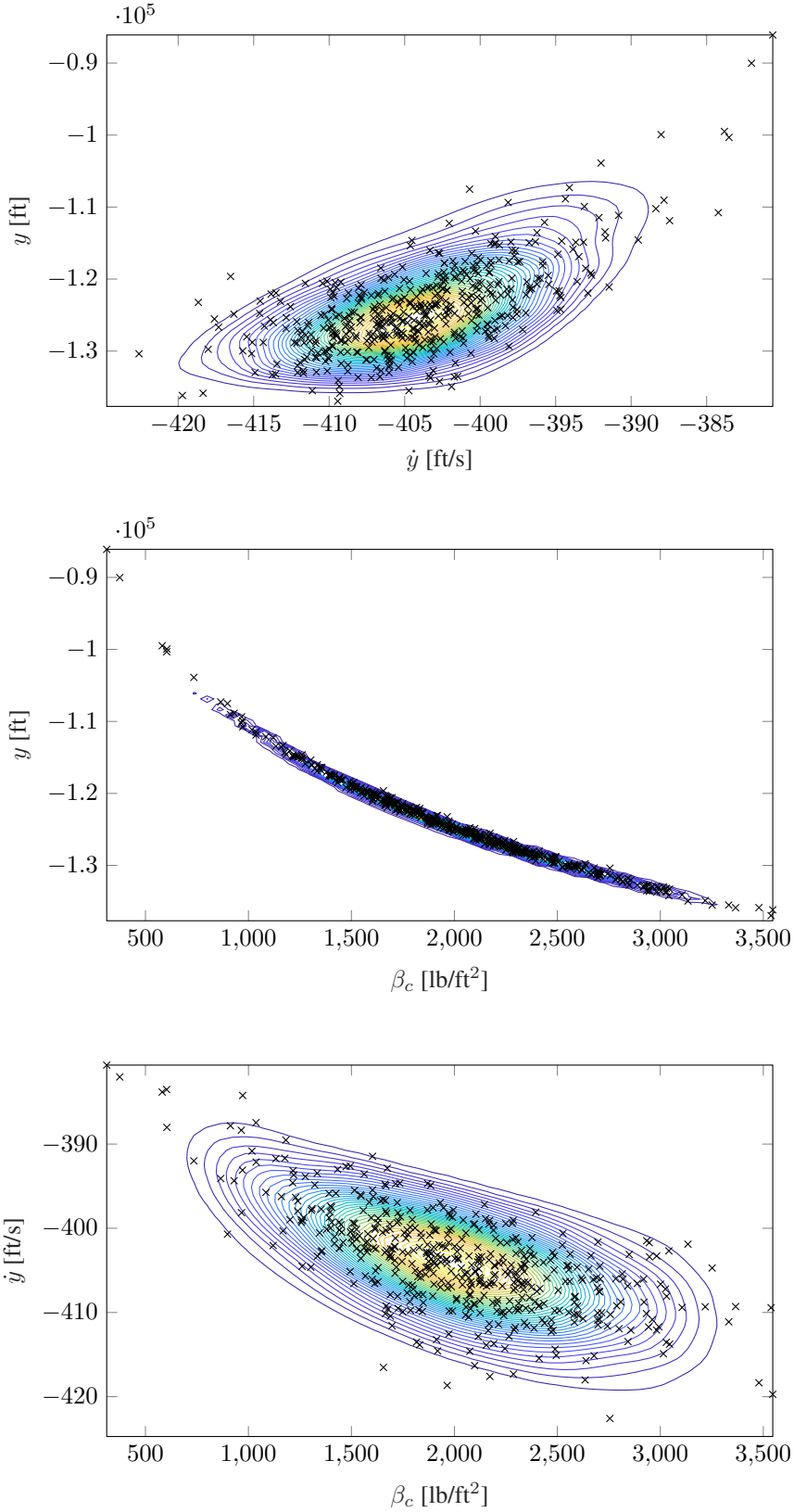


Figure 2.9 Contours of propagated GSF vs. 500 propagated MC samples drawn from $p(\boldsymbol{x}_0)$

3. NON-BAYESIAN FILTERING*

This chapter is devoted to the discussion and derivation of filters that perform information fusion using techniques other than Bayes' rule and how these alternative updates can lead to more robust algorithms. Since the major focus of this chapter, as well as those remaining, is centered almost entirely upon the filter update, the time indices k are omitted for brevity. Before introducing non-Bayesian filters, it is essential to first discuss Bayesian filtering to provide insight into why alternate solutions are needed in the first place. Note that a portion of this work is previously published as [58].

3.1 Bayesian Filtering

When it comes to the theory of estimation, there are few concepts more foundational than Bayes' rule, where many consider it the orthodox approach to fuse prior knowledge with incoming information [59]. Mathematically, Bayes' rule takes the form

$$p^{+}(\boldsymbol{x}) \propto p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})\,, \tag{3.1}$$

where $p^{+}(\boldsymbol{x})$ is the posterior distribution of the state $\boldsymbol{x}$ formed via a Bayes-optimal fusion of the prior distribution $p^{-}(\boldsymbol{x})$ and some likelihood $\ell(\boldsymbol{z}|\boldsymbol{x})$ that evaluates incoming information $\boldsymbol{z}$. While Bayes' rule is by far the most prevalent update used within the field of filtering [10], it does not preclude the potential for other classes of updates [60]. In fact, it can be argued that very few filter implementations are truly Bayesian, as practical applications often require approximations or extensions that prevent the update from strictly adhering to the Bayesian algorithm. For example, consider again the content of Eq. (3.1). Note that while no exact forms of the posterior, prior, and likelihood distributions have been defined, the Bayesian paradigm is predicated on the assumptions

*Part of the material reported in this chapter is reprinted with permission from "Adaptive Confidence Filter Update for High Uncertainty Environments" by Gunner S. Fritsch and Kyle J. DeMars, 2020. Proceedings of the AIAA SciTech Forum, Orbit Determination and Estimation Theory II, Copyright [2020] by Gunner S. Fritsch and Kyle J. DeMars.

that both the prior $p^-(\boldsymbol{x})$ and the likelihood model $\ell(\boldsymbol{z}|\boldsymbol{x})$ are correctly specified. This presents a problem; in reality, these distributions are very often approximated, as their precise forms make exact treatment either impossible or impractical. These issues are particularly prevalent in environments of high uncertainty, where little prior information is available and incoming information may be limited and sparse. Due to the lack of knowledge in such environments, deriving appropriately correct models can be more difficult. Therefore, while the Bayesian update remains the gold-standard for many filtering applications, this chapter explores alternative non-Bayesian (yet still probabilistic) filtering updates that are intended to be robust within environments of high uncertainty.

3.2 Generalized Variational Inference

Considering the natural limitations of Bayesian inference mentioned in Section 3.1, there is a clear demand for non-Bayesian updates. However, as many consider Bayes' theorem a foundational axiom within estimation theory, this prompts the question: without Bayes' rule, how can incoming data be incorporated into the estimated distribution? Fortunately, the field of generalized variational inference (GVI) is a substantiated candidate, offering a well-defined methodology for forming measurement updates posed as optimization problems [61]. In general, GVI establishes that the posterior distribution can be found as

$$p^+(\boldsymbol{x}) = \min_{p(\boldsymbol{x})\in\Pi} \left\{ \mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}(\boldsymbol{z},\boldsymbol{x})\big] + \mathcal{D}\big[p(\boldsymbol{x})\big\|p^-(\boldsymbol{x})\big] \right\}, \tag{3.2}$$

where Π constrains the distribution $p(\boldsymbol{x})$, and where $\mathcal{L}(\cdot,\cdot)$ is a loss function and $\mathcal{D}[\cdot\|\cdot]$ is a divergence. The behavior of the GVI update of Eq. (3.2) is entirely characterized through these three components, often referred to as the "rule of three", making GVI a flexible tool to investigate the practicality and behavior of non-Bayesian filters. Due to the optimization of Eq. (3.2), the convexity of functions is also of particular interest and is discussed frequently throughout this chapter as well as Section A.3.

3.2.1 The Rule of Three

The rule of three, as established in [61], is the notion that by selecting 1) a divergence, 2) a loss function, and 3) a family of feasible distributions, one is able to establish a GVI update of the form in Eq. (3.2). As such, this section seeks to define and investigate each of these three components.

3.2.1.1 Divergence

The general definition of a divergence, as it pertains to information theory, is a measure of directed distance between two probability distributions $p, q \in \mathbb{P}$ quantified by the operator $\mathcal{D}\big[\cdot\big\|\cdot\big] : \mathbb{P} \times \mathbb{P} \to \mathbb{R}^+$ pursuant to [62]

$$\mathcal{D}\big[p\big\|q\big] \geq 0 \;\; \forall\, p, q \in \mathbb{P}$$

and

$$\mathcal{D}\big[p\big\|q\big] = 0 \text{ if and only if } p = q\,.$$

As this definition is rather broad, it is often useful to refer to different classes of divergences that exist in certain forms. Two well-known classes are the f- and Bregman divergences.

f-Divergences

For any function $f(u)$ that is convex over $u \in (0, \infty)$ and satisfies $f(1) = 0$, the f-divergence is denoted as

$$\mathcal{D}_f\big[p\big\|q\big] = \int_{\mathbb{P}} q(\boldsymbol{x}) f\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}\,. \tag{3.3}$$

Note that $\mathcal{D}_f\big[p\big\|q\big]$ is jointly convex (convex with respect to both p and q) according to [63].

Bregman Divergences

For any strictly convex, real valued function $F : \mathbb{P} \to \mathbb{R}$, the Bregman divergence is defined as

$$\mathcal{D}_b\big[p\big\|q\big] = F(p) - F(q) - (p-q)F'(q)\,, \tag{3.4}$$

where $F'(q)$ is the derivative of F with respect to q, evaluated at q. According to [63], any valid Bregman divergence is guaranteed to be convex with respect to the first argument p.

The exact definition and classification aside, with regard to GVI, the divergence $\mathcal{D}$ quantifies how similar $p(\boldsymbol{x})$ is to the prior distribution $p^-(\boldsymbol{x})$, such that when the cost function of Eq. (3.2) is optimized, the prior information is appropriately accounted for. By appointing a divergence to Eq. (3.2), one is essentially specifying a function quantifying the importance of prior information when forming the posterior distribution. Therefore, three different divergences are introduced here, each possessing a unique behavior worthy of investigation.

3.2.1.1.1 Kullback-Leibler Divergence

Perhaps the most well-known statistical divergence, the Kullback-Leibler divergence (KLD) is given by [64]

$$\mathcal{D}_{KL}[p\|q] = \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}\,, \tag{3.5}$$

which belongs to the family of f-divergences of Eq. (3.3) for $f(u) = u\ln(u)$, the convexity of which is already established. For a fixed q, it can be further shown that the KLD is *strictly* convex with respect to p, a proof of which is provided in Section B.4. This is especially applicable to GVI, as it optimizes over the term $\mathcal{D}[p\|p^-]$, where p^- is the fixed prior pdf of the state, implying strict convexity relative to p.

To comment further on the nature of the KLD, note that Shannon entropy is given by [65]

$$\mathcal{H}_s[p] = -\int p(x) \ln \left[p(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}\,, \tag{3.6}$$

which is described best as the expected value of the self-information of $\boldsymbol{x}$ contained in $p(\boldsymbol{x})$. It is clear that replacing $\ln[p(\boldsymbol{x})]$ of Eq. (3.6) with $\ln[q(\boldsymbol{x})] - \ln[p(\boldsymbol{x})]$ directly recovers the KLD of Eq. (3.5), showing that the two are directly related, and that the KLD is essentially the amount of information "lost". The relationship between Eqs. (3.5) and (3.6) is why the KLD is often referred to as the relative entropy.

The characteristics of the KLD are well-established, with it being considered an extremely efficient and convenient divergence that is relatively sensitive to outliers [63, 66]. As such, it is included in this work to provide a performance baseline with which other, potentially more robust, divergences can be compared.

3.2.1.1.2 Rényi Divergence

The Rényi divergence (RD) is given by [67]

$$\mathcal{D}_R[p\|q] = \frac{1}{\alpha - 1} \ln \left(\int p^{\alpha}(\boldsymbol{x}) q^{1-\alpha}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right), \tag{3.7}$$

which is a divergence used widely throughout information theory, and, much like the KLD is related to Shannon entropy, Eq. (3.7) is closely related to Rényi entropy [68]. Here, the RD is found to be especially useful, as it is jointly convex in p and q for $\alpha \in [0, \infty]$, and as α approaches one, $\mathcal{D}_R[p\|q]$ recovers the KLD of Eq. (3.5) [67]. As such, the Rényi divergence is often referred to as a generalization to the KLD that is inherently capable of shifting exponential emphasis between p and q. For example, as $\alpha \in [0, 1]$ decreases, more emphasis is placed on q, with $\alpha = 0$ yielding a divergence entirely dependent on q. In this way, by carefully selecting α, the RD of Eq. (3.7) can be shifted away from the often-standard behavior of the KLD, which makes it of special interest to this work.

3.2.1.1.3 γ-Divergence

Lesser known than the KLD and the RD, the γ-divergence (γ-D) is given by [63, 69]

$$\mathcal{D}_\gamma[p\|q] = \frac{1}{\gamma(\gamma-1)} \ln\left(\frac{\left[\int p^\gamma(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right]\left[\int q^\gamma(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right]^{\gamma-1}}{\left[\int p(\boldsymbol{x})q^{\gamma-1}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right]^{\gamma}}\right) \tag{3.8}$$

and is specifically designed to be robust to outliers, something that the KLD is known to lack. As robust estimation is of special interest to this work, the γ-D is a good candidate for investigating non-Bayesian, GVI updates.

3.2.1.2 Loss Function

In general terms, the loss function $\mathcal{L}(\boldsymbol{z}, \boldsymbol{x})$ models a relationship between observation $\boldsymbol{z}$ and state $\boldsymbol{x}$ in such a way that scores how likely $\boldsymbol{z}$ is for a given state $\boldsymbol{x}$. This implies that the loss function must relate the measurements to the state in some way, but the exact form of $\mathcal{L}$ can vary depending on the desired behavior of Eq. (3.2) [70]. A more mathematically rigorous defintion is, for a state $\boldsymbol{x} \in \mathbb{X}$ of which observations $\boldsymbol{z}_i \in \mathbb{Z}$ are taken, a loss function is any operator $\mathcal{L} : \mathbb{X} \times \mathbb{Z} \to \mathbb{R}$ having the empirical risk minimizer

$$\hat{\boldsymbol{x}}_N = \min_{\boldsymbol{x}\in\mathbb{X}} \left[\sum_{i=1}^{N} \mathcal{L}(\boldsymbol{z}_i, \boldsymbol{x})\right]. \tag{3.9}$$

Note that while $\mathcal{L}(\boldsymbol{z}, \boldsymbol{x})$ is a function of $\boldsymbol{x}$, it should not be a function of its probability density function $p(\boldsymbol{x})$, but rather the measurement likelihood $\ell(\boldsymbol{z}|\boldsymbol{x})$—a concept discussed further in Section 4.1.

Recalling the GVI update of Eq. (3.2), it is reminded that GVI does not directly optimize over the loss function itself, but rather its expectation with respect to the posterior pdf. In this way it can be shown, at least in regard to $p(\boldsymbol{x})$, that the expected loss function is actually convex, making it ideal for convex optimization approaches [71]. A proof of this is provided in Section B.5.

There exist many functions that fit the definition of Eq. (3.9), each of which can produce unique behavior in the GVI posterior. To investigate various possible GVI updates, two different loss

functions are presented in this section.

3.2.1.2.1 Negative Log-Loss

The negative log loss (NLL) function is given by

$$\mathcal{L}_{NL}(\boldsymbol{z}, \boldsymbol{x}) = -\ln\big(\ell(\boldsymbol{z}|\boldsymbol{x})\big) \tag{3.10}$$

and is perhaps the most widely used loss function. It is popular mainly because the logarithm of the likelihood is more numerically stable than the likelihood itself, and by minimizing the negative log likelihood of Eq. (3.10), the likelihood itself is maximized. As such, it is no surprise that the NLL is, in fact, a staple of maximum likelihood estimation (MLE) [72]. Additionally, much like the KLD is used as a performance baseline for other divergences, the NLL is included here as the "standard" GVI loss function, with which other loss functions are compared.

3.2.1.2.2 γ-Loss

There are many other loss functions besides the NLL, one of which is the γ-loss function (γ-LF) presented in [61, 73]. It is mathematically given as

$$\mathcal{L}_{\gamma}(\boldsymbol{z}, \boldsymbol{x}) = \frac{\gamma}{1-\gamma}\left(\ell(\boldsymbol{z}|\boldsymbol{x})\Big[\int \ell(\boldsymbol{s}|\boldsymbol{x})^{\gamma}\mathrm{d}\boldsymbol{s}\Big]^{\frac{1}{\gamma}}\right)^{\gamma-1}, \tag{3.11}$$

and is selected here as it is of a form that does not include $p(\boldsymbol{x})$, as well as the availability of a numerically stable log form $-\ln[-\mathcal{L}_{\gamma}(\cdot,\cdot)]$ [61]. Of special note is that the γ-LF of Eq. (3.11) can be derived from the γ-D of Eq. (3.8) following the steps in [73], making the γ-LF a robust choice for GVI when $\gamma > 1$. For this reason, the γ-LF is a suitable loss function candidate for this work, as it will generally produce more conservative GVI posteriors than the NLL would.

3.2.1.3 Feasible Distributions

While GVI seeks to avoid defining a precise form for the posterior distribution, it is still useful, if not computationally necessary, to make some constraining assumptions on the pdf $p(\boldsymbol{x})$ of

Eq. (3.2). This is accomplished by designating a family of distributions, Π, that are considered feasible solutions to the GVI optimization problem. At the very least, Π must constrain $p(\boldsymbol{x})$ to

$$\int_{\mathbb{X}} p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1 \qquad \text{and} \qquad p(\boldsymbol{x}) \geq 0\,\forall \boldsymbol{x} \in \mathbb{X}\,, \tag{3.12}$$

which maintains $p(\boldsymbol{x})$ as a valid pdf.

3.2.1.3.1 Gaussian

While there are many possible distributions that can be used when appointing Π, many times the selection depends on the particular stochastic characteristics of the system being considered. However, as discussed in Section 2.2.2, the Gaussian distribution is known for its wide applicability and useful properties. In fact, many of the integrals found in the divergences and loss functions of Sections 3.2.1.1 and 3.2.1.2 are found to close under Gaussian assumptions, which will be shown in Section 3.2.2. Therefore, if the posterior is constrained to be Gaussian, then $p(\boldsymbol{x}) \in \Pi_g$, where

$$\Pi_g = p_g(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P})\,. \tag{3.13}$$

Note that the Gaussian constraint of Eq. (3.13) inherently enforces the requirements of Eq. (3.12) that ensures $p(\boldsymbol{x})$ is a valid pdf.

3.2.2 Robust GVI Updates

As the purpose of introducing GVI is to find robust alternatives to Bayes' rule for information fusion, this section seeks to develop cost functions for Eq. (3.2) by combining a divergence from Section 3.2.1.1 with a loss function from Section 3.2.1.2, under the constraint that posterior must be Gaussian subject to Eq. (3.13). While GVI is an incredibly useful tool, analytical solutions to Eq. (3.2) are often difficult, or even impossible to find. To this end, instead of seeking exact solutions to the GVI optimization, a numerical approach is deemed appropriate, as it allows for quick implementation and testing of different GVI updates by varying the rule of three. Unfortunately, even when the cost function of GVI is shown to be convex, the numerical stability of the

optimization process is not always sufficient. To address this, the Gaussian assumptions made in this section provide exact forms for the integrals of the divergences and loss functions, making numerical optimization more feasible. Additionally, these Gaussian assumptions also allow the GVI updates to be directly compared to the EKF, which is useful for performance analysis purposes.

3.2.2.1 Gaussian Divergences

For the divergences $\mathcal{D}\big[p\big\|p^-\big]$ presented in this section, it is assumed that the prior $p^-(\boldsymbol{x})$ is a Gaussian of the form

$$p^-(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^-, \boldsymbol{P}^-)\,. \tag{3.14}$$

which is identical to the prior of the EKF from Eq. (2.3c), but with the omission of the time index.

3.2.2.1.1 Gaussian KLD

Under the posterior constraint of Eq. (3.13) and the Gaussian prior assumption of Eq. (3.14), the KLD from Eq. (3.5) is shown to be [74]

$$\mathcal{D}_{KL}[p\|p^-] = \frac{1}{2}\left[\ln\left(\frac{|\boldsymbol{P}^-|}{|\boldsymbol{P}|}\right) - n_{\boldsymbol{x}} + \mathrm{tr}\big\{(\boldsymbol{P}^-)^{-1}\boldsymbol{P}\big\} + (\boldsymbol{m}^- - \boldsymbol{m})^T(\boldsymbol{P}^-)^{-1}(\boldsymbol{m}^- - \boldsymbol{m})\right], \tag{3.15}$$

where $n_{\boldsymbol{x}}$ is the dimension of the state vector. The exact derivation of Eq. (3.15) is shown in Section B.3, but is relatively well known.

3.2.2.1.2 Gaussian RD

Again, under the posterior constraint of Eq. (3.13) and prior assumption of Eq. (3.14), the RD from Eq. (3.7) is shown to be [75]

$$\mathcal{D}_R[p\|p^-] = \frac{\alpha}{2}(\boldsymbol{m} - \boldsymbol{m}^-)^T\tilde{\boldsymbol{P}}(\boldsymbol{m} - \boldsymbol{m}^-) - \frac{1}{2(\alpha-1)}\ln\left(\frac{|\tilde{\boldsymbol{P}}|}{|\boldsymbol{P}|^{1-\alpha}|\boldsymbol{P}^-|^{\alpha}}\right), \tag{3.16}$$

where

$$\tilde{\boldsymbol{P}} = \alpha \boldsymbol{P}^{-} + (1-\alpha)\boldsymbol{P}\,.$$

While not provided explicitly here, the derivation of Eq. (3.16) follows similarly to Section B.3.

3.2.2.1.3 Gaussian γ-D

Once again, under the posterior constraint of Eq. (3.13) and prior assumption of Eq. (3.14), the γ-D from Eq. (3.8) takes the form [61]

$$\begin{aligned}\mathcal{D}_{\gamma}[p\|p^{-}] = \frac{1}{2\gamma(1-\gamma)}\Bigg[& n_{\boldsymbol{x}}\gamma(\gamma-2)\ln(2\pi) + (\boldsymbol{m}-\boldsymbol{m}^{-})^{T}\tilde{\boldsymbol{P}}(\boldsymbol{m}-\boldsymbol{m}^{-}) \\ & + n_{\boldsymbol{x}}\ln\left(|\gamma|^{1+\gamma}|\gamma-1|^{-\gamma}(-\gamma)^{-1}\right) + \gamma^{2}\ln\left|\boldsymbol{P}^{-}\right| \\ & - \gamma\ln\left|[\boldsymbol{P}^{-}]^{2}(\boldsymbol{I} + \frac{1}{\gamma-1}\boldsymbol{P}^{-}\boldsymbol{P}^{-1})\right| + \ln\left|\boldsymbol{P}^{-}\boldsymbol{P}^{-1}\right|\Bigg]\,, \end{aligned} \tag{3.17}$$

where

$$\tilde{\boldsymbol{P}} = \left[\frac{1}{\gamma(1-\gamma)}\boldsymbol{P}^{-} - \frac{1}{\gamma}\boldsymbol{P}\right]^{-1}\,.$$

Due to the increased number of integrals in Eq. (3.8), the Gaussian form of the γ-D in Eq. (3.17) is more complex than the previous divergences.

3.2.2.2 Gaussian Expected Loss Functions

For the loss functions $\mathcal{L}(\boldsymbol{z},\boldsymbol{x})$ presented in this section, it is assumed that the measurement likelihood $\ell(\boldsymbol{z}|\boldsymbol{x})$ is a linear-Gaussian of the form

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_{g}(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\boldsymbol{R})\,. \tag{3.18}$$

These linear-Gaussian expected loss functions are then adapted to the nonlinear measurement model described in Eq. (2.1c) via first-order approximations.

3.2.2.2.1 Gaussian NLL

Under the posterior constraint of Eq. (3.13) and the linear-Gaussian measurement assumption of Eq. (3.18), the expected value of the NLL from Eq. (3.10) is

$$\begin{aligned}\mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}_{NL}(\boldsymbol{z},\boldsymbol{x})\big] &= \mathbb{E}_{p(\boldsymbol{x})}\big[-\ln\big(p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\boldsymbol{R})\big)\big] \\ &= -\mathbb{E}_{p(\boldsymbol{x})}\bigg[-\frac{1}{2}\ln|2\pi\boldsymbol{R}| - \frac{1}{2}\Big(\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}\Big)\bigg] \\ &= \frac{1}{2}\Big\{\ln|2\pi\boldsymbol{R}| + \boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] + \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}]\Big\}.\end{aligned} \tag{3.19}$$

Noting that $\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] = \boldsymbol{m}$ and $\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}] = \text{tr}\{\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}\} + \boldsymbol{m}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m}$, Eq. (3.19) becomes

$$\mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}_{NL}(\boldsymbol{z},\boldsymbol{x})\big] = \frac{1}{2}\bigg[\ln|2\pi\boldsymbol{R}| + (\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m}) + \text{tr}\big\{\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}\big\}\bigg]. \tag{3.20}$$

The form of Eq. (3.20) can be extended to the nonlinear measurement model of Eq. (2.1c) as

$$\mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}_{NL}(\boldsymbol{z},\boldsymbol{x})\big] = \frac{1}{2}\bigg[\ln|2\pi\boldsymbol{R}| + \big(\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{m})\big)^T\boldsymbol{R}^{-1}\big(\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{m})\big) + \text{tr}\big\{\boldsymbol{H}^T(\boldsymbol{m})\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\big\}\bigg]. \tag{3.21}$$

3.2.2.2.2 Gaussian γ-LF

Again, under the posterior constraint of Eq. (3.13) and the linear-Gaussian measurement assumption of Eq. (3.18), the expectation of the γ-LF from Eq. (3.11) can be shown to be [61]

$$\mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}_{\gamma}(\boldsymbol{z},\boldsymbol{x})\big] = -\gamma^{\frac{n_{\boldsymbol{z}}(\gamma-1)}{2\gamma}+1}(\gamma-1)^{-\frac{n_{\boldsymbol{z}}}{2}-1}\big|2\pi\boldsymbol{R}\big|^{\frac{(1-\gamma)^2-\gamma^2}{2\gamma}}p_g\Big(\boldsymbol{z}\Big|\boldsymbol{H}\boldsymbol{m},\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\frac{1}{\gamma-1}\boldsymbol{R}\Big), \tag{3.22}$$

assuming $\gamma \geq 1$. Note that $n_{\boldsymbol{z}}$ is the dimension of the measurement vector and the Gaussian term of Eq. (3.22) is an evaluated likelihood, not a pdf. Unfortunately, Eq. (3.22) is found to scale poorly when subjected to numerical optimization. As such, the log-stable form is introduced as

$$\mathbb{E}_{p(\boldsymbol{x})}\big[-\ln\big(-\mathcal{L}_\gamma(\boldsymbol{z},\boldsymbol{x})\big)\big] = \ln(\gamma-1) - \ln(\gamma) + \frac{\gamma-1}{2}\bigg[\frac{2\gamma-1}{\gamma}\ln|2\pi\boldsymbol{R}| + \frac{n_{\boldsymbol{z}}}{\gamma}\ln|\gamma| + (\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m}) + \mathrm{tr}\big\{\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}\big\}\bigg], \quad (3.23)$$

the derivation of which follows closely to that of Eq. (3.19). Again, Eq. (3.23) can also be adapted to the nonlinear measurement model of Eq. (2.1c) as

$$\mathbb{E}_{p(\boldsymbol{x})}\big[-\ln\big(-\mathcal{L}_\gamma(\boldsymbol{z},\boldsymbol{x})\big)\big] = \ln(\gamma-1) - \ln(\gamma) + \frac{\gamma-1}{2}\bigg[\frac{2\gamma-1}{\gamma}\ln|2\pi\boldsymbol{R}| + \frac{n_{\boldsymbol{z}}}{\gamma}\ln|\gamma| + \big(\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{m})\big)^T\boldsymbol{R}^{-1}\big(\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{m})\big) + \mathrm{tr}\big\{\boldsymbol{H}^T(\boldsymbol{m})\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\big\}\bigg]. \quad (3.24)$$

3.2.3 Simplified Falling Body Simulation

Much like Section 2.3.2, this simulation is designated to the falling body system of Fig. (2.4), yet it is simplified for the purposes of this analysis. More specifically, as the analyses of this section are principally concerned with the performance of the filter update, the dynamical model of Eq. (2.30) is altered to be linear by neglecting the effects of drag, such that the state vector $\boldsymbol{x}$ of Eq. (2.29) is reduced to

$$\boldsymbol{x} = \begin{bmatrix} y & \dot{y} \end{bmatrix}^T.$$

and the dynamical model for Eq. (2.30) becomes

$$\ddot{y} = -g\,,$$

GVI Update	$\mathcal{D}[p\|p^-]$	$\mathcal{L}(\boldsymbol{z},\boldsymbol{x})$	$p(\boldsymbol{x})\in\Pi$
1	$\mathcal{D}_{KL}$	$\mathcal{L}_{NL}$	Π_g
2	$\mathcal{D}_{KL}$	$\mathcal{L}_{\gamma}$	Π_g
3	$\mathcal{D}_{R}$	$\mathcal{L}_{NL}$	Π_g
4	$\mathcal{D}_{\gamma}$	$\mathcal{L}_{NL}$	Π_g

Table 3.1 Divergence, loss function, and feasible distributions for four GVI updates

where $g = 32.2\text{ft}/\text{s}^2$ is the acceleration due to gravity. The initial distribution corresponding to the reduced state vector is taken as

$$p^+(\boldsymbol{x}_0) = p_g(\boldsymbol{x}_0|\ \boldsymbol{m}_0, \boldsymbol{P}_0)\ , \tag{3.25}$$

where

$$\boldsymbol{m}_0 = \begin{bmatrix} 1000\text{ ft} \\ 0\ \text{ft}/\text{s} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{P}_0 = \begin{bmatrix} 2000\text{ft}^2 & 0 \\ 0 & 500(\text{ft}/\text{s})^2 \end{bmatrix}\ .$$

Measurement scans are generated every second, where valid sensor returns conform to the nonlinear, Gaussian model given by Eq. (2.33), with the measurement noise covariance for valid sensor returns taken to be $R = 10\text{ ft}^2$. The duration of the simulation is 10 seconds, and each analysis is performed over 500 MC trials, with MC statistics calculated as described in Section 2.3.1.

3.2.3.1 Filter Configurations

Five different filters are tested for the system, each with propagation algorithms identical to that of the EKF in Eqs. (2.3). In fact, the first filter is a true EKF, having the update of Eqs. (2.4), and is used as a point of reference when evaluating the GVI filters. The remaining four filters, shown in Table 3.1, have GVI updates corresponding to Eq. (3.2), with the appropriate Gaussian assumptions made such that the Gaussian forms of the divergences of Section 3.2.2.1 and expected loss functions of Section 3.2.2.2 are inserted directly into the GVI cost function.

To calculate the GVI posteriors, numerical optimization via Matlab's `fmincon` is performed

with the default interior-point method and a optimality tolerance of 10^{-15}. An inequality constraint enforces positive-definiteness of the posterior covariance by requiring the eigenvalues of $\boldsymbol{P}$ be positive. It is found that the dynamic and measurement models of the system play a large role in the behavior of the GVI updates, such that preliminary tuning of the GVI updates that require parameter selection of α and γ is advisable.

3.2.3.1.1 GVI Update 1: $\mathcal{D}_{KL}$ and $\mathcal{L}_{NL}$

The first GVI update (GVIU-1) consists of the $\mathcal{D}_{KL}$ of Eq. (3.15) and the expected $\mathcal{L}_{NL}$ of Eq. (3.20). This specific combination of $\mathcal{D}$ and $\mathcal{L}$ is selected as it is equivalent to a Bayesian update, a proof of which is presented in Section 3.3. Therefore, the results of this filter should closely match those of the EKF, as the EKF is (approximately) Bayes optimal under Gaussian assumptions.

3.2.3.1.2 GVI Update 2: $\mathcal{D}_{KL}$ and $\mathcal{L}_{\gamma}$

The second GVI update (GVIU-2) consists of the same $\mathcal{D}_{KL}$ of Eq. (3.15) as used in GVIU-1, but with the robust expected $\mathcal{L}_{\gamma}$ of Eq. (3.24). After trial and error, $\gamma \in [1, 2)$ is found to work well for the simplified falling body system, and $\gamma = 1.1$ is selected for the MC analyses.

3.2.3.1.3 GVI Update 3: $\mathcal{D}_{R}$ and $\mathcal{L}_{NL}$

The third GVI update (GVIU-3) consists of the same $\mathcal{L}_{NL}$ of Eq. (3.20) as used in GVIU-1, but with the $\mathcal{D}_{R}$ of Eq. (3.16). While this Rényi divergence-based update seems to work well for most values of $\alpha \in [0, 1)$, lower values of α are found to produce more conservative estimates. Thus, $\alpha = 0.2$ is selected as it shows significantly different behavior than the GVI update of Section 3.2.3.1.1, where the KLD is used.

3.2.3.1.4 GVI Update 4: $\mathcal{D}_{\gamma}$ and $\mathcal{L}_{NL}$

The fourth GVI update (GVIU-4) consists of the same $\mathcal{L}_{NL}$ of Eq. (3.20) as used in GVIU-1, but with the robust $\mathcal{D}_{\gamma}$ of Eq. (3.16). Again, preliminary tuning shows that this GVI update works

well for $\gamma \in (1, 2)$, but $\gamma = 1.001$ is selected for its specific behavior.

3.2.3.2 Analysis 1: Ideal Measurements

In this analysis, the performance of all five filters are compared to each other under ideal conditions; no spurious measurements are generated, and the filter is not misinformed concerning the system models. Beginning with the performance of the baseline EKF in Fig. (3.1a), the consistency of the MC and average filter uncertainties indicate a well-behaved filter, with the filter estimate accurately reflecting the true error in the system. Of course, the first result of note is the similarity between the EKF update of Fig. (3.1a) and the GVIU-1 of Fig. (3.1b). Since these are both (approximately) Bayesian updates, their identical behavior is expected and serves as a sign of correct numerical optimization. It is also worth noting that GVIU-1 is by far the quickest and most stable of all GVI updates, achieving optimality significantly faster and failing the least.

Examining the results of GVIU-2 in Fig. (3.1c) next, it is clear that this non-Bayesian update is far more conservative ($\sigma_{\text{filt},\iota} > \sigma_{\text{MC},\iota}$) than its Bayesian counterparts in Figs. (3.1a) and (3.1b). In fact, it does seem to be the most conservative of all the filters, which could possibly be a desired outcome, especially if measurements are mismodeled to a significant degree—a scenario that is studied further in the following analysis of Section 3.2.3.3. While the true average uncertainty (- - -) of GVIU-2 decreases more slowly than the other filters, it eventually reaches a comparable steady-state value as more measurements are processed.

Next, the Rényi-based GVIU-3 is examined, where it is shown to have a similar uncertainty profile to the Bayesian updates of Figs. (3.1a) and (3.1b), but with slightly more conservative estimates. Recalling that the RD reduces to the KLD as $\alpha \to 1$, this is precisely the type of behavior desired from such an update. It shows promise for systems where prior distributions of the state are misspecified, such as systems with imprecise dynamical models.

Lastly, the MC results of GVIU-4 are shown in Fig. (3.1e), where it also appears remarkably similar to the Bayesian updates of Figs. (3.1a) and (3.1b). The only clear difference occurs at the first time step, where, instead of the estimated uncertainty in the velocity decreasing as it does with the other filters, GVIU-4 initially becomes more uncertain of the velocity. In short, the posterior

formed by GVIU-4 causes less information gain in velocity until the uncertainty in the position is reduced at $t_k = 1$ sec. The exact mechanics behind this are hard to pinpoint, but it is interesting behavior, nonetheless.

3.2.3.3 Analysis 2: Measurement Model Mismatch

With Section 3.2.3.2 demonstrating successful estimation with the GVI filters under ideal conditions, their robustness is next to be tested. This second analysis accomplishes this by introducing faulty measurements into the system. While detailed discussion on faulty measurements is left for Section 4.2, they are invoked here to create a mismatch between the measurement model assumed by the filter and the model used to simulate measurement generation. To this end, there is a 10% that any given measurement received by the filter is faulty, being generated by the range measurement model of Eq. (2.33), but with an increased noise covariance of $R_f = 100R = 1000$ ft^2. Note that, as the filters are unaware of the existence of faulty measurements, all filters assume $R = 10$ ft^2.

Within this model mismatch analysis, two GVI updates are tested. First, GVIU-1 of Section 3.2.3.1.1 is selected as a performance reference for Bayesian updates, the MC results of which are shown in Fig. (3.2a). Upon inspection, it is immediately clear that GVIU-1 fails in the presence of model mismatch. Figure (3.2a) is clearly indicative of an overconfident filter, with average error standard deviations well outside the filter expectations. GVIU-1 clearly cannot compensate for the presence of faulty measurements, showing the inadequacy of Bayesian estimation under poor measurement modeling.

The second update is GVIU-2, which corresponds to the $\mathcal{L}_\gamma$-based update described in Section 3.2.3.1.2. GVIU-2 is a good candidate for measurement model mismatch as it optimizes over a robust loss function in place of the standard NLL. The MC results produced by GVIU-2 are provided in Fig. (3.2b), where it is clear that this more robust GVI update is remarkably well-behaved, even in the presence of faulty measurements. While slightly overconfident overall, the error is estimated far more accurately by GVIU-2 than GVIU-1. In short, the non-Bayesian GVIU-2 is more robust to the faulty measurements of this system than the Bayesian estimate of GVIU-1.

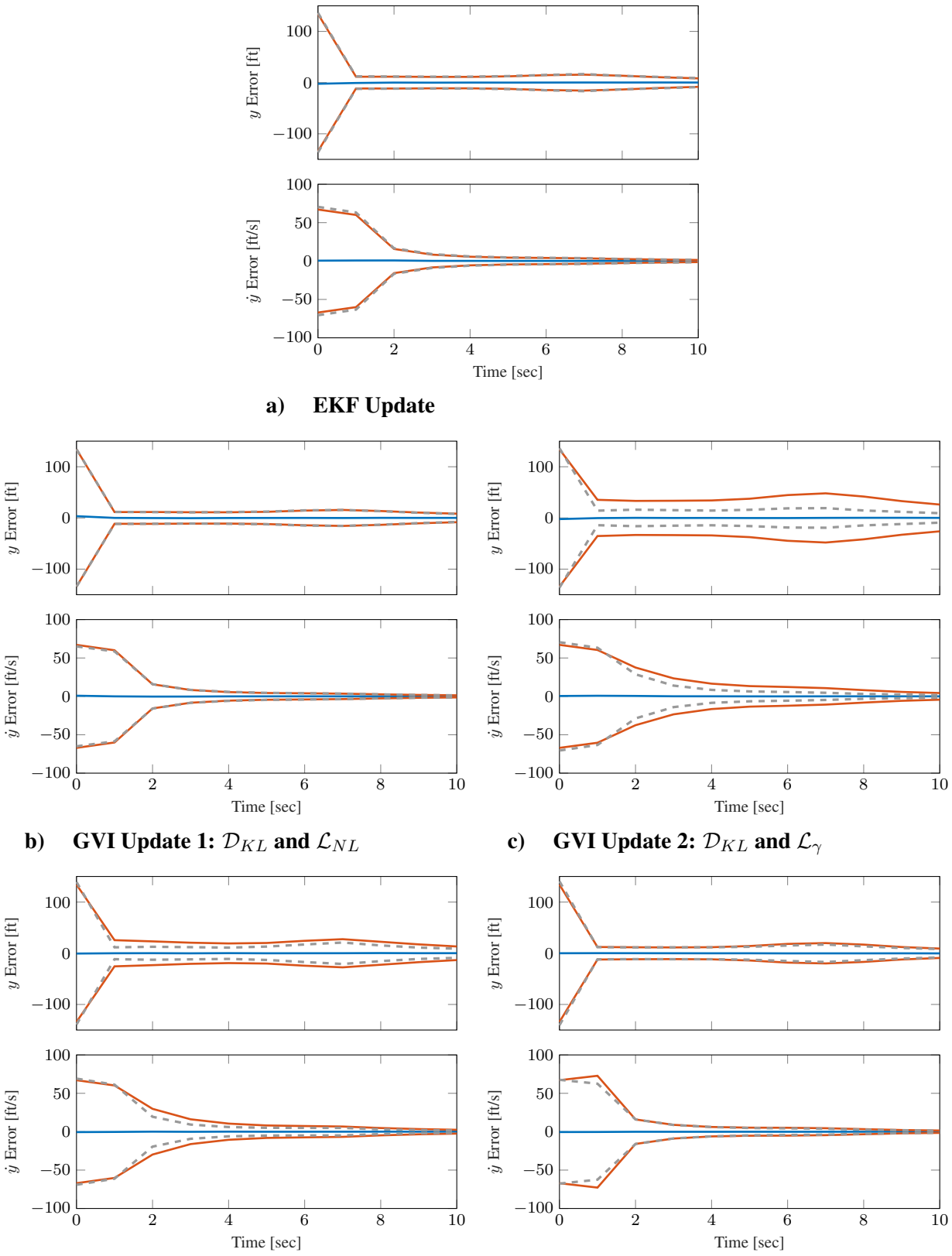
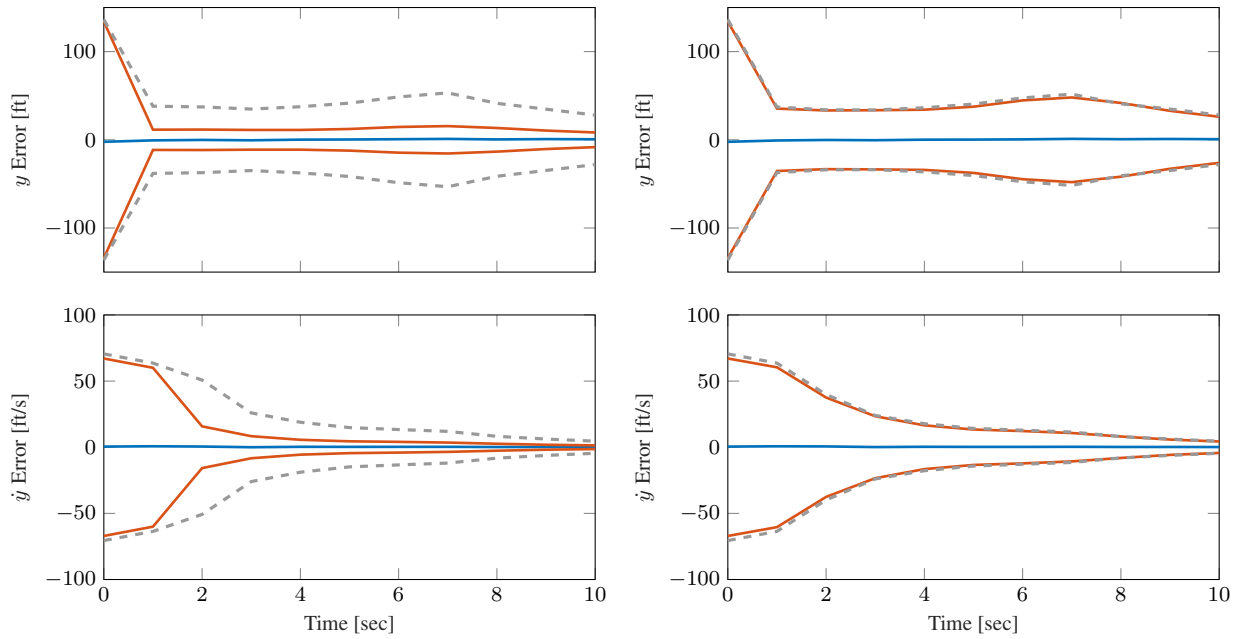


Figure 3.1 MC results of GVI falling body simulation plotted as $\overline{e}_{\iota}$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)



a) GVI Update 1: $\mathcal{D}_{KL}$ and $\mathcal{L}_{NL}$ **b) GVI Update 2: $\mathcal{D}_{KL}$ and $\mathcal{L}_{\gamma}$**

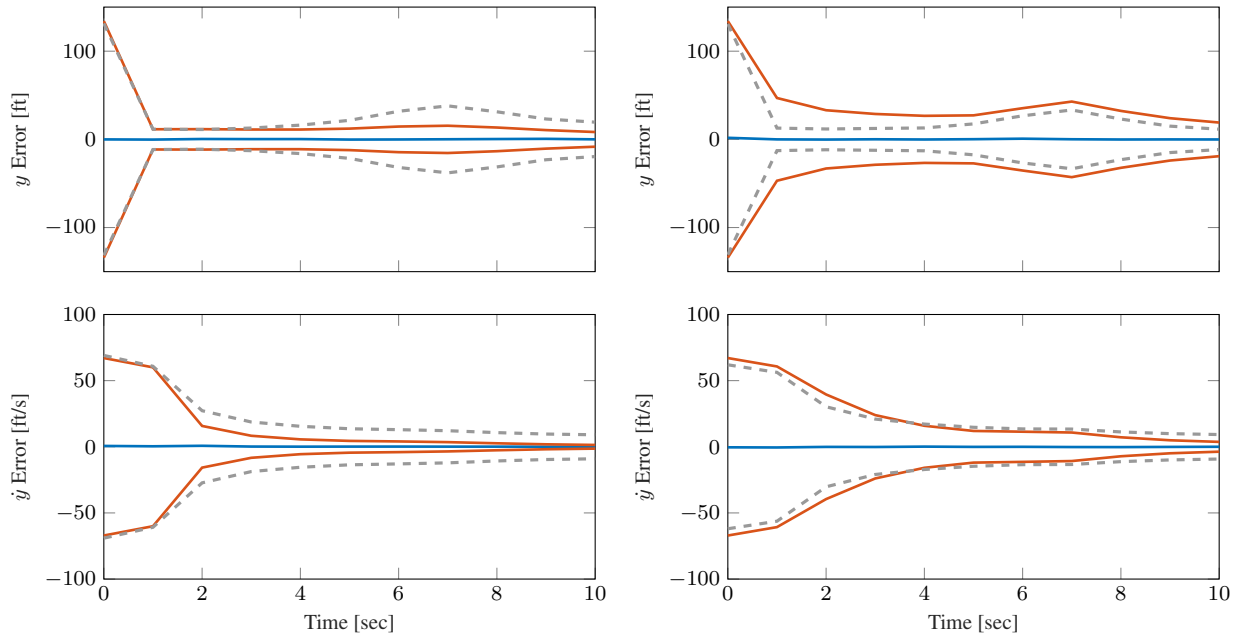
Figure 3.2 MC results of GVI falling body simulation with measurement model mismatch plotted as $\overline{e}_{\iota}$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

3.2.3.4 Analysis 3: Dynamics Model Mismatch

Much like the measurement model mismatch analysis of Section 3.2.3.3, this analysis evaluates the robust capabilities of GVI by mismatching the dynamics model. By providing the filter with a disingenuous dynamics model, each update is essentially supplied with an incorrect prior distribution $p^{-}(\boldsymbol{x})$. This dynamics mismatch is accomplished by imparting undisclosed process noise to the state propagation; at each time step, discrete process noise is sampled from a zero-mean Gaussian with covariance $\boldsymbol{Q}_k = \text{diag}\{75\text{ft}^2, 1(\text{ft}/\text{s})^2\}$, while the filter still assumes no process noise is present in the system.

Under these conditions, two GVI filters are evaluated, the comparative MC results of which are presented in Fig. (3.3). Again, in order to showcase Bayesian performance as a reference, the first filter of Fig. (3.3a) is the GVIU-1 of Section 3.2.3.1.1. The second filter in Fig. (3.3b) consists of the RD-based update of Section 3.2.3.1.3 (GVIU-3). While the actual errors of the two filters are about the same, the average estimated uncertainty differs significantly between the two. It is

clear that, as the simulation progresses, GVIU-3 remains mostly conservative, while the Bayesian GVIU-1 becomes increasingly overconfident. These results suggest that GVIU-3 is more resistant to dynamics model mismatch, making it a more robust alternative to Bayes' rule for such systems.



a) GVI Update 1: $\mathcal{D}_{KL}$ and $\mathcal{L}_{NL}$ **b) GVI Update 3: $\mathcal{D}_{R}$ and $\mathcal{L}_{NL}$**

Figure 3.3 MC results of GVI falling body simulation with dynamics model mismatch plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

3.3 Confidence-based Update

The utility of non-Bayesian updates is clearly showcased in Section 3.2, where multiple GVI updates are quickly introduced and tested. However, the posteriors are only calculated via direct numerical optimization, which can be computationally intensive and inappropriate for applications which require real-time estimation. Therefore, this section attempts to find practical, closed-form filters derived from GVI that can be implemented more efficiently. As the KLD and NLL are found to be exceptionally stable choices for GVI, they are used as the basis for the confidence-based update presented in this section, but are weighted to increase robustness.

At the core of most filters is the ability to construct a posterior distribution of the state $\boldsymbol{x}$ by fusing incoming data, $\boldsymbol{z}$, with some prior knowledge of the state. Predicating this discussion with the GVI posterior of Eq. (3.2), here it is assumed that incoming data is generated via $\ell(\boldsymbol{z}|\boldsymbol{x})$, and the prior knowledge of the state is contained within the distribution $p^-(\boldsymbol{x})$. If it is desired to construct a posterior distribution, $p^+(\boldsymbol{x})$, solely on prior information of $\boldsymbol{x}$, it can be shown that minimizing the divergence, via the relative entropy (i.e., Kullback-Leibler divergence) in this case, produces a posterior that minimizes the information gain from p^- to p^+ such that [61]

$$p^+(\boldsymbol{x}) = \min_{p(\boldsymbol{x})\in\Pi} \Big\{ \mathcal{D}_{KL}\big[p(\boldsymbol{x})\big\|p^-(\boldsymbol{x})\big] \Big\} = \min_{p(\boldsymbol{x})\in\Pi} \left\{ \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{p^-(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \right\}. \tag{3.26}$$

Note that in Eq. (3.26) the end result will be $p^+(\boldsymbol{x}) = p^-(\boldsymbol{x})$, which is equivalent to filtering processes where no update is performed—an outcome that holds for any valid divergence $\mathcal{D}$. On the other hand, if a posterior is to be constructed purely from the incoming data, the expected negative-log-likelihood can be minimized to yield

$$p^+(\boldsymbol{x}) = \min_{p(\boldsymbol{x})\in\Pi} \Big\{ \mathbb{E}_{p(\boldsymbol{x})}\big[\mathcal{L}_{NL}(\boldsymbol{z},\boldsymbol{x})\big] \Big\} = \min_{p(\boldsymbol{x})\in\Pi} \left\{ -\int p(\boldsymbol{x}) \ln\Big(\ell(\boldsymbol{z}|\boldsymbol{x})\Big) \mathrm{d}\boldsymbol{x} \right\}. \tag{3.27}$$

In the case of Eq. (3.27), $p^+(\boldsymbol{x}) = \delta_{\hat{\boldsymbol{x}}_{\mathrm{ML}}}(\boldsymbol{x})$, where $\hat{\boldsymbol{x}}_{\mathrm{ML}}$ is the maximum likelihood estimate of $\boldsymbol{x}$ [61]. To form a solution reliant upon both incoming information and prior knowledge, elements from both Eqs. (3.26) and (3.27) must be taken into consideration. This is accomplished through an optimization problem that defines the posterior distribution $p(\boldsymbol{x})$ as the solution that minimizes the cost function

$$J\big(p(\boldsymbol{x})\big) = \gamma \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{p^-(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} - \nu \int p(\boldsymbol{x}) \ln\Big(\ell(\boldsymbol{z}|\boldsymbol{x})\Big) \mathrm{d}\boldsymbol{x} \tag{3.28}$$

where γ and ν are scalar coefficients and $p(\boldsymbol{x})$ is restricted by Π to be a valid pdf such that

$$\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1 \quad \text{and} \quad p(\boldsymbol{x}) \geq 0 \; \forall \, \boldsymbol{x} \in \mathbb{X}. \tag{3.29}$$

For the optimization problem to be considered convex, the cost function of Eq. (3.28) must be convex, and the constraint of Eq. (3.29) must be affine. Fortunately, the forms of Eqs. (3.26) and (3.27) are chosen in part as they are both well known to be convex functions, such that they conform to the identity in Eq. (A.9). In fact, the KLD is *strictly convex* with respect to $p(\boldsymbol{x})$ for fixed $p^{-}(\boldsymbol{x})$, a characteristic proven in Eq. (B.55). As such, since Eq. (3.28) is a non-negative weighted sum of a convex and strictly convex functions, by the proof in Eq. (A.13), the cost function is strictly convex.

The constraint in Eq. (3.29), when restated such that

$$\begin{aligned} h\big(p(\boldsymbol{x})\big) &= \int p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - 1 \\ &= 0\,, \end{aligned}$$

can be shown to be affine such that

$$\begin{aligned} \theta h\big(p(\boldsymbol{x})\big) + (1-\theta)h\big(q(\boldsymbol{x})\big) &= \theta \int p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \theta + (1-\theta)\int q(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - 1 + \theta \\ &= \int \theta p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \int (1-\theta)q(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - 1 \\ &= \int \big(\theta p(\boldsymbol{x}) + (1-\theta)q(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x} - 1 \\ &= h\big(\theta p(\boldsymbol{x}) + (1-\theta)q(\boldsymbol{x})\big)\,. \end{aligned}$$

Thus, the constraint of Eq. (3.29) satisfies the definition of an affine function from Eq. (A.11). Note that $q(\boldsymbol{x})$ is simply an arbitrary function that exists on the same domain as $p(\boldsymbol{x})$. Therefore, the conditions for the optimization of Eqs. (3.28) and (3.29) to be convex are not only met, but as the cost function is strictly convex, there exists at most one unique solution that satisfies optimality. This is a convenient result, as any optimal solution found will be guaranteed to be unique.

Before minimizing, the Lagrange dual function is utilized such that the constraint of Eq. (3.29)

is inherently accounted for. Thus, the Lagrange dual function of Eq. (3.28) is

$$L(p(\boldsymbol{x}),\lambda) = \gamma \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} - \nu \int p(\boldsymbol{x}) \ln\Big(\ell(\boldsymbol{z}|\boldsymbol{x})\Big)\mathrm{d}\boldsymbol{x} + \lambda\Big(\int p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - 1\Big)\,, \tag{3.30}$$

which can be rearranged as

$$\begin{aligned} L(p(\boldsymbol{x}),\lambda) &= \int p(\boldsymbol{x}) \ln\left(\Big[\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x})}\Big]^{\gamma}\right)\mathrm{d}\boldsymbol{x} + \int p(\boldsymbol{x}) \ln\Big(\ell(\boldsymbol{z}|\boldsymbol{x})^{-\nu}\Big)\mathrm{d}\boldsymbol{x} + \int \lambda p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \lambda \\ &= \int p(\boldsymbol{x})\Bigg\{ \ln\left(\Big[\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x})}\Big]^{\gamma}\ell(\boldsymbol{z}|\boldsymbol{x})^{-\nu}\right) + \lambda\Bigg\}\mathrm{d}\boldsymbol{x} - \lambda \\ &= \int p(\boldsymbol{x})\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg]\mathrm{d}\boldsymbol{x} - \lambda\,. \end{aligned} \tag{3.31}$$

The distribution $p(\boldsymbol{x})$ that minimizes the Lagrange dual function is found by taking the derivative of Eq. (3.31) and equating it to zero. To do this, calculus of variations is most useful, as any local minima of a convex function will also be a global minima. Making use of the definition from Eq. (A.14), the first variation of Eq. (3.31) is

$$\begin{aligned} \delta L(p(\boldsymbol{x}),q(\boldsymbol{x}),\lambda) &= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Bigg\{ \int \Big(p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})\Big)\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg]\mathrm{d}\boldsymbol{x} - \lambda\Bigg\} \\ &= \int \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Bigg\{ \Big(p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})\Big)\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg]\Bigg\}\mathrm{d}\boldsymbol{x} - \frac{\mathrm{d}\lambda}{\mathrm{d}\epsilon} \\ &= \int \Bigg\{ \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big(p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})\Big)\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg] \\ &\qquad + \Big(p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})\Big)\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg]\Bigg\}\mathrm{d}\boldsymbol{x} \\ &= \int \Bigg\{ q(\boldsymbol{x})\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda\Bigg] \\ &\qquad + \gamma\Big(p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})\Big)\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Bigg[\ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right)\Bigg]\Bigg\}\mathrm{d}\boldsymbol{x} \\ &= \int q(\boldsymbol{x})\Bigg[\gamma \ln\left(\frac{p(\boldsymbol{x}) + \epsilon q(\boldsymbol{x})}{p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}\right) + \lambda + \gamma\Bigg]\mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{3.32}$$

Enforcing the requirement that $\epsilon = 0$ on Eq. (3.32) and setting $\delta L(p(\boldsymbol{x}), q(\boldsymbol{x}), \lambda) = 0$ yields

$$\int q(\boldsymbol{x}) \left[\gamma \ln \left(\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}} \right) + \lambda + \gamma \right] \mathrm{d}\boldsymbol{x} = 0 \,,$$

such that for any instantaneous support over $\boldsymbol{x}$

$$q(\boldsymbol{x}) \left[\gamma \ln \left(\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}} \right) + \lambda + \gamma \right] = 0 \,. \tag{3.33}$$

Here, it is noted that $q(\boldsymbol{x})$ is an arbitrary function. and that the expression in Eq. (3.33) must remain zero for all possible values of $q(\boldsymbol{x})$. Therefore, the remaining term must satisfy

$$\gamma \ln \left(\frac{p(\boldsymbol{x})}{p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}} \right) + \lambda + \gamma = 0 \,,$$

where algebraic manipulation yields

$$p(\boldsymbol{x}) = \exp \left\{ - \frac{\gamma + \lambda}{\gamma} \right\} p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}} \,. \tag{3.34}$$

From Eq. (3.34), it is simple to confirm that $p(\boldsymbol{x}) \geq 0$, thus satisfying one of the conditions of a valid pdf. To satisfy the remaining constraint from Eq. (3.29), for the distribution $p(\boldsymbol{x})$ to be a valid solution to the constrained optimization problem, it must obey

$$\begin{aligned} \int \exp \left\{ - \frac{\gamma + \lambda}{\gamma} \right\} p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}} \mathrm{d}\boldsymbol{x} = 1 \\ \exp \left\{ - \frac{\gamma + \lambda}{\gamma} \right\} = \frac{1}{\int p^{-}(\boldsymbol{x}) \ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}} \mathrm{d}\boldsymbol{x}} \,, \end{aligned} \tag{3.35}$$

which is also a result of $\frac{\partial L}{\partial \lambda} = 0$. Since $p(\boldsymbol{x})$ is the unique solution to the optimization problem being sought, it is here that $p^{+}(\boldsymbol{x}) = p(\boldsymbol{x})$ can be stated. Thus substitution of Eq. (3.35) into

Eq. (3.34) yields the optimal posterior as

$$p^+(\boldsymbol{x}) = \frac{p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{\nu}{\gamma}}}{\int p^-(\boldsymbol{s})\ell(\boldsymbol{z}|\boldsymbol{s})^{\frac{\nu}{\gamma}}\mathrm{d}\boldsymbol{s}}\,. \tag{3.36}$$

With the optimal solution found, it is useful to set $\gamma = \phi$ and $\nu = 1 - \phi$ such that the optimization problem can be restated as

$$\begin{aligned} p^+(\boldsymbol{x}) &= \min_{p(\boldsymbol{x})\in\Pi}\left\{(1-\phi)\mathbb{E}_{p(\boldsymbol{x})}\left[\mathcal{L}_{NL}(\boldsymbol{z},\boldsymbol{x})\right] + \phi\mathcal{D}_{KL}\left[p(\boldsymbol{x})\middle\|p^-(\boldsymbol{x})\right]\right\} \\ &= \min_{p(\boldsymbol{x})\in\Pi}\left\{-(1-\phi)\int p(\boldsymbol{x})\ln\Big(\ell(\boldsymbol{z}|\boldsymbol{x})\Big)\mathrm{d}\boldsymbol{x} + \phi\int p(\boldsymbol{x})\ln\left(\frac{p(\boldsymbol{x})}{p^-(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x}\right\}. \end{aligned} \tag{3.37}$$

The solution to Eq. (3.37) is therefore

$$p^+(\boldsymbol{x}) = \frac{\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{1-\phi}{\phi}}p^-(\boldsymbol{x})}{\int \ell(\boldsymbol{z}|\boldsymbol{s})^{\frac{1-\phi}{\phi}}p^-(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}\,. \tag{3.38}$$

Upon inspection, it is clear that when $\phi = 1/2$, Eq. (3.38) reduces to Bayes' rule of Eq. (3.1), which indicates that when both the prior and likelihood are considered equally important, the posterior is Bayes optimal. In some situations it may be useful to utilize a filter that is slightly more confident in the prior information than the incoming measurements, and in such a case, ϕ is set to some value greater than $1/2$. A filter update of this nature is more conservative than the typical Bayesian update, generally making it more robust to unmodeled effects and linearization errors. Furthermore, closed-form solutions under certain conditions produce a framework with varying confidence, where, under desired conditions, the filter can establish more or less confidence in its prior, or reduce back to a Bayesian update. With this in mind, the update of Eq. (3.38) is referred to as the confidence-based update.

3.3.1 Closed-Forms for Linear Filters

While the probabilistic form of the proposed update given in Eq. (3.38) is useful in terms of its generality, it should be refined into a more practical form before it can be directly compared

to existing filtering methods. Therefore, a linear-Gaussian form of the update is presented, which is then extended to nonlinear systems via first-order Taylor series linearization. These filters are well-suited to directly replace KFs or EKFs in existing estimation frameworks.

3.3.1.1 Linear-Gaussian Derivation

For the purposes of this derivation, a linear-Gaussian system is taken to govern the observations $\boldsymbol{z}$ that are generated regarding the state $\boldsymbol{x}$, such that the likelihood function is

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\, \boldsymbol{R})\,, \tag{3.39}$$

where $\boldsymbol{R}$ is the measurement noise covariance and $\boldsymbol{H}$ is the matrix corresponding to the linear mapping of the state into the measurement space. Additionally, it is also assumed that the prior distribution is available in the form of a Gaussian as

$$p^{-}(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^{-},\, \boldsymbol{P}^{-})\,, \tag{3.40}$$

where $\boldsymbol{m}^{-}$ and $\boldsymbol{P}^{-}$ are the mean and covariance, respectively. With the prior and likelihood as given, and using the power of a Gaussian identity from Eq. (A.5), it can be shown that

$$p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{1-\phi}{\phi}} = \left[\frac{\phi}{1-\phi}\right]^{\frac{n_{\boldsymbol{z}}}{2}} \left|2\pi\boldsymbol{R}\right|^{\frac{2\phi-1}{2\phi}} p_g\big(\boldsymbol{x}\big|\boldsymbol{m}^{+},\, \boldsymbol{P}^{+}\big) p_g\left(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m}^{-},\, \boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^{T} + \frac{\phi}{1-\phi}\boldsymbol{R}\right), \tag{3.41}$$

the components of which are defined in Eqs. (3.43), and where the Gaussian component corresponds to an evaluated likelihood, not a pdf. The identity of Eq. (3.41) is very similar to that found by Ho and Lee in [76], and a more detailed derivation can be found in Section B.2. If the expression in Eq. (3.41) is used in the general form of the proposed update from Eq. (3.38), then

$$p^{+}(\boldsymbol{x}) \propto \left[\frac{\phi}{1-\phi}\right]^{\frac{n_{\boldsymbol{z}}}{2}} \left|2\pi\boldsymbol{R}\right|^{\frac{2\phi-1}{2\phi}} p_g\big(\boldsymbol{x}\big|\boldsymbol{m}^{+},\, \boldsymbol{P}^{+}\big) p_g\left(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m}^{-},\, \boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^{T} + \frac{\phi}{1-\phi}\boldsymbol{R}\right). \tag{3.42}$$

Noting that any terms in Eq. (3.42) not dependent on $\boldsymbol{x}$ are eliminated when $p^+(\boldsymbol{x})$ is normalized, it is shown that the resulting posterior distribution is a Gaussian pdf of the form

$$p^+(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^+, \boldsymbol{P}^+)\,, \tag{3.43a}$$

where

$$\boldsymbol{m}^+ = \boldsymbol{m}^- + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}^-) \tag{3.43b}$$

$$\boldsymbol{P}^+ = \boldsymbol{P}^- - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}^- \tag{3.43c}$$

$$\boldsymbol{K} = \boldsymbol{P}^-\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{P}^-\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\right)^{-1}. \tag{3.43d}$$

It is also worth noting that the form of the gain $\boldsymbol{K}$ calculated in Eq. (3.43d) is similar to a form of underweighting mentioned in [77] and successfully used by [78]. This derivation shows possible confidence-based reasoning to support the use of this particular form of underweighting.

3.3.1.2 Scaling Factor for Adaptive Confidence

In order to allow for more autonomy within the filter update, a definition for the confidence factor ϕ is made such that the filter's overconfidence in the prior information vanishes as the estimate uncertainty becomes more compatible with the measurement uncertainty. In other words, ϕ is selected so that as the "difference" between prior and measurement uncertainties transitions from large to small, the update intelligently adapts from one that favors the prior information to one that approaches Bayesian optimality—a behavior that promotes more robust and conservative filtering operations in high uncertainty environments. Thus, ϕ should not be constant, but rather a function of the estimate and measurement uncertainties.

According to the ϕ presented in Eqs. (3.37) and (3.38), when the uncertainty "difference" between prior and measurement grows, the value of ϕ should approach 1^-. Conversely, as the "difference" between the two uncertainties becomes small, the value of ϕ should approach $(1/2)^+$. Under the Gaussian assumption utilized in Eq. (3.43), a useful tool to compare the prior and measurement

uncertainties is the residual covariance, given by

$$\boldsymbol{W} = \boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^{T} + \boldsymbol{R}\,. \tag{3.44}$$

As the "difference" between the prior and measurement uncertainties transitions from large to small, the "ratio" of the measurement noise covariance, $\boldsymbol{R}$, to the the residual covariance, $\boldsymbol{W}$, goes from "zero" to "identity", as $\boldsymbol{P}^{-}$ is symmetric, positive definite and $\boldsymbol{R}$ is symmetric, positive definite. Since ϕ is a scalar value, the ratio between $\boldsymbol{R}$ and $\boldsymbol{W}$ must also be reduced to a scalar. This can be done several ways, such as taking the traces of the matrices, but an appropriate approach should sufficiently capture all values of the covariances, not just the diagonals. Matrix norms are well-known operations that accomplish this, with the most popular being the spectral norm. However, for a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with eigenvalues $\boldsymbol{\lambda} = \lambda_1, \lambda_2, \ldots, \lambda_{\text{rank}(\boldsymbol{A})}$ the spectral norm is given by

$$\|\boldsymbol{A}\|_S = \max\{\boldsymbol{\lambda}\}\,,$$

such that only the largest eigenvalue is accounted for. A more appropriate selection is the Frobenius norm, given by [79]

$$\|\boldsymbol{A}\|_F = \sqrt{\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{\text{rank}(\boldsymbol{A})}^2}\,,$$

which accounts for all eigenvalues of the matrix. Indeed, it is found that taking the ratio of the Frobenius norm of $\boldsymbol{R}$ to the Frobenius norm of $\boldsymbol{W}$ is sufficient, and the ratio of the norms always produces values between zero and one. Therefore, if ϕ is defined as

$$\phi = 1 - \left(\frac{1}{2}\right)\left(\frac{\|\boldsymbol{R}\|_F}{\|\boldsymbol{W}\|_F}\right), \tag{3.45}$$

then ϕ will behave as

$$\lim_{\boldsymbol{W}\to\boldsymbol{\infty}} \phi = 1^{-} \tag{3.46a}$$

$$\lim_{\boldsymbol{W}\to\boldsymbol{R}^{+}} \phi = \left(1/2\right)^{+} . \tag{3.46b}$$

For brevity, the scalar ψ is defined to be

$$\psi = \frac{\|\boldsymbol{R}\|_F}{\|\boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^T + \boldsymbol{R}\|_F} . \tag{3.47}$$

While the properties of ϕ described in Eqs. (3.46) satisfy the conditions desired in the confidence-based update, it is found that the update operates better when ϕ is augmented to produce values closer to $1/2$, as values close to 1 prevent the filter from converging in a timely manner. Therefore, utilizing the definition of ψ in Eq. (3.47), a scaling factor denoted by β and an exponential term $(1-\psi)$ are introduced into Eq. (3.45) to produce

$$\phi = 1 - \frac{1}{2}\psi^{\beta(1-\psi)} , \tag{3.48}$$

where decreasing β drives the confidence-based update closer to Bayes' rule, and increasing β promotes a more conservative posterior. Plotting Eq. (3.48) yields Fig. (3.4), where $\beta = 0.80$ is plotted as (- - -) as it is found to be of particular use in many scenarios, but the exact value of β selected ultimately depends on the system at hand.

3.3.1.3 Extending Closed-Form Confidence-Based Update to Nonlinear Systems

While the closed-form update of Section 3.3.1.1 provides an insightful look into non-Bayesian filtering, as it is an exact filtering solution, it is strictly designed for systems with linear observation models pursuant to Eq. (3.39). Unfortunately, most observation models are nonlinear functions of

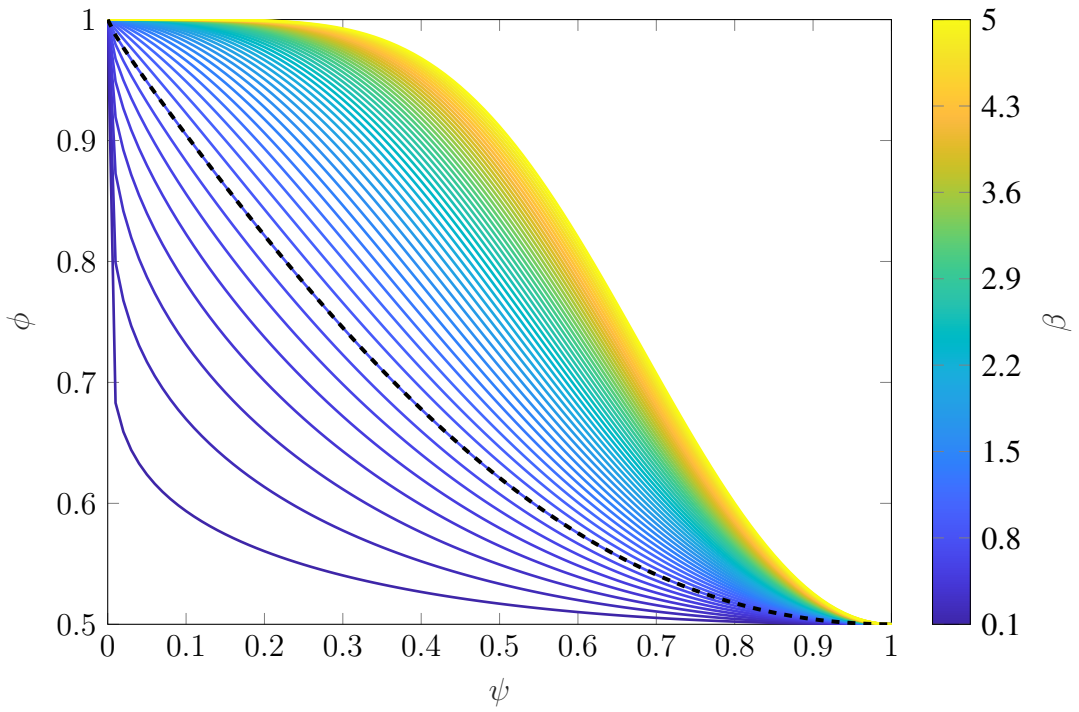


Figure 3.4 Visualization of Eq. (3.48) for varying values of β

$\boldsymbol{x}$ as shown in Eq. (2.1c), such that the measurement likelihood of Eq. (3.39) is replaced by

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{x}),\, \boldsymbol{R})\,. \tag{3.49}$$

In fact, the traditional (Bayesian) Kalman filter is built on the linear measurement distribution of Eq. (3.39), while the EKF assumes the nonlinear measurement distribution of Eq. (3.49), which is precisely the reason why the EKF is considered more useful for practical applications. In this way, just as the EKF is a linearized extension of the KF, a first-order Taylor series approximation can be used to extend the update of Eqs. (3.43) to nonlinear measurement models. Thus, given the prior of Eq. (3.40) and the likelihood of Eq. (3.49), the confidence-based update of Eq. (3.38) can be shown to be

$$p^+(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^+,\, \boldsymbol{P}^+)\,, \tag{3.50a}$$

where

$$\boldsymbol{m}^+ = \boldsymbol{m}^- + \boldsymbol{K}[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m}^-)] \tag{3.50b}$$

$$\boldsymbol{P}^+ = \boldsymbol{P}^- - \boldsymbol{K}\boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^- \tag{3.50c}$$

$$\boldsymbol{K} = \boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-)\boldsymbol{W}^{-1} \tag{3.50d}$$

$$\boldsymbol{W} = \boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-) + \frac{\phi}{1-\phi}\boldsymbol{R}\,, \tag{3.50e}$$

and where $\boldsymbol{H}(\boldsymbol{m}^-)$ is the Jacobian of $\boldsymbol{h}(\cdot)$ evaluated at $\boldsymbol{x} = \boldsymbol{m}^-$. Recalling the definition of ψ from Eq. (3.47), an alternate form of the residual covariance in Eq. (3.50e) is given by

$$\boldsymbol{W} = \boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-) + \big(2\psi^{\beta(\psi-1)} - 1\big)\boldsymbol{R}\,.$$

It should be noted that the identities of Eqs. (A.2) and (A.5) are essential when calculating the posterior of Eq. (3.50a).

3.3.1.4 Scalar Measurement Processing

When processing vector measurement data from sensors, it is not uncommon for there to exist a significant difference in the levels of measurement uncertainty between the scalar components. For instance, a sensor that outputs measurements of angle and range may possess a standard deviation of a fraction of a radian corresponding to the angle component, yet the range may have a standard deviation of thousands of meters or more. Such large disparities between measurement noise uncertainties can make the confidence-based gain of Eqs. (3.50d) and (3.50e) less effective, as it condenses information existing in matrices $\boldsymbol{R}$, $\boldsymbol{H}(\boldsymbol{m}^-)$, and $\boldsymbol{P}$ into the scalar ϕ, which can cause the update to be dominated by a single component of the measurement vector. This disparity is well quantified by the condition numbers of $\boldsymbol{R}$ and $\boldsymbol{W}$, which are given by

$$\kappa_{\text{cond}}(\boldsymbol{R}) = \frac{|\max(\boldsymbol{\lambda}_{\boldsymbol{R}})|}{|\min(\boldsymbol{\lambda}_{\boldsymbol{R}})|} \quad \text{and} \quad \kappa_{\text{cond}}(\boldsymbol{W}) = \frac{|\max(\boldsymbol{\lambda}_{\boldsymbol{W}})|}{|\min(\boldsymbol{\lambda}_{\boldsymbol{W}})|}\,,$$

where $\boldsymbol{\lambda_R}$ and $\boldsymbol{\lambda_W}$ are the vectors of eigenvalues for $\boldsymbol{R}$ and $\boldsymbol{W}$, respectively. If $\boldsymbol{R}$ and $\boldsymbol{W}$ are well conditioned matrices–such that κ_{cond} is close to 1–then the measurement noise and residual uncertainties for each scalar measurement are similarly sized, such that the Frobenius norm approach of Eq. (3.45) works well. However, in cases where $\kappa_{\text{cond}} \gg 1$, the Frobenius norm is mostly dictated by the behavior of the measurement components with larger eigenvalues, with the other measurement components having little effect on the adaptive confidence coefficient ϕ. High condition numbers can be attributed to several things, such as poorly scaled measurement noise covariances or large variations in prior state uncertainty between state components. Therefore, to appropriately calculate the confidence-based update of Section 3.3.1.3 in these conditions, the process of scalar measurement processing is leveraged.

One common trait of measurement models, especially within the field of spacecraft navigation, is that the scalar components of a measurement vector are often uncorrelated, resulting in a diagonal measurement noise covariance $\boldsymbol{R}$. Even in the few instances where this is not the case, it is often possible to formulate the measurement function in such a way that decorrelates the individual components. A frequent application of an uncorrelated $\boldsymbol{R}$ is the scalar processing of vector measurements within a filter [21]. Therefore, under the assumption of uncorrelated measurements, it is possible to apply the confidence-based update to the scalar components of the measurement vector independently by forming an alternate residual covariance $\boldsymbol{U}$ as

$$U_{i,i} = \left(\frac{\phi_i}{1-\phi_i}\right) R_{i,i}\,, \tag{3.51}$$

where $U_{i,i}$ and $R_{i,i}$ are the i^{th} diagonal elements of $\boldsymbol{U}$ and $\boldsymbol{R}$, respectively. In order to reflect the filtering behavior previously described in Section 3.3.1.2, ϕ_i is defined as the scalar counterpart to Eq. (3.48), such that

$$\phi_i = 1 - \frac{1}{2}\psi_i^{\beta(1-\psi_i)}\,, \tag{3.52}$$

where ψ_i is the scalar equivalent of Eq. (3.47) given by

$$\psi_i = \frac{R_{i,i}}{\boldsymbol{H}_i \boldsymbol{P}^- \boldsymbol{H}_i^T + R_{i,i}}\,, \tag{3.53}$$

and where $\boldsymbol{H}_i$ is the i^{th} row of the measurement Jacobian $\boldsymbol{H}(\boldsymbol{m}^-)$. Equations (3.51)–(3.53) are used to form an alternate version of the full confidence-based update—wherein the adaptive confidence treatment of Section 3.3.1.2 is applied to each scalar measurement component individually—with the gain given by

$$\boldsymbol{K} = \boldsymbol{P}^- \boldsymbol{H}^T(\boldsymbol{m}^-) \boldsymbol{W}^{-1} \tag{3.54a}$$

$$\boldsymbol{W} = \boldsymbol{H}(\boldsymbol{m}^-) \boldsymbol{P}^- \boldsymbol{H}^T(\boldsymbol{m}^-) + \boldsymbol{U}\,. \tag{3.54b}$$

Note that the Frobenius norm calculations found in Eq. (3.47) are not present in the calculation of ψ_i in Eq. (3.53), making this form of the confidence-based update more efficient, albeit limited by the assumption of uncorrelated measurements.

3.3.2 Closed-Forms for Nonlinear Filters

The two confidence-based filters derived in Section 3.3.1, while useful, still belong to the class of linear estimators described in Section 2.1. As one of the goals of this research is to promote the use of nonlinear filters—more specifically GM filters as described in Section 2.2.2—this section adapts the robust confidence-based update of Eq. (3.38), recalled here as

$$p^+(\boldsymbol{x}) = \frac{\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha p^-(\boldsymbol{x})}{\int \ell(\boldsymbol{z}|\boldsymbol{s})^\alpha p^-(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}}\,, \tag{3.55}$$

into a GM update, where $\alpha = \frac{1-\phi}{\phi}$ is substituted for brevity. To achieve a GM realization of Eq. (3.55), the prior $p^-(\boldsymbol{x})$ is assumed to take the form

$$p^-(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-,\ \boldsymbol{P}_\xi^-)\,, \tag{3.56}$$

where w_ξ^-, $\boldsymbol{m}_\xi^-$, and $\boldsymbol{P}_\xi^-$ correspond to the weight, mean, and covariance of the ξ^{th} GM component, respectively. Much like Section 3.3.1, two updates are derived in the remainder of this section; the first update assumes a linear-Gaussian measurement model, whereas the second update assumes the measurement model is nonlinear.

3.3.2.1 GM Update for Linear-Gaussian Systems

For this specific GM update, the measurement likelihood $\ell(\boldsymbol{z}|\boldsymbol{x})$ is taken to be the linear-Gaussian form of Eq. (3.39), recalled here as

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\ \boldsymbol{R})\,. \tag{3.57}$$

Therefore, taking the numerator of Eq. (3.55) and substituting in Eqs. (3.56) and (3.39), it can be shown that

$$\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha p^-(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\ \boldsymbol{R})^\alpha p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-,\ \boldsymbol{P}_\xi^-)\,. \tag{3.58}$$

Using Eq. (A.5) to rewrite the power of a Gaussian distribution, Eq. (3.58) can be reformed as

$$\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha p^-(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^- |\alpha|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi\boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g\left(\boldsymbol{z}\middle|\boldsymbol{H}\boldsymbol{x},\ \frac{1}{\alpha}\boldsymbol{R}\right) p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-,\ \boldsymbol{P}_\xi^-)\,. \tag{3.59}$$

Utilizing Ho's identity from Eqs. (A.2) and noting that a linear substitution can easily be made such that $\boldsymbol{h}(\boldsymbol{m}) = \boldsymbol{H}\boldsymbol{m}$, Eq. (3.59) can be rearranged to show

$$\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha p^-(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^- |\alpha|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi\boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g\left(\boldsymbol{z}\middle|\boldsymbol{H}\boldsymbol{m}_\xi^-, \boldsymbol{H}\boldsymbol{P}_\xi^-\boldsymbol{H}^T + \frac{1}{\alpha}\boldsymbol{R}\right) p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+,\ \boldsymbol{P}_\xi^+)\,, \tag{3.60}$$

where

$$\begin{aligned}
\boldsymbol{m}_\xi^+ &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}_\xi^-) \\
\boldsymbol{P}_\xi^+ &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H} \boldsymbol{P}_\xi^- \\
\boldsymbol{K}_\xi &= \boldsymbol{P}_\xi^- \boldsymbol{H}^T \Big(\boldsymbol{H} \boldsymbol{P}_\xi^- \boldsymbol{H}^T + \frac{1}{\alpha} \boldsymbol{R} \Big)^{-1} .
\end{aligned}$$

Now, evaluating the denominator, Eq. (3.60) can be integrated over the state $\boldsymbol{x}$ such that

$$\int \ell(\boldsymbol{z}|\boldsymbol{s})^\alpha p^-(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} \quad = \sum_{\xi=1}^{L^-} w_\xi^- |\alpha|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi \boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g\Big(\boldsymbol{z} \Big| \boldsymbol{H}\boldsymbol{m}_\xi^-, \boldsymbol{H}\boldsymbol{P}_\xi^- \boldsymbol{H}^T + \frac{1}{\alpha}\boldsymbol{R} \Big) . \quad (3.61)$$

It is worth mentioning that the factor α is not explicitly state-dependent, as it is removed from the integral.

Taking the results from Eqs. (3.60) and (3.61) and substituting back into the confidence-based update of Eq. (3.55) yields

$$p^+(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^+ p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+,\ \boldsymbol{P}_\xi^+) \,, \quad (3.62a)$$

where

$$w_\xi^+ = \frac{w_\xi^- \kappa_\xi}{\sum_{i=1}^{L^-} w_i^- \kappa_i} \quad (3.62b)$$

$$\boldsymbol{m}_\xi^+ = \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}_\xi^-) \quad (3.62c)$$

$$\boldsymbol{P}_\xi^+ = \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H} \boldsymbol{P}_\xi^- \quad (3.62d)$$

$$\boldsymbol{K}_\xi = \boldsymbol{P}_\xi^- \boldsymbol{H}^T \boldsymbol{W}_\xi^{-1} \quad (3.62e)$$

$$\boldsymbol{W}_\xi = \boldsymbol{H} \boldsymbol{P}_\xi^- \boldsymbol{H}^T + \frac{1}{\alpha} \boldsymbol{R} \quad (3.62f)$$

$$\kappa_\xi = |\alpha|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi \boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g\big(\boldsymbol{z} \big| \boldsymbol{H}\boldsymbol{m}_\xi^-, \boldsymbol{W}_\xi \big) \quad (3.62g)$$

$$\alpha = \frac{1-\phi}{\phi} \,. \quad (3.62h)$$

3.3.2.1.1 State-dependency of α

If it is desired to model α as a state-dependent value, such that $\alpha = \alpha(\boldsymbol{x})$, it should be noted that this will produce many complications when seeking a closed-form solution, as the integral of Eq. (3.61) is no longer be straight-forward to evaluate. In cases such as these, it is advised to use a zeroth-order approximation of the form $\alpha(\boldsymbol{x}) \approx \alpha(\boldsymbol{m})$, an approach that is discussed in greater detail in Section 4.4.2.3. For this reason, the term $|\alpha|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi \boldsymbol{R}\}^{\frac{1-\alpha}{2}}$ is kept within the definition of κ_ξ in Eq. (3.62g), as it may be defined to differ for each Gaussian component and measurement, making it dependent on the summation over ξ. In this case, the components of the update in Eqs. (3.62e)–(3.62g) become

$$\begin{aligned}
\boldsymbol{K}_\xi &= \boldsymbol{P}_\xi^- \boldsymbol{H}^T \boldsymbol{W}_\xi^{-1} \\
\boldsymbol{W}_\xi &= \boldsymbol{H} \boldsymbol{P}_\xi^- \boldsymbol{H}^T + \frac{1}{\alpha_\xi} \boldsymbol{R} \\
\kappa_\xi &= |\alpha_\xi|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi \boldsymbol{R}\}^{\frac{1-\alpha_\xi}{2}} p_g\big(\boldsymbol{z} \big| \boldsymbol{H} \boldsymbol{m}_\xi^-, \boldsymbol{W}_\xi\big) ,
\end{aligned}$$

where $\alpha_\xi = \alpha(w_\xi^-, \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-)$ denotes the relation between α_ξ and the ξ^{th} GM component. Recalling Eq. (3.62h), it is important to note that α_ξ can also be expressed as

$$\alpha_\xi = \frac{1 - \phi_\xi}{\phi_\xi} ,$$

which can be directly compared to the GVI form of the confidence-based update of Eq. (3.37), and for which adaptive forms have already been investigated by Section 3.3.1.2.

3.3.2.2 GM Update for Nonlinear Systems

Similar to the GM update for linear measurements of Section 3.3.2.1, the update presented in this section is also a confidence-based posterior $p^+(\boldsymbol{x})$ resulting from Eq. (3.55) subject to the GM prior $p^-(\boldsymbol{x})$ of Eq. (3.56). However, instead of the linear-Gaussian measurement model of

Eq. (3.57), the nonlinear measurement model of Eq. (3.49) is taken, recalled here as

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{x}),\, \boldsymbol{R})\,.$$

Utilizing Eqs. (A.5) and (A.2), and following steps similar to Section 3.3.2.1, it can be shown that

$$p^+(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^+ p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+,\, \boldsymbol{P}_\xi^+)\,, \tag{3.63}$$

where

$$\begin{aligned}
w_\xi^+ &= \frac{w_\xi^- k_\xi}{\sum_{i=1}^{L^-} w_i^- k_i} \\
\boldsymbol{m}_\xi^+ &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m}_\xi^-)] \\
\boldsymbol{P}_\xi^+ &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H}(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^- \\
\boldsymbol{K}_\xi &= \boldsymbol{P}_\xi^- \boldsymbol{H}^T(\boldsymbol{m}_\xi^-)\boldsymbol{W}_\xi^{-1} \\
\boldsymbol{W}_\xi &= \boldsymbol{H}(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^- \boldsymbol{H}^T(\boldsymbol{m}_\xi^-) + \frac{1}{\alpha_\xi}\boldsymbol{R} \\
\kappa_\xi &= |\alpha_\xi|^{-\frac{n_{\boldsymbol{z}}}{2}} \det\{2\pi \boldsymbol{R}\}^{\frac{1-\alpha_\xi}{2}} p_g\big(\boldsymbol{z}\big|\boldsymbol{h}(\boldsymbol{m}_\xi^-), \boldsymbol{W}_\xi\big) \\
\alpha_\xi &= \frac{1-\phi_\xi}{\phi_\xi}\,,
\end{aligned}$$

and where $\phi_\xi = \phi(w_\xi^-, \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-)$ can be defined to achieve some specific behavior, which is discussed in Section 3.3.1.2.

3.4 Application to Navigation

As previously discussed in Section 1.3.1, filtering has seen long-time use in the field of spacecraft navigation, with many practices being established specifically for the needs of spaceflight missions. One such well-established method is known as underweighting, which is presented in greater detail in this section, as one of the goals of this work is to improve upon current navigation

methodologies. The second-order EKF (SOEKF) is also included in this section, as it possesses a higher-order update that is useful for comparison purposes.

3.4.1 Navigation and the Underweighted EKF

Along with issues such as unmodeled effects and sensor failures, a common point of failure in linear estimators such as the EKF is the linearization itself. As mentioned previously, the EKF—emphasized in particular due to its popularity in spacecraft navigation—utilizes a first-order Taylor series approximation to derive the update of Eqs. (2.4), which is a valid approximation under the condition that any effects of second-order or higher remain negligible. And indeed, these higher-order effects are typically negligible under small updates: the residual covariance $\boldsymbol{W}$ of Eq. (2.4e) is comparable in size to the corresponding measurement noise covariance $\boldsymbol{R}$. Unfortunately, many applications frequently encounter situations where a large prior uncertainty is combined with a relatively precise measurement, resulting in a update that may produce substantial higher-order terms, such that neglecting them causes the filtering solution to become overconfident and, in many cases, fail by way of filter divergence [56]. To ensure robust operations, this issue cannot go unaddressed. As a solution, the spacecraft navigation community created underweighting to protect against these problems caused by large updates, and the technique has found regular use still today [6, 7, 80].

Underweighting is an *ad hoc*, procedure-first robustness technique, as it was developed as a direct solution to the aforementioned issues stemming from the EKF linearization. The entire methodology is based on the concept that during sufficiently large updates, such that higher-order effects may come into play, instead of performing the standard EKF update per Eqs. (2.4), the gain $\boldsymbol{K}$ of Eq. (2.4d) is "underweighted" which slows the rate by which information is gained by the filter. This effectively enlarges the updated covariance $\boldsymbol{P}^{+}$ of Eq. (2.4c) artificially, which generally makes the filtering solution under-confident (or conservative). However, conservative estimation is usually considered acceptable (within reason) provided the filter is less likely to become smug (or overconfident).

While underweighting can be accomplished in a variety of ways, there are certain variants

of underweighting considered to be "best practice" [21, 77, 80]. Common to most is a test on the components of the EKF update from Eqs. (2.4), which determines if underweighting need be applied. A preferred underweighting scheme has certain characteristics, which are typically

- the "size" of the update—or disparity between $\boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^{T}$ and $\boldsymbol{R}$—is inherently accounted for by the check,
- vector measurements are naturally treated by the operation, and
- the algorithms only require a single parameter be specified.

Fortunately, with respect to the EKF, these criteria can all be achieved simultaneously by replacing the residual covariance of Eq. (2.4e) with the underweighted alternative [19, 58]

$$\boldsymbol{W}_{\text{uw}} = \gamma_{\text{uw}}\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-}) + \boldsymbol{R}\,, \tag{3.64a}$$

where the time index k is omitted for brevity, and where

$$\gamma_{\text{uw}} = \begin{cases} 1\,, & \text{if } \|\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})\| > \frac{1-p_{\text{uw}}}{p_{\text{uw}}}\|\boldsymbol{R}\| \\ p_{\text{uw}}\,, & \text{if } \|\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})\| \leq \frac{1-p_{\text{uw}}}{p_{\text{uw}}}\|\boldsymbol{R}\| \end{cases}. \tag{3.64b}$$

The scalar p_{uw} is a user-defined design parameter that is typically set after some filter analysis and tuning. In order to enforce a "softer" update, the parameter p_{uw} should be assigned some value greater than one, as $p_{\text{uw}} = 1$ would yield an underweighted update that uses the Kalman gain, and $p_{\text{uw}} < 1$ would enforce an overconfident gain. For both Orion and the Shuttle program, $p_{\text{uw}} = 1.2$ was deemed a sufficient value such that most failures caused by large updates are avoided [7, 81].

3.4.2 Second-Order Extended Kalman Filter

While underweighting has indeed been shown to work well, instead of neglecting second-order (and higher) terms and artificially inflating the posterior covariance, a more informed procedure can be performed that includes second-order terms. This class of filter, appropriately referred to

as the second-order extended Kalman filter (SOEKF), neglects the Taylor series terms that are of the third-order and higher, thus allowing nonlinear effects up to the second-order to form the filtering solution. Accordingly, the SOEKF requires calculating second-order partial derivatives of the functions being approximated, which is often difficult and tedious.

In this work, the SOEKF equations are used as an additional benchmark against which the confidence-based update is compared. The propagation stage of the SOEKF is given by [10]

$$\begin{aligned}
\boldsymbol{m}_k^- &= \boldsymbol{f}(\boldsymbol{m}_{k-1}^+) + \frac{1}{2}\sum_{i=1}^{n_{\boldsymbol{x}}} \boldsymbol{e}_i \,\mathrm{tr}\Big\{ \boldsymbol{F}_{xx}^{(i)}(\boldsymbol{m}_{k-1}^+)\boldsymbol{P}_{k-1}^+ \Big\} \\
\boldsymbol{P}_k^- &= \boldsymbol{F}(\boldsymbol{m}_{k-1}^+)\boldsymbol{P}_{k-1}^+\boldsymbol{F}^T(\boldsymbol{m}_{k-1}^+) \\
&\quad + \frac{1}{2}\sum_{i=1}^{n_{\boldsymbol{x}}}\sum_{i'=1}^{n_{\boldsymbol{x}}} \boldsymbol{e}_i\boldsymbol{e}_{i'}^T \,\mathrm{tr}\Big\{ \boldsymbol{F}_{xx}^{(i)}(\boldsymbol{m}_{k-1}^+)\boldsymbol{P}_{k-1}^+ \Big(\boldsymbol{F}_{xx}^{(i')}(\boldsymbol{m}_{k-1}^+)\Big)^T \boldsymbol{P}_{k-1}^+ \Big\} + \boldsymbol{Q}_k \,,
\end{aligned} \tag{3.65}$$

where $n_{\boldsymbol{x}}$ is the number of states in $\boldsymbol{x}$, $\boldsymbol{e}_i$ is a column vector of zeros with a 1 in the i^{th} row, and $\boldsymbol{F}_{xx}^{(i)}(\cdot)$ is the Hessian matrix of the i^{th} row of the dynamic function, $f_i(\cdot)$, which is given by

$$\Big[\boldsymbol{F}_{xx}^{(i)}(\boldsymbol{m}_{k-1}^+)\Big]_{j,j'} = \left.\frac{\partial^2 f_i(\boldsymbol{x}_{k-1})}{\partial x_j \partial x_{j'}}\right|_{\boldsymbol{x}_{k-1}=\boldsymbol{m}_{k-1}^+} .$$

Note that $\Big[\boldsymbol{F}_{xx}^{(i)}(\cdot)\Big]_{j,j'}$ denotes the component corresponding to the j^{th} row and the $j^{\text{'th}}$th column of the $\boldsymbol{F}_{xx}^{(i)}(\cdot)$ matrix. The update portion of the SOEKF algorithm is given by [10]

$$\begin{aligned}
\boldsymbol{m}_k^+ &= \boldsymbol{m}_k^- + \boldsymbol{K}_k\left(\boldsymbol{z}_k - \boldsymbol{h}_k(\boldsymbol{m}_k^-) - \frac{1}{2}\sum_{i}^{n_{\boldsymbol{z}}} \boldsymbol{e}_i \,\mathrm{tr}\Big\{ \boldsymbol{H}_{xx}^{(i)}(\boldsymbol{m}_k^-)\boldsymbol{P}_k^- \Big\} \right) \\
\boldsymbol{P}_k^+ &= \boldsymbol{P}_k^- - \boldsymbol{K}_k\boldsymbol{S}_k\boldsymbol{K}_k^T \,,
\end{aligned} \tag{3.66}$$

where $n_{\boldsymbol{z}}$ is the number of scalar measurements in $\boldsymbol{z}_k$. The residual covariance and filter gain are

defined as

$$\boldsymbol{S}_k = \boldsymbol{H}(\boldsymbol{m}_k^-)\boldsymbol{P}_k^-\boldsymbol{H}^T(\boldsymbol{m}_k^-) + \frac{1}{2}\sum_{i=1}^{n_{\boldsymbol{z}}}\sum_{i'=1}^{n_{\boldsymbol{z}}} \boldsymbol{e}_i\boldsymbol{e}_{i'}^T \,\mathrm{tr}\Big\{\boldsymbol{H}_{xx}^{(i)}(\boldsymbol{m}_k^-)\boldsymbol{P}_k^-\Big(\boldsymbol{H}_{xx}^{(i')}(\boldsymbol{m}_k^-)\Big)^T\boldsymbol{P}_k^-\Big\} + \boldsymbol{R}_k \tag{3.67}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_k^-\boldsymbol{H}^T(\boldsymbol{m}_k^-)\boldsymbol{S}_k^{-1}\,, \tag{3.68}$$

where $\boldsymbol{H}_{xx}^{(i)}(\cdot)$ is the Hessian matrix of the i^{th} row of the measurement function, $h_i(\cdot)$, and is given by

$$\Big[\boldsymbol{H}_{xx}^{(i)}(\boldsymbol{m}_k^-)\Big]_{j,j'} = \left.\frac{\partial^2 h_i(\boldsymbol{x}_k)}{\partial x_j \partial x_{j'}}\right|_{\boldsymbol{x}_k=\boldsymbol{m}_k^-}.$$

Note that $\Big[\boldsymbol{H}_{xx}^{(i)}(\cdot)\Big]_{j,j'}$ denotes the component corresponding to the j^{th} row and the $j^{\text{'th}}$ column of the $\boldsymbol{H}_{xx}^{(i)}(\cdot)$ matrix.

3.5 Analysis of Linear Confidence-based Update

In order to test the capabilities of linear non-Bayesian updates, two simulations are designed to specifically compare the performance of an EKF equipped with the proposed, confidence-based update to both an underweighted EKF and a second-order EKF.

3.5.1 Falling Body Simulation

In this analysis, the falling body simulation of Fig. (2.4) is revisited. To be clear, all of the system models and constraints are as given in Section 2.3.2 with a few exceptions. As this analysis will look at the residual behavior of various filters, a single run is performed instead of multiple MC trials. Additionally, in an attempt to cause the filters to fail, a high level of uncertainty is generated by enforcing a measurement gap between 10 and 25 seconds, wherein it is assumed that the observer is unable to receive range measurements of the object. As this measurement gap significantly reduces the number of measurements in the simulation, the measurement frequency is increased to once every half second, and the simulation duration is extended to 40 seconds.

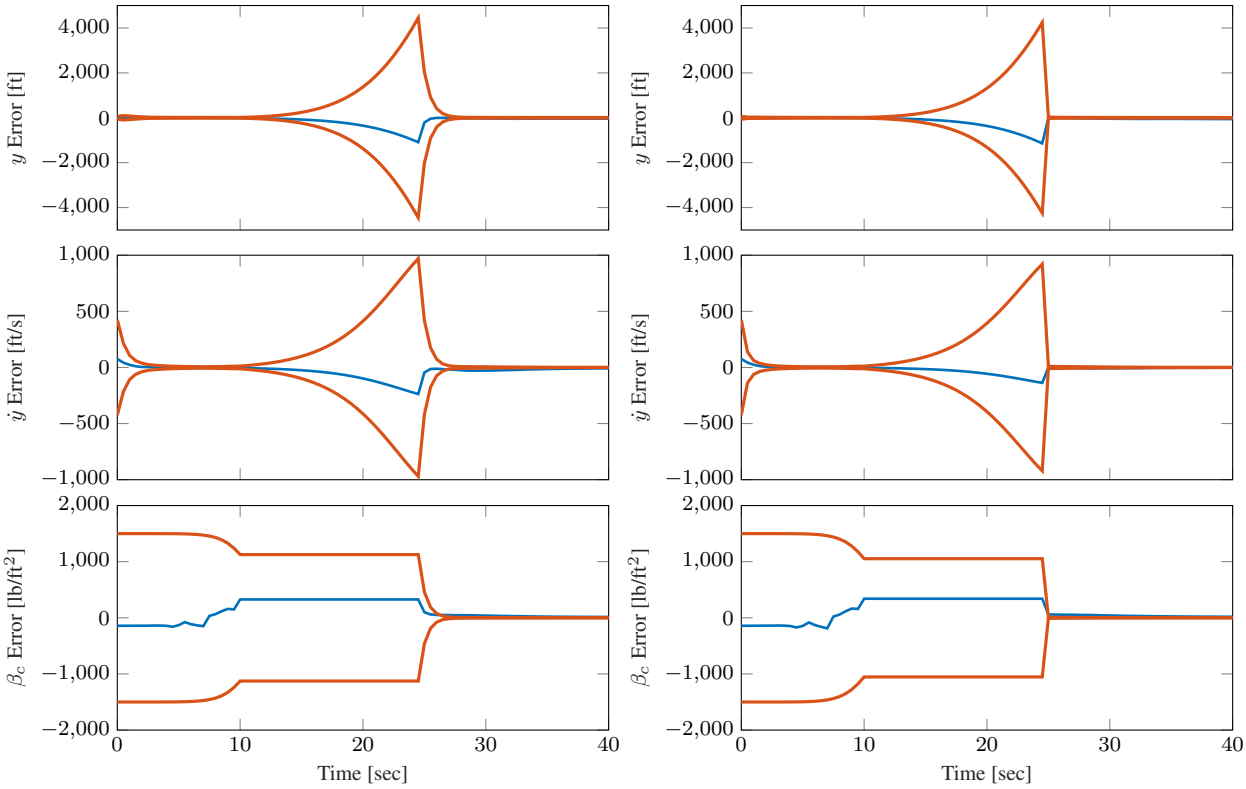
3.5.1.1 Filter Configurations

For this simulation, the performance of an EKF with the proposed update found in Eq. (3.50) is compared to that of an EKF with the underweighted update described in Eq. (3.64), as well as an SOEKF described in Eqs. (3.65)–(3.67). The scaling factor β—defined in Eq. (3.48) for the confidence-based update—is given a value of 0.8, and the underweighting factor p_{uw} from Eq. (3.64b) is set to 1.2. Each filter is supplied with the initial estimate given by Eq. (2.32) and the entire measurement set. A single simulation run is performed for all three filters, for which the initial conditions and measurement history is identical.

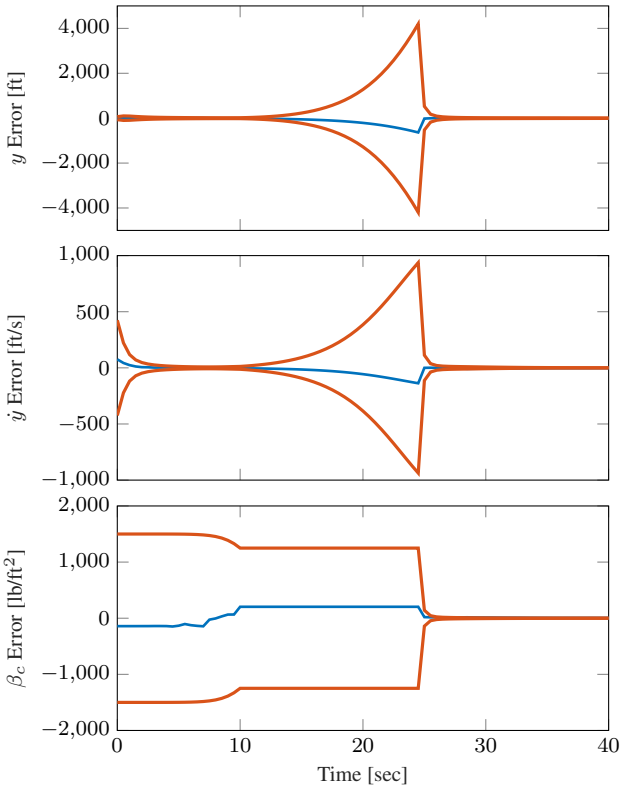
As the primary interest of this work is to compare the robustness of the proposed update to comparable, pre-established filters specifically built for handling environments of high uncertainty, it is somewhat difficult to find instances that cause the filters to fail, even with the excessive gap in measurement data. Note here that a filter is considered to fail if the estimate and measurement residuals leave a 3σ uncertainty interval for an appreciable amount of time. In general, it is found that only about one of every five instances of the simulation cause any filters to fail, and in about half of those cases, all of the filters fail.

3.5.1.2 Simulation Analysis

The estimation errors for a single run of the simulation are presented in Fig. (3.5) for the three filters, where a more detailed view is given in Fig. (3.6). The effect of the measurement gap between 10 and 25 seconds is clear in both the position and velocity channels, as the uncertainty grows quite large during this time frame. At 25 seconds, there is an immediate "snap-down" of the covariance in all three channels, which is the result of a sudden accurate measurement being received. Figures (3.5b) and (3.6b) show that the second-order update suffers the sharpest reduction in covariance at this time, with the underweighted filter appearing, at first, to have the most gradual uncertainty reduction. In terms of estimation performance, it is evident that while the errors for both the underweighted EKF and the SOEKF begin to drift outside the 3σ interval, the errors for the confidence-based filter do not.

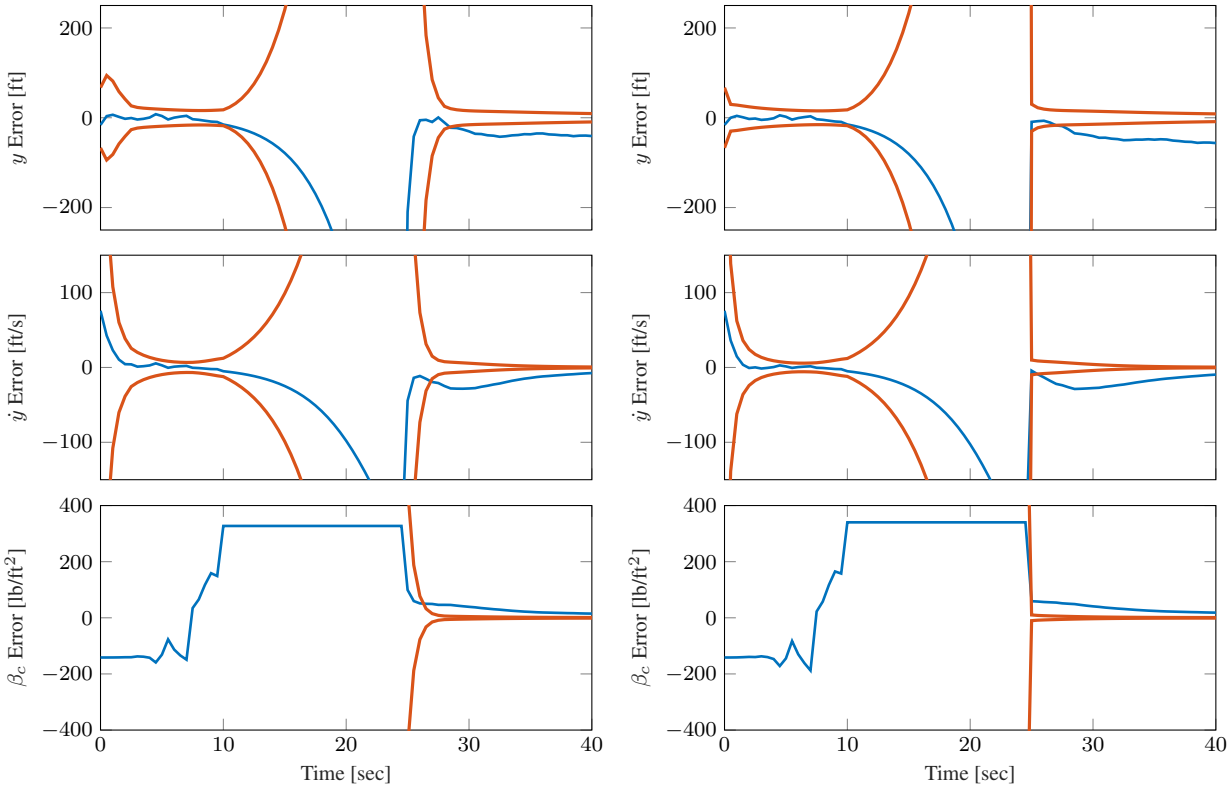


a) Underweighted update **b) Second-order update**

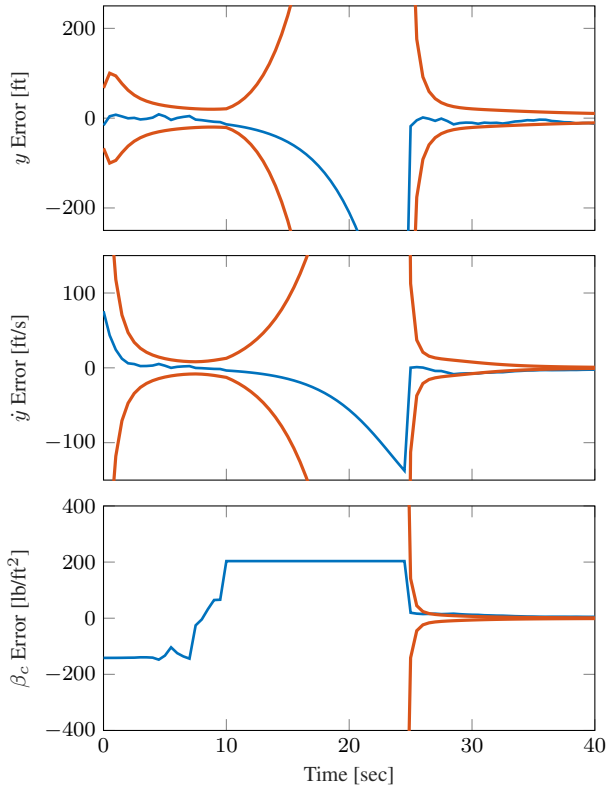


c) Confidence-based update

Figure 3.5 Comparison of filtering results, expressed as estimation error (——) and 3σ standard deviations of error covariance (——)



a) Underweighted update **b) Second-order update**



c) Confidence-based update

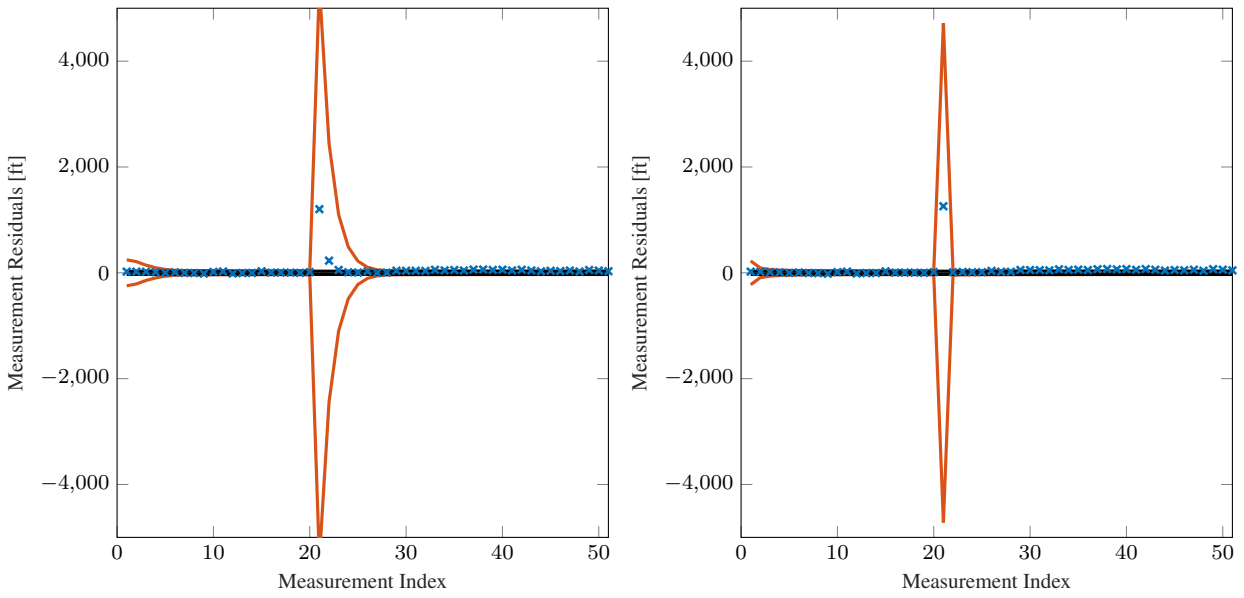
Figure 3.6 Zoomed-in view of the comparison of filtering results, expressed as estimation error (——) and 3σ standard deviations of error covariance (——)

Figures (3.7) and (3.8) contain the measurement residuals and the 3σ interval of the residual covariance, defined in Eq. (3.44). The residuals of both the underweighted EKF and the SOEKF are seen to drift outside of the 3σ interval in Figs. (3.8a) and (3.8b), explaining the poor performance observed in Figs. (3.6a) and (3.6b). In contrast, the residuals for the adaptive confidence filter in Fig. (3.8c) remain within the 3σ confidence intervals for the duration of the simulation. An especially interesting behavior is observed in Fig. (3.7), where the intensity of the covariance "snap-downs" is most prevalent. Here, it is clear from Fig. (3.7b) that the residual covariance of the SOEKF experiences the sharpest reduction in covariance—a trait of the second-order update previously observed in Fig. (3.6b).

A more unexpected result is found when comparing Figs. (3.7a) and (3.7c); the uncertainty of the adaptive confidence update experiences a faster reduction than that of the underweighted update. In general, a more conservative filter tends to be more robust than a less conservative one, which is the notion motivating the common practice of inflating the covariance via methods such as tuning noise and underweighting. However, for this select instance, it is apparent in Fig. (3.7) that the adaptive confidence update is actually less conservative than the underweighted update, even while producing a more accurate estimate. In short, the non-Bayesian update is *more* robust and *less* conservative.

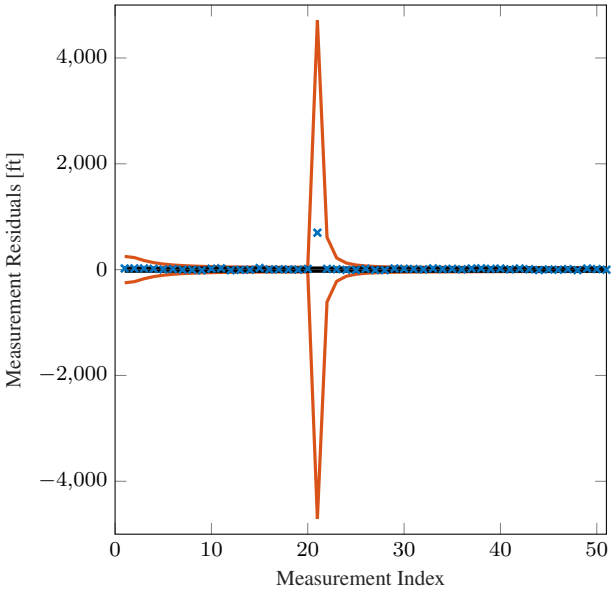
Figure (3.8) shows a more detailed view of the residual covariance behavior later on in the simulation. It is seen in Fig. (3.8c) that the residual covariance of the confidence-based update does become more conservative than that of the underweighted one, as it takes more time to reduce to a level of steady state. This is attributed to the fact that at this point in the simulation, the uncertainty in Fig. (3.8a) no longer surpasses the underweighting threshold, and, therefore, the update is no longer being underweighted. However, the status of the residuals in Fig. (3.8c) is significantly healthier than those of Figs. (3.8a) and (3.8b), which quickly fall outside the 3σ interval, a testament to the robust nature of the proposed update.

Here, it is worthwhile to make a brief discussion concerning the results of the SOEKF. Recall that second-order filters, by design, improve filtering operations by explicitly including second-



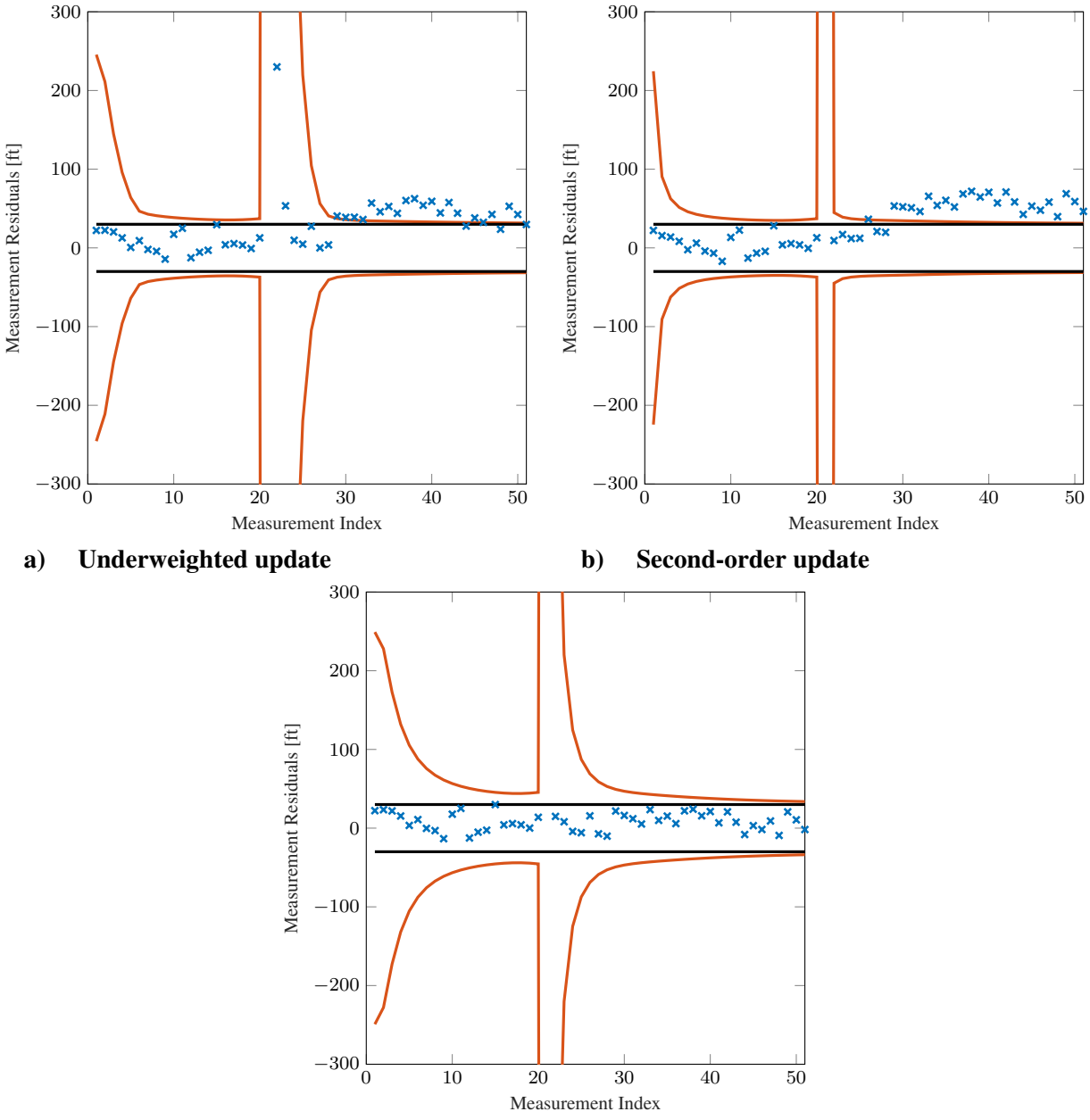
a) Underweighted update

b) Second-order update



c) Adaptive confidence update

Figure 3.7 Comparison of measurement residuals (×), plotted alongside 3σ representations of the residual covariance (——) and measurement noise covariance (——)



a) Underweighted update

b) Second-order update

c) Adaptive confidence update

Figure 3.8 Zoomed-in view of the comparison of measurement residuals (×), plotted alongside 3σ representations of the residual covariance (—) and measurement noise covariance (—)

order terms that are known to cause errors in filters built on first-order linearization, such as the EKF. In this way, it is generally more robust than a standard EKF, albeit at the cost of an increase in computational complexity. As expected, additional runs of the simulation show that the SOEKF is, in fact, the most robust filter overall—failing the least out of the three filters (the SOEKF fails in about 10% of runs) . However, there are cases where the SOEKF is outperformed, such as the run presented in Figs. (3.5)–(3.8). Being that second-order effects are supposedly accounted for by the SOEKF, its failure to maintain an accurate estimate must be attributed to some other cause. Based on this, it is a reasonable conjecture that there are errors in this specific run caused by third-order and higher terms such that the filtering solution of the SOEKF is degraded. As second-order filters themselves are generally avoided due to the tedious second-order derivatives and computational expense, it is unreasonable to solve this problem by deriving third-order or higher filters. Instead, it is more practical to implement methods such as underweighting or the proposed adaptive confidence update, as these circumvent the need for such complex filters while still enforcing robust estimation.

3.5.2 Relative Satellite Motion Simulation

In addition to the falling body simulation, this work also investigates the comparative performance of the confidence-based update via the well-known system of relative motion between two satellites, shown in Fig. (3.9). As the main aspect of interest in this paper is comparing the update stages of the filter, this particular system is specifically chosen for its linear dynamics, allowing for algorithmically identical estimate propagations between the different filters.

3.5.2.1 Configuration

The state vector consists of three-dimensional Cartesian position and velocity in the $\mathcal{E} = \{\hat{\boldsymbol{e}}_x, \hat{\boldsymbol{e}}_y, \hat{\boldsymbol{e}}_z\}$ frame of Fig. (3.9), written as

$$\boldsymbol{x} = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} \end{bmatrix}^T . \tag{3.69}$$

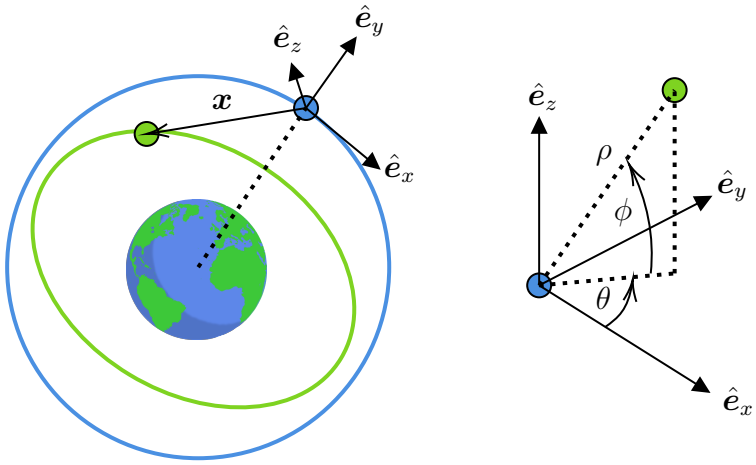


Figure 3.9 Reference frame for relative motion of two satellites

The system's dynamics are described by the Clohessy-Wiltshire equations, which can be found in [82], and there is no process noise included in the system. The observation model is taken to consist of two relative angle measurements (θ and ϕ) and a range ρ given by

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \theta \\ \phi \\ \rho \end{bmatrix} = \begin{bmatrix} \tan^{-1}\left(\frac{y}{x}\right) \\ \sin^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right) \\ \sqrt{x^2+y^2+z^2} \end{bmatrix}, \tag{3.70}$$

where measurements are generated according to the Gaussian model of Eq. (2.1c) with the corresponding measurement noise covariance being

$$\boldsymbol{R} = \begin{bmatrix} 10^{-3}\ \text{rad}^2 & 0 & 0 \\ 0 & 10^{-3}\ \text{rad}^2 & 0 \\ 0 & 0 & 0.5\ \text{m}^2 \end{bmatrix}. \tag{3.71}$$

The initial estimate is described by a multivariate Gaussian distribution with mean and covariance given by

$$\boldsymbol{m}_0 = \begin{bmatrix} \boldsymbol{m}_{0,\mathrm{p}} \\ \boldsymbol{m}_{0,\mathrm{v}} \end{bmatrix} = \begin{bmatrix} 55000 \text{ m} \\ 45000 \text{ m} \\ 40000 \text{ m} \\ -2149 \text{ m/s} \\ -1863 \text{ m/s} \\ -1600 \text{ m/s} \end{bmatrix} \tag{3.72}$$

and

$$\boldsymbol{P}_0 = \begin{bmatrix} \boldsymbol{P}_{0,\mathrm{p}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{0,\mathrm{v}} \end{bmatrix} = \begin{bmatrix} 10^7 \text{ m}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10^7 \text{ m}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10^7 \text{ m}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16031(\text{m/s})^2 & 0.5755(\text{m/s})^2 & 0 \\ 0 & 0 & 0 & 0.5755(\text{m/s})^2 & 16004(\text{m/s})^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 15991(\text{m/s})^2 \end{bmatrix}, \tag{3.73}$$

respectively.

In order to force rendezvous scenarios, only the initial positional components of the mean, $\boldsymbol{m}_{0,\mathrm{p}}$, and covariance, $\boldsymbol{P}_{0,\mathrm{p}}$, are user-specified. The velocity components of the initial mean, $\boldsymbol{m}_{0,\mathrm{v}}$, are calculated as a linear function of the initial positions such that $\boldsymbol{m}_0$ results in a rendezvous, where this linear function is then used to transform the positional covariance $\boldsymbol{P}_{0,\mathrm{p}}$ into the appropriate velocity covariance $\boldsymbol{P}_{0,\mathrm{v}}$ [82]. Accordingly, the positional elements of the initial true state are drawn from a multivariate Gaussian distribution given by

$$\boldsymbol{x}_{0,\mathrm{p}} \sim p_g(\boldsymbol{x}_{0,\mathrm{p}} | \boldsymbol{m}_{0,\mathrm{p}},\, \boldsymbol{P}_{0,\mathrm{p}}) \,,$$

and the initial true velocity states $\boldsymbol{x}_{0,\mathrm{v}}$ are calculated using the same linear function used to ensure rendezvous, such that the overall initial truth is $\boldsymbol{x}_0^T = [\boldsymbol{x}_{0,\mathrm{p}}^T \; \boldsymbol{x}_{0,\mathrm{v}}^T]$.

Each run lasts a duration of 20 seconds, where measurements are processed at a frequency of 2 Hz. The setup for the three filters used in the previous simulation of Section 3.5.1 is used here as well, with the exception that the confidence-based update is equipped with the scalar measurement treatment from Eqs. (3.51)–(3.54). In order to evaluate the consistency of the filters, a Monte Carlo simulation of 2500 trials is performed, where the Monte Carlo statistics are compared to the average estimates for each filter, per the discussion of Section 2.3.1.

3.5.2.2 Performance Comparison

The results from this simulation highlight the ability of the linear confidence-based update to perform more conservatively than both the SOEKF and the EKF with underweighting. Figures (3.10)–(3.12) show comparisons between the average filter standard deviations $\sigma_{\mathrm{filt},\iota}$ compared to the computed MC standard deviations $\sigma_{\mathrm{MC},\iota}$, where both are plotted as 3σ intervals. Recall that $\sigma_{\mathrm{filt},\iota} < \sigma_{\mathrm{MC},\iota}$ is indicative of filter overconfidence, whereas conservative behavior in a filter is denoted by $\sigma_{\mathrm{filt},\iota} > \sigma_{\mathrm{MC},\iota}$. While slightly conservative filters are generally acceptable in spacecraft navigation, overconfident filters are typically avoided at all costs.

Upon examining Fig. (3.10a), it is apparent that the underweighted EKF generates increasingly overconfident position estimates beginning around seven seconds, and both the SOEKF of Fig. (3.10b) and the confidence-based update of Fig. (3.10c) outperform underweighting by a large margin. The position estimate performance of the SOEKF and the proposed update can be seen in greater detail in Figs. (3.11b) and (3.11c), respectively. Here it is seen that the second-order update becomes slightly overconfident, even with the higher order calculations. The confidence-based update, however, does not suffer from overconfidence to the degree that the SOEKF does, and the average filter standard deviations end up matching the Monte Carlo statistics much more closely.

Similar behavior is examined in the velocity profiles of Fig. (3.12), where the underweighted EKF displays the most overconfidence, while the confidence-based update appears to be the least overconfident. Altogether, these results indicate that the confidence-based filter is less prone to

overconfidence than the traditional underweighted EKF.

3.6 Analysis of Nonlinear Confidence-based GM Update

In order to test the robustness of the nonlinear confidence-based update of Eq. (3.63), the relative motion simulation of Section 3.5.2 is adapted to incorporate model mismatch. The simulation parameters remain unchanged from Section 3.5.2.1 with the exception of measurements being generated as faulty 1% of the time according to Section 4.5.2.2, such that measurements are statisically distributed by

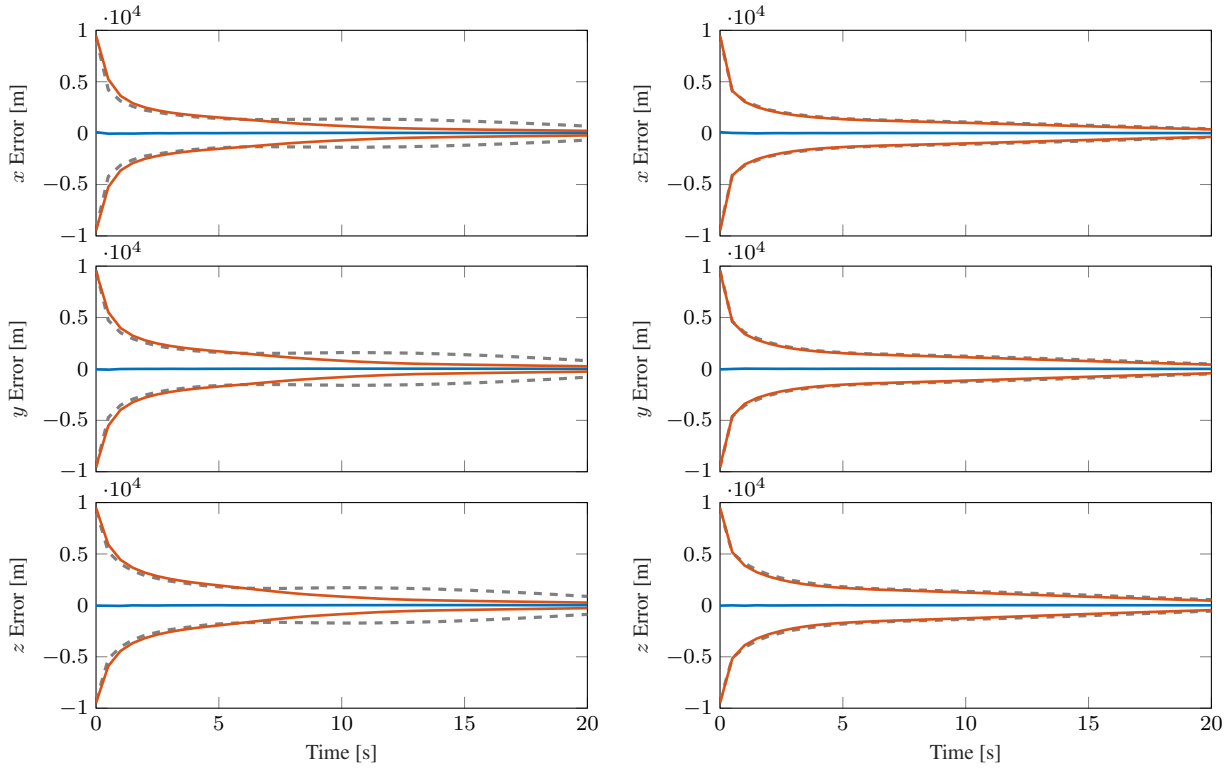
$$\boldsymbol{z}_k \sim v p_g(\boldsymbol{z}_k|\boldsymbol{h}(\boldsymbol{x}_k), \boldsymbol{R}) + [1-v] p_g(\boldsymbol{z}_k|\boldsymbol{h}(\boldsymbol{x}_k), \boldsymbol{R}_f)\,,$$

where $\boldsymbol{h}(\cdot)$ is the observational model of Eq. (3.70), $\boldsymbol{R}$ is the measurement noise covariance of Eq. (3.71), $\boldsymbol{R}_f = 100\boldsymbol{R}$ is the faulty measurement noise covariance, and $v = 0.99$ is the probability of validity, described in greater detail by Section 4.4.2. Two analyses are performed to evaluate the GM confidence-based filter under different forms of ϕ, where each analysis consists of 2500 MC trials, as described by Section 2.3.1.

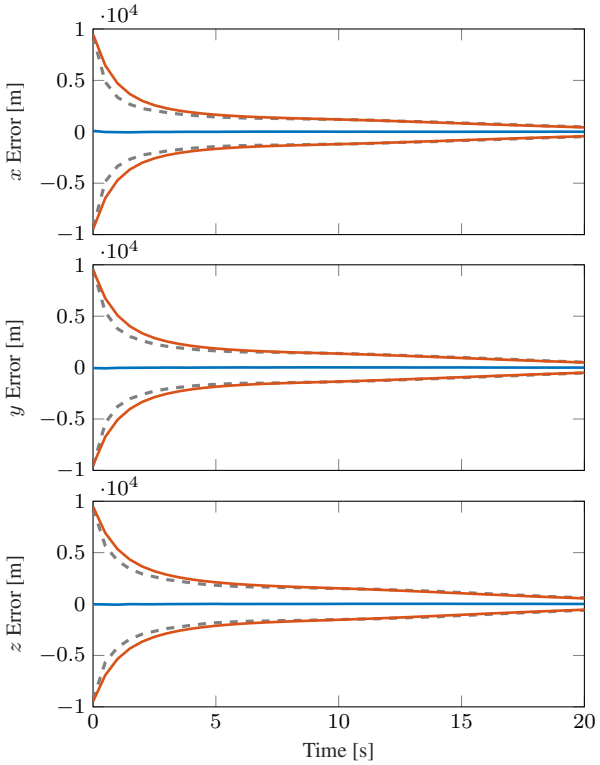
3.6.1 Analysis 1: Static ϕ

The two filters tested in this simulation are the second-order update and the nonlinear confidence-based updates, where the second-order update is configured identically to that in Section 3.5.2.1. The nonlinear confidence-based filter uses the GSF propagation of Section 2.2.2.4.1 and the GM update of Section 3.3.2.2, where $\phi_\xi = 0.7$ is selected to increase filter robustness in the presence of faulty measurements. To initialize the GM filter, the initial mean and covariance of Eqs. (3.72) and (3.73) are split five ways across the three positional dimensions according to Section 2.2.3.3, resulting in an initial GM of 125 components. Note that splitting is only used to initialize the filter and is not applied within the iterative update, such that the number of GM components at any given iteration is $L_k^+ \leq 125$. To avoid zero-weighted components, the pruning procedure of Section 2.2.3.1 is employed with $w_{\text{thresh}} = 10^{-9}$.

The MC results of this analysis are provided in Figs. (3.13) and (3.14) for the position and ve-

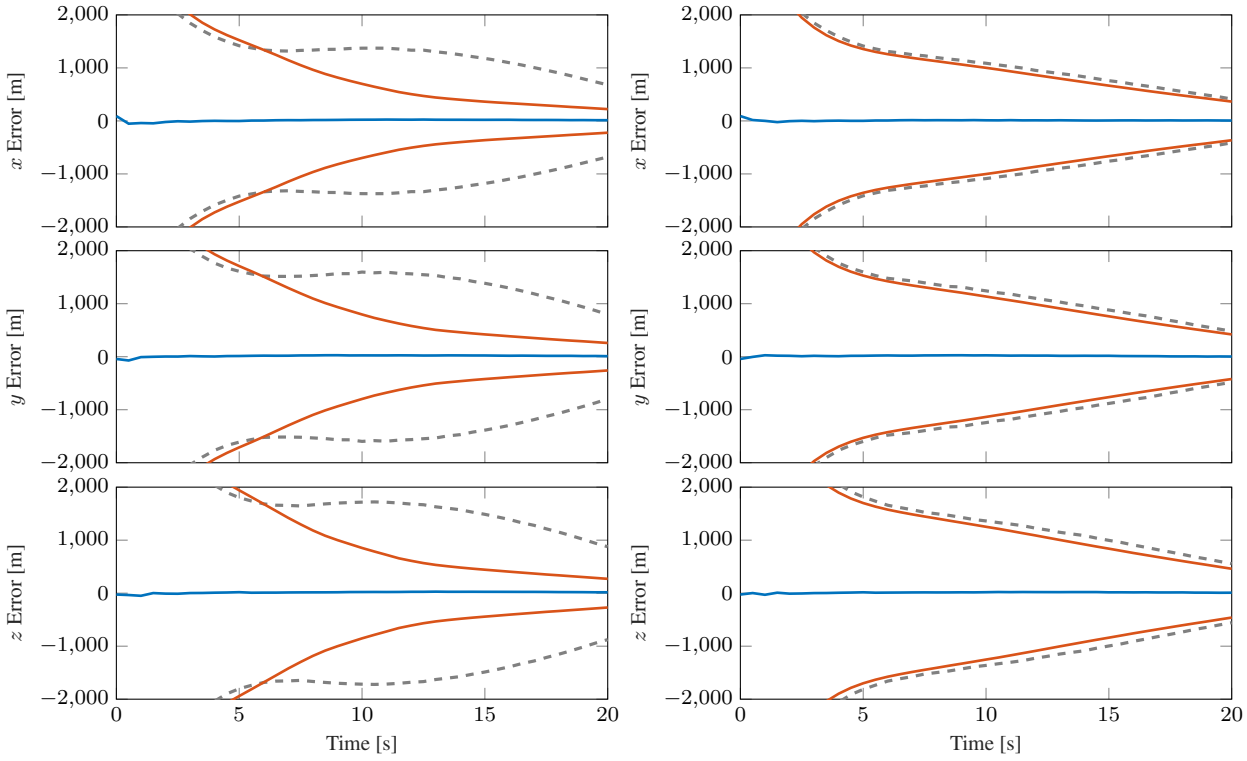


a) Underweighted update **b) Second-order update**

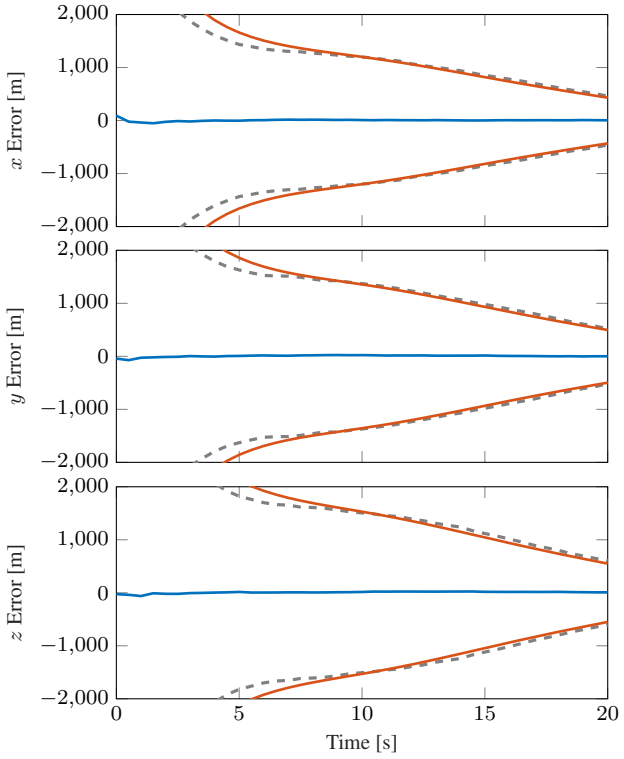


c) Linear confidence-based update

Figure 3.10 MC results for satellite rendevous plotted as $\bar{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of position components

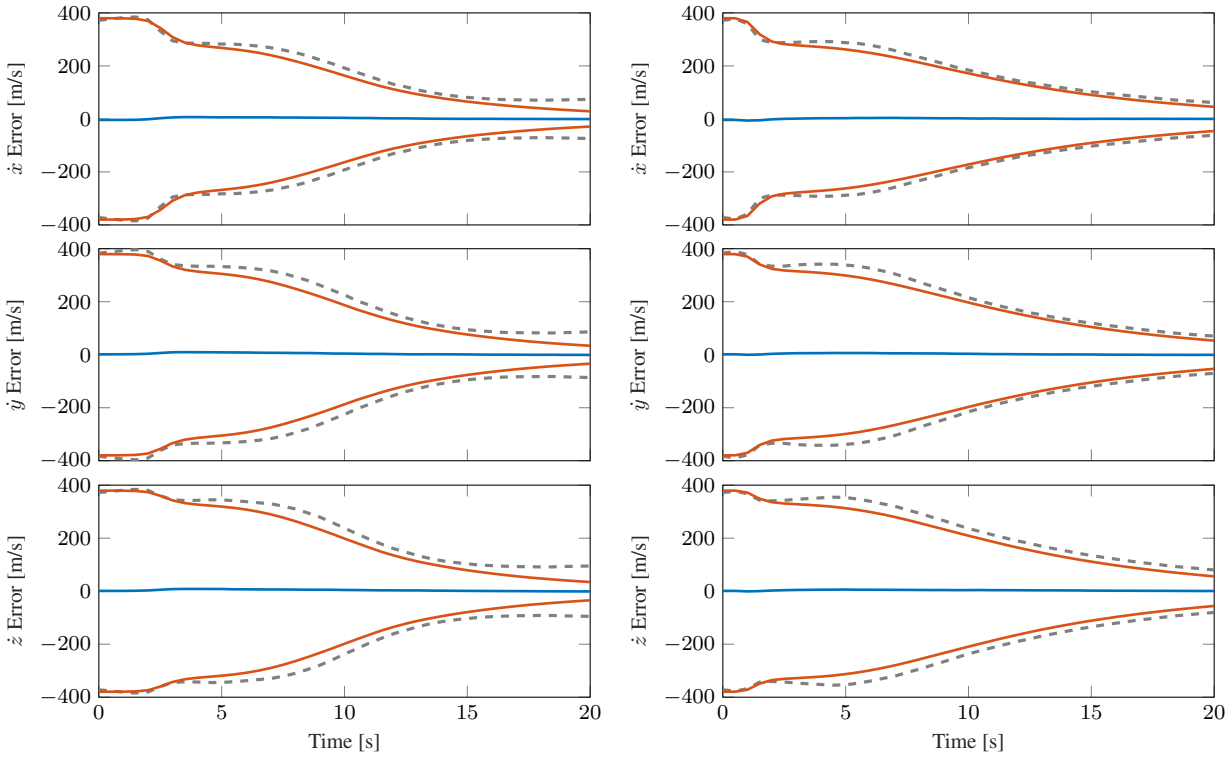


a) Underweighted update **b) Second-order update**

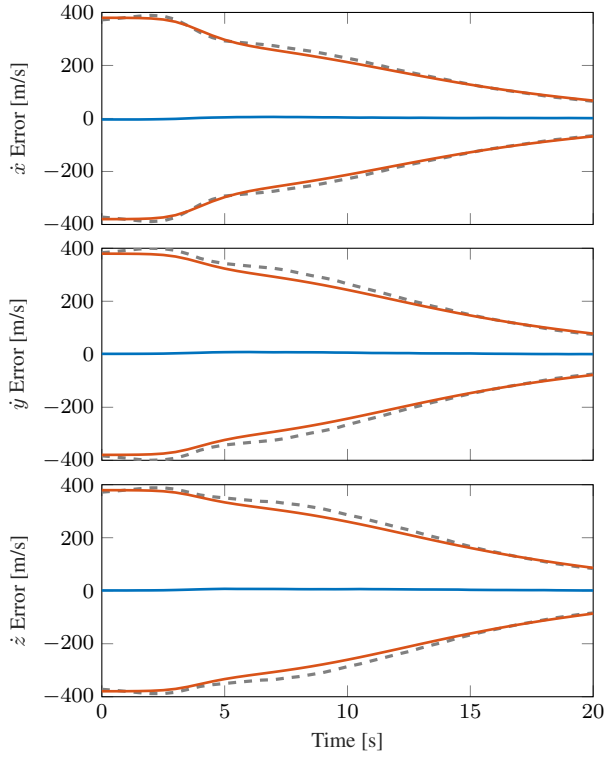


c) Linear confidence-based update

Figure 3.11 Zoomed-in view of MC results for satellite rendevous plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of position components



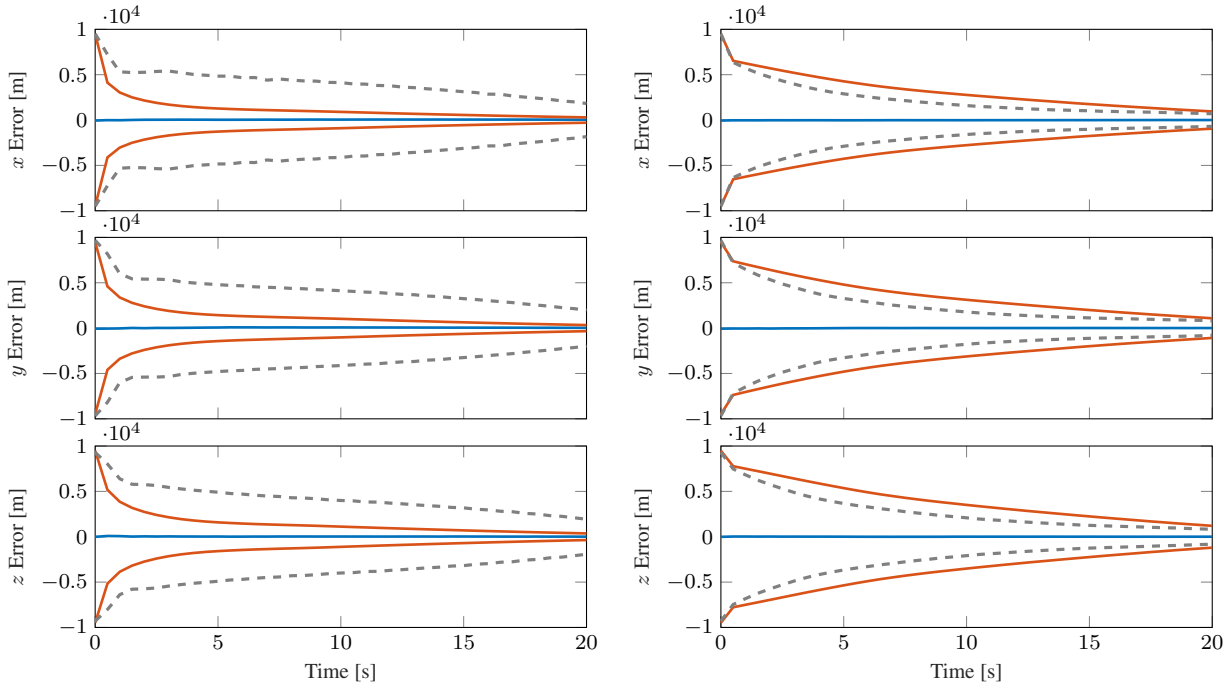
a) Underweighted update **b) Second-order update**



c) Linear confidence-based update

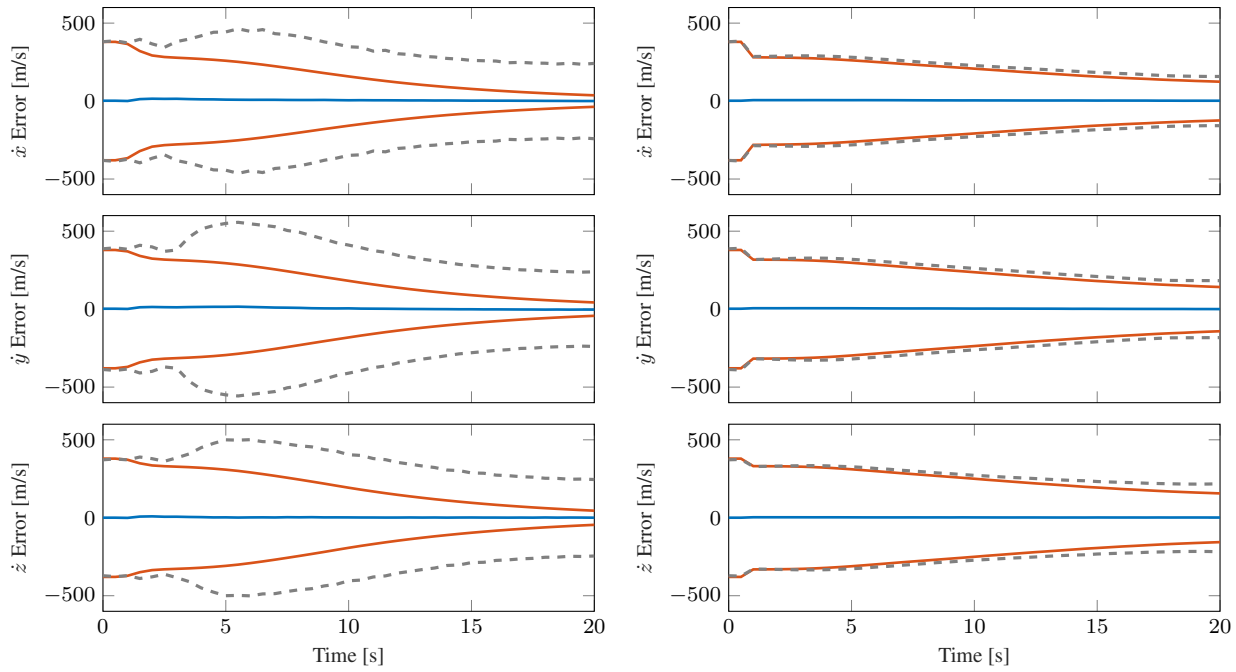
Figure 3.12 MC results for satellite rendevous plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of velocity components

locity components, respectively. Recalling the SOEKF performance of Figs. (3.10b) and (3.12b), upon the inclusion of measurement model mismatch, the performance of the second-order filter is degraded to that of Figs. (3.13a) and (3.14a), where the average MC standard deviations now far exceed those estimated by the SOEKF. While the average errors of Figs. (3.13a) and (3.14a) remain relatively low, the filter itself is extremely overconfident. The GM confidence-based filter (GM-CBF), on the other hand, exhibits healthier behavior, remaining conservative in the positional errors of Fig. (3.13b) and exhibiting slightly overconfident behavior in the velocity errors of Fig. (3.14b). Note that increasing the severity of the model mismatch—which in this case is done by decreasing v–results in higher errors that may force the GMCBF into overconfident estimation. To prevent this, it is advisable to tune the filter, which is relatively straightforward for a static value of ϕ. By doing so, a desired level of conservatism can be imparted to the GMCBF.



a) Second-order update **b) GM confidence-based update**

Figure 3.13 Model mismatch MC results for relative motion simulation with static ϕ plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of position components



a) Second-order update **b) GM confidence-based update**

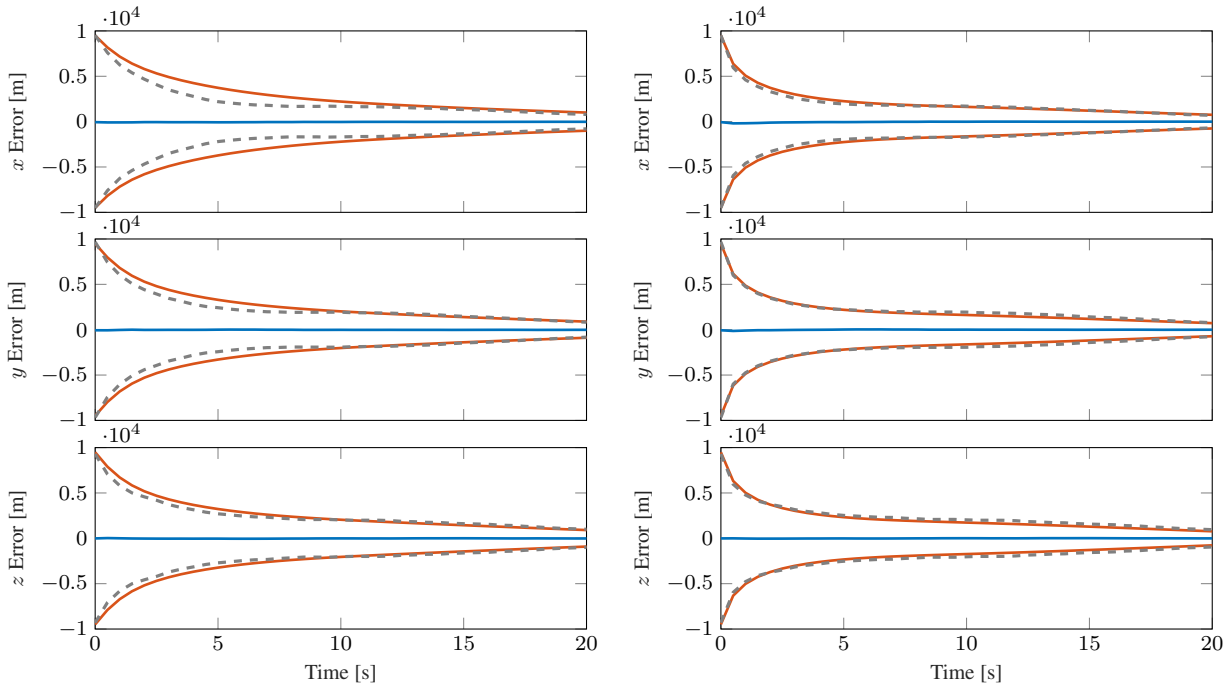
Figure 3.14 Model mismatch MC results for relative motion simulation with static ϕ plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of velocity components

3.6.2 Analysis 2: Adaptive ϕ

While the preceding analysis of Section 3.6.1 evaluates the GMCBF under a set value of ϕ, much work is done in Section 3.3.1.2 to develop an adaptive form of ϕ. Therefore, this analysis extends the adaptive ϕ to the same GMCBF as Section 3.6.1 and simultaneously evaluates the linear confidence-based update of Section 3.3.1.3 for comparison under model mismatch. Both filters utilize the ϕ defined in Eq. (3.48) with $\beta = 1.1$, the effect of which is seen in Fig. (3.4). The MC results of this analysis are provided in Figs. (3.15) and (3.16), which show that both confidence-based filters perform well when subjected to measurements that are faulty 1% of the time. In fact, the performance of both filters is remarkably similar, with the main difference between the two being a faster drop in positional uncertainty of the GM filter of Fig. (3.15b) when compared to the linear filter of Fig. (3.15a). The GM filter is also found to have slightly lower MC errors overall, which is most evident in the final velocity standard deviations of Fig. (3.16b) when compared

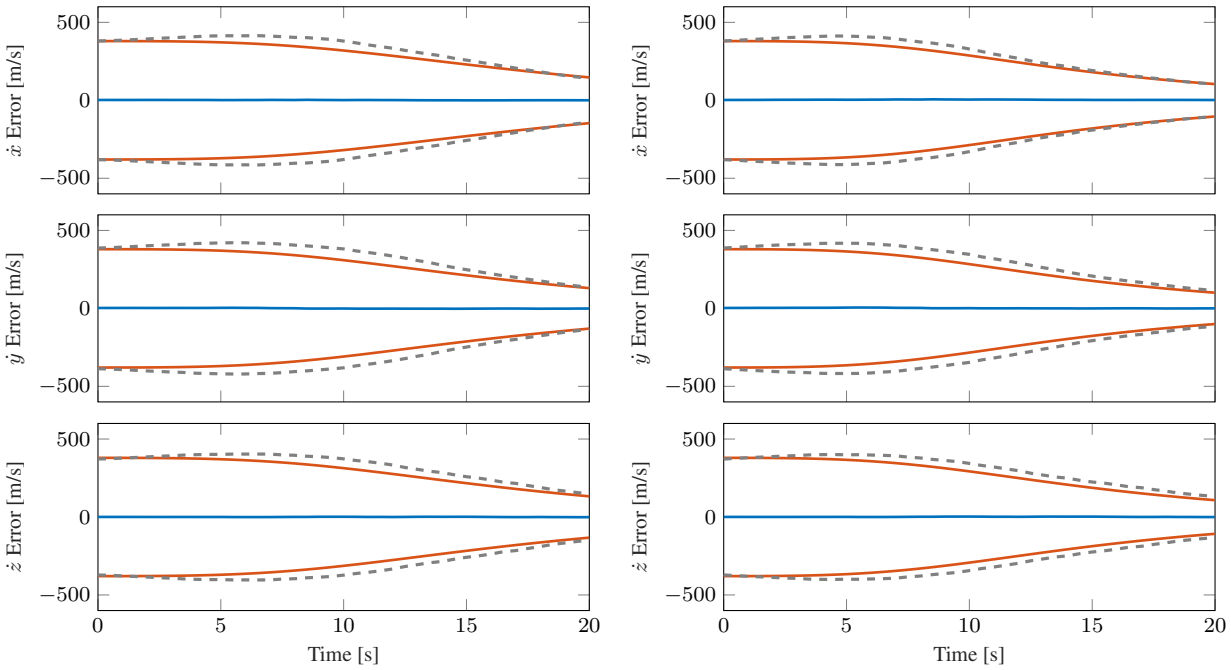
to Fig. (3.16a). These slight improvements to the filter estimate are attributed to the benefits of nonlinear filtering; unlike the linear confidence-based filter, the GMCBF is capable of decreased linearization error and non-Gaussian pdfs of the state. Therefore, while both filters perform well, the improved performance of the GMCBF is to be expected.

Aside from the superior performance of the GMCBF, of additional note is its change in behavior when constructed with an adaptive ϕ rather than the static ϕ of Section 3.6.1. Examining the positional elements first, Fig. (3.13b) indicates that the static ϕ GMCBF is much more conservative than its adaptive counterpart in Fig. (3.15b). However, the adaptive ϕ produces positional estimates with $\sigma_{\text{filt},\iota} \approx \sigma_{\text{MC},\iota}$ that are more accurate overall. The behavior of the velocity channel estimates of the GMCBF also differs significantly between Figs. (3.14b) and (3.16b). Whereas the static ϕ becomes more overconfident as the simulation progresses, the adaptive ϕ begins reducing overconfidence around 10 seconds. Regardless of which behavior is deemed preferable, these comparisons highlight the importance that ϕ has on the nature of the confidence-based posterior.



a) Linear confidence-based update **b) GM confidence-based update**

Figure 3.15 Model mismatch MC results for relative motion simulation with adaptive ϕ plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of position components



a) Linear confidence-based update **b) GM confidence-based update**

Figure 3.16 Model mismatch MC results for relative motion simulation with adaptive ϕ plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -) of velocity components

4. FAULT-COGNIZANT FILTERING*

Whereas Chapter 3 investigates how removing inherent Bayesian assumptions can result in more robust filters, this chapter is concerned with the level of fidelity and overall approach by which measurement generation is modeled. Instead of simply ignoring the manner in which faulty measurements are generated—eliminating their effects via faulty measurement screening—what benefits are there in attempting to model these faults? This chapter answers this question by developing filters that are fault-cognizant, beginning by reconsidering how the measurement likelihood is traditionally handled. It is quickly apparent that, depending on the system at hand, the measurement likelihood can be customized to account for a wide variety of sensors and faults, such that, by appropriately adjusting the assumptions of the sensor model, multiple fault-cognizant measurement models (FCMMs) are produced, each with a corresponding likelihood. In turn, each likelihood results in a unique fault-cognizant update (FCU) when evaluated via Bayes' rule, closed-forms of which are readily realized through the GMMs of Section 2.2.2. The performance of these proposed filters is then tested through multiple simulations, where they are compared to other well-established filters. Note that a portion of this work is previously published as [83].

4.1 Traditional Measurement Modeling

Commonly referred to as the measurement likelihood, $\ell(\boldsymbol{z}|\boldsymbol{x})$ of Eqs. (1.6) and (3.1) describes the probability of a given state $\boldsymbol{x}$ based on information observed from measurement $\boldsymbol{z}$. In fact, from a statistical standpoint, the measurement $\boldsymbol{z}$ is a vector of random variables that is stochastically generated as

$$\boldsymbol{z} \sim \ell(\boldsymbol{z}|\boldsymbol{x})\,. \tag{4.1}$$

*Part of the material reported in this chapter is reprinted with permission from "Nonlinear Gaussian Mixture Filtering with Intrinsic Fault Resistance" by Gunner S. Fritsch and Kyle J. DeMars, 2021. Journal of Guidance, Control, and Dynamics, 44, 2172–2185, Copyright [2021] by Gunner Fritsch and Kyle J. DeMars.

The precise form that the likelihood of Eq. (4.1) takes is entirely determined by the modeling assumptions inferred from the system at hand. For example, considering the notation of $\ell(\boldsymbol{z}|\boldsymbol{x})$ alone, the assumptions that $\boldsymbol{z}$ is a single measurement corresponding to $\boldsymbol{x}$, a single state (or single target), has already been established. A more general form of the measurement likelihood could be $\ell(\boldsymbol{Z}|\boldsymbol{X})$, where $\boldsymbol{Z}$ is some set (or scan) of measurements, and $\boldsymbol{X}$ is some set of targets/states. In many applications, including spacecraft navigation, the primary interest is a single vector of states (i.e., $\boldsymbol{X} \rightarrow \boldsymbol{x}$) that generate observations via measurement models that have usually been structured such that single measurements can be sequentially processed by the filter (i.e., $\boldsymbol{Z} \rightarrow \boldsymbol{z}$). There are also plenty of applications that do not operate within these specific limitations, such as the field of multi-target multi-sensor fusion [84], which operates using measurement likelihoods and algorithms built on different assumptions, such as the inclusion of missed detections and false alarms [48]. Regardless of the specific application or sensor suite at hand, correctly administering these assumptions is the core element to deriving filtering algorithms, which begins with defining the measurement likelihood in this chapter.

Correct Likelihood Terminology

A likelihood $\ell(A|B)$ describes the joint probability of A as a function of B. While the term $\ell(\boldsymbol{z}|\boldsymbol{x})$ is dubbed the measurement likelihood in this work, it is important to note that $\ell(\boldsymbol{z}|\boldsymbol{x})$ is technically a likelihood of the state $\boldsymbol{x}$ and a pdf of the measurement $\boldsymbol{z}$.

While still acknowledging the existence of a wide variety of possible assumptions, it is apparent that a suprisingly large number of filtering methodologies are built upon a relatively narrow set of assumptions. Much like the frequent use of Bayes' rule to construct filter updates, most measurement models, especially in the realm of navigation, are constructed upon two core assumptions. Firstly, that all incoming measurements are assumed valid and generated by a single valid likelihood distribution as

$$\boldsymbol{z}_v \sim \ell_v(\boldsymbol{z}_v|\boldsymbol{x})\,, \tag{4.2}$$

and secondly that the filter ingests a single measurement per update (possibly via preprocessing of the sensor output). Figure (4.1) is a visualization of the described traditional measurement model, which is useful as a reference when more advanced measurement models are proposed later on. Whether realistic or not, Fig. (4.1) emphasizes the fact that, according to this traditional model, the measurement space consists of a single valid measurement $\boldsymbol{z}_v$, which is assumed always detected and reported to the filter. If it is additionally assumed that the sensor noise is additive-

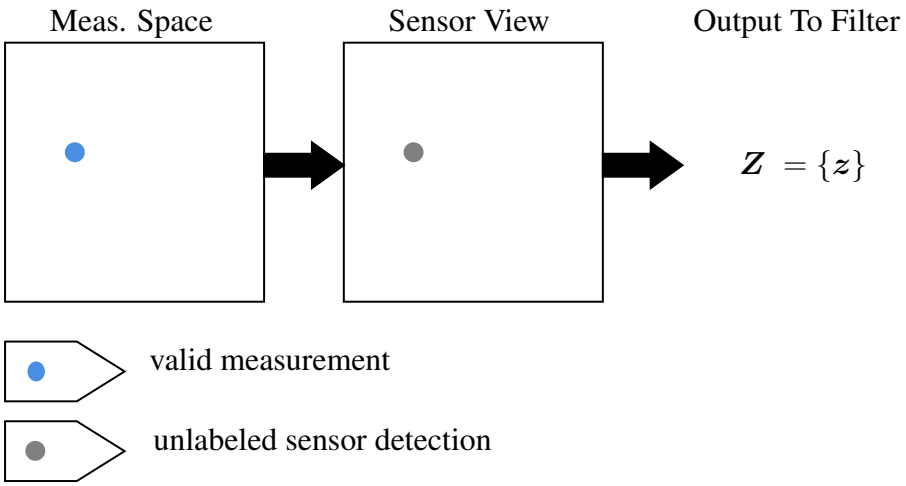


Figure 4.1 Visualization of a traditional measurement model

Gaussian, this entire collection of assumptions necessarily constrains the measurement model to a form similar to Eq. (2.1c), which is fine in ideal sensing conditions when measurements are perfectly described by their models. Unfortunately, this is rarely the case, as unmodeled effects and computational errors frequently produce sensor returns that do not align with the simplistic model of Eq. (2.1c)—measurements that are classified as "faulty" by this work.

4.2 The Reality of Faulty Measurements

Faulty measurements, by definition, are difficult to identify and diagnose. This is a direct result of good sensor development; sensors are designed to produce consistent and accurate measurements, keeping faults and erroneous returns to a minimum whenever possible. In general, the more

infrequently faulty measurements occur, the more difficult they are to model. Of course, certain sensors may be more susceptible to faulty measurements than others, such as vision-based cameras which can be subject to large amounts of clutter across their sensor space, or low-cost sensors that are designed to be cheaply manufactured at the potential cost to measurement consistency. Regardless of the specific sensor, however, it is always bad practice not to account for faulty measurements to some degree; indeed, as previously mentioned in Section 1.2, most practical filtering applications employ some level of operational robustness to protect against these faults, though it is typically a procedure-first approach such as residual editing. Most of these procedure-first methods are disadvantaged by binary decisions: either a measurement is included or it is not. This binary operation presents a couple of theoretical issues for the filter, as it is essentially informing the filter that any approved measurement is guaranteed to be valid and that any edited measurement is guaranteed to be false. For instance, consider a faulty measurement that, for whatever reason, is not flagged as faulty by the filter. In this case, as the likelihood of Eq. (4.2) is assumed, the filter wrongly classifies this data as a valid measurement, and from that point forward, the filtering solution has no means of amending this fallacious measurement history. Similarly, valid sensor returns may be incorrectly classified as faulty, such that no information whatsoever is gained from that sensor return. Due to cases such as these, it can be argued that such procedure-first robustness techniques only work well in situations where there is a clear difference between faulty data and valid data, and even then, the performance is completely dependent upon tuning parameters instead of statistical models built specifically for the sensor/system at hand. As such, this chapter seeks a model-first approach to fault-resistance, which begins at the level of measurement modeling. For the purposes of this work, faulty measurements are spatially independent and identically distributed (IID) as

$$\boldsymbol{z}_f \sim \ell_f(\boldsymbol{z}_f|\boldsymbol{x})\,, \tag{4.3}$$

the exact nature of which will be discussed further in Section 4.5.

4.3 Fault-Cognizant Measurement Models

Recalling the typical sensor modeling assumptions mentioned in Section 4.1, which results in algorithms that typically rely on screening incoming measurements to ensure robustness, what if the assumptions were changed such that faulty measurements were modeled directly? This would require extra effort when developing sensor models, but would permit a more nuanced approach to the treatment of faulty returns and could potentially decrease the amount of filter tuning required by procedure-first robustness while simultaneously improving filtering accuracy. Ideally, faulty data handling needs to possess two characteristics: (1) soft decisions that allow the filter to process the complete set of incoming measurements [85], and (2) intrinsic existence within the filtering framework, potentially requiring less human interference and thus promoting greater levels of autonomy. The probability hypothesis density (PHD) filter is an excellent example of such a technique, as it employs models that include, but down-weight, measurements with low probability of being correct [41, 86]. This significantly reduces the damaging effects of faulty data, while retaining the information to be processed by the filter and reducing the time and cost associated with filter tuning. In response, this section proposes a fault-cognizant measurement model that inherently accounts for erroneous data. More precisely, each time the filter receives data, the model will account for the possibilities that (1) the incoming data is valid and (2) the incoming data is faulty, where the updated estimate is comprised of a weighted combination of the two possibilities, retaining all incoming information while remaining protected from faulty data.

It should be noted that the traditional method of multiple-model filtering can be considered, under specific assumptions, as closely related to the development of the FCMMs of this section. However, most multiple-model filters are described as "banks" of Kalman filters (or extended Kalman filters), necessarily enforcing Gaussian assumptions upon the measurement models to achieve closed-form Bayesian updates [33, 55]. Furthermore, these traditional multiple-model filters are frequently constructed as linear estimators, the drawbacks of which are discussed in Section 2.1, with inputs/outputs of the update existing as a single mean and covariance pair. Some work has been done to create nonlinear multiple-model filters, but focus has mainly been on particle

filtering implementations [87], where this work focuses on GM realizations per Section 2.2.2.

The remainder of this section derives four different fault-cognizant measurement models, each with a different application in mind. As a precursor to the derivations, it is useful to remember the foundational concept of probability that

$$\Pr\{\text{Outcome Y}\} = \frac{\text{All Events With Outcome Y}}{\text{All Events}},$$

where $\Pr\{\cdot\}$ denotes probability. This simple concept is presented here to emphasize that correctly accounting for all possible outcomes is crucial when developing probabilistic models, and that it may or may not be trivial to do so. As such, the derivations of this section mostly involve the consideration and classification of different possible events. It will be seen that by making small changes in the assumptions concerning these events, various measurement likelihoods can be determined, each useful for different sensor-system pairs. The first FCMM proposed is the simplest, as it is intended to directly replace the traditional measurement model of Fig. (4.1) and fit easily into pre-existing filtering frameworks. Each subsequent FCMM becomes more complex, but also more general, with the third and fourth models being distinct generalizations of the first two FCMMs. Specifically, the FCMM of Section 4.3.1 introduces faulty measurement modeling, but still reports scans of single measurements to the filter. The following FCMM derived in Section 4.3.2 allows for measurement scans of varying size, yet assumes at most one valid measurement can be generated. The remaining FCMMs of Sections 4.3.3 and 4.3.4 further allow for the possibility of multiple valid measurements within a measurement scan, where Section 4.3.3 assumes all valid measurements are IID, and Section 4.3.4 assumes the valid measurements are generated from unique distributions.

4.3.1 FCMM-1: Single Measurement Returns

As mentioned in Section 4.3, it is ill-advised not to account for faulty measurements at some level. However, it is advantageous to design a FCMM that can easily be implemented into existing sensor/filter architectures. To accomplish this, the traditional measurement model of Fig. (4.1) can

be expanded upon to include the possibility of faulty measurements, yet still produce measurement scans consisting of a single measurement [83]. Thus, FCMM-1 is established, where, at each instance of a filtering update, two mutually exclusive events exist:

1. the sensor produces a single valid measurement or
2. the sensor produces a single invalid measurement.

It is clear that if no measurement is reported by sensor, the filter is unable to update the prior distribution with new information; thus, in this case, the filtering architecture skips the update stage. Therefore, FCMM-1 only considers cases where some type of measurement data is received by the filter. Accordingly, the newly proposed FCMM is visually represented by Fig. (4.2). Recalling the traditional measurement model of Fig. (4.1), the core difference of FCMM-1 is that the measurement space now includes a single faulty measurement, z_f, which is spatially generated according to Eq. (4.3).

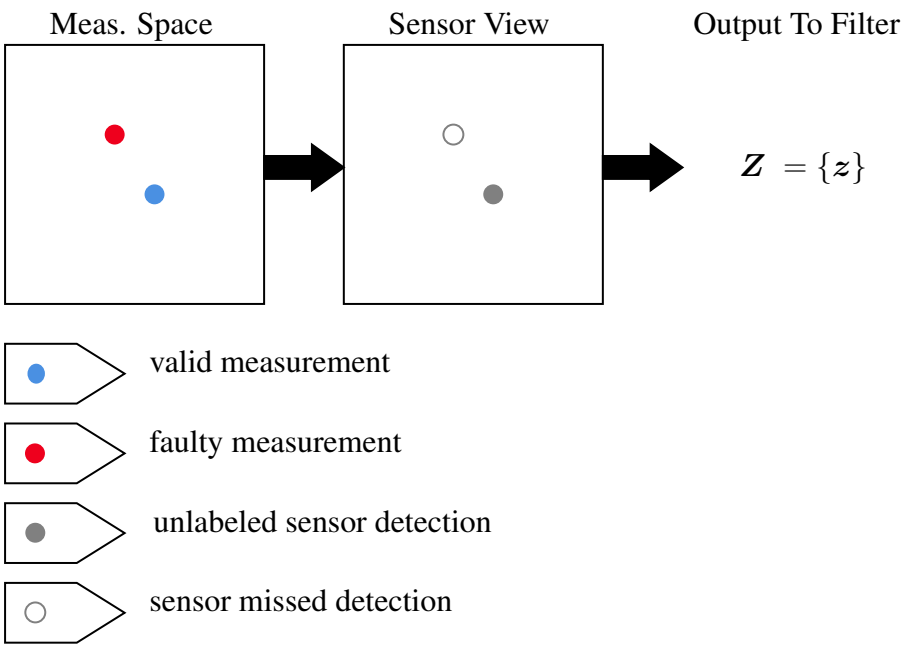


Figure 4.2 Visualization of FCMM-1 with single measurement returns

In order to be properly accounted for within an update, the fault-cognizant measurement model must be represented by a likelihood function. In this first FCMM derivation, it is useful to define a probability mass function (pmf) as

$$\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) = \Pr\left\{\boldsymbol{z} \in \boldsymbol{\theta}|\boldsymbol{x}\right\}, \tag{4.4}$$

such that Eq. (4.4) is the probability that the random vector $\boldsymbol{z}$ exists in the domain of $\boldsymbol{\theta}$ when conditioned on $\boldsymbol{x}$. Probability mass functions are directly related to likelihood functions via

$$\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) = \int_{\boldsymbol{\theta}} \ell(\boldsymbol{z}|\boldsymbol{x})\mathrm{d}\boldsymbol{z}\ . \tag{4.5}$$

Following techniques similar to Mahler [48], it can be significantly easier to first identify a specific pmf through straightforward probability rules and then solve for the likelihood via differentiation, such that Eq. (4.5) becomes

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{z}}\Big\{\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x})\Big\}\ . \tag{4.6}$$

As such, this is the solution method utilized for deriving FCMM-1. Recall that $\boldsymbol{z}$ can either correspond to a valid or an invalid measurement and that these events are mutually exclusive. Since these events are disjoint, the law of total probability dictates that the pmf of Eq. (4.4) is equivalently stated as

$$\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) = \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta}) \cap (\boldsymbol{z} = \boldsymbol{z}_v)|\boldsymbol{x}\right\} + \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta}) \cap (\boldsymbol{z} = \boldsymbol{z}_f)|\boldsymbol{x}\right\}, \tag{4.7}$$

where $\Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta}) \cap (\boldsymbol{z} = \boldsymbol{z}_v)|\boldsymbol{x}\right\}$ is the probability that $\boldsymbol{z}$ exists in $\boldsymbol{\theta}$ and is a valid measurement for a given $\boldsymbol{x}$, while $\Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta}) \cap (\boldsymbol{z} = \boldsymbol{z}_f)|\boldsymbol{x}\right\}$ is the conditional probability for the same $\boldsymbol{x}$ that $\boldsymbol{z}$ exists in $\boldsymbol{\theta}$ but is an invalid measurement. As the events $\boldsymbol{z} = \boldsymbol{z}_v$ and $\boldsymbol{z} = \boldsymbol{z}_f$ are dependent upon

$\boldsymbol{z} \in \boldsymbol{\theta}$, by the definition of conditional probability, Eq. (4.7) becomes

$$\begin{aligned}\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) = \Pr\left\{(\boldsymbol{z} = \boldsymbol{z}_v)|\boldsymbol{x}\right\} \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta})|(\boldsymbol{z} = \boldsymbol{z}_v), \boldsymbol{x}\right\} \\ + \Pr\left\{(\boldsymbol{z} = \boldsymbol{z}_f)|\boldsymbol{x}\right\} \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta})|(\boldsymbol{z} = \boldsymbol{z}_f), \boldsymbol{x}\right\}. \end{aligned} \tag{4.8}$$

Next, each term of Eq. (4.8) is examined individually. For brevity, the probability that a measurement is valid subject to a given state $\boldsymbol{x}$ is taken to be

$$\Pr\left\{(\boldsymbol{z} = \boldsymbol{z}_v)|\boldsymbol{x}\right\} = v(\boldsymbol{x}) \,, \tag{4.9}$$

where $v(\boldsymbol{x})$ is constrained by

$$0 \leq v(\boldsymbol{x}) \leq 1 \,. \tag{4.10}$$

Clearly, the probability that random measurement $\boldsymbol{z}$ is faulty must be the complement of Eq. (4.9), such that

$$\begin{aligned}\Pr\left\{(\boldsymbol{z} = \boldsymbol{z}_f)|\boldsymbol{x}\right\} &= 1 - \Pr\left\{(\boldsymbol{z} = \boldsymbol{z}_v)|\boldsymbol{x}\right\} \\ &= 1 - v(\boldsymbol{x}) \,. \end{aligned} \tag{4.11}$$

Using the newly defined probability of validity $v(\boldsymbol{x})$ of Eqs. (4.9) and (4.11), Eq. (4.8) takes on the form

$$\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) = v(\boldsymbol{x}) \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta})|(\boldsymbol{z} = \boldsymbol{z}_v), \boldsymbol{x}\right\} + \left[1 - v(\boldsymbol{x})\right] \Pr\left\{(\boldsymbol{z} \in \boldsymbol{\theta})|(\boldsymbol{z} = \boldsymbol{z}_f), \boldsymbol{x}\right\} \,. \tag{4.12}$$

Noting that the definitions of the pmf in Eqs. (4.4) and (4.5) can be equated, the pmfs of $\boldsymbol{z} \in \boldsymbol{\theta}$ in

Eq. (4.12) can be represented as integrated likelihoods, such that Eq. (4.12) becomes

$$\begin{aligned}\rho_{\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{x}) &= v(\boldsymbol{x})\int_{\boldsymbol{\theta}} \ell\big(\boldsymbol{z}|(\boldsymbol{z}=\boldsymbol{z}_v),\boldsymbol{x}\big)\mathrm{d}\boldsymbol{z} + \big[1-v(\boldsymbol{x})\big]\int_{\boldsymbol{\theta}} \ell\big(\boldsymbol{z}|(\boldsymbol{z}=\boldsymbol{z}_f),\boldsymbol{x}\big)\mathrm{d}\boldsymbol{z} \\ &= v(\boldsymbol{x})\int_{\boldsymbol{\theta}} \ell_v(\boldsymbol{z}|\boldsymbol{x})\mathrm{d}\boldsymbol{z} + \big[1-v(\boldsymbol{x})\big]\int_{\boldsymbol{\theta}} \ell_f(\boldsymbol{z}|\boldsymbol{x})\mathrm{d}\boldsymbol{z}\,, \end{aligned} \tag{4.13}$$

where the valid and faulty measurement likelihoods of Eqs. (4.2) and (4.3) are inserted. Now that an adequately complete pmf is found, the identity of Eq. (4.6) indicates that by differentiating Eq. (4.13) with respect to $\boldsymbol{z}$, the likelihood for FCMM-1 is solved for as

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = v(\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + \big[1-v(\boldsymbol{x})\big]\ell_f(\boldsymbol{z}|\boldsymbol{x})\,. \tag{4.14}$$

Note that no precise models for $v(\boldsymbol{x})$, $\ell_v(\boldsymbol{z}|\boldsymbol{x})$, or $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ are administered in the derivation of Eq. (4.14), aside from the assumption that valid and faulty measurements are spatially independent from one another. Possible forms for these models are investigated in Sections 4.4 and 4.5, which must be done before the closed-form updates of Section 4.6.2 can be reached.

4.3.2 FCMM-2: Single-Valid Measurement

In this second FCMM (FCMM-2), the concepts behind the derivation in Section 4.3.1 are expanded upon, wherein a fault-cognizant likelihood is derived that accounts for both faulty and valid returns. The key difference in this model is that, instead of allowing only a single measurement per sensor scan, as is done in Section 4.3.1, FCMM-2 permits multiple measurements to be included in each sensor scan. While FCMM-1 is simpler, as it is intended to be a direct replacement for existing architecture, the temporal distributions for valid and faulty measurements are necessarily interdependent when a scan is restricted to a single return, such that the frequencies of $\boldsymbol{z}_v$ and $\boldsymbol{z}_f$ are both quantified by a single function—the probability of validity $v(\boldsymbol{x})$. This interdependency prevents the independent modeling of valid and faulty detection occurrences, which in turn disallows the use of the probabilities of detection and false alarm. More discussion on this topic is found in Section 4.4.2.1, where it is shown that this interdependency can be obviated by permitting multi-

ple returns per sensor scan, as the receipt of valid and faulty measurements are no longer mutually exclusive events. Thus, FCMM-2 assumes sensor operations abide by the following constraints:

1. a scan can contain, at most, one valid measurement;
2. a scan can contain any number of invalid, or faulty, measurements; and
3. a scan can contain no measurements.

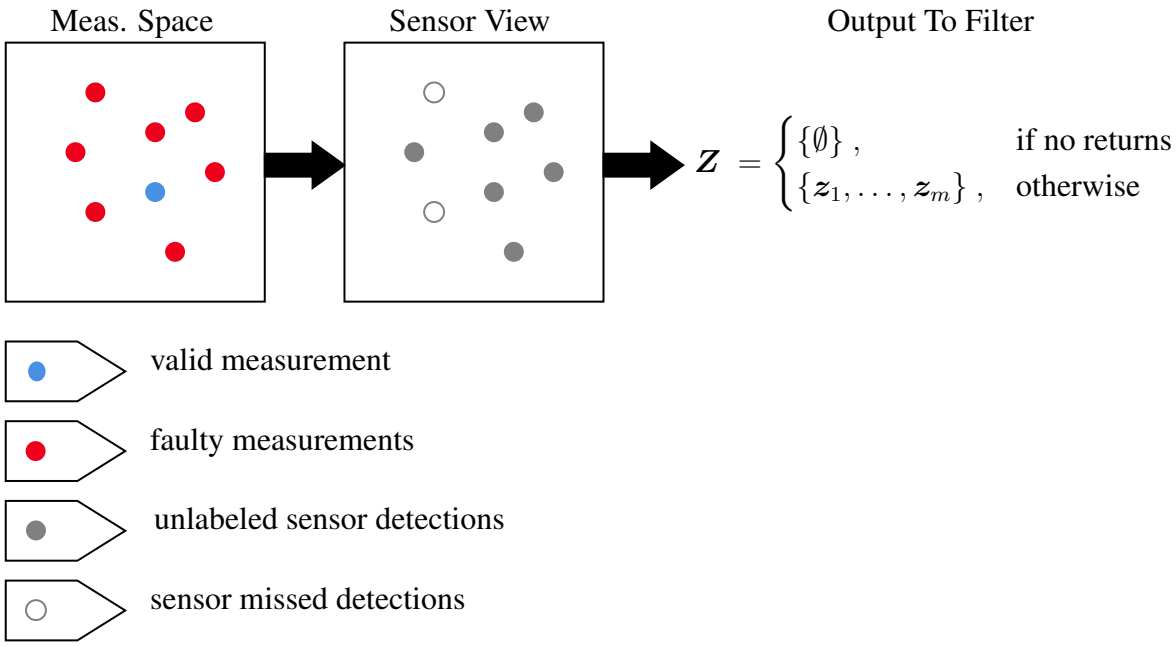


Figure 4.3 Visualization of FCMM-2 with measurement scans containing, at most, one valid measurement

These sensor operating conditions ultimately result in the FCMM-2 visualization of Fig. (4.3). This model is well-suited for applications where a sensor returns a measurement scan with little pre-processing, resulting in multiple possible faulty measurements, yet still only producing a single valid measurement. For example, consider a ground-based observer tracking the state $\boldsymbol{x}$ of a satellite via imagery, which might produce images similar to Fig. (4.4) taken from [1]. In such a case, m clusters of bright pixels are returned as measurements from the sensor. However, at most

one of these returns will correspond to a valid measurement generated by the satellite of interest, with the possibility of no valid measurement being generated, especially if the satellite is outside of the sensor's view. Therefore, all remaining $m-1$ measurements (m if no valid detection) are generated by the surrounding star-field and image noise, all of which are classified as faulty measurements in regard to FCMM-2. It is with systems such as Fig. (4.4) in mind that FCMM-2 is developed.

Figure 4.4 Typical telescope imagery for satellite tracking, where satellite of interest is marked by green rectangle—reprinted from [1]

Consider that, upon each activation, a sensor returns a measurement set of m measurements as $\boldsymbol{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m\}$. Furthermore, it is assumed that a single target is present, and the target can, at most, produce a single valid measurement return $\boldsymbol{z}_v \in \mathbb{V}$ pursuant to the likelihood function $\ell_v(\boldsymbol{z}|\boldsymbol{x})$ of Eq. (4.2). Similarly, faulty measurements are generated via likelihood $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ of Eq. (4.3), are denoted by $\boldsymbol{z}_f \in \mathbb{F}$, and are assumed to be independent of one another. Thus, the total likelihood of FCMM-2 (in the event of m measurements) is

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \Pr\left(\boldsymbol{Z} \cap m \cap \left[(\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cup (\boldsymbol{z}_v \nsubseteq \boldsymbol{Z})\right] \middle| \boldsymbol{x}\right), \tag{4.15}$$

where $(\boldsymbol{z}_v \subseteq \boldsymbol{Z})$ denotes the event that the measurement set contains a valid target generated measurement, and the event where the detection is missed is $(\boldsymbol{z}_v \nsubseteq \boldsymbol{Z})$. As the target is either detected or is not, these are mutually exclusive events, such that Eq. (4.15) becomes

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \Pr\left(\boldsymbol{Z} \cap m \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right) + \Pr\left(\boldsymbol{Z} \cap m \cap (\boldsymbol{z}_v \nsubseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right). \tag{4.16}$$

As a direct result of conditional probability, Eq. (4.16) becomes

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) = {} & \Pr\left(\boldsymbol{Z} \cap m\middle|(\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) \Pr\left((\boldsymbol{z}_v \subseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right) \\ & + \Pr\left(\boldsymbol{Z} \cap m\middle|(\boldsymbol{z}_v \nsubseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) \Pr\left((\boldsymbol{z}_v \nsubseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right),\end{aligned} \tag{4.17}$$

where

$$\Pr\left((\boldsymbol{z}_v \subseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right) = p_D(1, \boldsymbol{x}) \tag{4.18}$$

is the state-dependent probability that a single valid detection is generated by the sensor. Therefore, the probability that the target is not detected must be complementary, such that

$$\begin{aligned}\Pr\left((\boldsymbol{z}_v \nsubseteq \boldsymbol{Z})\middle|\boldsymbol{x}\right) &= p_D(0, \boldsymbol{x}) \\ &= 1 - p_D(1, \boldsymbol{x}).\end{aligned} \tag{4.19}$$

Therefore, by the definitions of Eqs. (4.18) and (4.19), Eq. (4.17) can be expressed as

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = p_D(1, \boldsymbol{x}) \Pr\left(\boldsymbol{Z} \cap m|(\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) + [1 - p_D(1, \boldsymbol{x})] \Pr\left(\boldsymbol{Z} \cap m|(\boldsymbol{z}_v \nsubseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right). \tag{4.20}$$

Attention is now directed towards $\Pr\left(\boldsymbol{Z} \cap m\middle|(\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right)$ of Eq. (4.20), which is the probability that (in this case) exactly one valid measurement is returned by the target. As a result, $m - 1$ measurements must be faulty. For now, consider the event where the i^{th} measurement is valid,

denoted by $(\boldsymbol{z}_i = \boldsymbol{z}_v)$, where the collection of events for $i = 1, 2, \ldots, m$ is mutually exclusive, such that the probability becomes

$$\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \sum_{i=1}^{m} \Pr\left(\boldsymbol{Z} \cap m \cap (\boldsymbol{z}_i = \boldsymbol{z}_v) \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right). \tag{4.21}$$

The properties of conditional events can again be utilized such that Eq. (4.21) becomes

$$\begin{aligned}\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \sum_{i=1}^{m} &\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_i = \boldsymbol{z}_v) \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) \\ &\Pr\left((\boldsymbol{z}_i = \boldsymbol{z}_v) \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right). \end{aligned} \tag{4.22}$$

Here, if it is assumed that measurements are unordered and that any $\boldsymbol{z}_i$ has an equal probability of being the valid measurement, then the probability of the event $(\boldsymbol{z}_i = \boldsymbol{z}_v)$, given $\boldsymbol{Z}$ contains a valid measurement, is uniformly distributed over the number of measurements as

$$\Pr\left((\boldsymbol{z}_i = \boldsymbol{z}_v) \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \begin{cases} \frac{1}{m}\,, & \text{if } m > 0 \\ 0\,, & \text{otherwise} \end{cases}, \tag{4.23}$$

such that Eq. (4.22) becomes

$$\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \frac{1}{m} \sum_{i=1}^{m} \Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_i = \boldsymbol{z}_v) \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right). \tag{4.24}$$

Yet again making use of conditional probabilities, Eq. (4.24) takes the form

$$\begin{aligned}\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \frac{1}{m} \sum_{i=1}^{m} &\Pr\left(\boldsymbol{Z} \middle| m \cap (\boldsymbol{z}_i = \boldsymbol{z}_v) \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right) \\ &\times \Pr\left(m \middle| (\boldsymbol{z}_i = \boldsymbol{z}_v) \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right). \end{aligned} \tag{4.25}$$

The term $\Pr\left(m \middle| (\boldsymbol{z}_i = \boldsymbol{z}_v) \cap (\boldsymbol{z}_v \subseteq \boldsymbol{Z}) \cap \boldsymbol{x}\right)$ is the probability that m measurements are generated, provided one of the measurements is valid. Equivalently, this can be considered the probability

that $m-1$ measurements are faulty, or

$$\Pr\left(m\middle|(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)=p_F(m-1,\boldsymbol{x})\,, \tag{4.26}$$

where $p_F(m-1,\boldsymbol{x})$ signifies the state-dependent probability of $m-1$ false alarms occurring. Thus, Eq. (4.25) becomes

$$\Pr\left(\boldsymbol{Z}\cap m\middle|(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)=\frac{p_F(m-1,\boldsymbol{x})}{m}\sum_{i=1}^{m}\Pr\left(\boldsymbol{Z}\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right). \tag{4.27}$$

At this point, it is recalled that faulty measurements are independently and identically distributed according to the faulty likelihood function $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ and valid target-generated measurements are distributed according to the valid likelihood function $\ell_v(\boldsymbol{z}|\boldsymbol{x})$. Therefore, it can be shown that the remaining term inside the summation of Eq. (4.27) becomes

$$\begin{aligned}
&\Pr\left(\boldsymbol{Z}\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)\\
&\quad=\Pr\left(\boldsymbol{z}_1\cap\boldsymbol{z}_2\cap\cdots\cap\boldsymbol{z}_m\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)\\
&\quad=\prod_{j=1}^{m}\Pr\left(\boldsymbol{z}_j\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)\\
&\quad=\Pr\left(\boldsymbol{z}_i\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)\prod_{\substack{j=1\\j\neq i}}^{m}\Pr\left(\boldsymbol{z}_j\middle|m\cap(\boldsymbol{z}_i=\boldsymbol{z}_v)\cap(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)\\
&\quad=\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\prod_{\substack{j=1\\j\neq i}}^{m}\ell_f(\boldsymbol{z}_j|\boldsymbol{x})\,.
\end{aligned} \tag{4.28}$$

Substituting the result of Eq. (4.28) into Eq. (4.27) yields

$$\Pr\left(\boldsymbol{Z}\cap m\middle|(\boldsymbol{z}_v\subseteq\boldsymbol{Z})\cap\boldsymbol{x}\right)=\frac{p_F(m-1,\boldsymbol{x})}{m}\sum_{i=1}^{m}\left\{\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\prod_{\substack{j=1\\j\neq i}}^{m}\ell_f(\boldsymbol{z}_j|\boldsymbol{x})\right\}. \tag{4.29}$$

Next, Eq. (4.20) is revisited to address the term $\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right)$. By conditional probability

$$\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) = \Pr\left(\boldsymbol{Z} \middle| m \cap (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) \Pr\left(m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right), \tag{4.30}$$

where $\Pr\left(m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right)$ is the probability that m measurements are generated, provided that none of them are valid. This is also the probability that all m measurements are faulty, denoted by

$$\Pr\left(m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) = p_F(m, \boldsymbol{x}). \tag{4.31}$$

Similar to the developments of Eq. (4.28), the expression $\Pr\left(\boldsymbol{Z} \middle| m \cap (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right)$ can be transformed as

$$\begin{aligned}
\Pr\left(\boldsymbol{Z} \middle| m \cap (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) &= \Pr\left(\boldsymbol{z}_1 \cap \boldsymbol{z}_2 \cap \cdots \cap \boldsymbol{z}_m \middle| m \cap (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) \\
&= \prod_{j=1}^{m} \Pr\left(\boldsymbol{z}_j \middle| m \cap (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) \\
&= \prod_{j=1}^{m} \ell_f(\boldsymbol{z}_j | \boldsymbol{x}),
\end{aligned} \tag{4.32}$$

assuming, again, that the faulty measurements are independently and identically distributed according to the faulty likelihood of Eq. (4.3). Equations (4.31) and (4.32) are sufficient to show that Eq. (4.30) becomes

$$\Pr\left(\boldsymbol{Z} \cap m \middle| (\boldsymbol{z}_v \notin \boldsymbol{Z}) \cap \boldsymbol{x}\right) = p_F(m, \boldsymbol{x}) \prod_{j=1}^{m} \ell_f(\boldsymbol{z}_j | \boldsymbol{x}). \tag{4.33}$$

Accordingly, the results of Eqs. (4.29) and (4.33) allow the FCMM-2 likelihood of Eq. (4.20) to

ultimately be expressed as

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = p_D(1,\boldsymbol{x})\frac{p_F(m-1,\boldsymbol{x})}{m}\sum_{i=1}^{m}\left\{\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\prod_{\substack{j=1\\j\neq i}}^{m}\ell_f(\boldsymbol{z}_j|\boldsymbol{x})\right\} + [1-p_D(1,\boldsymbol{x})]p_F(m,\boldsymbol{x})\prod_{j=1}^{m}\ell_f(\boldsymbol{z}_j|\boldsymbol{x})\,. \tag{4.34}$$

4.3.2.1 Special Case 1 [$m = 0$]:

It should be noted that due to a division by m, the case where there are no measurements seems to produce a singularity, at a glance. While many filters will not perform an update unless at least one measurement is received, the fact that no measurement is returned can provide information in and of itself. Therefore, in situations where $\boldsymbol{Z} = \{\emptyset\}$, Eq. (4.34) becomes

$$\ell(\emptyset|\boldsymbol{x}) = [1-p_D(1,\boldsymbol{x})]p_F(0,\boldsymbol{x})\,, \tag{4.35}$$

as the first term vanishes—since the probability of -1 false alarms is zero as well as the probability of Eq. (4.23)—and by definition $\prod_{j=1}^{0}\ell_f(\boldsymbol{z}_j|\boldsymbol{x}) = 1$. In typical single-target filtering applications that do not account for state-dependent probabilities, when no measurements are produced, no update is performed. Leveraging the developments here, however, an update can be performed due to the presence of the probabilities of detection and false alarm.

4.3.2.2 Special Case 2 [$m = 1$]:

In the instance where exactly a single measurement is received, or $\boldsymbol{Z} = \{\boldsymbol{z}\}$, it can be shown that the likelihood of Eq. (4.34) becomes

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_D(1,\boldsymbol{x})p_F(0,\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + [1-p_D(1,\boldsymbol{x})]p_F(1,\boldsymbol{x})\ell_f(\boldsymbol{z}|\boldsymbol{x})\,, \tag{4.36}$$

which looks very similar to the FCMM-1 likelihood derived using the probability of validity in Section 4.3.1. As mentioned previously, the main difference between the two likelihoods is that

the probabilities of detection and false alarm are modeled independently of one another as two distinct parameters, whereas the probability of validity attempts to achieve a similar likelihood using only a single parameter. Additional discussion on this is provided in Section 4.4.2.1.

4.3.3 FCMM-3: Multiple-Valid Measurement - IID

The FCMM-2 of Section 4.3.2 is useful for scans of multiple measurements collected in cluttered environments. However, it inherently assumes that at most one valid measurement can be generated. This section derives a third fault-cognizant measurement model (FCMM-3) under the conditions that multiple valid measurements are IID, such that they correspond to a single measurement distribution pursuant to Eq. (4.2). The commonly used methodology of extended target tracking is used for similar situations, where a single object is able to generate multiple valid measurements [88]. This procedure can be thought of as similar to this and is visualized in Fig. (4.5). As an example, consider a simplistic LiDAR system, wherein the location of a car is to be tracked, such as in Fig. (4.6). A typical LiDAR return consists of a point cloud corresponding to the object surfaces, so it can be expected for a car to return multiple valid measurements. Various environmental features will also return measurements, but since they do not correspond to the target, they are considered non-valid (or faulty in this case) returns by the FCMM.

As an additional example, consider once more the satellite tracking system of Fig. (4.4). There may be situations where the optical observer returns two or more valid measurements, caused by some unmodeled effect such as multiple, separated reflective surfaces of the satellite. If two or more valid measurements are generated by the target, the previous FCCMs of Sections 4.3.1 and 4.3.2 have no way to reconcile such an event, as their architecture only allows for a single possible valid measurement. Therefore, this section explores a solution to such a problem by developing an appropriate measurement model. This model can be considered a more general form of FCMM-2 and will be seen to reduce to Eq. (4.34) under the correct conditions.

Consider that, upon each activation, a sensor returns a set of m measurements as $\boldsymbol{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m\}$. Furthermore, it is assumed that a single target $\boldsymbol{x}$ is present. However, this target can produce multiple valid measurements given by $\boldsymbol{z}_v \in \mathbb{V}$ pursuant to some yet unspecified likelihood function

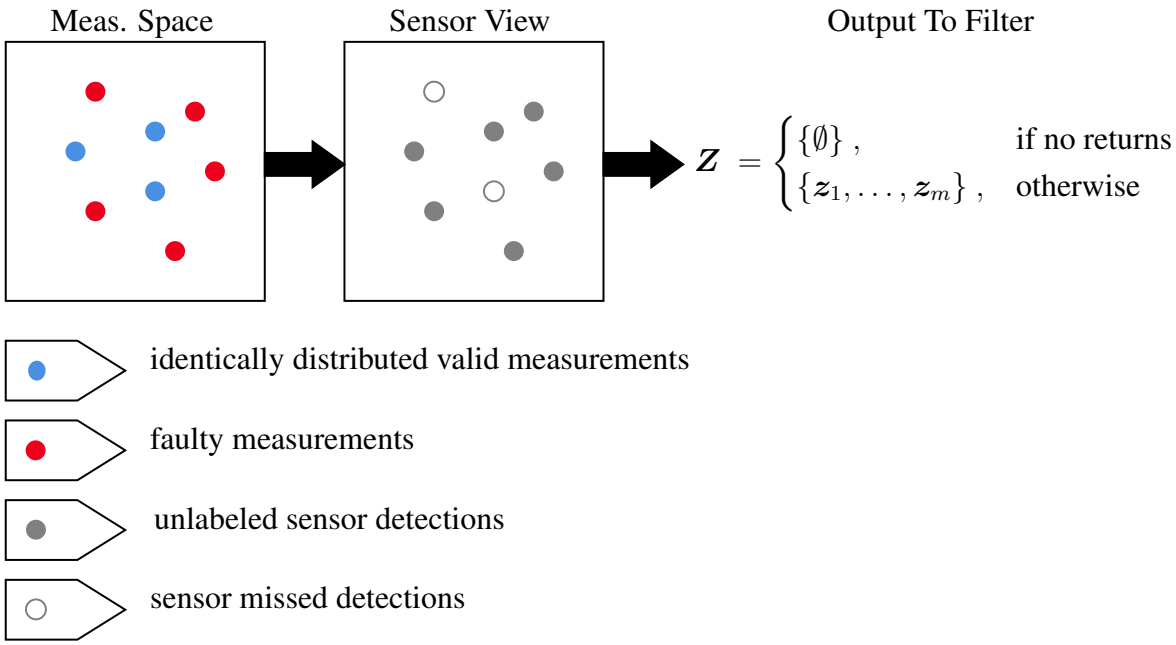


Figure 4.5 Visualization of FCMM-3 with multiple IID valid measurements

$\ell_v(\boldsymbol{z}|\boldsymbol{x})$. The subset of these valid measurements is denoted by $\boldsymbol{Z}_v = \{\boldsymbol{z}_{1,v}, \boldsymbol{z}_{2,v}, \ldots, \boldsymbol{z}_{n,v}\} \subseteq \boldsymbol{Z}$ such that $0 \leq n \leq m$. Accordingly, multiple faulty measurements are also possible, are generated via likelihood $\ell_f(\boldsymbol{z}|\boldsymbol{x})$, and are denoted by $\boldsymbol{z}_f \in \mathbb{F}$. The subset of these faulty measurements is given by $\boldsymbol{Z}_f = \{\boldsymbol{z}_{1,f}, \boldsymbol{z}_{2,f}, \ldots, \boldsymbol{z}_{m-n,f}\} \subseteq \boldsymbol{Z}$, such that $\boldsymbol{Z} = \boldsymbol{Z}_v \cup \boldsymbol{Z}_f$ and $\boldsymbol{Z}_v \cap \boldsymbol{Z}_f = \{\emptyset\}$. The total likelihood of FCMM-3 is therefore

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \Pr\left(\boldsymbol{Z} \cap m \bigcup_{n=0}^{m} n|\boldsymbol{x}\right), \tag{4.37}$$

where $\bigcup_{n=0}^{m} n$ represents all the possible events of the number of valid measurements contained in $\boldsymbol{Z}$. Clearly the events $(n=0)$, $(n=1)$, $\ldots$, $(n=m)$ are all mutually exclusive, as $\boldsymbol{Z}_v$ can not be of two different sizes at the same time. Thus, Eq. (4.37) becomes

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{m} \Pr\left(\boldsymbol{Z} \cap m \cap n|\boldsymbol{x}\right). \tag{4.38}$$

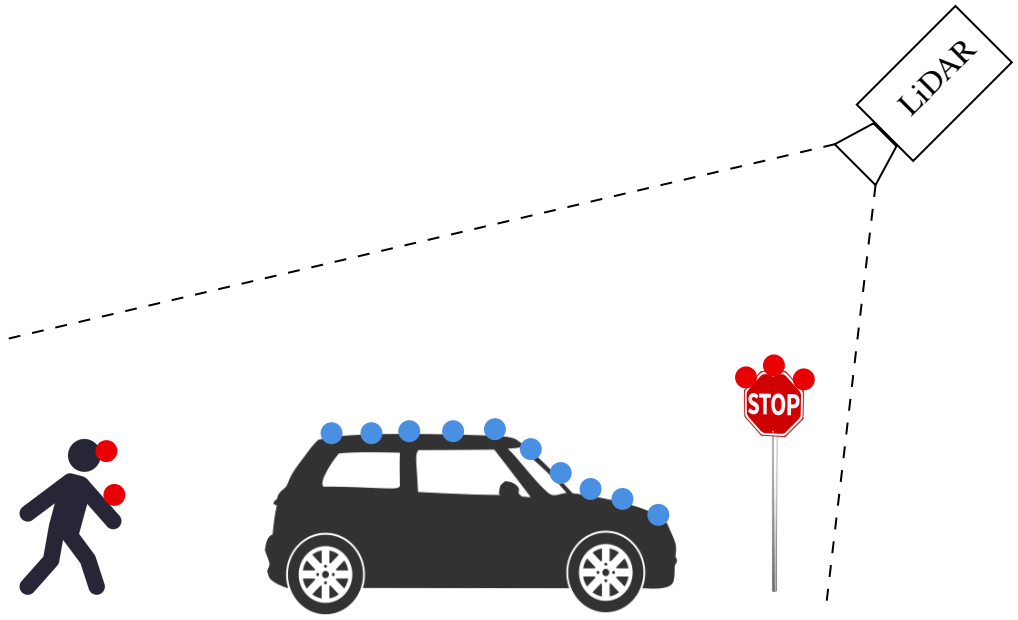


Figure 4.6 Simplistic LiDAR example, where blue dots correspond to valid target returns of the vehicle, and red dots correspond to faulty (or non-valid) returns created by various environmental objects

By the rules of conditional probability, Eq. (4.38) is given by

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{m} \Pr\left(\boldsymbol{Z} \cap m|n \cap \boldsymbol{x}\right) \Pr\left(n|\boldsymbol{x}\right), \tag{4.39}$$

where $\Pr\left(n|\boldsymbol{x}\right)$ is the probability that, for some given state $\boldsymbol{x}$, measurement set $\boldsymbol{Z}$ will contain n valid measurements, or that n detections will occur. For this, we can recall the notation of Eq. (4.18), or that the probability of n valid detections is

$$\Pr\left(n|\boldsymbol{x}\right) = p_D(n, \boldsymbol{x}), \tag{4.40}$$

where $p_D(\cdot, \boldsymbol{x})$ denotes that the probability is (potentially) state-dependent. With the definition of

Eq. (4.40), the form of Eq. (4.39) becomes

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{m} p_D(n, \boldsymbol{x}) \Pr\left(\boldsymbol{Z} \cap m | n \cap \boldsymbol{x}\right). \tag{4.41}$$

Now, the term $\Pr\left(\boldsymbol{Z} \cap m | n \cap \boldsymbol{x}\right)$ of Eq. (4.41) is examined specifically. By conditional probability, it becomes

$$\Pr\left(\boldsymbol{Z} \cap m | n \cap \boldsymbol{x}\right) = \Pr\left(m | n \cap \boldsymbol{x}\right) \Pr\left(\boldsymbol{Z} | m \cap n \cap \boldsymbol{x}\right). \tag{4.42}$$

Based on the fact that $\boldsymbol{Z}_v \cup \boldsymbol{Z}_f = \boldsymbol{Z}$ and that $\boldsymbol{Z}_v \cap \boldsymbol{Z}_f = \{\emptyset\}$, it is clear that if there are n valid measurements, then there must be $m - n$ faulty measurements. So, the probability that $\boldsymbol{Z}$ contains m measurements conditioned on the event n is equivalent to the probability that $\boldsymbol{Z}$ contains $m - n$ false alarms. Recalling the notation of Eq. (4.26), the probability of $m - n$ faulty measurements is mathematically expressed as

$$\Pr\left(m | n \cap \boldsymbol{x}\right) = p_F(m - n, \boldsymbol{x}),$$

such that Eq. (4.42) becomes

$$\Pr\left(\boldsymbol{Z} \cap m | n \cap \boldsymbol{x}\right) = p_F(m - n, \boldsymbol{x}) \Pr\left(\boldsymbol{Z} | m \cap n \cap \boldsymbol{x}\right). \tag{4.43}$$

Next, the term $\Pr\left(\boldsymbol{Z} | m \cap n \cap \boldsymbol{x}\right)$ of Eq. (4.43) is inspected. Since this probability is conditional upon the number of valid measurements being n, out of a measurement set $\boldsymbol{Z}$ of size m, it must now account for all of the different combinations in which n measurements of the set $\boldsymbol{Z}$ can be assigned as valid. From a combinatorics perspective, it is known that there are $\binom{m}{n}$ unique combinations (or hypotheses), with each event being mutually exclusive, such that the probability can be written as

$$\Pr\left(\boldsymbol{Z} | m \cap n \cap \boldsymbol{x}\right) = \sum_{j=1}^{\binom{m}{n}} \Pr\left(\boldsymbol{Z} \cap H_j | m \cap n \cap \boldsymbol{x}\right), \tag{4.44}$$

Note that $H_{v,j} = \{\boldsymbol{Z}_v \subseteq \boldsymbol{Z}\}_j$ denotes the hypothesis for the j^{th} unique assignment of n valid measurements within the set $\boldsymbol{Z}$ of m measurements. Note that for every unique assignment of valid measurements $H_{v,j}$ there exists an assignment of faulty measurements that is also unique, which is defined as $H_{f,j} = \boldsymbol{Z} \backslash H_{v,j} = \{\boldsymbol{Z}_f \subseteq \boldsymbol{Z}\}_j$. Ultimately, this indicates that specifying the hypothesis $H_{v,j} = \{\boldsymbol{Z}_v \subseteq \boldsymbol{Z}\}_j$ is sufficient to create a unique assignment combination for all m measurements of $\boldsymbol{Z}$ and includes all possible assignment hypotheses, which will be denoted by $H_j = H_{v,j} \cup H_{f,j}$. With this in mind, and again utilizing the rules of conditional probability, Eq. (4.44) becomes

$$\Pr\left(\boldsymbol{Z}|m \cap n \cap \boldsymbol{x}\right) = \sum_{j=1}^{\binom{m}{n}} \Pr\left(H_j|m \cap n \cap \boldsymbol{x}\right) \Pr\left(\boldsymbol{Z}|H_j \cap m \cap n \cap \boldsymbol{x}\right), \tag{4.45}$$

Noting that all of the hypotheses H_j for $j = 1,\ 2,\ \ldots,\ \binom{m}{n}$ are equally likely, the probability of each (conditional on some given state $\boldsymbol{x}$ and m measurements, n of which are valid) takes on a discrete uniform distribution as

$$\Pr\left(H_j\middle|n \cap \boldsymbol{x}\right) = \frac{1}{\binom{m}{n}},$$

such that Eq. (4.45) becomes

$$\Pr\left(\boldsymbol{Z}|m \cap n \cap \boldsymbol{x}\right) = \frac{1}{\binom{m}{n}} \sum_{j=1}^{\binom{m}{n}} \Pr\left(\boldsymbol{Z}\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right). \tag{4.46}$$

Next, consider the term $\Pr\left(\boldsymbol{Z}\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right)$ of Eq. (4.46) and recall that valid and faulty measurements are independently and identically spatially distributed according to the likelihood functions $\ell_v(\boldsymbol{z}|\boldsymbol{x})$ and $\ell_f(\boldsymbol{z}|\boldsymbol{x})$, respectively. As this probability is already conditioned upon the

j^{th} unique assignment hypothesis H_j, it can be shown that

$$\begin{aligned}\Pr\left(\boldsymbol{Z}\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right) &= \Pr\left(\boldsymbol{z}_1 \cap \boldsymbol{z}_2 \cap \cdots \cap \boldsymbol{z}_m\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right) \\ &= \prod_{\gamma=1}^{m} \Pr\left(\boldsymbol{z}_\gamma\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right) \\ &= \prod_{\substack{k=1 \\ \boldsymbol{z}_k \in H_{v,j}}}^{n} \ell_v(\boldsymbol{z}_k|\boldsymbol{x}) \prod_{\substack{\ell=1 \\ \boldsymbol{z}_\ell \in H_{f,j}}}^{m-n} \ell_f(\boldsymbol{z}_\ell|\boldsymbol{x})\,. \end{aligned} \tag{4.47}$$

Taking the results of Eqs. (4.43), (4.46), and (4.47), it can be shown that the likelihood of Eq. (4.41) becomes

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) &= \sum_{n=0}^{m} p_D(n,\boldsymbol{x}) \Pr\left(\boldsymbol{Z} \cap m|n \cap \boldsymbol{x}\right) \\ &= \sum_{n=0}^{m} p_D(n,\boldsymbol{x}) p_F(m-n,\boldsymbol{x}) \Pr\left(\boldsymbol{Z}|m \cap n \cap \boldsymbol{x}\right) \\ &= \sum_{n=0}^{m} \left\{ \frac{p_D(n,\boldsymbol{x}) p_F(m-n,\boldsymbol{x})}{\binom{m}{n}} \sum_{j=1}^{\binom{m}{n}} \Pr\left(\boldsymbol{Z}\middle|H_j \cap m \cap n \cap \boldsymbol{x}\right) \right\} \\ &= \sum_{n=0}^{m} \left\{ \frac{p_D(n,\boldsymbol{x}) p_F(m-n,\boldsymbol{x})}{\binom{m}{n}} \sum_{j=1}^{\binom{m}{n}} \left[\prod_{\substack{k=1 \\ \boldsymbol{z}_k \in H_{v,j}}}^{n} \ell_v(\boldsymbol{z}_k|\boldsymbol{x}) \prod_{\substack{\ell=1 \\ \boldsymbol{z}_\ell \in H_{f,j}}}^{m-n} \ell_f(\boldsymbol{z}_\ell|\boldsymbol{x}) \right] \right\}. \end{aligned} \tag{4.48}$$

It is clear that this measurement model of Eq. (4.48) is fairly complex. However, if it is limited to cases where $n = \{0, 1\}$—such that only the first two terms of the summation over n in Eq. (4.48) are taken—it does reduce to the likelihood of Eq. (4.34), indicating that it is a more generalized form of the expression. This simplification can be done simply by setting the probability of detection $p_D(n, \boldsymbol{x}) = 0\ \forall n \geq 2$, such that the probability of more than one valid detection is zero.

4.3.4 FCMM-4: Multiple Unique Valid Measurements

The model FCMM-3 of Section 4.3.3 is designed specifically to allow for multiple valid measurements. A key constraint of FCMM-3, however, is that all valid measurements are assumed to

belong to the same valid measurement distribution of Eq. (4.2), such that they are IID. This section attempts another approach, wherein all valid measurements are assumed to be uniquely spatially distributed according to

$$\boldsymbol{z}_{v,i} \sim \ell_{v,i}(\boldsymbol{z}_{v,i}|\boldsymbol{x}) \quad \forall\, i = 1,\, 2,\, \ldots,\, q\,. \tag{4.49}$$

The motivation behind this fourth fault-cognizant measurement model (FCMM-4) is to allow for the definition of individual valid measurement models, should the need arise. Consider, for instance, vision-based terrain relative navigation (TRN), wherein an optical camera acquires images of some fixed terrain with q topographical features generating unlabeled sensor returns corresponding to the locations of those features, a simplified example of which is shown in Fig. (4.7). These feature locations are known *a priori* and logged in a map, where each feature is expected to generate sensor returns that are unique from the others, requiring the individual measurement likelihoods of Eq. (4.49). Any sensor returns contained in the image that do not correspond to one of the q features are considered faulty by FCMM-4. Typically in TRN applications, the image is pre-processed to match measurements to their most likely feature locations, such that the filter receives pairs of feature locations and feature labels, a process known as data association. FCMM-4, however, assumes that no data association is performed by the image pre-processing, such that the filter assumes any measurement could be generated by any of the q valid measurement distributions (i.e. features) or be a faulty return. In this way, the update resulting from FCMM-4 inherently performs data association itself. An appropriate visualization of FCMM-4 is provided in Fig. (4.8), which accentuates the uniqueness of valid measurments when compared to FCMM-3 of Fig. (4.5).

To begin, the generation and nature of a measurement scan must be characterized. Assume that upon a single measurement scan, the filter receives measurement set $\boldsymbol{Z}$ comprised of m different measurements as

$$\boldsymbol{Z} = \{\boldsymbol{z}_1,\, \boldsymbol{z}_2,\, \ldots,\, \boldsymbol{z}_m\}\,. \tag{4.50}$$

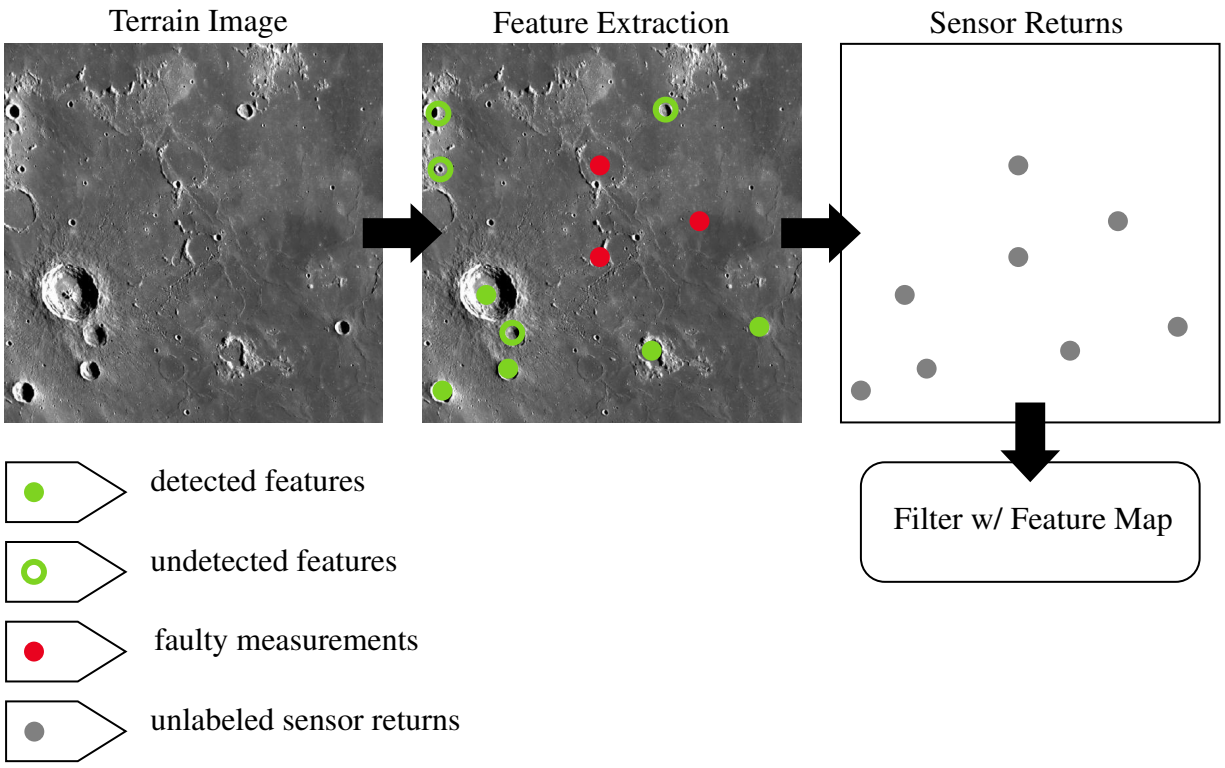


Figure 4.7 Example of possible crater-based TRN application for FCMM-4

Furthermore, assume that there exist q different valid measurement subsets $\boldsymbol{Z}_{v,i} \subseteq \boldsymbol{Z}$ that can contain at most a single measurement, such that

$$\boldsymbol{Z}_{v,i} = \{\emptyset\} \qquad \text{or} \qquad \boldsymbol{Z}_{v,i} = \{\boldsymbol{z}_{v,i}\} \quad \forall i = 1, 2, \ldots q \, . \tag{4.51}$$

This directly corresponds to the frequent reality that TRN features can only generate a single measurement, and where in this case, each valid measurement subset $\boldsymbol{Z}_{v,i}$ corresponds to a different mapped feature.

Additionally, assume that there exists a single faulty measurement subset $\boldsymbol{Z}_f \subseteq \boldsymbol{Z}$ that can include any number (up to m) of faulty measurements. Note that the intersections and unions of

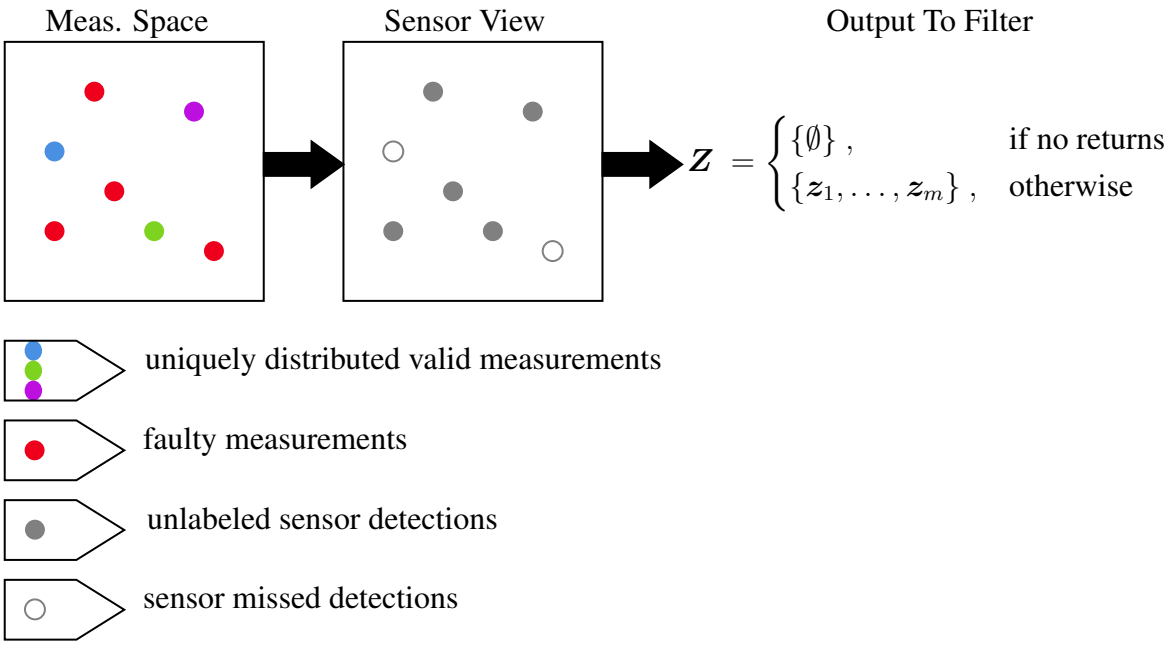


Figure 4.8 Visualization of FCMM-4 with multiple non-ID valid measurements

the measurement subsets lead to

$$\left[\bigcup_{i=1}^{q} \boldsymbol{Z}_{v,i}\right] \cup \boldsymbol{Z}_f = \boldsymbol{Z} \qquad \left[\bigcap_{i=1}^{q} \boldsymbol{Z}_{v,i}\right] \cap \boldsymbol{Z}_f = \{\emptyset\}\,. \tag{4.52}$$

Now consider the event where n measurements of $\boldsymbol{Z}$ are valid and are contained by one of the measurement subsets $\boldsymbol{Z}_{v,i}$. In this case, the faulty measurement subset must be comprised of $m-n$ measurements, such that

$$\boldsymbol{Z}_f = \{\boldsymbol{z}_{f,1},\ \boldsymbol{z}_{f,2},\ \ldots,\ \boldsymbol{z}_{f,m-n}\}. \tag{4.53}$$

Note that the number of valid measurements cannot be greater than the number of valid measurement subsets or the number of received measurements, such that $0 \leq n \leq \min(m, q)$.

Furthermore, it is obvious that n different valid measurement subsets $\boldsymbol{Z}_{v,i}$ must contain a single measurement, while $q-n$ must not contain any. From a combinatorics perspective, it can be shown that there are $\binom{q}{n}$ unique ways to select n of the q valid measurement subsets. Therefore, allow the

hypothesis $H_{n,j}$ to denote the j^{th} unique selection of n choices of $\boldsymbol{Z}_{v,i}$ as considered to generate a valid measurement, where $j = 1,\, 2,\, \ldots,\, \binom{q}{n}$ accounts for all possible unique selections.

It is important to note that the hypotheses $H_{n,j}$ only specify which of the subsets $\boldsymbol{Z}_{v,i}$ contain a measurement and which ones do not. It does not make any assumption as to which measurement belongs to which subset. Therefore, in order to consider all possible events, each unique measurement assignment must be considered. Therefore, let $H^{\text{UA}}_{n,j,k}$ denote the k^{th} unique assignment of m measurements to n subsets, given that $H_{n,j}$ has already been specified.

Taking into consideration all of the possible events, the likelihood of FCMM-4 can be fashioned as

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \Pr\left(\boldsymbol{Z} \cap m \cap \Big[\bigcup_{n=0}^{\min(m,q)} n \cap \big[\bigcup_{j=1}^{\binom{q}{n}} H_{n,j} \cap \big[\bigcup_{k=1}^{m\text{P}n} H^{\text{UA}}_{n,j,k}\big]\big]\Big] \middle| \boldsymbol{x}\right). \tag{4.54}$$

It is clear that the events $n = 0, 1, \ldots, \min(m, q)$ are mutually exclusive, such that

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \Pr\left(\boldsymbol{Z} \cap m \cap n \cap \big[\bigcup_{j=1}^{\binom{q}{n}} H_{n,j} \cap \big[\bigcup_{k=1}^{m\text{P}n} H^{\text{UA}}_{n,j,k}\big]\big] \middle| \boldsymbol{x}\right).$$

Furthermore, since the valid measurement subset hypotheses $H_{n,j}$ are also mutually exclusive, it can also be shown that

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \Pr\left(\boldsymbol{Z} \cap m \cap n \cap H_{n,j} \cap \big[\bigcup_{k=1}^{m\text{P}n} H^{\text{UA}}_{n,j,k}\big] \middle| \boldsymbol{x}\right).$$

Similarly, since the unique measurement assignments of hypothesis $H^{\text{UA}}_{n,j,k}$ are mutually exclusive as well,

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\text{P}n} \Pr\left(\boldsymbol{Z} \cap m \cap n \cap H_{n,j} \cap H^{\text{UA}}_{n,j,k} | \boldsymbol{x}\right). \tag{4.55}$$

Following the rules of conditional probability, Eq. (4.55) becomes

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \Pr\left(m \cap n|\boldsymbol{x}\right) \Pr\left(\boldsymbol{Z} \cap H_{n,j} \cap H_{n,j,k}^{\mathrm{UA}}|m \cap n \cap \boldsymbol{x}\right). \tag{4.56}$$

Here, it is noted that the term $\Pr\left(m \cap n|\boldsymbol{x}\right)$ of Eq. (4.56) refers to the probability of having m measurements, where n of them are valid. This is the same as considering the probability of the event where there are $m-n$ faulty measurements, which, recalling the notation of Eq. (4.26), is expressed as

$$\Pr\left(m \cap n|\boldsymbol{x}\right) = p_F(m-n, \boldsymbol{x}),$$

which is the state-dependent probability of receiving $m-n$ faulty measurements. Thus, it can be shown that Eq. (4.56) becomes

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) &= \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} p_F(m-n, \boldsymbol{x}) \Pr\left(\boldsymbol{Z} \cap H_{n,j} \cap H_{n,j,k}^{\mathrm{UA}}|m \cap n \cap \boldsymbol{x}\right) \\ &= \sum_{n=0}^{\min(m,q)} p_F(m-n, \boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \Pr\left(\boldsymbol{Z} \cap H_{n,j} \cap H_{n,j,k}^{\mathrm{UA}}|m \cap n \cap \boldsymbol{x}\right),\end{aligned}$$

where once again, conditional probability yields

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} p_F(m-n, \boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \Pr\left(H_{n,j}|m \cap n \cap \boldsymbol{x}\right) \Pr\left(\boldsymbol{Z} \cap H_{n,j,k}^{\mathrm{UA}}|H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right), \tag{4.57}$$

where the term $\Pr\left(H_{n,j}|m \cap n \cap \boldsymbol{x}\right)$ is the probability that hypothesis $H_{n,j}$ describes the unique selection of valid measurement subsets, given the fact that n measurements are assumed to be valid. At this point, it is useful to redefine the probability of detection, given originally by Eq. (4.18). For a given valid measurement subset $\boldsymbol{Z}_{v,i}$, the state-dependent probability that the sensor detects a

feature, and thus generates a single valid measurement, is denoted as

$$\Pr(\boldsymbol{Z}_{v,i} = \{\boldsymbol{z}_{v,i}\}|\boldsymbol{x}) = p_{D,i}(1, \boldsymbol{x})\,.$$

Conversely, the probability that a given $\boldsymbol{Z}_{v,i}$ will not generate a detection is the complement

$$\begin{aligned}\Pr(\boldsymbol{Z}_{v,i} = \{\emptyset\}|\boldsymbol{x}) &= p_{D,i}(0, \boldsymbol{x}) \\ &= 1 - p_{D,i}(1, \boldsymbol{x})\,.\end{aligned}$$

It then follows that the term $\Pr\left(H_{n,j}|m \cap n \cap \boldsymbol{x}\right)$ of Eq. (4.57) is directly interpreted as the probability that n valid measurement subsets ($\boldsymbol{Z}_{v,i} \in H_{n,j}$) generate a detection while $q - n$ valid measurement subsets do not, or

$$\begin{aligned}&\Pr\left(H_{n,j}|m \cap n \cap \boldsymbol{x}\right) \\ &\quad= \Pr\left(\Big[\bigcap_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} \boldsymbol{Z}_{v,i} = \{\boldsymbol{z}_{v,i}\}\Big] \cap \Big[\bigcap_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} \boldsymbol{Z}_{v,i} = \{\emptyset\}\Big] \,\Big|\, m \cap n \cap \boldsymbol{x}\right) \\ &\quad= \prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} \Pr\left(\boldsymbol{Z}_{v,i} = \{\boldsymbol{z}_{v,i}\}|m \cap n \cap \boldsymbol{x}\right) \prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} \Pr\left(\boldsymbol{Z}_{v,i} = \{\emptyset\}|m \cap n \cap \boldsymbol{x}\right) \\ &\quad= \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1, \boldsymbol{x})\Big]\Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1, \boldsymbol{x})\Big]\,. \qquad (4.58)\end{aligned}$$

Equation (4.58) can substituted back into the likelihood of Eq. (4.57) to yield

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) = & \sum_{n=0}^{\min(m,q)} p_F(m-n,\boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \Big[\prod_{\boldsymbol{Z}_{v,i}\in H_{n,j}}^{n} p_{D,i}(1,\boldsymbol{x}) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1,\boldsymbol{x}) \Big] \\ & \times \Pr\left(\boldsymbol{Z} \cap H_{n,j,k}^{\mathrm{UA}} | H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right) \\ = & \sum_{n=0}^{\min(m,q)} p_F(m-n,\boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i}\in H_{n,j}}^{n} p_{D,i}(1,\boldsymbol{x}) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1,\boldsymbol{x}) \Big] \\ & \times \sum_{k=1}^{m\mathrm{P}n} \Pr\left(\boldsymbol{Z} \cap H_{n,j,k}^{\mathrm{UA}} | H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right),\end{aligned}$$

which again, by conditional probability, becomes

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) = & \sum_{n=0}^{\min(m,q)} p_F(m-n,\boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i}\in H_{n,j}}^{n} p_{D,i}(1,\boldsymbol{x}) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1,\boldsymbol{x}) \Big] \\ & \times \sum_{k=1}^{m\mathrm{P}n} \Pr\left(H_{n,j,k}^{\mathrm{UA}} | H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right) \Pr\left(\boldsymbol{Z} | H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right). \quad (4.59)\end{aligned}$$

The term $\Pr\left(H_{n,j,k}^{\mathrm{UA}} | H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right)$ of Eq. (4.59) denotes the probability that the unique measurement assignment hypothesis $H_{n,j,k}^{\mathrm{UA}}$ occurs, given the valid subset selection hypothesis $H_{n,j}$ and the fact that n of the m measurements are valid. If it is assumed that the order of the measurements is purely random, then all of the unique measurement assignments are equally likely. Therefore, the probability is taken to be a uniform distribution as

$$\Pr\left(H_{n,j,k}^{\mathrm{UA}} | H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right) = \frac{1}{m\mathrm{P}n} = \frac{(m-n)!}{m!},$$

which, when inserted back into Eq. (4.59), yields

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} p_F(m-n,\boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1,\boldsymbol{x}) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1,\boldsymbol{x}) \Big] \times \sum_{k=1}^{m\mathrm{P}n} \frac{(m-n)!}{m!} \Pr\left(\boldsymbol{Z}|H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right). \tag{4.60}$$

Now, upon examining the final term $\Pr\left(\boldsymbol{Z}|H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right)$ of Eq. (4.60), it is clear that this probability can be represented by the collective intersection of m measurements, or

$$\begin{aligned}\Pr\left(\boldsymbol{Z}|H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right) &= \Pr\Big(\bigcap_{r=1}^{m} \boldsymbol{z}_r | H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\Big) \\ &= \prod_{r=1}^{m} \Pr\left(\boldsymbol{z}_r | H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right),\end{aligned}$$

which becomes a product as the measurements are assumed to be generated independently.

Here it is useful to recall the individual valid measurement likelihoods of Eq. (4.49). For instance, if it is known that valid measurement subset $\boldsymbol{Z}_{v,i}$ generated a particular measurement $\boldsymbol{z}_r$, then the likelihood of $\boldsymbol{z}_r$ given the state $\boldsymbol{x}$ would be defined as

$$\ell_{v,i}(\boldsymbol{z}_r|\boldsymbol{x}) = \Pr(\boldsymbol{z}_r|(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}),$$

where the notation $(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}$ is used to further clarify the event that the measurement assignment of $\boldsymbol{z}_r$ to valid measurement subset $\boldsymbol{Z}_{v,i}$ exists in the unique measurement assignment hypothesis $H_{n,j,k}^{\mathrm{UA}}$. Similarly, the likelihood of some measurement $\boldsymbol{z}_r$ (given state $\boldsymbol{x}$), if it is known to belong the faulty measurement subset $\boldsymbol{Z}_f$, is defined as

$$\ell_f(\boldsymbol{z}_r|\boldsymbol{x}) = \Pr(\boldsymbol{z}_r|(\boldsymbol{z}_r \in \boldsymbol{Z}_f) \in H_{n,j,k}^{\mathrm{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}).$$

Therefore, it is possible to show that

$$\Pr\left(\boldsymbol{Z}|H_{n,j,k}^{\text{UA}} \cap H_{n,j} \cap m \cap n \cap \boldsymbol{x}\right) = \Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\text{UA}}}^{n} \ell_{v,i}(\boldsymbol{z}_r|\boldsymbol{x})\Big]\Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_f) \in H_{n,j,k}^{\text{UA}}}^{m-n} \ell_f(\boldsymbol{z}_r|\boldsymbol{x})\Big],$$

which can be reinserted into the FCMM-4 likelihood of Eq. (4.60) to yield

$$\begin{aligned}\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} p_F(m-n,\boldsymbol{x}) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1,\boldsymbol{x})\Big]\Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1,\boldsymbol{x})\Big] \\ \times \sum_{k=1}^{m\text{P}n} \frac{(m-n)!}{m!}\Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\text{UA}}}^{n} \ell_{v,i}(\boldsymbol{z}_r|\boldsymbol{x})\Big]\Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_f) \in H_{n,j,k}^{\text{UA}}}^{m-n} \ell_f(\boldsymbol{z}_r|\boldsymbol{x})\Big]. \end{aligned} \tag{4.61}$$

It is quite clear that FCMM-4 of Eq. (4.61) is considerably more complex than the preceding models. However, if the constraint that $q = 1$ is introduced, such that only a single valid measurement distribution is allowed, Eq. (4.61) reduces back to FCMM-2 of Eq. (4.35).

4.4 Valid Measurement Modeling

"Valid" measurements, or measurements corresponding to ideal sensor returns as in Eq. (4.2), are briefly discussed in Section 4.1. Essentially, these are the measurements that a sensor is intended to produce by design. Both the temporal statistics (frequency of measurements) and spatial statistics (value of $\boldsymbol{z}_v$) are considered in this section.

4.4.1 Valid Spatial Distribution

The spatial distribution of valid measurements is

$$\boldsymbol{z}_v \sim \ell_v(\boldsymbol{z}_v|\boldsymbol{x}), \tag{4.62}$$

which is the "traditional measurement model" mentioned in Section 4.1. For many sensors, the spatial statistics of Eq. (4.62) are readily available, as many efforts in the design, manufacturing, and calibration of sensors prioritize the correct characterization of Eq. (4.62). As such, modeling the spatial distribution of valid measurements is not a focal point of this work, and the discussion

of this section will remain brief.

4.4.1.1 Spatially Gaussian Valid Measurements

While not all sensors are subject to Gaussian noise—such as the pulsar-based navigation used by SEXTANT [89]—as previously stated in Section 2.2.2.1, most of the time assuming measurements are subject to additive, zero-mean, Gaussian noise is an acceptable assumption, such as in the EKF assumptions of Eq. (2.1c). Therefore, while admitting other noise distributions for measurements are possible, this work often assumes that valid measurements are spatially distributed as

$$\ell_v(\boldsymbol{z}_v|\boldsymbol{x}) = p_g(\boldsymbol{z}_v|\boldsymbol{h}_v(\boldsymbol{x}), \boldsymbol{R}_v)\,, \tag{4.63}$$

where $\boldsymbol{h}_v(\cdot)$ is the valid measurement function and $\boldsymbol{R}_v$ is the valid noise covariance. Note that this will be the primary valid measurement model for the remainder of the fault-cognizant filtering derivations.

4.4.1.2 Spatially Modeling Valid Measurements as GMMs

If, for some reason, sensor noise is unable to be modeled as a single Gaussian distribution, it may be prudent to model the noise via the Gaussian mixture models previously discussed in Section 2.2.2.2. Again, the main benefits of using a GMM are that it allows for a relatively flexible measurement noise (i.e. multi-modal, heavy-tails, etc.) while still utilizing the properties of Gaussians to enact closed-form GMF updates. Thus, if a GMM representation is elected, the valid measurement likelihood of Eq. (4.62) becomes

$$\ell_v(\boldsymbol{z}_v|\boldsymbol{x}) = \sum_{k=1}^{L^v} w_k^v p_g\big(\boldsymbol{z}_v|\boldsymbol{h}_k^v(\boldsymbol{x}), \boldsymbol{R}_k^v\big)\,,$$

where w_k^v, $\boldsymbol{h}_k^v(\cdot)$, and $\boldsymbol{R}_k^v$ are the weight, mean, and covariance of the k^{th} GM component.

4.4.2 Valid Temporal Distribution

The temporal distribution of valid measurements refers to the statistical frequency at which valid measurement(s) are generated, and is characterized by probability of validity for single measurement returns, such as for FCMM-1 of Section 4.3.1, and probability of detection for the FCMMs with variable sized measurement scans of Sections 4.3.2–4.3.4.

4.4.2.1 Probability of Validity vs Probability of Detection

First, it is useful to distinguish between the values of probability of detection $p_D(\cdot, \boldsymbol{x})$, probability of false alarm $p_F(\cdot, \boldsymbol{x})$, and probability of validity $v(\boldsymbol{x})$. Both probabilities $p_D(\cdot, \boldsymbol{x})$ and $p_F(\cdot, \boldsymbol{x})$ (in all their forms) are created specifically for likelihoods that permit multiple sensor returns, while $v(\boldsymbol{x})$ is intended for an FCMM limited to a single sensor return (i.e., FCMM-1). Therefore, in order to relate $v(\boldsymbol{x})$ to $p_D(\cdot, \boldsymbol{x})$ and $p_F(\cdot, \boldsymbol{x})$, it must be done within purview of scans of single measurement returns, such that

$$p_D(i, \boldsymbol{x}) = 0\,, \quad \forall\, i \geq 2 \tag{4.64a}$$

$$p_F(i, \boldsymbol{x}) = 0\,, \quad \forall\, i \geq 2\,, \tag{4.64b}$$

where it is recalled that $p_D(i, \boldsymbol{x})$ is the probability that i valid measurements are generated, and $p_F(i, \boldsymbol{x})$ is the probability that i faulty measurements are generated. By total probability, it must hold that

$$\sum_{i=0}^{\infty} p_D(i, \boldsymbol{x}) = 1 \tag{4.65a}$$

$$\sum_{i=0}^{\infty} p_F(i, \boldsymbol{x}) = 1\,, \tag{4.65b}$$

such that combining Eqs. (4.64) and (4.65) results in

$$p_D(0, \boldsymbol{x}) = 1 - p_D(1, \boldsymbol{x}) \tag{4.66a}$$

$$p_F(0, \boldsymbol{x}) = 1 - p_F(1, \boldsymbol{x}) \,. \tag{4.66b}$$

As it is desired to relate $p_D(\cdot, \boldsymbol{x})$ and $p_F(\cdot, \boldsymbol{x})$ to the probability of validity, the constraint that only one measurement can be generated must be enforced. In other words, if a valid measurement is generated, then a faulty measurement is not, and vice versa. This can be done by representing $v(\boldsymbol{x})$ as

$$v(\boldsymbol{x}) = p_F(0, \boldsymbol{x}) p_D(1, \boldsymbol{x}) \tag{4.67a}$$

$$1 - v(\boldsymbol{x}) = p_F(1, \boldsymbol{x})[1 - p_D(1, \boldsymbol{x})] \,. \tag{4.67b}$$

Solving Eqs. (4.67a) and (4.67b) for $v(\boldsymbol{x})$ and equating the expressions, it directly follows that

$$p_F(0, \boldsymbol{x}) p_D(1, \boldsymbol{x}) = 1 - p_F(1, \boldsymbol{x})[1 - p_D(1, \boldsymbol{x})] \,,$$

from which it is seen that

$$p_D(1, \boldsymbol{x}) = \frac{p_F(0, \boldsymbol{x})}{2p_F(0, \boldsymbol{x}) - 1} \,,$$

or, that the probability of detection is reliant upon the probability of false alarm when the FCMM is limited to single sensor returns. Whether or not this interdependency is desirable must be considered on a case by case basis, but if it is desired to model $p_D(\cdot, \boldsymbol{x})$ and $p_F(\cdot, \boldsymbol{x})$ independently of one another, an FCMM that allows for variable sized measurement scans must be selected.

Since probability of validity closely aligns with the probability of detection, much of the upcoming discussion on the modeling of $p_D(\cdot, \boldsymbol{x})$ also applies to $v(\boldsymbol{x})$, making discussion of both rather redundant. Therefore, most temporal modeling of valid measurements will be done in re-

gard to the probability of detection, while acknowledging that the proposed models also apply to $v(\boldsymbol{x})$.

4.4.2.2 Low Fidelity Model: Neglecting p_D

It is not uncommon to implement filters that, instead of accounting for events like missed-detections, elect to ignore such sensor behavior, resulting in updates that are only performed in the presence of some return from the sensor. In such filters, no model for probability of detection is needed; instead, these filters commonly rely upon *ad hoc* extensions like residual editing to screen out erroneous sensor data in order to protect filtering operations [21]. While these filters tend to be computationally efficient enough for most real-time navigation applications, they often exchange the burden of constructing more accurate sensor models for an increase in manual parameter tuning and the requirement for extra machinery to ensure robust filtering operations. Nevertheless, these filters, wherein probabilities such as $p_D(\cdot, \boldsymbol{x})$ are completely disregarded, remain some of the most proven estimation architectures in existence and should not be dismissed outright [5, 7, 81]. To this end, a nonlinear GM filter with residual editing is introduced in Eq. (4.7.2) as a fair representative of a filter with a so-called low fidelity model for the temporal distribution of valid measurements.

4.4.2.3 Medium Fidelity Model: Zeroth-Order Approximation

There are many cases where estimation processes must be performed on a system where the probability of detection is state-dependent, and neglecting this reality outright—as in Section 4.4.2.2—is deemed overly detrimental to filtering operations. In cases such as these, the most common practice is to repeatedly approximate $p_D(\cdot, \boldsymbol{x})$ during each update so that it may be treated as state-independent, which this work considers to be a medium fidelity approach to modeling the temporal distribution of valid measurements.

Considering that the probability of i detections $p_D(i, \boldsymbol{x})$ is some (possibly nonlinear) function of state $\boldsymbol{x}$, it can be expressed via a Taylor series as

$$p_D(i, \boldsymbol{x}) = \sum_{k=0}^{\infty} \frac{(\boldsymbol{x} - \boldsymbol{m})^k}{k!} \left[\frac{\mathrm{d}^k p_D(i, \boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^k} \right]_{\boldsymbol{x}=\boldsymbol{m}}.$$

A zeroth-order approximation of $p_D(i, \boldsymbol{x})$ is made by evaluating the $k = 0$ term of the Taylor series and neglecting higher-order terms such that

$$\begin{aligned} p_D(i, \boldsymbol{x}) &= p_D(i, \boldsymbol{m}) + \mathcal{O}(\boldsymbol{x}) \\ &\approx p_D(i, \boldsymbol{m}) \,, \end{aligned} \tag{4.68}$$

where $\boldsymbol{m}$ can be selected in a number of ways, depending on the desired behavior of the filter [90]. As GMMs are of primary interest to this work, two potential, appropriate choices for $\boldsymbol{m}$ are considered.

4.4.2.3.1 Zeroth-Order Approximation About Overall Prior

This first method is considered to be a more conservative approach, in that it approximates $p_D(\cdot, \boldsymbol{x})$ before the measurement update actually occurs. It is done by evaluating $p_D(\cdot, \boldsymbol{x})$ at the single point $\boldsymbol{x} = \boldsymbol{m}^-$, which is the overall mean of Eq. (2.22) extracted from a prior GM $p^-(\boldsymbol{x})$ such as the one of Eq. (2.9). This method produces an approximation of $p_D(\cdot, \boldsymbol{x})$ before the processing of measurements, while also treating the GM as a single distribution by operating on the overall mean instead of the means of individual components. However, this does require the extra step of merging the GM prior, and thus is not used much in this work.

4.4.2.3.2 Zeroth-Order Approximation About Individual GM Components

Alternative to Section 4.4.2.3.1, noting that $p_D(\cdot, \boldsymbol{x})$ is often directly multiplied with individual GM components, the probability of detection may be approximated such that

$$\sum_{\ell=1}^{L} w_\ell p_D(\cdot, \boldsymbol{x}) p_g\big(\boldsymbol{x}\big|\boldsymbol{m}_\ell, \boldsymbol{P}_\ell\big) \approx \sum_{\ell=1}^{L} w_\ell p_D(\cdot, \boldsymbol{m}_\ell) p_g\big(\boldsymbol{x}\big|\boldsymbol{m}_\ell, \boldsymbol{P}_\ell\big) \,. \tag{4.69}$$

Note that this method allows for the calculation of $p_D(\boldsymbol{x})$ following measurement processing, while also evaluating $p_D(\cdot, \boldsymbol{x})$ at each GM component across the entire distribution. As such, this second method is selected as the main manner by which the zeroth-order approximation of $p_D(\cdot, \boldsymbol{x})$ is

carried out.

When is a Zeroth-Order Approximation Acceptable?

The approximation of Eq. (4.68) is generally considered acceptable in cases where the probability of detection is relatively constant compared to the state $\boldsymbol{x}$ [91]. In reality, this criteria is frequently violated; the probability of detection often switches between high and low values immediately as an object transitions over a sensor's field-of-view. Furthermore, no state information can be gleaned directly from the probability of detection during the measurement update. As such, estimation operations in some systems may fail under this treatment of $p_D(\cdot, \boldsymbol{x})$.

4.4.2.4 High Fidelity Model: Gaussian Modeled Probabilities

In order to derive a state-dependent probability of detection that lends itself to a closed-form update, the beneficial properties of Gaussians mentioned in Section 2.2.2.1 are recalled. Consider the case where the probability of detection is modeled as a Gaussian distribution of the form

$$p_D(\cdot, \boldsymbol{x}) = p_g\big(p_{D_z} \big| g(\boldsymbol{x}), R_D\big)\,, \tag{4.70}$$

where $g(\cdot)$ is some nonlinear function of the state, R_D is an associated variance, and p_{D_z} is a random variable drawn from the distribution of $p_D(\cdot, \boldsymbol{x})$. Immediately, it is clear that some restrictions should be placed upon the elements of Eq. (4.70), as the probability of detection must exist within the interval $[0, 1]$ to be a valid probability. However, as the support of a Gaussian extends into infinity, the assumption of Eq. (4.70) poses a possible violation to this condition. To address this incompatibility, note that most sensor models are founded on similar fallacies; most sensor returns, while assumed to have Gaussian noise, are in reality bounded to exist within some finite interval pursuant to either some physical constraint of the system or behavior enforced upon the sensor. Therefore, pending careful treatment of Eq. (4.70), the Gaussian assumption on $p_D(\cdot, \boldsymbol{x})$ should be comfortable enough for practical use. Even so, it is highly recommended to define $g(\cdot)$ and

R_D such that the vast majority of the pdf $p_D(\cdot, \boldsymbol{x})$ lies within $[0, 1]$, as well as enforcing all p_{D_z} realizations to remain within that interval when simulated, even if the filtering algorithms treat p_{D_z} as unbounded.

In regard to FCMM-2, specifically, consider the case of single valid measurements, such that the constraints of Eqs. (4.64a) and (4.66a) hold. Furthermore, assume that $p_D(1, \boldsymbol{x})$ is Gaussian distributed as in Eq. (4.70). Recalling Section 2.2.2.1, where it is noted that a Gaussian random variable will remain a Gaussian under a linear transformation, the probability of missed detection $p_D(0, \boldsymbol{x})$ is known to be distributed as

$$\begin{aligned} p_D(0, \boldsymbol{x}) &= \big[1 - p_D(1, \boldsymbol{x})\big] \\ &= p_g\big(\big[1 - p_{D_z}\big] \big| \big[1 - g(\boldsymbol{x})\big], R_D\big) \,. \end{aligned} \tag{4.71}$$

The transformation of Eq. (4.71) is especially useful when developing the high fidelity FCU corresponding to FCMM-2 in Section 4.6.3.2, as it reduces the number of GM components created with each update, as well as directly enforces non-negativity upon the posterior pdf. If, alternatively, the "1" and "$-p_D(1, \boldsymbol{x})$" of Eq. (4.71) are treated independently of one another, linearization errors can cause the pdf to go negative, while also producing more GM components within the posterior pdf.

4.5 Faulty Measurement Modeling

At this point, the existence of faulty measurements has been discussed in Section 4.2, which justified the derivations of four fault-cognizant measurement models in Section 4.3. However, the manner by which these faulty measurements are actually distributed, both spatially and temporally, is yet to be clarified—a necessary step to produce filtering solutions. As such, this section explores possible statistical distributions for faulty measurements.

4.5.1 Faulty Temporal Distribution

The temporal distribution of faulty measurements is completely characterized by the probability of validity $v(\boldsymbol{x})$ for FCMM-1 of Section 4.3.1, and the probability of false alarm $p_F(\cdot, \boldsymbol{x})$ for the other FCMMs. This section contains discussions on both.

4.5.1.1 Example: Physical Interpretation of Probability of False Alarm

In order to better understand how false alarms due to background sensor noise can be created, a basic concept in signal processing is presented, similar to the hypothesis testing example of [55]. Consider the case where a scalar signal z is received, neglecting dependence on state $\boldsymbol{x}$ for simplicity. In the event that the signal is a false alarm ($z = z_f$), it is distributed according to some density $\ell_f(z|z \in \mathbb{F})$. Conversely, if the signal that is generated is valid ($z = z_v$), then it is distributed according to $\ell_v(z|z \in \mathbb{V})$. It is further assumed that a valid signal generally produces a "stronger" signal, where "stronger" is taken as

$$\mathbb{E}\{z_v\} > \mathbb{E}\{z_f\}\,,$$

where $\mathbb{E}\{\cdot\}$ denotes expectation.

Next, the concept of a detection threshold γ is introduced, where signal z is only returned by the sensor when $z \geq \gamma$. This detection threshold can be set in a variety of ways, but is generally intended to limit the number of false alarms (faulty measurements) returned while still including the majority of the valid signals. Figures (4.9) and (4.10) give a visual representation of a possible signal processing scenario, where $\ell_f(z|z \in \mathbb{F})$ and $\ell_v(z|z \in \mathbb{V})$ are pdfs, and γ is a possible detection threshold. Again omitting state-dependencies for simplicity, the probability that a valid signal is returned (i.e., probability of detection) is essentially calculated via

$$p_D(1) = \int_{\gamma}^{\infty} \ell_v(z|z \in \mathbb{V})\mathrm{d}z\,,$$

which is the shaded region of Fig. (4.9). Similarly, the probability that a single false alarm signal

is returned is

$$p_F(1) = \int_{\gamma}^{\infty} \ell_f(z|z \in \mathbb{F})\mathrm{d}z\,,$$

which is the shaded region of Fig. (4.10).

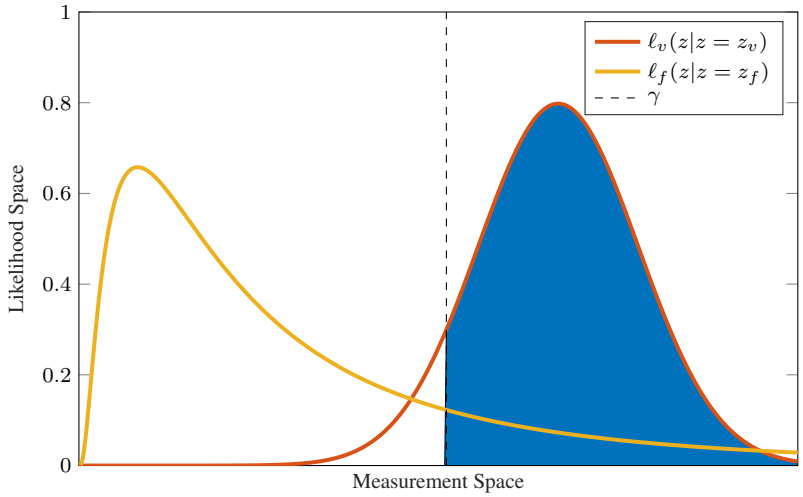


Figure 4.9 Signal likelihoods with shaded probability of detection for threshold γ

While not representative of all sensor models, this method of addressing the probabilities of detection/false alarm is relatively common within the field of estimation. A direct comparison can be made between this example and the voltage thresholding of photodiode photoreceivers in [92] used for LiDAR and laser rangefinders .

4.5.1.2 Probability of False Alarm vs. False Alarm Rate

When developing temporal faulty measurement models for a wide breadth of sensors, it is not unusual for a sensor to provide a false alarm rate (FAR) rather than probability of false alarm $p_F(\cdot, \boldsymbol{x})$ [93]. While the reasoning behind this may be different from sensor to sensor, the use of FAR is usually attributed to the method by which a sensor is calibrated. For example, laser rangefinders (LRFs) are able to set a detection threshold based on the number of returns received when

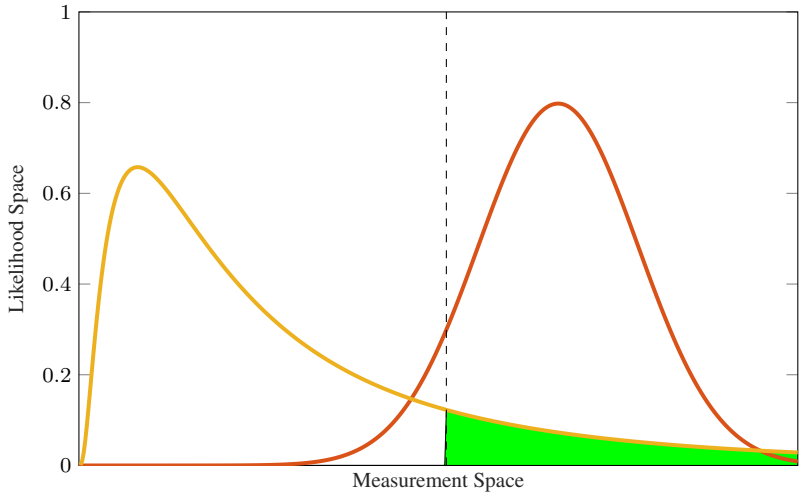


Figure 4.10 Signal likelihoods with shaded probability of false alarm for threshold γ

no valid returns are expected [94]. If no optical signal is produced by the LRF, then clearly all m sensor returns received are false alarms. Therefore, the FAR is the ratio of the number of returns to the time interval of the calibration process, which is

$$\text{FAR} = \frac{m}{\Delta t_{\text{cal}}} \, .$$

A more restrictive detection threshold will decrease the number of m false alarms and consequently lower the FAR. Accordingly, the detection threshold can be tuned such that a desired FAR is reached.

As filtering algorithms are more readily built on probabilities of m false alarms rather than FARs it is useful to be able to construct $p_F(m, \boldsymbol{x})$ from FAR. The concept behind this is simple, as it is clear that the FAR must be multiplied by some amount of time to recover a probability. In fact, the probability that a single false alarm will occur is

$$p_F(1, \boldsymbol{x}) = \text{FAR} \times \Delta t_{\text{op}}(\boldsymbol{x}) \, ,$$

where $\Delta t_{\text{op}}(\boldsymbol{x})$ is a sensor operation window that may be dependent upon the state. If it is assumed

that the false alarms are independently generated, then the probability of m false alarms is simply

$$p_F(m, \boldsymbol{x}) = \prod_{i=1}^{m} p_F(1, \boldsymbol{x}) = p_F(1, \boldsymbol{x})^m \,.$$

For any measurement set $\boldsymbol{Z}$ containing m measurement returns, the probability that none are false alarms is given by

$$p_F(0, \boldsymbol{x}) = 1 - p_F(m, \boldsymbol{x}) \,.$$

It is noted that typical FARs provided by sensors will be state-independent, and thus the probability of false alarm is only dependent upon the state if $\Delta t_{\text{op}} = \Delta t_{\text{op}}(\boldsymbol{x})$. Furthermore, FARs usually correspond to faulty measurements generated during "idle" sensor calibration, and as such, may only attribute for faulty measurements generated by background signal noise.

4.5.1.3 State-Dependency of Probability of False Alarm

The probability of false alarm in the FCMMs of Sections 4.3.2–4.3.4 is assumed to be state-dependent, as $p_F(\cdot, \boldsymbol{x})$ is a function of the state. While it is useful to allow the temporal distribution of false alarms to change with respect to the state vector, it is a major inconvenience when attempting to find a closed-form Bayesian update, a problem that also arises for probability of detection. As such, the same approaches to deal with a state-dependent probability of detection in Sections 4.4.2.2–4.4.2.4 can be used for probability of false alarm. Specifically, the zeroth-order approximation of Eq. (4.68) can be directly applied to $p_F(\cdot, \boldsymbol{x})$ as well and is the most advisable approach should state-dependency be deemed non-negligible.

4.5.1.4 Temporally Poisson Faulty Measurements

Even with the modeling discussions of Sections 4.5.1.1 and 4.5.1.2, models for probability of false alarm are typically difficult to find and can vary significantly between sensors and systems. In lieu of deriving a realistic, state-dependent model, a typical approach is to model faulty measurements as temporally state-independent events that occur at a rate that is described by a Poisson

distribution, such that probability of i false alarms is

$$p_F(i) = \frac{\lambda^i}{i! \mathrm{e}^{\lambda}}, \tag{4.72}$$

where λ is the average number of false alarms expected to be contained in sensor return $\boldsymbol{Z}$. As the credibility of this assumption is well-established in [48], this work is justified in utilizing this assumption for the majority of the fault-cognizant update derivations. Note that the Poisson probability of false alarm is used frequently in other clutter-related literature [19, 95]. The probability of Eq. (4.72) is not only applicable to a variety of systems, but also eliminates the troublesome state-dependency.

4.5.2 Faulty Spatial Distribution

Previously, Section 4.2 assumed that faulty measurements are spatially IID according to Eq. (4.3) but did not further prescribe any specific statistical distributions. As such, this section explores different possible spatial distributions that may be used when modeling faulty measurements.

4.5.2.1 Spatially Uniform Faulty Measurements

Consider the potential behavior and traits of a physical sensor that returns scalar measurements z_i. In most cases, there exists a bounded domain of possible sensor returns that can be generated, such that all measurements (valid and faulty) must lie within some field-of-view of the sensor. Practically, this can either be (1) the natural result of a sensor's impossibility to generate returns outside of a certain region of the measurement space, or (2) strictly enforced by only permitting sensor returns that occur within some specified domain of allowable operation. For example, a typical range sensor will not report ranges that are negative, as that is outside the sensor's natural field-of-view, or that are greater than a maximum effective distance, which is an enforced bound beyond which any returns are considered untrustworthy [96]. Thus, all incoming range measurements, both valid and faulty, will occur within the region $[0, \text{Max Range}]$. Due to this, the first spatial faulty measurement model considered is one where all scalar faulty measurements $z_{f,i}$ must

be sensor returns that exist within a bounded domain as

$$z_{f,i} \in [a_i, b_i] \, . \tag{4.73}$$

Furthermore, while a valid measurement is more likely to occur at some values as opposed to others subject to the likelihood $\ell_v(\boldsymbol{z}|\boldsymbol{x})$, it is sometimes sensible to assume that any faulty measurement has an equally likely chance to exist at all possible values within the sensor domain [97]. In accordance, the faulty likelihood function for a sensor with scalar returns can be assumed to be generated by a uniform distribution as

$$\ell_{f,i}(z_i) = \begin{cases} \frac{1}{b_i - a_i} & , \quad \forall \; z_i \in [a_i, b_i] \\ 0 & , \quad \text{otherwise} \end{cases} \, , \tag{4.74}$$

for $b_i > a_i$. This notion of uniformly distributed scalar faulty measurements can be extended to m-dimensional vector sensor returns of the form

$$\boldsymbol{z}_f = \begin{bmatrix} z_{f,1} & z_{f,2} & \cdots & z_{f,m} \end{bmatrix}^T \, . \tag{4.75}$$

Note that, due to the FCMMs being derived under the assumption that $(\boldsymbol{z} = \boldsymbol{z}_v)$ and $(\boldsymbol{z} = \boldsymbol{z}_f)$ are mutually exclusive events, it follows that the entire measurement vector is either completely valid or completely faulty; there cannot exist valid and invalid scalar elements within the same vector measurement. If it is desired that such an effect be modeled, it is recommended to process the individual scalar components of the measurement vector as individual measurements.

The likelihood corresponding to the vector measurement of Eq. (4.75) can be expressed as

$$\ell_f(\boldsymbol{z}) = \ell_f(z_1 \cap z_2 \cap \cdots \cap z_m) \, , \tag{4.76}$$

or, the overall faulty spatial distribution for the vector measurement is the likelihood of the intersection of the events given all of the individual scalar components are faulty. This essentially

enforces the aforementioned condition, that either all or none of the scalar measurements are faulty for a given sensor return. If it is assumed that the scalar components of the faulty measurements are generated independently of one another, the multivariate likelihood function of Eq. (4.76) can be expressed as

$$\begin{aligned}\ell_f(\boldsymbol{z}) &= \ell_{f,1}(z_1)\ell_{f,2}(z_2)\cdots\ell_{f,m}(z_m) \\ &= \prod_{i=1}^{m}\ell_{f,i}(z_i) \\ &= \prod_{i=1}^{m}\begin{cases}\frac{1}{b_i-a_i} & , \quad \forall\ z_i \in [a_i, b_i] \\ 0 & , \qquad \text{otherwise}\end{cases} .\end{aligned} \tag{4.77}$$

Based on the form of Eq. (4.77), it is clear that any time a scalar measurement z_i falls outside of its respective bounds, such that either $z_i > b_i$ or $z_i < a_i$, the result will be the zero likelihood event $\ell_f(\boldsymbol{z}) = 0$. However, as the condition that physical sensor returns, valid or faulty, will always exist inside the domain $[a_i, b_i]$ is enforced, such an event is impossible, and the likelihood of Eq. (4.77) can be restated simply as

$$\ell_f(\boldsymbol{z}) = \prod_{i=1}^{m}\frac{1}{b_i - a_i} . \tag{4.78}$$

If the size of the domain for each scalar measurement is regarded as a one-dimensional sensor "length", then an m-dimensional sensor "volume" can be defined as

$$V = \prod_{i=1}^{m}\left(b_i - a_i\right) ,$$

such that Eq. (4.78) can be expressed in terms of this so-called sensor volume as

$$\ell_f(\boldsymbol{z}) = \frac{1}{V} . \tag{4.79}$$

It is important to note that the uniformly distributed faulty likelihood of Eq. (4.79) is independent

of the state, signifying that no information concerning the explicit condition of the state $\boldsymbol{x}$ can be inferred from faulty measurements of this type.

4.5.2.2 Spatially Normal Faulty Measurements

Instead of assuming uniformly distributed faulty measurements as in Section 4.5.2.1, consider the case where each individual faulty measurement is generated according to some state-dependent Gaussian distribution, such that

$$\boldsymbol{z}_f = \boldsymbol{h}_f(\boldsymbol{x}) + \boldsymbol{w}_f \ , \tag{4.80}$$

where $\boldsymbol{h}_f(\cdot)$ is some (potentially) nonlinear function and $\boldsymbol{w}_f \sim p_g(\boldsymbol{w}_f|0, \boldsymbol{R}_f)$. Therefore, the state-dependent faulty likelihood subject to this zero-mean Gaussian noise becomes

$$\ell_f(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{h}_f(\boldsymbol{x}), \boldsymbol{R}_f) \ . \tag{4.81}$$

When the faulty spatial distribution takes the form of Eq. (4.81), it can be shown that the FCMM-1 of Eq. (4.14) becomes

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = v(\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + \big[1 - v(\boldsymbol{x})\big] p_g(\boldsymbol{z}|\boldsymbol{h}_f(\boldsymbol{x}), \boldsymbol{R}_f) \ . \tag{4.82}$$

Operating on the assumption that the faulty measurements are normally distributed is advantageous, as, again, Gaussian distributions have many preferred mathematical properties. Unlike the uniformly distributed likelihood of Eq. (4.79), the Gaussian likelihood no longer requires bounds on sensor returns, instead requiring a mean and covariance. Furthermore, it also permits dependency upon the state, allowing for information gain explicating $\boldsymbol{x}$ from the spatial distribution of faulty measurements. While aspects of this faulty measurement model seem favorable relative to the uniform measurements of Section 4.5.2.1, there may be doubt as to whether faulty measurements actually occur in this fashion. To assuage legitimacy concerns, a possible Gaussian model for faulty measurements is explained in terms of the FCMM-1 likelihood of Eq. (4.82).

Consider a valid scalar measurement generated as

$$z_v = h_v(\boldsymbol{x}) + w_v \ , \tag{4.83}$$

where w_v is a zero-mean white noise sequence with covariance R_v. Based on Eq. (4.83), the valid likelihood takes on the form

$$\ell_v(z|\boldsymbol{x}) = p_g\big(z|h_v(\boldsymbol{x}), R_v\big) \ , \tag{4.84}$$

which is the scalar counterpart to the fairly common assumption made in Eq. (2.1c). Now, it is proposed that faulty measurements occur in a similar fashion, such that they are normally distributed around the mean $h_v(\boldsymbol{x})$. However, the faulty measurement noise covariance will differ as $R_f > R_v$, such that

$$z_f = h_v(\boldsymbol{x}) + w_f \tag{4.85}$$

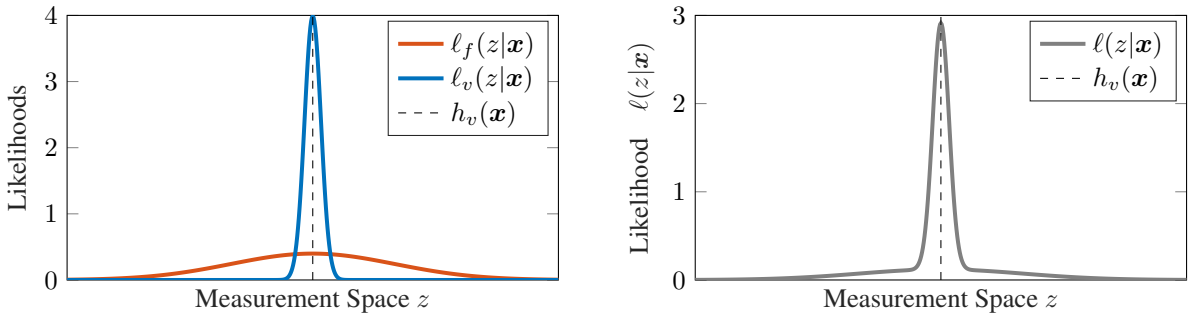
and

$$\ell_f(z|\boldsymbol{x}) = p_g\big(z|h_v(\boldsymbol{x}), R_f\big) \ . \tag{4.86}$$

The rationale for this normally distributed faulty measurement model arises from the assumption that invalid measurements are more likely to occur near the actual measurement mean [97]. If both of these likelihoods are plotted over one another, Fig. (4.11a) shows a comparison for the case of $\sqrt{R_f} = 10\sqrt{R_v}$.

Taking the Gaussian likelihoods of Eqs. (4.84) and (4.86), the FCMM-1 likelihood of Eq. (4.82) becomes

$$\ell(z|\boldsymbol{x}) = v(\boldsymbol{x})p_g\big(z|h_v(\boldsymbol{x}), R_v\big) + \big[1 - v(\boldsymbol{x})\big]p_g\big(z|h_v(\boldsymbol{x}), R_f\big) \ . \tag{4.87}$$



a) Valid vs. Faulty Likelihoods **b) Fault-Cognizant Likelihood**

Figure 4.11 Example likelihoods for normally distributed faulty measurements for $v(\boldsymbol{x}) = 0.7$

Notice that Eq. (4.87) is essentially a two component Gaussian mixture (GM) with state-dependent weights $v(\boldsymbol{x})$ and $1 - v(\boldsymbol{x})$. For simplicity, the probability of validity is taken to be constant such that $v(\boldsymbol{x}) = 0.7$, and the resulting FCMM-1 likelihood of Eq. (4.87) is plotted in Fig. (4.11b). The vast majority of measurements will exist within the main central peak, yet this distribution also appreciates the small chance that measurements are generated farther from the mean than $\ell_v(z|\boldsymbol{x})$ alone would expect. Increasing the variance R_f has the effect of enlarging the portion of the measurement space where faulty measurements are expected. This increase does not change the expected frequency of faulty measurements, as that is attributed to $v(\boldsymbol{x})$ alone, but instead "flattens" the faulty likelihood profile of Fig. (4.11a) and elongates the "tails" of Fig. (4.11b).

4.5.2.3 Approximating Spatial Distribution of Faulty Measurements with GMMs

As previously noted in Section 4.5.2.2, there are many benefits to using Gaussians when modeling measurement distributions. Yet to assume that a distribution is Gaussian, no matter how broadly it may apply, still enforces a restrictive assumption upon the update. In order to take advantage of Gaussian properties while avoiding restricting the faulty likelihood to a specific distribution, the popular concept of approximating distributions via Gaussian mixture models is again introduced, but this time for the spatial distribution of faulty measurements.

Recalling the discussion on GMMs from Section 2.2.2.2, this section presumes that the actual

faulty likelihood $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ is able to be approximated by an L_f-component GM as

$$\ell_f(\boldsymbol{z}|\boldsymbol{x}) \approx \sum_{j=1}^{L_f} w_j^f p_g\big(\boldsymbol{z}|\boldsymbol{h}_j^f(\boldsymbol{x}), \boldsymbol{R}_j^f\big) \ , \tag{4.88}$$

where w_j^f, $\boldsymbol{h}_j^f(\boldsymbol{x})$, and $\boldsymbol{R}_j^f$ are the weight, mean, and covariance of the j^{th} GM component, respectively. In general, an increase in the number of components enables a more accurate approximation of $\ell_f(\boldsymbol{z}|\boldsymbol{x})$. With this in mind, the key to this method is to find an appropriate distribution that models faulty measurement returns. Two notions occur immediately: first, that each faulty measurement distribution will be sensor specific, and second, that an accurate model will most likely require in-depth knowledge of the sensor behavior. However, as long as some GM representation of the likelihood can be produced, a closed-form solution to the update equation (assuming proper treatment of the state-dependencies in $p_D(\cdot, \boldsymbol{x})$, $p_F(\cdot, \boldsymbol{x})$, and $v(\boldsymbol{x})$) is possible.

While the GM likelihood of Eq. (4.88) may be used to only approximate the actual faulty measurement spatial distribution, examining the manner in which the GM likelihood model would generate physical faulty measurements is still insightful. Consider a special case where a faulty measurement is guaranteed to be returned by a sensor. Then, there exist L_f mutually independent events denoted by

$$\boldsymbol{z}_f = \boldsymbol{z}_{f,j} \ ,$$

indicating the event where the faulty measurement is generated according to the j^{th} model, or

$$\boldsymbol{z}_f = \boldsymbol{h}_j^f(\boldsymbol{x}) + \boldsymbol{w}_{f,j} \quad \text{if} \quad \boldsymbol{z}_f = \boldsymbol{z}_{f,j} \ ,$$

where $\boldsymbol{w}_{f,j} \sim p_g(\boldsymbol{w}_{f,j}|\boldsymbol{0},\ \boldsymbol{R}_j^f)$. The probability of event $\boldsymbol{z}_f = \boldsymbol{z}_{f,j}$ is equivalent to the j^{th} GM weight as

$$\Pr(\boldsymbol{z}_f = \boldsymbol{z}_{f,j}) = w_j^f \ .$$

Assuming faulty measurements are distributed in this manner, the FCMM-1 likelihood of Eq. (4.14) yields

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = v(\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + \big[1 - v(\boldsymbol{x})\big] \sum_{j=1}^{L_f} w_j^f p_g\big(\boldsymbol{z}|\boldsymbol{h}_j^f(\boldsymbol{x}), \boldsymbol{R}_j^f\big) \,. \tag{4.89}$$

4.6 Fault-Cognizant Updates

While much discussion is contained in the previous sections of this chapter concerning fault-cognizant modeling, no operable filter forms have been presented as of yet. As such, this section details several closed-form updates by selecting 1) a FCMM from Section 4.3, 2) a valid measurement model from Section 4.4, and 3) a faulty measurement model from Section 4.5, the collection of which are inserted into Eq. (3.1) to form a Bayesian update.

4.6.1 Closed-Form Related GMM Assumptions

For the sake of practical, recursive algorithms, the filters derived in this section are specifically designed to have closed forms; that is, the form of the inputs to the update should match the form of its outputs. To accomplish this, the prior distribution $p^-(\boldsymbol{x})$ is taken to be a GMM of the form

$$p^-(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^- \,,\, \boldsymbol{P}_\xi^-) \,, \tag{4.90}$$

which is the same form as a prior that is propagated by a GSF in Eq. (2.9) (with the omission of time index k). Therefore, not only will Eq. (4.90) grant closed-form filtering solutions, but each filter will be nonlinear by default.

4.6.2 FCU-1: Scans of Single Measurements

The first type of fault-cognizant update (FCU-1) corresponds to FCMM-1 of Section 4.3.1, where it is assumed that sensor scans contain exactly one measurement, the nature of which can either be valid or faulty. As such, the general fault-cognizant measurement likelihood of each filter

corresponds to Eq. (4.14), such that a Bayesian update takes the form

$$\begin{aligned} p^+(\boldsymbol{x}) &\propto p^-(\boldsymbol{x})\ell(\boldsymbol{Z}|\boldsymbol{x}) \\ &\propto v(\boldsymbol{x})p^-(\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + \big[1 - v(\boldsymbol{x})\big]p^-(\boldsymbol{x})\ell_f(\boldsymbol{z}|\boldsymbol{x})\,. \end{aligned} \tag{4.91}$$

Additionally, all valid measurements are assumed to be Gaussian distributed as per Eq. (4.63), and since this update type is limited to single measurement returns, the probabilities of detection and false alarm are replaced by the probability of validity per the discussion of Section 4.4.2.1, which, for the sake of simplicity, is assumed state-independent by the zeroth-order approximation of Section 4.4.2.3, such that $v(\boldsymbol{x}) = v$. Note that investigations into state-dependent probabilities of detection are addressed in Section 4.6.3. Under these conditions, the update of Eq. (4.91) becomes

$$p^+(\boldsymbol{x}) \propto \sum_{\xi=1}^{L^-} \Big[v w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) p_g(\boldsymbol{z}|\boldsymbol{h}_v(\boldsymbol{x}), \boldsymbol{R}_v)\Big] + \sum_{\xi=1}^{L^-} \Big[[1-v] w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-)\ell_f(\boldsymbol{z}|\boldsymbol{x})\Big]\,. \tag{4.92}$$

Therefore, the only modeling assumption that remains unspecified is the spatial distribution of faulty measurements $\ell_f(\boldsymbol{z}|\boldsymbol{x})$, three of which are presented in Section 4.5.2. Thus, three distinct FCU-1 updates are derived in this section, each one corresponding to a different faulty spatial distribution.

4.6.2.1 FCU-1 with Spatially Uniform Faulty Measurements

Here, the algorithms for a FCU-1 are derived by taking the update of Eq. (4.92) and assuming that faulty measurements are uniformly distributed across the sensor space, pursuant to Section 4.5.2.1. In this case, when $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ is defined according to Eq. (4.79), the posterior of

Eq. (4.92) becomes

$$p^+(\boldsymbol{x}) \propto \sum_{\xi=1}^{L^-} \left[v w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) p_g(\boldsymbol{z}|\boldsymbol{h}_v(\boldsymbol{x}), \boldsymbol{R}_v) \right] + \sum_{\xi=1}^{L^-} \left[\frac{1-v}{V} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \right]. \quad (4.93)$$

Making use of the identity of Eqs. (A.2), it can be shown that Eq. (4.93) becomes

$$p^+(\boldsymbol{x}) \propto v \sum_{\xi=1}^{L^-} \left[\kappa_\xi w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+) \right] + \frac{1-v}{V} \sum_{\xi=1}^{L^-} \left[w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \right], \quad (4.94)$$

where

$$\boldsymbol{m}_\xi^+ = \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi[\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}_\xi^-)] \quad (4.95a)$$

$$\boldsymbol{P}_\xi^+ = \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H}_v(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \quad (4.95b)$$

$$\boldsymbol{W}_\xi = \boldsymbol{H}_v(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \boldsymbol{H}_v^T(\boldsymbol{m}_\xi^-) + \boldsymbol{R}_v \quad (4.95c)$$

$$\boldsymbol{K}_\xi = \boldsymbol{P}_\xi^- \boldsymbol{H}_v^T(\boldsymbol{m}_\xi^-) \boldsymbol{W}_\xi^{-1} \quad (4.95d)$$

$$\kappa_\xi = p_g\big(\boldsymbol{z}\big|\boldsymbol{h}_v(\boldsymbol{m}_\xi^-), \boldsymbol{W}_\xi\big), \quad (4.95e)$$

and where $\boldsymbol{H}_v(\boldsymbol{m})$ is the Jacobian of the valid measurement function $\boldsymbol{h}_v(\cdot)$ evaluated at $\boldsymbol{x} = \boldsymbol{m}$. Therefore, the normalized posterior is

$$p^+(\boldsymbol{x}) = \frac{v \sum_{\xi=1}^{L^-} \left[\kappa_\xi w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+) \right] + \frac{1-v}{V} \sum_{\xi=1}^{L^-} \left[w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \right]}{v \sum_{\xi=1}^{L^-} \left[\kappa_\xi w_\xi^- \int p_g(\boldsymbol{s}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+) \mathrm{d}\boldsymbol{s} \right] + \frac{1-v}{V} \sum_{\xi=1}^{L^-} \left[w_\xi^- \int p_g(\boldsymbol{s}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \mathrm{d}\boldsymbol{s} \right]}. \quad (4.96)$$

As $p_g(\boldsymbol{s}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+)$ and $p_g(\boldsymbol{s}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-)$ are both pdfs of the state, they integrate to unity over the support of $\boldsymbol{s}$. Therefore, Eq. (4.96) becomes

$$p^+(\boldsymbol{x}) = \sum_{\xi=1}^{L^-} \left[w_\xi^v p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+) \right] + \sum_{\xi=1}^{L^-} \left[w_\xi^f p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \right], \quad (4.97)$$

where

$$w_\xi^v = \frac{v\kappa_\xi w_\xi^-}{\eta}\,, \qquad w_\xi^f = \frac{(1-v)w_\xi^-}{V\eta}\,, \qquad \text{and } \eta = v\sum_{\xi=1}^{L^-}\left[\kappa_\xi w_\xi^-\right] + \frac{1-v}{V}\,, \tag{4.98}$$

such that the posterior distribution of Eq. (4.97) is a GM with $2L^-$ number of components. As a result, each application of the update doubles the number of components within the GM description of the estimate distribution.

4.6.2.2 FCU-1 with Spatially GM Faulty Measurements

This section provides a description of an FCU-1 update where the faulty spatial distribution is taken to be a Gaussian mixture, following the model presented in Section 4.5.2.3. Assuming that $\ell_f(\boldsymbol{z}|\boldsymbol{x})$ does indeed equate to Eq. (4.88), the posterior distribution of Eq. (4.92) becomes

$$p^+(\boldsymbol{x}) \propto v\sum_{\xi=1}^{L^-}\left[w_\xi^-\kappa_\xi^v p_g(\boldsymbol{x}|\boldsymbol{\mu}_\xi^v,\boldsymbol{\Pi}_\xi^v)\right] + (1-v)\sum_{\xi=1}^{L^-}\sum_{j=1}^{L^f}\left[w_\xi^- w_j^f\kappa_{\xi,j}^f p_g(\boldsymbol{x}|\boldsymbol{\mu}_{\xi,j}^f,\boldsymbol{\Pi}_{\xi,j}^f)\right]\,, \tag{4.99}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_\xi^v &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi^v\left[\boldsymbol{z}-\boldsymbol{h}_v(\boldsymbol{m}_\xi^-)\right] & \boldsymbol{\mu}_{\xi,j}^f &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_{\xi,j}^f\left[\boldsymbol{z}-\boldsymbol{h}_j^f(\boldsymbol{m}_\xi^-)\right]\\
\boldsymbol{\Pi}_\xi^v &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi^v\boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^- & \boldsymbol{\Pi}_{\xi,j}^f &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_{\xi,j}^f\boldsymbol{H}_j^f(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\\
\boldsymbol{W}_\xi^v &= \boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\left[\boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\right]^T + \boldsymbol{R}_v & \boldsymbol{W}_{\xi,j}^f &= \boldsymbol{H}_j^f(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\left[\boldsymbol{H}_j^f(\boldsymbol{m}_\xi^-)\right]^T + \boldsymbol{R}_j^f\\
\boldsymbol{K}_\xi^v &= \boldsymbol{P}_\xi^-\left[\boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\right]^T\left[\boldsymbol{W}_\xi^v\right]^{-1} & \boldsymbol{K}_{\xi,j}^f &= \boldsymbol{P}_\xi^-\left[\boldsymbol{H}_j^f(\boldsymbol{m}_\xi^-)\right]^T\left[\boldsymbol{W}_{\xi,j}^f\right]^{-1}\\
\kappa_\xi^v &= p_g(\boldsymbol{z}|\boldsymbol{h}_v(\boldsymbol{m}_\xi^-),\boldsymbol{W}_\xi^v) & \kappa_{\xi,j}^f &= p_g(\boldsymbol{z}|\boldsymbol{h}_j^f(\boldsymbol{m}_\xi^-),\boldsymbol{W}_{\xi,j}^f)\,,
\end{aligned}$$

and where $\boldsymbol{H}_j^f(\boldsymbol{m})$ is the Jacobian of $\boldsymbol{h}_j^f(\cdot)$ evaluated at $\boldsymbol{x}=\boldsymbol{m}$. Note that two applications of the identity in Eqs. (A.2) are used to produce Eq. (4.99). The normalizing denominator of Bayes' rule

is found by integrating Eq. (4.99) as

$$\begin{aligned}\eta &= \int p^{-}(\boldsymbol{s})\ell(\boldsymbol{z}|\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &= v\sum_{\xi=1}^{L^{-}}\left[w_{\xi}^{-}\kappa_{\xi}^{v}\right] + (1-v)\sum_{\xi=1}^{L^{-}}\sum_{j=1}^{L^{f}}\left[w_{\xi}^{-}w_{j}^{f}\kappa_{\xi,j}^{f}\right]. \end{aligned} \tag{4.100}$$

Combining the results of Eqs. (4.99) and (4.100), this FCU-1 posterior becomes

$$p^{+}(\boldsymbol{x}) = \sum_{\xi=1}^{L^{-}}\left[w_{\xi}^{v}p_{g}(\boldsymbol{x}|\boldsymbol{\mu}_{\xi}^{v},\boldsymbol{\Pi}_{\xi}^{v})\right] + \sum_{\xi=1}^{L^{-}}\sum_{j=1}^{L^{f}}\left[w_{\xi,j}^{f}p_{g}(\boldsymbol{x}|\boldsymbol{\mu}_{\xi,j}^{f},\boldsymbol{\Pi}_{\xi,j}^{f})\right], \tag{4.101}$$

where

$$w_{\xi}^{v} = \frac{vw_{\xi}^{-}\kappa_{\xi}^{v}}{\eta} \qquad \text{and} \qquad w_{\xi,j}^{f} = \frac{(1-v)w_{\xi}^{-}w_{j}^{f}\kappa_{\xi,j}^{f}}{\eta},$$

which is a GM with $L^{-}\times(L^{f}+1)$ components. Thus, for each iteration of the filter, the update produces $L^{-}\times L^{f}$ more components.

4.6.2.3 FCU-1 with Spatially Normal Faulty Measurements

Instead of developing this FCU-1 variant directly, it is more pragmatic to observe that when the GM of Eq. (4.88) is taken to consist of only a single component, it reduces to the expression of Eq. (4.81), demonstrating that the normally distributed faulty measurement model is a special case of the GM faulty measurement model. Therefore, under the conditions that the faulty likelihood function is given by the Gaussian model of Section 4.5.2.2, it is relatively straightforward to show that the updated posterior of Eq. (4.101) reduces to

$$p^{+}(\boldsymbol{x}) = \sum_{\xi=1}^{L^{-}}\left[w_{\xi}^{v}p_{g}(\boldsymbol{x}|\boldsymbol{\mu}_{\xi}^{v},\boldsymbol{\Pi}_{\xi}^{v})\right] + \sum_{\xi=1}^{L^{-}}\left[w_{\xi}^{f}p_{g}(\boldsymbol{x}|\boldsymbol{\mu}_{\xi}^{f},\boldsymbol{\Pi}_{\xi}^{f})\right], \tag{4.102}$$

where

$$
\begin{aligned}
w_\xi^v &= \frac{v w_\xi^- \kappa_\xi^v}{\eta} & w_\xi^f &= \frac{(1-v) w_\xi^- \kappa_\xi^f}{\eta} \\
\boldsymbol{\mu}_\xi^v &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi^v \left[\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}_\xi^-)\right] & \boldsymbol{\mu}_\xi^f &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi^f \left[\boldsymbol{z} - \boldsymbol{h}_f(\boldsymbol{m}_\xi^-)\right] \\
\boldsymbol{\Pi}_\xi^v &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi^v \boldsymbol{H}_v(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- & \boldsymbol{\Pi}_\xi^f &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi^f \boldsymbol{H}_f(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \\
\boldsymbol{W}_\xi^v &= \boldsymbol{H}_v(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \left[\boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\right]^T + \boldsymbol{R}_v & \boldsymbol{W}_\xi^f &= \boldsymbol{H}_f(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \left[\boldsymbol{H}_f(\boldsymbol{m}_\xi^-)\right]^T + \boldsymbol{R}_f \\
\boldsymbol{K}_\xi^v &= \boldsymbol{P}_\xi^- \left[\boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\right]^T \left[\boldsymbol{W}_\xi^v\right]^{-1} & \boldsymbol{K}_\xi^f &= \boldsymbol{P}_\xi^- \left[\boldsymbol{H}_f(\boldsymbol{m}_\xi^-)\right]^T \left[\boldsymbol{W}_\xi^f\right]^{-1} \\
\kappa_\xi^v &= p_g(\boldsymbol{z} | \boldsymbol{h}_v(\boldsymbol{m}_\xi^-), \boldsymbol{W}_\xi^v) & \kappa_\xi^f &= p_g(\boldsymbol{z} | \boldsymbol{h}_f(\boldsymbol{m}_\xi^-), \boldsymbol{W}_\xi^f) \,,
\end{aligned}
$$

and where

$$
\eta = \sum_{\xi=1}^{L^-} w_\xi^- \left[v \kappa_\xi^v + (1-v) \kappa_\xi^f \right] .
$$

Equation (4.102) indicates the posterior pdf of this update is a GM with $2L^-$ components, such that each iteration of the filter doubles the number of GM components.

Merging FCU-1 Components for Inclusion in Traditional Filter

The updates of Section 4.6.2 are attractive, by design, in that they are well-suited to replace a preexisting EKF update with relative ease. For instance, if a (Bayesian) EKF is already implemented in a system, then the single valid Gaussian measurement model should already be available, and the prior is necessarily a Gaussian distribution—or a GM of a single component. Then, the only obstacle to a closed-form solution, where all algorithms of the EKF remain the same with the exception of the update, is the fact that the FCU-1 updates produce GM posteriors, when the EKF requires single Gaussian priors. Fortunately, this discrepancy is readily addressed by the addition of a single step following the FCU-1 equations of either Eq. (4.97), Eq. (4.101), or Eq. (4.102). By immediately merging the components of the GM posteriors into a single Gaussian distribution, per Section 2.2.3.2, the posterior of the FCU-1 becomes identical in form to that of the EKF. In this way, as long as a suitable faulty measurement model is specified, the merged FCU-1 becomes a straightforward substitution for the update within the traditional EKF of Section 2.1.4.

4.6.3 FCU-2: Single-Valid Measurement

The second type of fault-cognizant update (FCU-2) corresponds to FCMM-2 of Section 4.3.2, where it is assumed that sensor scans contain at most one valid return and any number of faulty returns. As such, the general fault-cognizant measurement likelihood $\ell(\boldsymbol{Z}|\boldsymbol{x})$ of each FCU-2 filter corresponds to Eq. (4.34). Similar to the FCU-1 derivation of Section 4.6.3, assumptions are strategically enforced until a closed-form solution is achieved.

First, the assumptions that faulty measurements are spatially uniform and temporally Poisson,

pursuant to Sections 4.5.2.1 and 4.5.1.4, are applied such that the FCMM of Eq. (4.34) becomes

$$\begin{aligned}
\ell(\boldsymbol{Z}|\boldsymbol{x}) &= p_D(1,\boldsymbol{x})\frac{\lambda^{m-1}}{m(m-1)!\mathrm{e}^{\lambda}}\sum_{i=1}^{m}\left\{\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\prod_{j=1,j\neq i}^{m}\frac{1}{V}\right\}+[1-p_D(1,\boldsymbol{x})]\frac{\lambda^m}{m!\mathrm{e}^{\lambda}}\prod_{j=1}^{m}\frac{1}{V}\\
&= p_D(1,\boldsymbol{x})\frac{\lambda^{m-1}}{m!\mathrm{e}^{\lambda}}\sum_{i=1}^{m}\left\{\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\frac{1}{V^{m-1}}\right\}+[1-p_D(1,\boldsymbol{x})]\frac{\lambda^m}{m!\mathrm{e}^{\lambda}}\frac{1}{V^m}\\
&= \frac{\lambda^m}{m!\mathrm{e}^{\lambda}}\frac{1}{V^m}\left\{p_D(1,\boldsymbol{x})\frac{V}{\lambda}\sum_{i=1}^{m}\left[\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\right]+[1-p_D(1,\boldsymbol{x})]\right\}\\
&= \frac{p_F(m)}{V^m}\left\{\frac{p_D(1,\boldsymbol{x})}{\kappa_c}\sum_{i=1}^{m}\left[\ell_v(\boldsymbol{z}_i|\boldsymbol{x})\right]+[1-p_D(1,\boldsymbol{x})]\right\},
\end{aligned}\tag{4.103}$$

where κ_c is a commonly used parameter referred to as the clutter intensity, which relates the temporal and spatial distribution of false measurements as [48]

$$\kappa_c = \lambda\ell_f(\boldsymbol{z})\,,\tag{4.104a}$$

which, in this case, is

$$\kappa_c = \frac{\lambda}{V}\,.\tag{4.104b}$$

If the valid measurements are assumed to be spatially Gaussian according to Eq. (4.63), then Eq. (4.103) further reduces to

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \frac{p_F(m)}{V^m}\left\{\frac{1}{\kappa_c}\sum_{i=1}^{m}\left[p_D(1,\boldsymbol{x})p_g\big(\boldsymbol{z}_i\big|\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{R}\big)\right]+[1-p_D(1,\boldsymbol{x})]\right\}.\tag{4.105}$$

The likelihood of Eq. (4.105) can be multiplied with the GM prior of Eq. (4.90), such that a

Bayesian posterior is formed as

$$p^+(\boldsymbol{x}) \propto \frac{1}{\kappa_c} \sum_{i=1}^{m} \sum_{\xi=1}^{L^-} \left\{ p_D(1,\boldsymbol{x}) w_\xi^- p_g\big(\boldsymbol{z}_i \big| \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{R}\big) p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-\big) \right\} + \sum_{\xi=1}^{L^-} \left\{ [1 - p_D(1,\boldsymbol{x})] w_\xi^- p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-\big) \right\}. \tag{4.106}$$

Here, the identity of Eq. (A.2) is applied to the product of Gaussians in Eq. (4.106) to produce

$$p^+(\boldsymbol{x}) \propto \frac{1}{\kappa_c} \sum_{i=1}^{m} \sum_{\xi=1}^{L^-} \left\{ p_D(1,\boldsymbol{x}) \kappa_{\xi,i}^+ w_\xi^- p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_{\xi,i}^+, \boldsymbol{P}_{\xi,i}^+\big) \right\} + \sum_{\xi=1}^{L^-} \left\{ [1 - p_D(1,\boldsymbol{x})] w_\xi^- p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-\big) \right\}. \tag{4.107}$$

where

$$\boldsymbol{m}_{\xi,i}^+ = \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi\big(\boldsymbol{z}_i - \boldsymbol{h}(\boldsymbol{m}_\xi^-)\big) \tag{4.108a}$$

$$\boldsymbol{P}_{\xi,i}^+ = \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H}(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \tag{4.108b}$$

$$\boldsymbol{W}_\xi = \boldsymbol{H}(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \boldsymbol{H}^T(\boldsymbol{m}_\xi^-) + \boldsymbol{R} \tag{4.108c}$$

$$\boldsymbol{K}_\xi = \boldsymbol{P}_\xi^- \boldsymbol{H}^T(\boldsymbol{m}_\xi^-) \boldsymbol{W}_\xi^{-1} \tag{4.108d}$$

$$\kappa_{\xi,i}^+ = p_g\big(\boldsymbol{z}_i \big| \boldsymbol{h}(\boldsymbol{m}_\xi^-), \boldsymbol{W}_\xi\big), \tag{4.108e}$$

and where $\boldsymbol{H}(\cdot)$ is the Jacobian of $\boldsymbol{h}(\cdot)$. Note that no precise model for the probability of detection has been established. As such, the remainder of the section derives two different FC updates, one with a zeroth-order approximation of $p_D(1,\boldsymbol{x})$ and the other with Gaussian model for $p_D(1,\boldsymbol{x})$. The update of Eq. (4.107) and the corresponding components of Eqs. (4.108) are the basis for each variation of the proposed FCU-2 to be presented.

4.6.3.1 FCU-2 with Zeroth-Order Approximated p_D:

If a zeroth-order approximation of $p_D(1, \boldsymbol{x})$ is made, as per Section 4.4.2.3 and Eq. (4.69), the FCU-2 from Eqs. (4.107) and (4.108) becomes

$$\begin{aligned} p^+(\boldsymbol{x}) \propto \frac{1}{\kappa_c} \sum_{i=1}^{m} \sum_{\xi=1}^{L^-} \left\{ p_D(1, \boldsymbol{m}_{\xi,i}^+) \kappa_{\xi,i}^+ w_\xi^- p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_{\xi,i}^+, \boldsymbol{P}_{\xi,i}^+\big) \right\} \\ + \sum_{\xi=1}^{L^-} \left\{ [1 - p_D(1, \boldsymbol{m}_\xi^-)] w_\xi^- p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-\big) \right\}, \end{aligned} \tag{4.109}$$

which, once normalization is performed, yields

$$p^+(\boldsymbol{x}) = \sum_{i=1}^{m} \sum_{\xi=1}^{L^-} \left\{ w_{\xi,i}^+ p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_{\xi,i}^+, \boldsymbol{P}_{\xi,i}^+\big) \right\} + \sum_{\xi=1}^{L^-} \left\{ w_\xi^+ p_g\big(\boldsymbol{x} \big| \boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-\big) \right\}, \tag{4.110a}$$

where

$$\begin{aligned} w_{\xi,i}^+ &= \frac{p_D(1, \boldsymbol{m}_{\xi,i}^+) \kappa_{\xi,i}^+ w_\xi^-}{\kappa_c \eta} \\ w_\xi^+ &= \frac{\big[1 - p_D(1, \boldsymbol{m}_\xi^-)\big] w_\xi^-}{\eta} \\ \eta &= \sum_{\xi=1}^{L^-} w_\xi^- \left\{ \big[1 - p_D(1, \boldsymbol{m}_\xi^-)\big] + \frac{1}{\kappa_c} \sum_{i=1}^{m} \left[p_D(1, \boldsymbol{m}_{\xi,i}^+) \kappa_{\xi,i}^+ \right] \right\}. \end{aligned} \tag{4.110b}$$

The posterior of Eq. (4.110a) is a GM of $L^- \times (m + 1)$ components, such that an additional $L^- \times m$ components are generated with each filter iteration. Note that in the trivial case where $p_D(1, \boldsymbol{x})$ is taken to be a constant, the resulting update can simply be stated as simplified version of Eqs. (4.110a) and (4.110b) with $p_D(1, \boldsymbol{m}) = p_D(1)$.

4.6.3.2 FCU-2 with Gaussian p_D:

This variant of the FCU-2 assumes that the probability of detection can be modeled as a Gaussian distribution with a mean that is a function of the state $\boldsymbol{x}$, per the discussion of Section 4.4.2.4.

In this case, the update of Eqs. (4.107) and (4.108) becomes

$$p^+(\boldsymbol{x}) \propto \frac{1}{\kappa_c}\sum_{i=1}^{m}\sum_{\xi=1}^{L^-}\Big\{\kappa_{\xi,i}^+ w_\xi^- p_g\big(p_{D_z}\big|g(\boldsymbol{x}),R_D\big)p_g\big(\boldsymbol{x}\big|\boldsymbol{m}_{\xi,i}^+,\boldsymbol{P}_{\xi,i}^+\big)\Big\} + \sum_{\xi=1}^{L^-}\Big\{w_\xi^- p_g\big(\big[1-p_{D_z}\big]\big|\big[1-g(\boldsymbol{x})\big],R_D\big)p_g\big(\boldsymbol{x}\big|\boldsymbol{m}_\xi^-,\boldsymbol{P}_\xi^-\big)\Big\} \tag{4.111}$$

following the proper substitutions of Eqs. (4.70) and (4.71). Next, two simultaneous applications of Eq. (A.2) are used to evaluate both Gaussian products to give

$$p^+(\boldsymbol{x}) \propto \frac{1}{\kappa_c}\sum_{\xi=1}^{L^-}\sum_{i=1}^{m}\Big\{w_\xi^-\kappa_{\xi,i}^+\kappa_{\xi,i}^{D^+}p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi,i}^{D^+},\boldsymbol{P}_{\xi,i}^{D^+})\Big\} + \sum_{\xi=1}^{L^-}\Big\{w_\xi^-\kappa_\xi^{D^-}p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^{D^-},\boldsymbol{P}_\xi^{D^-})\Big\}, \tag{4.112}$$

where

$$\begin{aligned}
\boldsymbol{m}_{\xi,i}^{D^+} &= \boldsymbol{m}_{\xi,i}^+ + \boldsymbol{K}_{\xi,i}^{D^+}\big(p_{D_z}-g(\boldsymbol{m}_{\xi,i}^+)\big) & \boldsymbol{m}_\xi^{D^-} &= \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi^{D^-}\big(p_{D_z}-g(\boldsymbol{m}_\xi^-)\big)\\
\boldsymbol{P}_{\xi,i}^{D^+} &= \boldsymbol{P}_{\xi,i}^+ - \boldsymbol{K}_{\xi,i}^{D^+}\boldsymbol{G}(\boldsymbol{m}_{\xi,i}^+)\boldsymbol{P}_{\xi,i}^+ & \boldsymbol{P}_\xi^{D^-} &= \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi^{D^-}\boldsymbol{G}(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\\
W_{\xi,i}^{D^+} &= \boldsymbol{G}(\boldsymbol{m}_{\xi,i}^+)\boldsymbol{P}_{\xi,i}^+\boldsymbol{G}^T(\boldsymbol{m}_{\xi,i}^+) + R_D & W_\xi^{D^-} &= \boldsymbol{G}(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\boldsymbol{G}^T(\boldsymbol{m}_\xi^-) + R_D\\
\boldsymbol{K}_{\xi,i}^{D^+} &= \boldsymbol{P}_{\xi,i}^+\boldsymbol{G}^T(\boldsymbol{m}_{\xi,i}^+)\big[W_{\xi,i}^{D^+}\big]^{-1} & \boldsymbol{K}_\xi^{D^-} &= \boldsymbol{P}_\xi^-\boldsymbol{G}^T(\boldsymbol{m}_\xi^-)\big[W_\xi^{D^-}\big]^{-1}\\
\kappa_{\xi,i}^{D^+} &= p_g\Big(p_{D_z}\big|g(\boldsymbol{m}_{\xi,i}^+),W_{\xi,i}^{D^+}\Big) & \kappa_\xi^{D^-} &= p_g\Big(\big[1-p_{D_z}\big]\big|\big[1-g(\boldsymbol{m}_\xi^-)\big],W_\xi^{D^-}\Big).
\end{aligned}$$

Note that $\boldsymbol{m}_{\xi,i}^+$ and $\boldsymbol{P}_{\xi,i}^+$ are from Eqs. (4.110), and that $\boldsymbol{G}(\cdot)$ is the Jacobian of $g(\cdot)$ such that the Jacobian of $1-g(\boldsymbol{x})$ is simply

$$\frac{\partial}{\partial\boldsymbol{x}}\Big(1-g(\boldsymbol{x})\Big) = -\boldsymbol{G}(\boldsymbol{x})\,.$$

Therefore, given the update of Eq. (4.112), the normalized Bayesian posterior is found to be

$$p^+(\boldsymbol{x}) = \sum_{\xi=1}^{L^-}\sum_{i=1}^{m}\left\{w_{\xi,i}^{D^+}\, p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi,i}^{D^+}, \boldsymbol{P}_{\xi,i}^{D^+})\right\} + \sum_{\xi=1}^{L^-}\left\{w_{\xi}^{D^-}\, p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi}^{D^-}, \boldsymbol{P}_{\xi}^{D^-})\right\}, \tag{4.113a}$$

where

$$w_{\xi,i}^{D^+} = \frac{w_\xi^- \kappa_{\xi,i}^+ \kappa_{\xi,i}^{D^+}}{\kappa_c \eta}, \quad w_\xi^{D^-} = \frac{w_\xi^- \kappa_\xi^{D^-}}{\eta}, \quad \text{and } \eta = \sum_{\xi=1}^{L^-} w_\xi^- \left\{\kappa_\xi^{D^-} + \frac{1}{\kappa_c}\sum_{i=1}^{m}\left[\kappa_{\xi,i}^+ \kappa_{\xi,i}^{D^+}\right]\right\}. \tag{4.113b}$$

which corresponds to a Gaussian mixture of $L^- \times (m+1)$ components. In order for this update to be accomplished, some value of p_{D_z} must be reported to the filter by some external source.

4.6.4 FCU-3: Multiple-Valid Measurement – IID

Proceeding similarly to Sections 4.6.2 and 4.6.3, this section derives a fault-cognizant filter update (FCU-3) subject to the FCMM-3 of Section 4.3.3 and Eq. (4.48), where a sensor produces measurement scans that can contain multiple IID valid measurements. While the discussion of Section 4.6.1 forces the prior $p^-(\boldsymbol{x})$ into the form of the GM in Eq. (4.90), FCU-3 additionally assumes that the spatial distribution of valid measurements $\ell_v(\boldsymbol{z}|\boldsymbol{x})$ also takes the form of a GM according to Section 4.4.1.2, one with L^v number of components, such that

$$\ell_v(\boldsymbol{z}|\boldsymbol{x}) = \sum_{\gamma=1}^{L^v} w_\gamma^v p_g(\boldsymbol{z}|\boldsymbol{h}_\gamma(\boldsymbol{x}), \boldsymbol{R}_\gamma)\,, \tag{4.114}$$

where w_γ^v, $\boldsymbol{h}_\gamma(\boldsymbol{x})$, and $\boldsymbol{R}_\gamma$ are the weight, mean, and covariance of the γ^{th} GM component respectively. Note that if $L_v = 1$, Eq. (4.114) reduces to the traditional Gaussian measurement model of Eq. (2.1c), but taking it to be a GMM allows for a valid spatial distribution that can adapt to different sensor models, should the need arise.

As various faulty spatial distributions are already surveyed by FCU-1 of Section 4.6.2, for simplicity, FCU-3 assumes that faulty measurements occur uniformly across the sensor space ac-

cording to Eq. (4.79) such that the faulty spatial distribution becomes

$$\ell_f(\boldsymbol{z}|\boldsymbol{x}) = \ell_f(\boldsymbol{z}) = \frac{1}{V}\,, \tag{4.115}$$

where V is the sensor volume. With the inclusion of Eqs. (4.114) and (4.115), the FCMM-3 of Eq. (4.48) becomes

$$\begin{aligned}
\ell(\boldsymbol{Z}|\boldsymbol{x}) &= \sum_{n=0}^{m}\left\{\frac{p_D(n,\boldsymbol{x})p_F(m-n,\boldsymbol{x})}{\binom{m}{n}}\sum_{j=1}^{\binom{m}{n}}\left[\prod_{\substack{k=1\\ \boldsymbol{z}_k\in H_{v,j}}}^{n}\ell_v(\boldsymbol{z}_k|\boldsymbol{x})\prod_{\substack{\ell=1\\ \boldsymbol{z}_\ell\in H_{f,j}}}^{m-n}\ell_f(\boldsymbol{z}_\ell|\boldsymbol{x})\right]\right\}\\
&= \sum_{n=0}^{m}\left\{\frac{p_D(n,\boldsymbol{x})p_F(m-n,\boldsymbol{x})}{\binom{m}{n}}\sum_{j=1}^{\binom{m}{n}}\left[\prod_{\substack{k=1\\ \boldsymbol{z}_k\in H_{v,j}}}^{n}\sum_{\gamma=1}^{L^v}w_\gamma^v p_g(\boldsymbol{z}_k|\boldsymbol{h}_\gamma(\boldsymbol{x}),\boldsymbol{R}_\gamma)\prod_{\substack{\ell=1\\ \boldsymbol{z}_\ell\in H_{f,j}}}^{m-n}\frac{1}{V}\right]\right\}\\
&= \sum_{n=0}^{m}\left\{\frac{p_D(n,\boldsymbol{x})p_F(m-n,\boldsymbol{x})}{\binom{m}{n}V^{m-n}}\sum_{j=1}^{\binom{m}{n}}\left[\prod_{\substack{k=1\\ \boldsymbol{z}_k\in H_{v,j}}}^{n}\sum_{\gamma=1}^{L^v}w_\gamma^v p_g(\boldsymbol{z}_k|\boldsymbol{h}_\gamma(\boldsymbol{x}),\boldsymbol{R}_\gamma)\right]\right\}.
\end{aligned} \tag{4.116}$$

The Bayesian posterior of FCU-3 is proportional to the product of Eqs. (4.90) and (4.116), which is

$$\begin{aligned}
&p^+(\boldsymbol{x})\\
&\propto \sum_{\xi=1}^{L^-}w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-,\boldsymbol{P}_\xi^-)\sum_{n=0}^{m}\left\{\frac{p_D(n,\boldsymbol{x})p_F(m-n,\boldsymbol{x})}{\binom{m}{n}V^{m-n}}\sum_{j=1}^{\binom{m}{n}}\left[\prod_{\substack{k=1\\ \boldsymbol{z}_k\in H_{v,j}}}^{n}\sum_{\gamma=1}^{L^v}w_\gamma^v p_g(\boldsymbol{z}_k|\boldsymbol{h}_\gamma(\boldsymbol{x}),\boldsymbol{R}_\gamma)\right]\right\}\\
&\propto \sum_{n=0}^{m}\left\{\frac{p_D(n,\boldsymbol{x})p_F(m-n,\boldsymbol{x})}{\binom{m}{n}V^{m-n}}\sum_{j=1}^{\binom{m}{n}}\left[\sum_{\xi=1}^{L^-}w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-,\boldsymbol{P}_\xi^-)\prod_{\substack{k=1\\ \boldsymbol{z}_k\in H_{v,j}}}^{n}\sum_{\gamma=1}^{L^v}w_\gamma^v p_g(\boldsymbol{z}_k|\boldsymbol{h}_\gamma(\boldsymbol{x}),\boldsymbol{R}_\gamma)\right]\right\}.
\end{aligned} \tag{4.117}$$

In order to further reduce the expression of Eq. (4.117), the Gaussian mixture terms must be directly multiplied together. This is difficult, however, as Ho's equation of Eq. (A.2) requires a direct multiplication of the terms $p_g(\boldsymbol{x}|\cdot,\cdot)$ and $p_g(\boldsymbol{z}_k|\cdot,\cdot)$ only. If the square bracketed portion of

Eq. (4.117) is expanded, it is found that

$$\begin{aligned}
&\sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \prod_{\substack{k=1 \\ \boldsymbol{z}_k \in H_{v,j}}}^{n} \sum_{\gamma=1}^{L^v} w_\gamma^v p_g(\boldsymbol{z}_k|\boldsymbol{h}_\gamma(\boldsymbol{x}), \boldsymbol{R}_\gamma) \\
&\quad = \sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \left(\sum_{\gamma_1=1}^{L^v} w_{\gamma_1}^v p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \right) \cdots \left(\sum_{\gamma_n=1}^{L^v} w_{\gamma_i}^v p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \right) \\
&\quad = \sum_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \left(\sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} w_{\gamma_1}^v p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \cdots w_{\gamma_n}^v p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_i}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \right) \\
&\quad = \sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w_{\gamma_1}^v p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \cdots w_{\gamma_n}^v p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \\
&\quad = \sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w_{\gamma_1}^v \cdots w_{\gamma_n}^v w_\xi^- p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \ ,
\end{aligned} \tag{4.118}$$

where it is recalled that all $\boldsymbol{z}_k \in H_{v,j}\ \forall k = 1, 2, \ldots, n$, as this is for one unique combination of valid assignments. It is seen that this expression in Eq. (4.118) has a $n+1$ summations that sum over of a product of n Gaussians of the form $p_g(\boldsymbol{z}_k|\boldsymbol{h}_{\gamma_k}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_k})$ for $k = 0,\ 1,\ \ldots,\ n$ with a single Gaussian of the form $p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-)$. At this stage, it is noted that each measurement must be processed and incorporated into the distribution some way. This can be done via a sequential implementation of Ho's rule of Eqs. (A.2) with n steps. Starting with the product of Gaussians in Eq. (4.118), the first iteration takes the form

$$\begin{aligned}
&p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \\
&\qquad = \kappa_1 p_g(\boldsymbol{x}|\boldsymbol{\mu}_1^\xi, \boldsymbol{\Pi}_1^\xi) p_g(\boldsymbol{z}_2|\boldsymbol{h}_{\gamma_2}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_2}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \ ,
\end{aligned} \tag{4.119}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_1^{\xi} &= \boldsymbol{m}_{\xi}^{-} + \boldsymbol{K}_1\big(\boldsymbol{z}_1 - \boldsymbol{h}_{\gamma_1}(\boldsymbol{m}_{\xi}^{-})\big) \\
\boldsymbol{\Pi}_1^{\xi} &= \boldsymbol{P}_{\xi}^{-} - \boldsymbol{K}_1\boldsymbol{H}_{\gamma_1}(\boldsymbol{m}_{\xi}^{-})\boldsymbol{P}_{\xi}^{-} \\
\boldsymbol{W}_1 &= \boldsymbol{H}_{\gamma_1}(\boldsymbol{m}_{\xi}^{-})\boldsymbol{P}_{\xi}^{-}\boldsymbol{H}_{\gamma_1}^{T}(\boldsymbol{m}_{\xi}^{-}) + \boldsymbol{R}_{\gamma_1} \\
\boldsymbol{K}_1 &= \boldsymbol{P}_{\xi}^{-}\boldsymbol{H}_{\gamma_1}^{T}(\boldsymbol{m}_{\xi}^{-})\boldsymbol{W}_1^{-1} \\
\kappa_1 &= p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{m}_{\xi}^{-}), \boldsymbol{W}_1) \ .
\end{aligned}$$

The notation change to $\boldsymbol{\mu}_{\theta}^{\xi}$ and $\boldsymbol{\Pi}_{\theta}^{\xi}$ is to better differentiate them from the predicted ($\boldsymbol{m}_{\xi}^{-}$, $\boldsymbol{P}_{\xi}^{-}$) and updated ($\boldsymbol{m}_{\xi}^{+}$, $\boldsymbol{P}_{\xi}^{+}$) means and covariances of the ξ^{th} GM component of the estimated distribution, as $\boldsymbol{\mu}_{\theta}^{\xi}$ and $\boldsymbol{\Pi}_{\theta}^{\xi}$ denote the mean and covariance of the GM distribution before it has incorporated all of the measurement data into the estimated distribution. The process can be iterated to some θ^{th} step such that Eq. (4.119) becomes

$$\begin{aligned}
&\kappa_1 \cdots \kappa_{\theta-1} p_g(\boldsymbol{x}|\boldsymbol{\mu}_{\theta-1}^{\xi}, \boldsymbol{\Pi}_{\theta-1}^{\xi}) p_g(\boldsymbol{z}_{\theta}|\boldsymbol{h}_{\gamma_{\theta}}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_{\theta}}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \\
&\qquad = \kappa_1 \cdots \kappa_{\theta} p_g(\boldsymbol{x}|\boldsymbol{\mu}_{\theta}^{\xi}, \boldsymbol{\Pi}_{\theta}^{\xi}) p_g(\boldsymbol{z}_{\theta+1}|\boldsymbol{h}_{\gamma_{\theta+1}}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_{\theta+1}}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \ , \quad (4.120a)
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{\theta}^{\xi} &= \boldsymbol{\mu}_{\theta-1}^{\xi} + \boldsymbol{K}_{\theta}\big(\boldsymbol{z}_{\theta} - \boldsymbol{h}_{\gamma_{\theta}}(\boldsymbol{\mu}_{\theta-1}^{\xi})\big) && (4.120b) \\
\boldsymbol{\Pi}_{\theta}^{\xi} &= \boldsymbol{\Pi}_{\theta-1}^{\xi} - \boldsymbol{K}_{\theta}\boldsymbol{H}_{\gamma_{\theta}}(\boldsymbol{\mu}_{\theta-1}^{\xi})\boldsymbol{\Pi}_{\theta-1}^{\xi} && (4.120c) \\
\boldsymbol{W}_{\theta} &= \boldsymbol{H}_{\gamma_{\theta}}(\boldsymbol{\mu}_{\theta-1}^{\xi})\boldsymbol{\Pi}_{\theta-1}^{\xi}\boldsymbol{H}_{\gamma_{\theta}}^{T}(\boldsymbol{\mu}_{\theta-1}^{\xi}) + \boldsymbol{R}_{\gamma_{\theta}} && (4.120d) \\
\boldsymbol{K}_{\theta} &= \boldsymbol{\Pi}_{\theta-1}^{\xi}\boldsymbol{H}_{\gamma_{\theta}}^{T}(\boldsymbol{\mu}_{\theta-1}^{\xi})\boldsymbol{W}_{\theta}^{-1} && (4.120e) \\
\kappa_{\theta} &= p_g(\boldsymbol{z}_{\theta}|\boldsymbol{h}_{\gamma_{\theta}}(\boldsymbol{\mu}_{\theta-1}^{\xi}), \boldsymbol{W}_{\theta}) \ . && (4.120f)
\end{aligned}$$

The process of Eqs. (4.119) and (4.120a) is performed n times, where the final iteration is

$$\begin{aligned}\kappa_1 \cdots \kappa_{n-1} p_g(\boldsymbol{x}|\boldsymbol{\mu}_{n-1}^{\xi}, \boldsymbol{\Pi}_{n-1}^{\xi}) p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) \\ = \kappa_1 \cdots \kappa_n p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi}^{+}, \boldsymbol{P}_{\xi}^{+}) \,,\end{aligned} \tag{4.121}$$

where

$$\begin{aligned}\boldsymbol{m}_{\xi}^{+} &= \boldsymbol{\mu}_{n-1}^{\xi} + \boldsymbol{K}_n\big(\boldsymbol{z}_n - \boldsymbol{h}_{\gamma_n}(\boldsymbol{\mu}_{n-1}^{\xi})\big) \\ \boldsymbol{P}_{\xi}^{+} &= \boldsymbol{\Pi}_{n-1}^{\xi} - \boldsymbol{K}_n \boldsymbol{H}_{\gamma_n}(\boldsymbol{\mu}_{n-1}^{\xi}) \boldsymbol{\Pi}_{n-1}^{\xi} \\ \boldsymbol{W}_n &= \boldsymbol{H}_{\gamma_n}(\boldsymbol{\mu}_{n-1}^{\xi}) \boldsymbol{\Pi}_{n-1}^{\xi} \boldsymbol{H}_{\gamma_n}^{T}(\boldsymbol{\mu}_{n-1}^{\xi}) + \boldsymbol{R}_{\gamma_n} \\ \boldsymbol{K}_n &= \boldsymbol{\Pi}_{n-1}^{\xi} \boldsymbol{H}_{\gamma_n}^{T}(\boldsymbol{\mu}_{n-1}^{\xi}) \boldsymbol{W}_n^{-1} \\ \kappa_n &= p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{\mu}_{n-1}^{\xi}), \boldsymbol{W}_n) \,.\end{aligned}$$

After completing the final iteration of Eq. (4.121), the summation of Eq. (4.118) can be represented as

$$\begin{aligned}&\sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w_{\gamma_1}^{v} \cdots w_{\gamma_n}^{v} w_{\xi}^{-} p_g(\boldsymbol{z}_1|\boldsymbol{h}_{\gamma_1}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_1}) \cdots p_g(\boldsymbol{z}_n|\boldsymbol{h}_{\gamma_n}(\boldsymbol{x}), \boldsymbol{R}_{\gamma_n}) p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi}^{-}, \boldsymbol{P}_{\xi}^{-}) \\ &\quad = \sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w_{\gamma_1}^{v} \cdots w_{\gamma_n}^{v} w_{\xi}^{-} \kappa_1 \cdots \kappa_n p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi}^{+}, \boldsymbol{P}_{\xi}^{+}) \,.\end{aligned} \tag{4.122}$$

It should be noted that this process can become computationally burdensome, as the iterative application of Ho's rule described must be performed $L^- \times [L^v]^n$ times, each requiring n sequential updates. This also produces $L^- \times [L^v]^n$ GM components. As such, it becomes important to explore possible methods of curbing computation costs.

Taking the result of Eq. (4.122), it can be shown that Eq. (4.117) becomes

$$p^+(\boldsymbol{x}) = \sum_{n=0}^{m} \Bigg\{ \frac{p_D(n,\boldsymbol{x}) p_F(m-n,\boldsymbol{x})}{\binom{m}{n} V^{m-n}} \times \sum_{\substack{j=1 \\ H_{v,j}}}^{\binom{m}{n}} \Bigg[\sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w^v_{\gamma_1} \cdots w^v_{\gamma_n} w^-_{\xi} \kappa_1 \cdots \kappa_n p_g(\boldsymbol{x} | \boldsymbol{m}^+_{\xi}, \boldsymbol{P}^+_{\xi}) \Bigg] \Bigg\}, \tag{4.123}$$

where the hypothesis notation $H_{v,j}$ has been added to the summation over j to indicate the selection of the j^{th} unique assignment hypothesis of valid measurements. The denominator of Bayes' rule can be found by taking the integral of Eq. (4.123) over the support of $\boldsymbol{x}$, which yields

$$\int p^-(\boldsymbol{s}) \ell(\boldsymbol{Z}|\boldsymbol{s}) \mathrm{d}\boldsymbol{s} = \int \sum_{n=0}^{m} \Bigg\{ \frac{p_D(n,\boldsymbol{s}) p_F(m-n,\boldsymbol{s})}{\binom{m}{n} V^{m-n}} \times \sum_{\substack{j=1 \\ H_{v,j}}}^{\binom{m}{n}} \Bigg[\sum_{\gamma_1=1}^{L^v} \cdots \sum_{\gamma_n=1}^{L^v} \sum_{\xi=1}^{L^-} w^v_{\gamma_1} \cdots w^v_{\gamma_n} w^-_{\xi} \kappa_1 \cdots \kappa_n p_g(\boldsymbol{s} | \boldsymbol{m}^+_{\xi}, \boldsymbol{P}^+_{\xi}) \Bigg] \Bigg\} \mathrm{d}\boldsymbol{s}\,, \tag{4.124}$$

where it is clear that having state-dependent probability of detection and false alarm makes integration problematic. To address this issue, two different approaches may be taken. Either the probabilities of $p_D(n,\boldsymbol{x})$ and $p_F(m-n,\boldsymbol{x})$ can be forced to be state-independent (via a zeroth-order approximation or otherwise), or they can be modeled as Gaussian distributions. As Gaussian models for $p_D(\cdot,\boldsymbol{x})$ are previously surveyed in the FCU-2 of Section 4.6.3.2, state-dependency is neglected here for simplicity. Therefore, if the probabilities of detection and false alarm are taken to be state-independent such that $p_D(n,\boldsymbol{x}) = p_D(n)$ and $p_F(m-n,\boldsymbol{x}) = p_F(m-n)$, then the only term in Eq. (4.124) dependent on the state is the Gaussian distribution, which integrates to unity,

such that

$$\int p^{-}(\boldsymbol{s})\ell(\boldsymbol{Z}|\boldsymbol{s})\mathrm{d}\boldsymbol{s}$$
$$= \sum_{n=0}^{m}\left\{\frac{p_D(n)p_F(m-n)}{\binom{m}{n}V^{m-n}}\sum_{\substack{j=1\\H_{v,j}}}^{\binom{m}{n}}\left[\sum_{\gamma_1=1}^{L^v}\cdots\sum_{\gamma_n=1}^{L^v}\sum_{\xi=1}^{L^-}w_{\gamma_1}^{v}\cdots w_{\gamma_n}^{v}w_{\xi}^{-}\kappa_1\cdots\kappa_n\right]\right\}. \quad (4.125)$$

The results of Eqs. (4.123) and (4.125) can be combined into the complete closed-form update of FCU-3 as

$$p^{+}(\boldsymbol{x}) = \frac{\sum_{n=0}^{m}\left\{\frac{p_D(n)p_F(m-n)}{\binom{m}{n}V^{m-n}}\sum_{\substack{j=1\\H_{v,j}}}^{\binom{m}{n}}\left[\sum_{\gamma_1=1}^{L^v}\cdots\sum_{\gamma_n=1}^{L^v}\sum_{\xi=1}^{L^-}w_{\gamma_1}^{v}\cdots w_{\gamma_n}^{v}w_{\xi}^{-}\kappa_1\cdots\kappa_n p_g(\boldsymbol{x}|\boldsymbol{m}_{\xi}^{+},\boldsymbol{P}_{\xi}^{+})\right]\right\}}{\sum_{n=0}^{m}\left\{\frac{p_D(n)p_F(m-n)}{\binom{m}{n}V^{m-n}}\sum_{\substack{j=1\\H_{v,j}}}^{\binom{m}{n}}\left[\sum_{\gamma_1=1}^{L^v}\cdots\sum_{\gamma_n=1}^{L^v}\sum_{\xi=1}^{L^-}w_{\gamma_1}^{v}\cdots w_{\gamma_n}^{v}w_{\xi}^{-}\kappa_1\cdots\kappa_n\right]\right\}}, \quad (4.126)$$

which is a GM of $L^{-}\times\sum_{n=0}^{m}\{\binom{m}{n}\times[L^v]^n\}$ components, making merging and pruning highly advisable.

4.6.5 FCU-4: Multiple Uniquely-Distributed Valid Measurement

Using the FCMM-4 of Section 4.3.4, wherein measurement scans can contain multiple valid measurements corresponding to unique distributions, this section derives a corresponding fault-cognizant update (FCU-4). To immediately simplify the update, the spatial distribution of faulty measurements is taken to be uniform according to Eq. (4.79), and the probabilities of detection and false alarm are assumed to be state-independent, such that FCMM-4 of Eq. (4.61) becomes

$$
\begin{aligned}
\ell(\boldsymbol{Z}|\boldsymbol{x}) = & \sum_{n=0}^{\min(m,q)} p_F(m-n) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \\
& \times \sum_{k=1}^{m \mathrm{P} n} \frac{(m-n)!}{m!} \Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}}^{n} \ell_{v,i}(\boldsymbol{z}_r|\boldsymbol{x}) \Big] \Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_f) \in H_{n,j,k}^{\mathrm{UA}}}^{m-n} \frac{1}{V} \Big] \\
= & \sum_{n=0}^{\min(m,q)} p_F(m-n) \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \\
& \times \sum_{k=1}^{m \mathrm{P} n} \frac{(m-n)!}{m! V^{m-n}} \Big[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}}^{n} \ell_{v,i}(\boldsymbol{z}_r|\boldsymbol{x}) \Big] . \qquad (4.127)
\end{aligned}
$$

Next, since Gaussian-distributed valid measurements are so common, it is assumed here that all valid measurements correspond to Gaussian distributions per Section 4.4.1.1. That is to say, consider a valid measurement generated by subset $\boldsymbol{Z}_{v,i}$ to follow the equation

$$
\boldsymbol{z}_{v,i} = \boldsymbol{h}_{v,i}(\boldsymbol{x}) + \boldsymbol{w}_{v,i} \,,
$$

where $\boldsymbol{h}_{v,i}(\cdot)$ is a nonlinear observation function specific to $\boldsymbol{Z}_{v,i}$, and $\boldsymbol{w}_{v,i}$ is a zero-mean Gaussian white noise process defined by

$$
\boldsymbol{w}_{v,i} \sim p_g(\boldsymbol{w}_{v,i}|\boldsymbol{0}, \boldsymbol{R}_{v,i}) \,,
$$

where $\boldsymbol{R}_{v,i}$ is the corresponding measurement noise covariance. It can be shown that the single measurement likelihood for $\boldsymbol{Z}_{v,i}$ is

$$
\ell_{v,i}(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{h}_{v,i}(\boldsymbol{x}), \boldsymbol{R}_{v,i}) \,. \qquad (4.128)
$$

Thus, applying this model to the likelihood of Eq. (4.127) yields

$$\ell(\boldsymbol{Z}|\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} p_F(m-n) \left[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \right] \left[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \right] \times \frac{(m-n)!}{m! V^{m-n}} \left[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}}^{n} p_g(\boldsymbol{z}_r | \boldsymbol{h}_{v,i}(\boldsymbol{x}), \boldsymbol{R}_{v,i}) \right]. \tag{4.129}$$

Under the current assumptions, the form of the likelihood of Eq. (4.129) lends itself to a Bayesian filter that produces a multiple component Gaussian-mixture posterior. Therefore, in order to derive a closed-form update, the GM prior assumption of Section 4.6.1 is reasserted such that $p^{-}(\boldsymbol{x})$ again takes the form of Eq. (4.90). Examining the Bayesian posterior directly, it is clear that

$$p^{+}(\boldsymbol{x}) \propto \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \sum_{\xi=1}^{L^{-}} p_F(m-n) \left[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \right] \left[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \right] \times \frac{(m-n)!}{m! V^{m-n}} w_{\xi}^{-} p_g(\boldsymbol{x} | \boldsymbol{m}_{\xi}^{-}, \boldsymbol{P}_{\xi}^{-}) \left[\prod_{(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}}^{n} p_g(\boldsymbol{z}_r | \boldsymbol{h}_{v,i}(\boldsymbol{x}), \boldsymbol{R}_{v,i}) \right] \tag{4.130}$$

results in n sequential applications of Ho's rule once the product over $(\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}) \in H_{n,j,k}^{\mathrm{UA}}$ is expanded, similar to Eq. (4.118); however, the explanation of the process here is shortened for brevity. Consider the multiplication of the Gaussians in Eq. (4.130) one at a time. Thus, these n multiplications can be carried out iteratively as

$$\kappa_{\theta}^{\xi} p_g(\boldsymbol{x} | \boldsymbol{\mu}_{\theta}^{\xi}, \boldsymbol{\Pi}_{\theta}^{\xi}) = \kappa_{\theta-1}^{\xi} p_g(\boldsymbol{x} | \boldsymbol{\mu}_{\theta-1}^{\xi}, \boldsymbol{\Pi}_{\theta-1}^{\xi}) p_g(\boldsymbol{z}_{\theta} | \boldsymbol{h}_{v,\theta}(\boldsymbol{x}), \boldsymbol{R}_{v,\theta}), \quad \forall\, \theta = 1, 2, \ldots, n, \tag{4.131a}$$

where $\boldsymbol{z}_{\theta}$, $\boldsymbol{h}_{v,\theta}(\boldsymbol{x})$, and $\boldsymbol{R}_{v,\theta}$ are selected as the θ^{th} unique assignment of $\boldsymbol{z}_r \in \boldsymbol{Z}_{v,i}$ of hypothesis

$H^{\text{UA}}_{n,j,k}$. The components of Eq. (4.131a) are generated via Eqs. (A.2) as

$$\boldsymbol{\mu}^{\xi}_{\theta} = \boldsymbol{\mu}^{\xi}_{\theta-1} + \boldsymbol{K}_{\theta}\big(\boldsymbol{z}_{\theta} - \boldsymbol{h}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1})\big) \tag{4.131b}$$

$$\boldsymbol{\Pi}^{\xi}_{\theta} = \boldsymbol{\Pi}^{\xi}_{\theta-1} - \boldsymbol{K}_{\theta}\boldsymbol{H}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1})\boldsymbol{\Pi}^{\xi}_{\theta-1} \tag{4.131c}$$

$$\boldsymbol{W}_{\theta} = \boldsymbol{H}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1})\boldsymbol{\Pi}^{\xi}_{\theta-1}\boldsymbol{H}^{T}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1}) + \boldsymbol{R}_{v,\theta} \tag{4.131d}$$

$$\boldsymbol{K}_{\theta} = \boldsymbol{\Pi}^{\xi}_{\theta-1}\boldsymbol{H}^{T}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1})\boldsymbol{W}^{-1}_{\theta} \tag{4.131e}$$

$$\kappa^{\xi}_{\theta} = \kappa^{\xi}_{\theta-1}p_g(\boldsymbol{z}_{\theta}|\boldsymbol{h}_{v,\theta}(\boldsymbol{\mu}^{\xi}_{\theta-1}), \boldsymbol{W}_{\theta})\,, \tag{4.131f}$$

where $\boldsymbol{H}_{v,\theta}(\cdot)$ is the Jacobian of $\boldsymbol{h}_{v,\theta}(\cdot)$. The recursion of Eq. (4.131a) is initialized with $\kappa^{\xi}_{0} = w^{-}_{\xi}$, $\boldsymbol{\mu}^{\xi}_{0} = \boldsymbol{m}^{-}_{\xi}$, and $\boldsymbol{\Pi}^{\xi}_{0} = \boldsymbol{P}^{-}_{\xi}$, which, after n applications, results in $\kappa^{\xi}_{n}p_g(\boldsymbol{x}|\boldsymbol{\mu}^{\xi}_{n}, \boldsymbol{\Pi}^{\xi}_{n})$. Since this Gaussian is formed within the purview of hypothesis $H^{\text{UA}}_{n,j,k}$, it is important that the notation reflect this, which is done by defining

$$\kappa^{+}_{n,j,k,\xi}p_g(\boldsymbol{x}|\boldsymbol{m}^{+}_{n,j,k,\xi}, \boldsymbol{P}^{+}_{n,j,k,\xi}) = \Big\{\kappa^{\xi}_{n}p_g(\boldsymbol{x}|\boldsymbol{\mu}^{\xi}_{n}, \boldsymbol{\Pi}^{\xi}_{n})\Big|H^{\text{UA}}_{n,j,k}\Big\} \tag{4.132a}$$

$$\kappa^{+}_{n,j,k,\xi} = \kappa^{\xi}_{n} \tag{4.132b}$$

$$\boldsymbol{m}^{+}_{n,j,k,\xi} = \boldsymbol{\mu}^{\xi}_{n} \tag{4.132c}$$

$$\boldsymbol{P}^{+}_{n,j,k,\xi} = \boldsymbol{\Pi}^{\xi}_{n}\,, \tag{4.132d}$$

such that Eq. (4.130) can be expressed in terms of Eqs. (4.132) as

$$\begin{aligned} p^{+}(\boldsymbol{x}) \propto \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\text{P}n} \sum_{\xi=1}^{L^{-}} p_F(m-n)\Big[\prod_{\boldsymbol{Z}_{v,i}\in H_{n,j}}^{n} p_{D,i}(1)\Big]\Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1 - p_{D,i}(1)\Big] \\ \times \frac{(m-n)!}{m!V^{m-n}}\kappa^{+}_{n,j,k,\xi}p_g(\boldsymbol{x}|\boldsymbol{m}^{+}_{n,j,k,\xi}, \boldsymbol{P}^{+}_{n,j,k,\xi})\,. \end{aligned} \tag{4.133}$$

At this point, statistical distributions have been assumed for most models with the exception of $p_{D,i}(1)$ and $p_F(m-n)$ (beyond being state-independent). If specialized models for both probabilities are required for a specific system, it is advised that the form of FCU-4 in Eq. (4.133) be used. In

fact, the probabilities of detection $p_{D,i}(1)$ will, most likely, require individually modeled functions for most applications, so no further specification of their models are made. However, following the discussion of Section 4.5.1.4, faulty measurements can frequently be temporally modeled by the Poisson distribution of Eq. (4.72). Enforcing the Poisson assumption results in

$$p_F(m-n) = \frac{\lambda^{m-n}}{(m-n)!\mathrm{e}^{\lambda}}\,, \tag{4.134}$$

which results in the update of Eq. (4.133) becoming

$$\begin{aligned} p^+(\boldsymbol{x}) \propto & \frac{\lambda^m}{\mathrm{e}^{\lambda} m! V^m} \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \sum_{\xi=1}^{L^-} \frac{V^n}{\lambda^n} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \\ & \times \kappa^+_{n,j,k,\xi} p_g(\boldsymbol{x} | \boldsymbol{m}^+_{n,j,k,\xi}, \boldsymbol{P}^+_{n,j,k,\xi}) \\ \propto & \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \sum_{\xi=1}^{L^-} \frac{1}{\kappa_c^n} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \\ & \times \kappa^+_{n,j,k,\xi} p_g(\boldsymbol{x} | \boldsymbol{m}^+_{n,j,k,\xi}, \boldsymbol{P}^+_{n,j,k,\xi})\,. \end{aligned} \tag{4.135}$$

The term $\frac{\lambda^m}{\mathrm{e}^{\lambda} m! V^m}$ can be removed, as it is constant for all $\boldsymbol{x}$, and the definition of clutter intensity from Eqs. (4.104) is recalled as

$$\kappa_c^n = \left(\frac{\lambda}{V} \right)^n .$$

After proper normalization of Eq. (4.135), the closed-form update of FCU-4 is

$$p^+(\boldsymbol{x}) = \sum_{n=0}^{\min(m,q)} \sum_{j=1}^{\binom{q}{n}} \sum_{k=1}^{m\mathrm{P}n} \sum_{\xi=1}^{L^-} w^+_{n,j,k,\xi} p_g(\boldsymbol{x} | \boldsymbol{m}^+_{n,j,k,\xi}, \boldsymbol{P}^+_{n,j,k,\xi})\,, \tag{4.136a}$$

where

$$w^{+}_{n,j,k,\xi} = \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \frac{\kappa^{+}_{n,j,k,\xi}}{\kappa_c^n \, \eta} \,, \tag{4.136b}$$

and where the normalization constant η is calculated as

$$\eta = \sum_{n=0}^{\min(m,q)} \frac{1}{\kappa_c^n} \sum_{j=1}^{\binom{q}{n}} \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{n,j}}^{n} p_{D,i}(1) \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{n,j}}^{q-n} 1 - p_{D,i}(1) \Big] \sum_{k=1}^{m\mathrm{P}n} \sum_{\xi=1}^{L^-} \kappa^{+}_{n,j,k,\xi} \,. \tag{4.136c}$$

Note that the FCU-4 posterior of Eq. (4.136) is a GM of size $L^+ = L^- \times \sum_{n=0}^{\min(m,q)} \{ \binom{q}{n} \times m\mathrm{P}n \}$, which quickly becomes computationally intractable if component management is not enforced. In fact, in addition to the pruning and merging methods described in Sections 2.2.3.1 and 2.2.3.2, in this case it is advisable to enforce an upper component limit L^+_{limit}, where, after each update's typical merging and pruning, only the L^+_{limit} components with the largest weights $w^+_{n,j,k,\xi}$ are carried forward to the next iteration of the filter. Additional discussion on tractability concerns are made in the following section.

4.6.6 Tractable Implementations

While the FCUs presented in Section 4.6 are exact, closed-form solutions to Bayes' rule, under some circumstances, the number of GM components generated by each update may become unwieldy, especially in the cases of FCU-3 and FCU-4. In such cases, merging and pruning as described in Sections 2.2.3.1 and 2.2.3.2 may not be sufficient to ensure computational tractability. Thus, this section investigates additional approaches to managing computations when designing fault-cognizant filters. Note that the appropriate manner by which tractability is achieved is heavily system dependent; factors such as size of the prior, forms of the valid and faulty likelihoods, temporal distribution of measurements, etc., determine which tractability method is best. Since these factors vary between the eight different updates of Section 4.6, the tractability procedures of this section are limited to FCU-4, as it one of the more computationally complex updates presented. Similar methods can be developed for FCU-1, FCU-2, and FCU-3 if the need arises.

The posterior pdf of each FCU update is a GM of the same form as Eq. (2.10), where w_ξ^+ is the weight of the ξ^{th} GM component. Generally speaking, a larger weight corresponds to a component contributing more to the overall pdf, while components with smaller weights contribute less. This is the very relation that motivates pruning—where components with the smallest weights are removed from the GM—and is also the main consideration in this section; any posterior weights that are exceedingly small are less useful to the GM description of the pdf, and, thus, the computational power required to calculate these low-weighted GM components can be avoided at little cost to the accuracy of the posterior.

4.6.6.1 Approximating Posterior Weights

To avoid as many computations as possible, it is good practice to calculate all of the posterior GM weights before performing the full update, as low-weighted GM components can be neglected, thus saving computations. Essentially, these components are pruned from the pdf before they are fully calculated, which is referred to here as preliminary pruning. For example, recall FCU-1 of Section 4.6.2.1. The posterior weights of Eq. (4.98) can be calculated with equations Eqs. (4.95c) and (4.95e) alone, skipping the calculations of Eq. (4.95a), Eq. (4.95b), and Eq. (4.95d), which involve a relatively burdensome inversion of the residual covariance. FCU-1 therefore benefits greatly from pre-calculation of the posterior weights, but this is not the case for every FCU update. FCU-3 and FCU-4, in particular, require iterative applications of Ho's rule that cannot be avoided when calculating the exact posterior weights, which involves the majority of calculations required by the full update. Therefore, instead of computing the precise posterior weights, it is more useful to approximate the posterior weights in a computationally efficient manner and predict which posterior GM components are worth their computational cost.

Since the exact manner by which weights are approximated is up to the user and varies between updates, weight approximation for FCU-4 is done specifically, where similar approaches can be designed for the other FCUs. Consider the (unnormalized) posterior weights of FCU-4

from Eq. (4.136b), given by

$$w^+_{n,j,k,\xi} = \frac{1}{\kappa_c^n}\Big[\prod_{\boldsymbol{Z}_{v,i}\in H_{n,j}}^{n} p_{D,i}(1)\Big]\Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1)\Big]\kappa^+_{n,j,k,\xi}\,, \tag{4.137}$$

the complete calculation of which is relatively intensive, as $\kappa^+_{n,j,k,\xi}$ requires n iterative applications of Ho's rule. Thus, a less burdensome approach is to approximate Eq. (4.137) by

$$\begin{aligned} w^+_{n,j,k,\xi} \approx \frac{w^-_\xi}{\kappa_c^n}\Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1,\boldsymbol{m}^-_\xi)\Big] \\ \times\Big[\prod_{(\boldsymbol{z}_r\in\boldsymbol{Z}_{v,i})\in H^{\mathrm{UA}}_{n,j,k}}^{n} p_{D,i}(1,\boldsymbol{m}^-_\xi)p_g\Big(\boldsymbol{z}_r|\boldsymbol{h}_{v,i}(\boldsymbol{m}^-_\xi),\,\boldsymbol{H}_{v,i}(\boldsymbol{m}^-_\xi)\boldsymbol{P}^-_\xi\boldsymbol{H}^T_{v,i}(\boldsymbol{m}^-_\xi)+\boldsymbol{R}_{v,i}\Big)\Big]\,, \end{aligned} \tag{4.138}$$

where w^-_ξ is the ξ^{th} weight of the prior GM pdf that is present in $\kappa^+_{n,j,k,\xi}$, such that the iterative calculations of Eq. (4.131b), Eq. (4.131c), and Eq. (4.131e) are avoided. Preliminary pruning can be performed on the weights of Eq. (4.138) before the full calculation of the update.

While approximating the posterior weights via Eq. (4.138) does indeed save computations, it still requires approximating all of the weights. In cases where there is a large number of prior GM components in $p^-(\boldsymbol{x})$, it may be more practical to further reduce computations by

$$\begin{aligned} \omega^+_{n,j,k} &\approx \sum_{\xi=1}^{L^-} w^+_{n,j,k,\xi} \\ &\approx \frac{1}{\kappa_c^n}\Big[\prod_{\boldsymbol{Z}_{v,i}\notin H_{n,j}}^{q-n} 1-p_{D,i}(1,\hat{\boldsymbol{m}}^-)\Big] \\ &\quad\times\Big[\prod_{(\boldsymbol{z}_r\in\boldsymbol{Z}_{v,i})\in H^{\mathrm{UA}}_{n,j,k}}^{n} p_{D,i}(1,\hat{\boldsymbol{m}}^-)p_g\Big(\boldsymbol{z}_r|\boldsymbol{h}_{v,i}(\hat{\boldsymbol{m}}^-),\,\boldsymbol{H}_{v,i}(\hat{\boldsymbol{m}}^-)\hat{\boldsymbol{P}}^-\boldsymbol{H}^T_{v,i}(\hat{\boldsymbol{m}}^-)+\boldsymbol{R}_{v,i}\Big)\Big]\,, \end{aligned} \tag{4.139}$$

where $\hat{\boldsymbol{m}}^-$ and $\hat{\boldsymbol{P}}^-$ are the estimated mean and covariance extracted from the prior GM via Eqs. (2.22) and (2.24), respectively. Compared to Eq. (4.138), Eq. (4.139) reduces the number

of required calculations by a factor of $L^- - 1$. Then, if preliminary pruning is performed on the weights of Eq. (4.139), any $\omega^+_{n,j,k}$ that is selected to be pruned will result in the omission of $w^+_{n,j,k,\xi}$, $\boldsymbol{m}^+_{n,j,k,\xi}$, and $\boldsymbol{P}^+_{n,j,k,\xi}$ for $\xi = 1, 2, \ldots, L^-$ from the calculation of the update. Similar approximations to the posterior weights can be made for FCU-1, FCU-2, and FCU-3.

4.6.6.2 Subset of Feasible Features

Recalling the optical TRN system described by Fig. (4.7), note that this application of fault-cognizant filtering is predicated on a pre-existing map of feature locations. In cases where this map contains thousands of features (or more), which is certainly possible in many optical navigation applications, the number of GM components generated by FCU-4 becomes computationally unreasonable. Of course, many features, especially those outside the estimated FOV of the sensor, have extremely low probabilities of detection, as illustrated by Fig. (4.12).

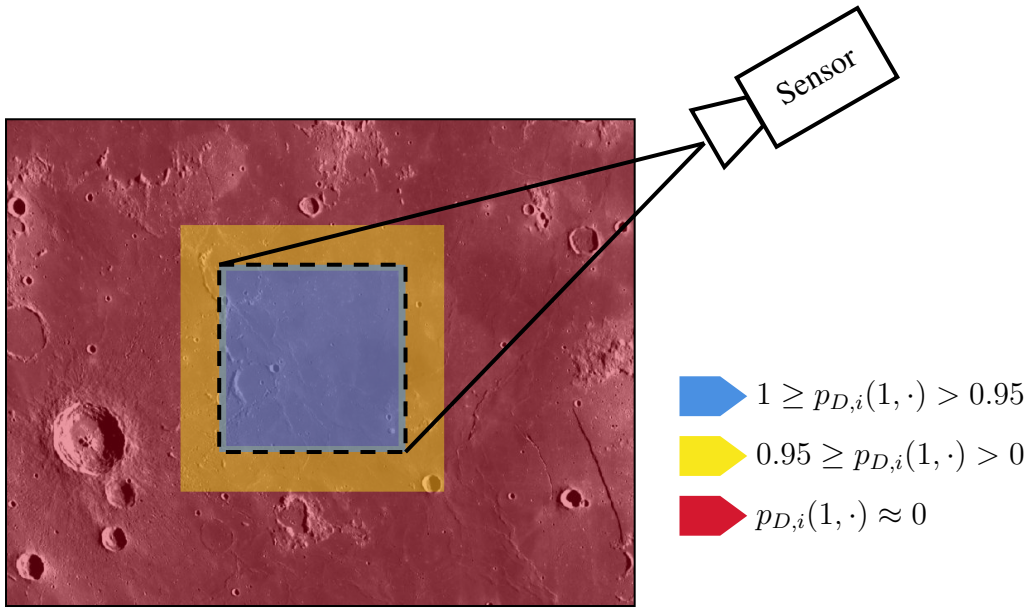


Figure 4.12 Example of possible probability of detection behavior for optical TRN

To limit the number of features considered in FCU-4, a subset of feasible features can be formed by simply neglecting any features that occur in the red region within the feature map. To

quickly accomplish this, the state estimate $\hat{\boldsymbol{m}}^-$ can be extracted from the prior pdf via Eq. (2.22), which is used to evaluate the probabilities of detection. It then follows that any i^{th} feature having

$$p_{D,i}(1, \hat{\boldsymbol{m}}^-) \leq \Upsilon_{\text{neglect}}$$

should be excluded from the subset of feasible features, where $\Upsilon_{\text{neglect}}$ is some user-specified, near-zero threshold value. Note that this process quickly decreases the number of features q, which subsequently limits the number of unique selection hypotheses $H_{n,j}$ for $j = 1, 2, \ldots, \binom{q}{n}$ that must be considered.

4.6.6.3 Ranked Assignment Approach

The posterior of FCU-4 described in Section 4.6.5 considers all possible hypotheses matching measurements to features and clutter. While this provides a complete probabilistic description of all possible events, many of these hypotheses are highly unlikely, yielding low posterior GM weights. It is common practice in fields such as multi-target tracking (MTT) to identify and neglect these unlikely hypotheses and therefore only consider the most probable unique measurement assignments. For example, measurement gating is typically employed in multi-target filters such as joint probabilistic data association (JPDA) [98] and multiple hypothesis tracking (MHT) [99], where only measurements that occur within a bounded region of probability about a target are considered possible returns from that target.

Another method when evaluating measurement hypotheses is to solve the ranked assignment problem, which seeks the K most probable measurement association hypotheses via optimization of a cost matrix. First solved by Murty in [100], this approach is used to great effect in FISST filters such as the δ-generalized labeled multi-Bernoulli (δ-GLMB) filter, which provides an exact, tractable solution to the multi-target Bayes' rule [101]. The process begins by defining a cost matrix $\boldsymbol{C} \in \mathbb{R}^{q \times m}$, the entries of which can be calculated as

$$C_{ij} = -\ln \left\{ \frac{\int p^-(\boldsymbol{x}) p_{D,i}(1, \boldsymbol{x}) \ell_{v,i}(\boldsymbol{z}_j | \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}}{\kappa_c \int p^-(\boldsymbol{x}) \big[1 - p_{D,i}(1, \boldsymbol{x})\big] \, \mathrm{d}\boldsymbol{x}} \right\}, \tag{4.140}$$

where Eq. (4.140) is the "cost" of assigning measurement $\boldsymbol{z}_j \in \boldsymbol{Z}$ to feature i corresponding to $\ell_{v,i}(\cdot|\cdot)$ and $p_{D,i}(1,\cdot)$. Making note of the GM prior of Eq. (4.90), the Gaussian likelihoods of Eq. (4.128), and enforcing a zeroth-order approximation of $p_{D,i}(1,\boldsymbol{x})$ about the ξ^{th} GM mean, Eq. (4.140) can be shown to be

$$C_{ij} = -\ln\left\{\frac{\sum_{\xi=1}^{L^-} w_\xi^- p_{D,i}(1,\boldsymbol{m}_\xi^-) p_g\big(\boldsymbol{z}_j|\boldsymbol{h}_{v,i}(\boldsymbol{m}_\xi^-),\, \boldsymbol{H}_{v,i}(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^-\boldsymbol{H}_{v,i}^T(\boldsymbol{m}_\xi^-)+\boldsymbol{R}_{v,i}\big)}{\kappa_c \sum_{\xi=1}^{L^-} w_\xi^- \big[1-p_{D,i}(1,\boldsymbol{m}_\xi^-)\big]}\right\}. \tag{4.141}$$

Once the cost matrix is calculated, the optimal assignment matrix can be solved for via

$$\boldsymbol{S}^* = \min_{\boldsymbol{S}} \Big\{\boldsymbol{S}^T\boldsymbol{C}\Big\},$$

where $\boldsymbol{S} \in \mathbb{R}^{q\times m}$ is the assignment matrix whose rows and columns must sum to 1 or 0. Each entry is defined as

$$S_{ij} = \begin{cases} 1\,, & \text{if measurement } j \text{ is assigned to feature } i \\ 0\,, & \text{otherwise} \end{cases}.$$

Instead of only seeking the optimal assignment, Murty's algorithm [100] can be used to solve for the K assignments with the lowest cost function, which are represented by hypotheses H_ζ^M, where $\zeta = 1,\, 2,\, \ldots,\, K$. In some cases, dummy variables may need to be introduced, as Murty's algorithm requires square matrices.

The end result is a posterior that is no longer of the form of Eqs. (4.136), which has multiple summations over all possible hypotheses. Instead, the new posterior sums over the H_ζ^M hypothesis as

$$p^+(\boldsymbol{x}) \propto \sum_{\zeta=1}^{K}\sum_{\xi=1}^{L^-} w_{\zeta,\xi}^+ p_g(\boldsymbol{x}|\boldsymbol{m}_{\zeta,\xi}^+, \boldsymbol{P}_{\zeta,\xi}^+)\,, \tag{4.142}$$

where

$$w_{\zeta,\xi}^{+} = \Big[\prod_{\boldsymbol{Z}_{v,i} \in H_{\zeta}^{M}} \frac{p_{D,i}(1)}{\kappa_c} \Big] \Big[\prod_{\boldsymbol{Z}_{v,i} \notin H_{\zeta}^{M}} 1 - p_{D,i}(1) \Big] \kappa_{\zeta,\xi}^{+} \,, \tag{4.143}$$

and where

$$\kappa_{\zeta,\xi}^{+} = \kappa_{n}^{\xi} \tag{4.144a}$$

$$\boldsymbol{m}_{\zeta,\xi}^{+} = \boldsymbol{\mu}_{n}^{\xi} \tag{4.144b}$$

$$\boldsymbol{P}_{\zeta,\xi}^{+} = \boldsymbol{\Pi}_{n}^{\xi} \,. \tag{4.144c}$$

Note that κ_n^{ξ}, $\boldsymbol{\mu}_n^{\xi}$, and $\boldsymbol{\Pi}_n^{\xi}$ are solved for via the iterative procedure of Eqs. (4.131) under the unique assignment of hypothesis H_{ζ}^{M}.

4.6.7 Order of Measurement Processing

It is well known that the order in which measurements are processed can affect the estimate quality of the EKF [21]. In short, due to the linearization of the EKF, measurements resulting in large updates—i.e. $\boldsymbol{W}$ of Eq. (2.4e) is much larger than the corresponding measurement noise covariance $\boldsymbol{R}$—can accrue large linearization errors that may negatively effect the processing of subsequent measurements from other sensors. While the FCU updates presented in this section are not EKFs but GM filters, they are still constructed using first-order Taylor series approximations via Eqs. (A.2), so the same concerns exist here, especially for FCU-3 and FCU-4 of Sections 4.6.4 and 4.6.5, where posterior Gaussian components are formed via successive applications of Eqs. (A.2). Therefore, it is a good practice to set the order by which measurements are ingested into FCU-3 and FCU-4 as "smallest" to "largest" via

$$\frac{\|\boldsymbol{R}_{\gamma_1}\|}{\|\boldsymbol{W}_1\|} \geq \cdots \geq \frac{\|\boldsymbol{R}_{\gamma_\theta}\|}{\|\boldsymbol{W}_\theta\|} \geq \cdots \geq \frac{\|\boldsymbol{R}_{\gamma_n}\|}{\|\boldsymbol{W}_n\|} \tag{4.145a}$$

$$\frac{\|\boldsymbol{R}_{v,1}\|}{\|\boldsymbol{W}_1\|} \geq \cdots \geq \frac{\|\boldsymbol{R}_{v,\theta}\|}{\|\boldsymbol{W}_\theta\|} \geq \cdots \geq \frac{\|\boldsymbol{R}_{v,n}\|}{\|\boldsymbol{W}_n\|} \,, \tag{4.145b}$$

where Eq. (4.145a) is the ordering for FCU-3 of Eqs. (4.120), and Eq. (4.145b) is the ordering for FCU-4 of Eqs. (4.131). In this way, the measurement updates most likely to have the smallest linearization error will be processed first, which will generally reduce the linearization errors for the "larger" updates later on.

4.7 Application to Navigation

As discussed in Section 1.3.1, filters are considered the backbone of spaceflight navigation systems, as they perform the crucial task of estimating a vehicle's pose by processing incoming observation data from various sensors and fusing it with prior knowledge of the vehicle's state. Since a navigation filter must be capable of accurate and robust operation before being considered flight-ready, it is a best practice to implement a level of fault tolerance within the architecture of a navigation system to account for different faults that may occur. One of the most common faults is the acquisition of erroneous measurements where, due to some issue such as sensor malfunction or unmodeled effects, some sensor returns do not match the observation model of the filter, as already discussed in Section 4.2. These faulty measurements are detrimental to navigation performance and quickly debase a filtering solution if not addressed. It has already been mentioned that, as a response, most filters adopt some method of screening out invalid sensing data. In the field of navigation specifically, where the vast majority of filters are variants of the extended Kalman filter (EKF) [32], the most prominent method is known as residual editing, which has been used in spaceflight since the Apollo missions and remains just as relevant today in large-scale missions, such as Orion [7, 102].

In general terms, residual editing is an auxiliary construct that acts as an intermediary between the sensors and the filter update [21]. If the difference between an incoming measurement and the filter's prediction of that measurement is found to exceed a specified threshold, then the navigation filter rejects the measurement completely, and the filter update is not performed. Otherwise, when the residual is within the threshold, the filter update proceeds normally, processing the measurement and assuming it valid. While the efficacy of residual editing has been demonstrated numerous times in practical applications, it is not a guaranteed safeguard against the adverse effects of faulty

measurements, as was seen during NASA's DART mission, where the navigation solution was corrupted by erroneous data, unfortunately resulting in a collision during close proximity operations and the premature retirement of the mission [85, 103].

In order to better prevent mishaps such as these in future missions, the short-comings of residual editing must be examined. First and foremost, residual editing is a supplementary procedure that employs binary decision making, i.e., either a measurement is included and considered valid by the filter or it is edited and removed from the measurement set completely. This binary operation presents a couple of theoretical issues for the filter, as it is essentially informing the filter that any approved measurement is guaranteed to be valid and that any edited measurement is guaranteed to be false.

For a long time, such drawbacks were considered unavoidable, as the top priority of onboard navigation systems is computational efficiency, as evidenced by the continued use of the EKF—a linear filter formatted for a nonlinear system [32]. However, advancements in high performance spaceflight computing (HPSC) are promising to ease the traditional computational restrictions that limited the navigation architecture of past spaceflight missions [104, 105]. Leading organizations in the spaceflight industry, such as NASA, are now expressing more interest in alternatives to the EKF and residual editing that are more suitable for the environment of space and its inherent complexities [106]. The fault-cognizant filters presented in this chapter are one possible alternative to residual editing, just as the GSF is one possible alternative to the EKF. In fact, FCU-1 of Section 4.6.2 is designed specifically to replace existing residual editing methods, and, as such, the performance of the FCU-1 is compared to that of residual editing to test its efficacy as a replacement. In fact, residual editing is included in all of the simulations of this chapter, as it provides an excellent baseline that illustrates how all of the fault-cognizant filters perform relative to current practice. As such, the next section of this chapter discusses residual editing in greater technical detail, even deriving a form of residual editing for GM filtering.

4.7.1 Residual Editing

As one of the main responsibilities of a filter is to incorporate new information via measurements generated by the real measurement model of Fig. (1.1), ensuring these sensor returns behave as expected by the filter—which is dictated by the assumed measurement model of Fig. (1.1)—is perhaps the most obvious approach to improve filter robustness. Again, to accomplish this, many linear estimators, such as the EKF, employ a procedure-first robustness technique known as residual editing [102, 107]. With regard to the EKF update of Eqs. (2.4), this process involves computing the measurement residual of Eq. (2.4f), which quantifies the difference between the predicted measurement $\boldsymbol{h}_k(\boldsymbol{m}_k^-)$ and the actual measurement $\boldsymbol{z}_k$. This residual is then compared to some specified threshold, where, if it is found to exceed said threshold, the measurement is omitted from the update entirely, and the filter update is skipped. Conversely, if the residual exists within the set threshold, then the filter ingests the measurement and the update proceeds as normal per Eqs. (2.4).

While all residual editing schemes screen measurements more or less as described, the manner in which the editing threshold is calculated can vary considerably between applications. The simplest threshold is one that is set to a specific measurement value, such that the residual of Eq. (2.4f) (or the norm of the residual in the case of vector measurements) is compared to that constant threshold value directly. Such a method is ill-advised, as it does not automatically adapt as the filter uncertainty changes. A more appropriate method, and one by which many filters operate [7, 19], is specifying a residual editing threshold relative to the filter's uncertainty in the residual, which, with respect to the EKF, is quantified by the residual covariance $\boldsymbol{W}$ of Eq. (2.4e). Furthermore, an ideal residual editing threshold not only accounts for the filtering uncertainty, but also

- requires a single, scalar value be specified to set the editing threshold,
- naturally accounts for vector measurements and the cross-correlations between their components, and
- can be selected intuitively according to some established statistical principle.

Consequently, all of these objectives can be satisfied simultaneously to a certain degree. Since the measurement residual $\Delta \boldsymbol{z}_k$ is taken to be unbiased, it can be shown under linear-Gaussian assumptions that the squared Mahalanobis distance satisfies a χ^2 distribution [108], which inspires a residual editing threshold check of

$$\Delta \boldsymbol{z}_k^T \boldsymbol{W}^{-1} \Delta \boldsymbol{z}_k > \Upsilon \,, \tag{4.146}$$

where, if true, the measurement is rejected by the filter. Due to the linearizations inherent within the EKF, the (squared) Mahalanobis distance of Eq. (4.146) only approximately satisfies the χ^2 distribution. Nevertheless, it is considered good practice to assign the threshold value Υ according to a χ^2 probability gate [21].

4.7.2 Extending Residual Editing to GM Filters

To emphasize this work's relevance to navigation, a baseline filter with residual editing is constructed, with which comparison analyses can be performed on the proposed FCU filters of Section 4.6. However, the proposed fault-cognizant filters are actualized as nonlinear filters and therefore have a clear advantage over the traditional EKF-based residual editing described in Section 4.7.1 before the benefits of fault-cognizance are even considered. In order to sustain equality when comparing the FCUs to the baseline filter, residual editing must be applied to a comparable nonlinear filter, which, in this case, is the GSF of Section 2.2.2.4. As little to no previous work outside of [83] exists that outfits the GSF with a residual editing scheme, this work seeks to construct such an extension. It is important to note that the exact manner in which residual editing should be accomplished is up to interpretation, but the discussion herein attempts to justify the proposed residual editing architecture.

At some point, the question must be asked if residual editing should be applied separately to each individual GM component, or if an entire measurement should be edited completely. The answer to this depends on how one views the role of residual editing. If its purpose is to diminish the effects caused by large residuals during the update, then applying editing to each GM compo-

nent may seem the answer. However, it is arguable that the ultimate goal of residual editing is to eliminate faulty data from being processed, as it does not adhere to the observation model. Therefore, if a measurement is deemed faulty, then it should be edited globally from the update, such that each GM component receives identical treatment. Additionally, while the GM components are individual distributions themselves, they collectively describe a single distribution, and updating certain portions of a distribution differently than others is a questionable practice, at best. In short, a measurement's validity must be categorized relative to the entire distribution; it cannot be treated as valid for some components and faulty for others. As such, the update presented here operates upon the assumption that residual editing should be a global application within the realm of GM filtering.

As discussed in Section 4.7.1, traditional residual editing within the EKF commonly utilizes the squared Mahalanobis distance of the measurement residual to perform a threshold check such that if [19, 21]

$$[\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}^-)]^T \boldsymbol{W}^{-1} [\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}^-)] > \Upsilon \tag{4.147}$$

is true, where Υ is some specified residual editing threshold and

$$\boldsymbol{W} = \boldsymbol{H}_v(\boldsymbol{m}^-) \boldsymbol{P}^- \boldsymbol{H}_v^T(\boldsymbol{m}^-) + \boldsymbol{R}_v$$

is the residual covariance, then the measurement is rejected. Otherwise, if the residual falls within the desired threshold, the usual EKF update takes place. Note that Eq. (4.147) is a restatement of Eq. (4.146), but omits the time index k while emphasizing that residuals are computed relative to the valid distribution. It can be shown that when valid measurements $\boldsymbol{z}_v$ are generated by the Gaussian model of Eq. (2.1c), the evaluated likelihood of $\boldsymbol{z}$ given some prior mean $\boldsymbol{m}^-$ and residual covariance $\boldsymbol{W}$ is calculated as

$$\kappa_{\text{EKF}}^{\text{RE}} = \left| 2\pi \boldsymbol{W} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}^-)]^T \boldsymbol{W}^{-1} [\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}^-)] \right\}, \tag{4.148}$$

where it is clear that the exponential term contains the (squared) Mahalanobis distance of Eq. (4.147). This allows the inequality of Eq. (4.147) to be expressed in terms of the evaluated likelihood $\kappa_{\text{EKF}}^{\text{RE}}$ of Eq. (4.148), such that

$$\kappa_{\text{EKF}}^{\text{RE}} < \left|2\pi \boldsymbol{W}\right|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\Upsilon \Big\} \tag{4.149}$$

is an equivalent expression where, if true, indicates that the measurement should be rejected. This treatment of residual editing can be extended to the update of a GSF, where the prior pdf of $\boldsymbol{x}$ must be available in the GM form of Eq. (4.90). For this GM representation, the evaluated likelihood equivalent of $\kappa_{\text{EKF}}^{\text{RE}}$ from Eq. (4.148) can be expressed as

$$\kappa_{\text{GSF}}^{\text{RE}} = \sum_{\xi=1}^{L^-} w_\xi^- \kappa_\xi \,,$$

where

$$\begin{aligned} \kappa_\xi &= \left|2\pi \boldsymbol{W}_\xi\right|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}[\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}_\xi^-)]^T \boldsymbol{W}_\xi^{-1} [\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}_\xi^-)] \Big\} \\ \boldsymbol{W}_\xi &= \boldsymbol{H}_v(\boldsymbol{m}_\xi^-) \boldsymbol{P}_\xi^- \boldsymbol{H}_v^T(\boldsymbol{m}_\xi^-) + \boldsymbol{R}_v \,. \end{aligned}$$

As a result, the residual thresholding inequality of Eq. (4.149) when subjected to a GM prior becomes

$$\kappa_{\text{GSF}}^{\text{RE}} < \sum_{\xi=1}^{L^-} w_\xi^- \left|2\pi \boldsymbol{W}_\xi\right|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\Upsilon \Big\} \,, \tag{4.150}$$

such that the updated pdf for a GSF with a global application of residual editing is given as

$$p^+(\boldsymbol{x}) = \begin{cases} \sum\limits_{\xi=1}^{L^-} w_\xi^- p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^-, \boldsymbol{P}_\xi^-) \,, & \text{if } \kappa_{\text{GSF}}^{\text{RE}} < \sum\limits_{\xi=1}^{L^-} w_\xi^- \left|2\pi \boldsymbol{W}_\xi\right|^{-\frac{1}{2}} \exp\big\{ -\frac{1}{2}\Upsilon \big\} \\ \sum\limits_{\xi=1}^{L^+} w_\xi^+ p_g(\boldsymbol{x}|\boldsymbol{m}_\xi^+, \boldsymbol{P}_\xi^+) \,, & \text{otherwise} \end{cases} , \tag{4.151}$$

where

$$L^+ = L^- \tag{4.152a}$$

$$w_\xi^+ = \frac{w_\xi^- \kappa_\xi}{\sum_{\xi=1}^{L^-} w_\xi^- \kappa_\xi} \tag{4.152b}$$

$$\boldsymbol{m}_\xi^+ = \boldsymbol{m}_\xi^- + \boldsymbol{K}_\xi[\boldsymbol{z} - \boldsymbol{h}_v(\boldsymbol{m}_\xi^-)] \tag{4.152c}$$

$$\boldsymbol{P}_\xi^+ = \boldsymbol{P}_\xi^- - \boldsymbol{K}_\xi \boldsymbol{H}_v(\boldsymbol{m}_\xi^-)\boldsymbol{P}_\xi^- \tag{4.152d}$$

$$\boldsymbol{K}_\xi = \boldsymbol{P}_\xi^- \boldsymbol{H}_v^T(\boldsymbol{m}_\xi^-)\boldsymbol{W}_\xi^{-1}\,. \tag{4.152e}$$

4.8 Simulation and Results

This section implements and tests some of the fault-cognizant filters derived in this chapter. This is accomplished within the scope of two different simulated systems, the first being the familiar falling body simulation and the second being an orbit determination scenario. As the filter update is the element being tested, all filters are equipped with identical propagation algorithms equivalent to Eqs. (2.7)–(2.9) of the GSF.

4.8.1 Simplified Falling Body Simulation Revisited

Once again, the falling body system of Fig. (2.4) is revisited. As the primary interest of this analysis is the performance of the filter update, as opposed to the filter propagation, the effects of nonlinear dynamics are of less interest. Thus, the dynamics are made linear by neglecting drag, such that the simplified falling body system of Section 3.2.3 already provides the appropriate system configuration, with a few exceptions. Measurement scans are generated every second, where valid sensor returns conform to the nonlinear, Gaussian model given by Eq. (2.33), where the measurement noise covariance for valid sensor returns is taken to be $R_v = 10$ ft^2. Faulty measurements are produced using models specified in each analysis, which will change based on the type of fault-cognizant filter used; however, all filters are equipped with the GSF propagation algorithms of Eqs. (2.7)–(2.9), such that the propagation of the state estimate for each filter is functionally equivalent. The duration of the simulation is 10 seconds, and each analysis is performed

over 10,000 Monte Carlo (MC) trials.

As the fault-cognizant filters and the GSF are both nonlinear filters constructed on GM approximations of the estimated distribution, the accuracy of their solutions improve with an increase in the number of components [47]. The baseline GSF does not naturally produce additional components in the update as the fault-cognizant filter does, and consequently does not benefit from the nonlinear architecture if initialized with only a single Gaussian. To address this issue, the Gaussian distribution of Eq. (3.25) is split into a GM of 9 components via a three-component splitting library applied across both position and velocity as described in Section 2.2.3.3. Additionally, to bypass any approximation error that the splitting process may incur, the initial state is drawn directly from the 9 component GM instead of Eq. (3.25). To regulate the number of GM components that the fault-cognizant filters generate, they are equipped with the merging and pruning procedures described in Sections 2.2.3.1 and 2.2.3.2, where the pruning and merging thresholds are set to $w_{\text{thresh}} = 10^{-5}$ and $d_{\text{thresh}} = 10^{-4}$, respectively. It is found that under these thresholds, components are reduced primarily by pruning, with only a few components being merged in each analysis, and that, on average, the maximum number of components accumulated within a single simulation run is only 41.

4.8.2 Falling Body Analysis of FCU-1

This first set of analyses of the simplified falling body problem is used to test the FCU-1 updates derived in Section 4.6.2, where performance is compared to a baseline filter equipped with the residual editing update of Section 4.7.2. As such, all measurement scans will consist of single measurement returns, where the state-independent probability of validity is taken to be $v = 0.99$, such that any given measurement has a 1% chance of being generated as faulty. Addtionally, these analyses are also used to survey the various faulty spatial distributions presented in Section 4.5.2 in order to better understand how faulty measurement distributions affect residual editing and the fault-cognizant filters themselves.

In an effort to enforce impartiality when comparing the baseline and proposed filters, the performance of the GSF with residual editing is tested under the varying conditions of the simulation.

The results of these initial tests are then used to tune the residual editing threshold parameter, subsequently improving the overall performance of the baseline filter. Referring to the threshold check of Eq. (4.150), it is found that a threshold of $\Upsilon = 16.45$ produces a well performing filter, and it is also noted that this value corresponds to a χ^2 probability gate of 99.995% for one dimensional range measurements.

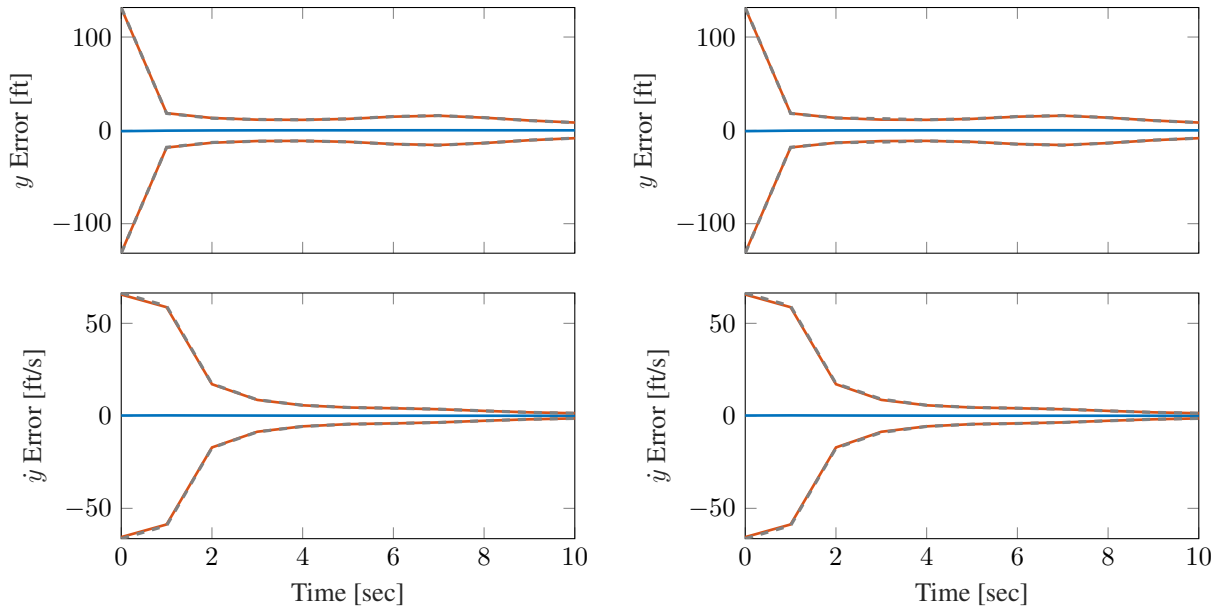
4.8.2.1 Performance Comparison Analysis Under Ideal Conditions

To show that all filters are operational, this analysis evaluates the performance of the FCU-1 filter when the exact faulty measurement distribution is known—an ideal condition. This is useful as a performance reference for the later analyses. Two specific cases are examined, the first being when the faulty measurements are uniformly distributed and the second being when they are normally distributed.

4.8.2.1.1 Uniformly Distributed Faulty Range Measurements

In this case, the faulty measurements are generated according to a uniform distribution as outlined in Section 4.5.2.1. It is assumed that the FCU-1 filter is informed of the correct faulty measurement model and is equipped with the uniform FCU-1 update of Eq. (4.97). For this simulation, the range sensor can only return measurements in the domain $z \in [0 \text{ ft}, \ 100000 \text{ ft}]$, such that the sensor volume is $V = 100,000$ ft. The results of this analysis are provided in Fig. (4.13), which include average filter error results as well as MC statistics pursuant to Eqs. (2.25) and (2.26).

Again, the sign of a well-behaving filter is one in which the average filter error is nearly zero and where the MC statistics closely match those of the filter average. Immediately upon inspection, it is clear that the outcome of the FCU-1 filter of Fig. (4.13a) is nearly identical to that of the residual editing filter of Fig. (4.13b), where they both produce ideal behavior in the presence of uniform faulty measurements. In fact, the overconfidence of residual editing is imperceptible enough to be considered practically negligible. While this level of performance is expected of the fault-cognizant filter, as it precisely models the faulty measurements, the exceptional performance of residual editing is credited to the large size of the sensor volume ($V = 100,000$ ft) relative to

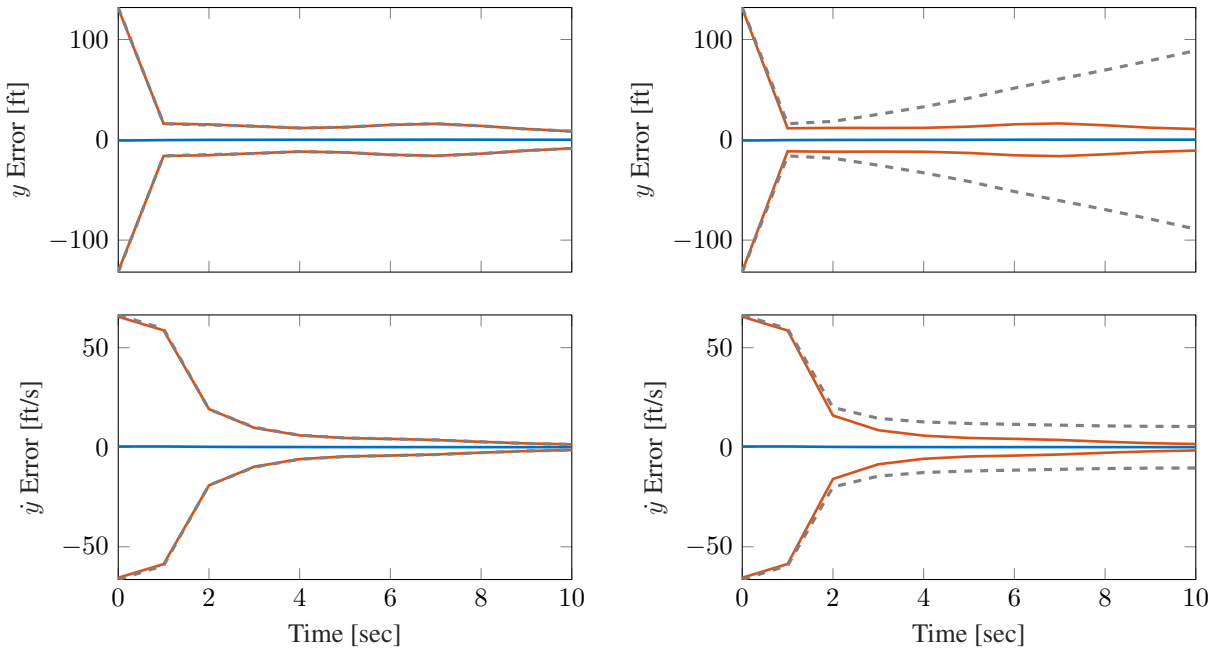


a) FCU-1 Update **b) Residual Editing Update**

Figure 4.13 Comparison results for uniformly distributed faulty measurements, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

the covariance of the estimate. In such a situation, few, if any, faulty measurements occur within the editing threshold, and thus filtering operations are adequately protected by residual editing. Consequently, as the sensor volume V becomes smaller, it is anticipated that the performance of residual editing will degrade.

Recalling attention to the near identical nature of Figs. (4.13a) and (4.13b), this is attributed to the fact that the FCU-1 update via a uniform faulty likelihood operates similarly to residual editing. Since the uniform faulty likelihood is state-independent, the means and covariances of the components are not updated when accounting for the possibility of faulty data. However, the weights of the components are still updated, which reflects the possibility that the measurement may or may not be faulty, such that, on average, the FCU-1 update can lead to lower errors than the residual editing update.



a) FCU-1 Update **b) Residual Editing Update**

Figure 4.14 Comparison of results for normally distributed faulty measurements, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

4.8.2.1.2 Normally Distributed Faulty Range Measurements

In this case, the faulty measurements are produced by a single normal distribution according to the model provided in Section 4.5.2.2. The faulty observation function $\boldsymbol{h}_f(\cdot)$ is taken to be identical to the valid observation function $h_v(\cdot)$ of Eq. (2.33), but the faulty measurement noise covariance differs as $R_f = 100R_v = 10,000$ ft^2. Again, it is assumed that the FCU-1 filter is given the exact true measurement model and, as such, is equipped with the normal FCU-1 update of Eq. (4.102). The average and Monte Carlo results of the analysis are given in Fig. (4.14), where Fig. (4.14a) presents the results of the proposed FCU-1 update, and Fig. (4.14b) presents the results of the baseline GSF with residual editing.

Here, the results are more promising than the uniform distribution results of Section 4.8.2.1.1, as the performance of the FCU-1 update in Fig. (4.14a) far exceeds that of residual editing in Fig. (4.14b). While the results of the FCU-1 update are nearly the same between the analyses of Figs. (4.13) and (4.14), the results of the baseline filter with residual editing become much worse,

as seen in Fig. (4.14b). The MC statistics for both the position and velocity estimates now exceed the average filter error covariance to the point that the GSF with residual editing can be considered a divergent filter. This poor performance is caused by the remarkably similar nature of the faulty and valid measurement distributions. While the covariances of each may differ significantly in size, the means of the distributions are the same, which leads to more faulty measurements falling inside of the residual editing threshold. Subsequent analyses further confirm that cases such as this make residual editing much less effective.

4.8.2.2 Model Mismatch Analysis

This analysis is designed to test how robust the FCU-1 update is to unknown faulty measurement models, which can be compared to the baseline filter's performance in Figs. (4.14b) and (4.17b). As residual editing makes no explicit modeling assumptions as to how faulty data is generated, baseline results strictly unique to this analysis are not attainable and thus are presented only in other analyses to avoid redundant figures.

Two different cases are analyzed, the first when faulty measurements are normally distributed and the second when they are exponentially distributed. In both cases, the FCU-1 filter utilizes the uniform FCU-1 update with the filter probability of validity set as $v_{\text{filt}} = 0.90$, an erroneous assumption as faulty measurements are generated according to the true probability $v = 0.99$. While not strictly true, setting $v_{\text{filt}} < v$ is generally a conservative assumption, as there is typically less state information gained from faulty measurements than from valid measurements. Since conservative filter behavior is customarily more desirable than overconfidence, it is advised that $v_{\text{filt}} \leq v$ for most practical applications.

The motivation behind this section is based on the notion that many data screening methods, including residual editing, are appealing in part because no extra modeling considerations need be made for invalid returns. This application of the uniform FCU-1 update is intended to mimic a scenario in which, instead of expending effort to find an appropriate model, faulty measurements are simply treated as uniformly distributed.

4.8.2.2.1 Normally Distributed Faulty Range Measurements

This case is configured similarly to the previous analysis of Section 4.8.2.1.2, as the faulty measurements are produced by the same normal distribution. However, instead of an FCU-1 filter equipped with the correct update, results are taken from a filter with the uniform FCU-1 update of Eq. (4.97) where the sensor volume is again taken to be $V = 100,000$ ft. The data resulting from this model mismatch analysis is provided in Fig. (4.15a).

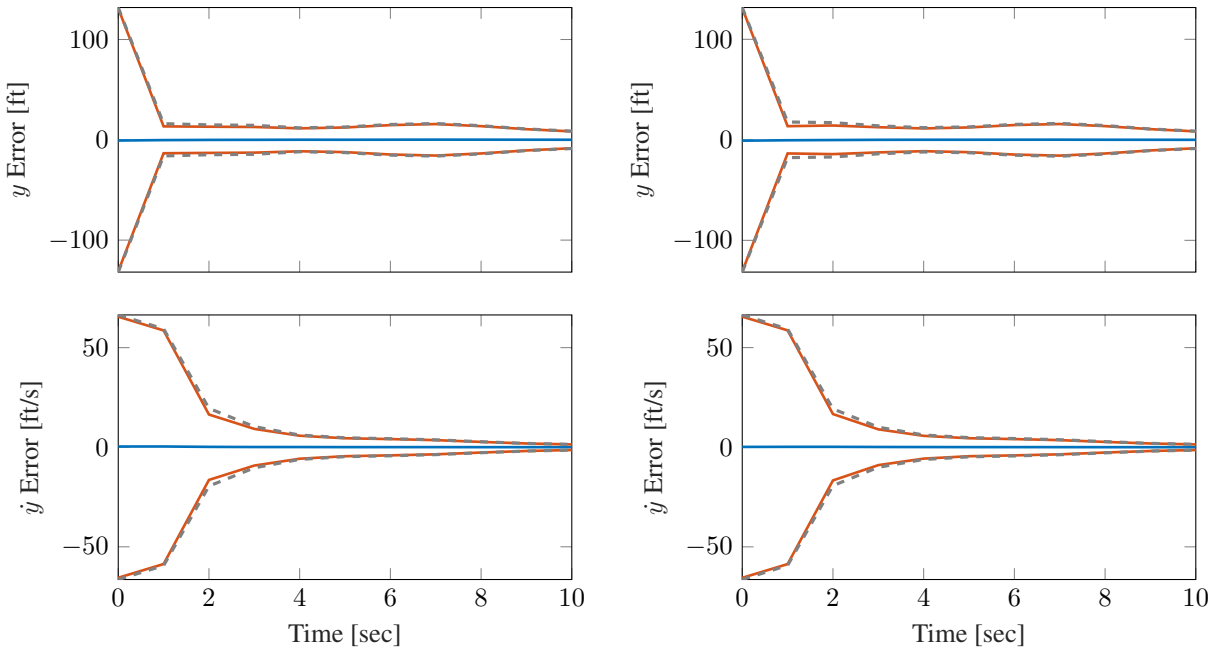
First, it is important to mention that Fig. (4.14b) contains the residual editing results for normally distributed faulty measurements, which should be referenced for comparison. When inspecting the FCU-1 update alone, the results in Figs. (4.14a) and (4.15a) are also important to compare, as they both operate on measurement sets where the faulty measurements are generated by a normal distribution. By introducing model mismatch via the uniform FCU-1 update, the filter performance is degraded slightly, as indicated by the minor overconfidence of Fig. (4.15a) when compared to Fig. (4.14a). However, the FCU-1 filter still performs exceptionally well considering it is subjected to the burden of model mismatch. In fact, assuming a uniform distribution and implementing an FCU-1 filter produces results that handily outperform the baseline filter with residual editing of Fig. (4.14b), despite the FCU-1 filter fallaciously adopting $v_{\text{filt}} = 0.90$ when, in truth, $v = 0.99$.

4.8.2.2.2 Exponentially Distributed Faulty Range Measurements

To further test the effects of model mismatch, this case again utilizes a filter outfitted with the uniform FCU-1 update of Eq. (4.97), with a sensor volume of $V = 100,000$ ft. The actual faulty measurements are generated according to an exponential distribution as

$$z_f = h_v(\boldsymbol{x}) - d_{\text{offset}} + w_{\text{exp}} \,, \tag{4.153}$$

where $h_v(\cdot)$ is the valid measurement model from Eq. (2.33), $d_{\text{offset}} = 25$ ft is a parameter that offsets the mean of the faulty measurement distribution from the mean of the valid distribution,



a) Normally Distributed Faulty Measurements

b) Exponentially Distributed Faulty Measurements

Figure 4.15 Model mismatch results for proposed FCU-1 filter, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

and w_{exp} is a random noise drawn from an exponential distribution as

$$w_{\text{exp}} \sim p_e(\lambda) \ , \tag{4.154}$$

where $\lambda = 50$ ft is the rate parameter of the exponential distribution. This model is intended to imitate a scenario where some unmodeled effect of the target produces faulty range measurements that indicate the object is closer than it actually is, and is especially dubious by design. As the faulty measurement model is state-dependent via the function $h_v(\boldsymbol{x})$, and the faulty distribution is concentrated relatively close to the valid distribution, it is more difficult for the filters to differentiate between valid and faulty measurements. Plotting the average filter error and Monte Carlo statistics for the FCU-1 filter yields Fig. (4.15b).

Once again, when comparing the results of the proposed filter in Fig. (4.15b) to the baseline filter of Fig. (4.17b), the FCU-1 update undoubtedly outperforms residual editing. There do exist

similarities between the results of Figs. (4.15a) and (4.15b), where the FCU-1 update is slightly overconfident earlier in the simulation, but regains a healthier estimate later on. However, as this analysis is subjected to model mismatch, the exact source of the overconfidence is challenging to diagnose. Regardless, even with mismodeled faults, the proposed filter is certainly more robust than the baseline filter, as Fig. (4.17b) is, again, characteristic of divergent filter behavior. It is clear that unmodeled faulty measurements existing in close proximity to the valid measurements is not as much of an issue for the FCU-1 filter as it is for residual editing.

4.8.2.3 FCU-1 via GM Approximation Analysis

This analysis examines the abilities of the proposed FCU-1 filter when the actual faulty measurement distribution is approximated with a GM. Here, it is assumed that faulty sensor returns conform to the same exponential model presented in Eqs. (4.153) and (4.154). The proposed filter, however, is limited to a 5 component GM approximation of the exponential distribution, utilizing the FCU-1 update described in Section 4.6.2.2. The approximation is made using a variant of the expectation-maximization (EM) algorithm [45], and Fig. (4.16) shows how closely the two likelihoods resemble each other when plotted with respect to the measurement space. Clearly there are

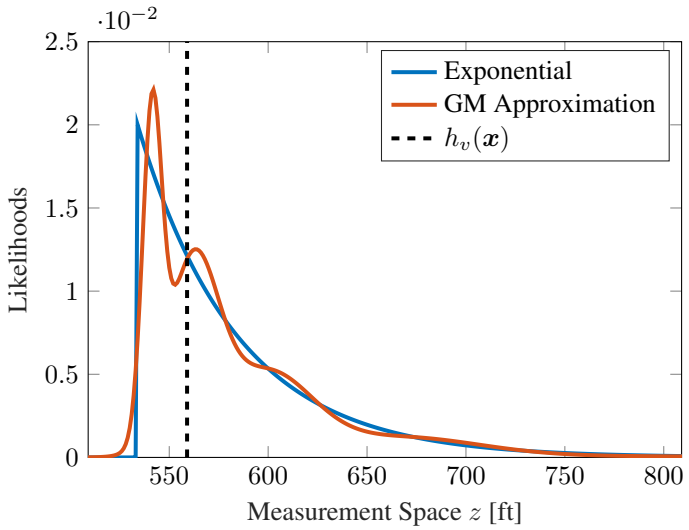
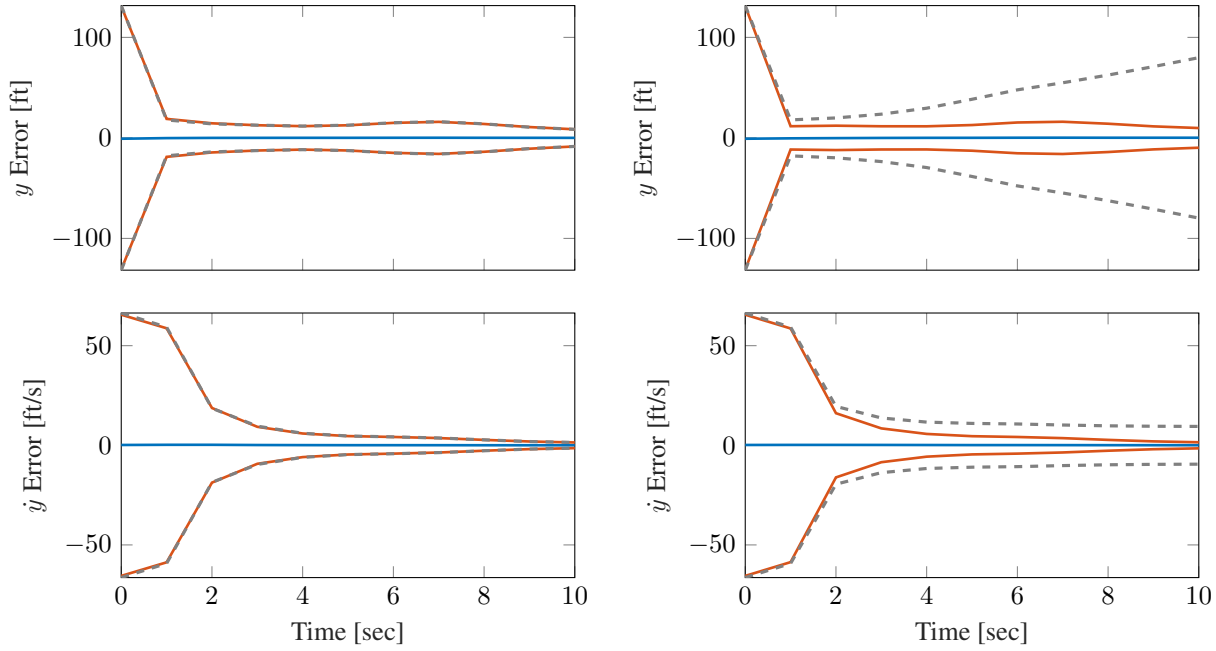


Figure 4.16 Comparison of exponential faulty measurement distribution to GM approximation for $y = 0$

dissimilarities between the two distributions of Fig. (4.16), but equipping the FCU-1 update with a rough approximation such as this tests its capabilities under an extreme case. Additionally, more components compound the number of computations in the update, making fewer components more desirable when seeking to limit the computational burden. The results of the Monte Carlo trials for both the proposed and baseline filters are found in Fig. (4.17).



a) FCU-1 Update **b) Residual Editing Update**

Figure 4.17 GM approximation results for exponentially distributed faulty measurements, expressed as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

For a filter possessing only an approximate representation of the faulty measurement distribution, the FCU-1 update performs remarkably well, as evidenced by Fig. (4.17a). As expected, there is a large decrease in the filter's overconfidence when compared to Fig. (4.17b) where a GSF with residual editing is used. What is more interesting is that there are only marginal improvements between Fig. (4.15b)—where the FCU-1 is supplied with the incorrect faulty spatial distribution—and Fig. (4.17a)—where the FCU-1 filter directly approximates the faulty distribution. While this

indicates the FCU filter with a GM approximate of the faulty measurement model produces estimates that are more accurate than simply assuming the distribution is uniform, the enhancement may not be substantial enough to justify the extra modeling. It is therefore worth mentioning that as the probability of validity v increases, such that faulty measurements become less influential, modeling the faulty distribution—either exactly or approximately—within the FCU-1 update becomes less beneficial. This is supported by the performance of the FCU-1 filters in Figs. (4.13a), (4.14a), (4.15a), (4.15b), and (4.17a), where, despite being subjected to various faulty distributions and outfitted with different updates, the uncertainty profiles are extremely similar. As $v = 0.99$ in these simulations, the posterior estimate is almost entirely composed of the valid likelihood, and thus there is little information gain across the state-dependent FCU-1 updates. Conversely, as v decreases, the information gain from the state-dependent faulty measurements increase, and the uncertainty profiles of the FCU-1 filter will differ more between analyses, warranting the effort required to model the faulty distribution.

The decision to accurately model the faulty distribution notwithstanding, it is apparent that the FCU-1 filter performs as well as, if not considerably better than, residual editing in all cases, signifying that there exist clear advantages to utilizing a filter with soft decisions as opposed to one constructed on strict categorizations.

4.8.3 Falling Body Analysis of FCU-2

While the analyses of Section 4.8.2 are concerned with testing FCU-1 under various faulty measurement spatial distributions, this section uses the simplified falling body simulation to analyze FCU-2 of Section 4.6.3 with varying models for probability of detection. As the results from Section 4.8.2.2 support the choice of uniform faulty measurements as a highly robust model, faulty measurements are taken to be spatially uniform across a sensor field-of-view on the interval $[0 \text{ ft}, 5000 \text{ ft}]$—such that the sensor volume is $V_{\text{range}} = 5000$ ft—and are temporally Poisson according to Eq. (4.72) with rate parameter specified in each analysis. In these analyses, the pruning threshold is raised to $w_{\text{thresh}} = 10^{-8}$, such that the number of components remains tractable at all times, with the maximum number of components accumulated in each trial, averaged across

all MC trials, being 191 for the FCU-2 filter with Gaussian model and 70 for the FCU-2 filter with zeroth-order approximation. Since the GSF with residual editing does not naturally generate additional components, the number of components never exceeds 9.

The probability of detection is randomly sampled from the Gaussian model of Section 4.4.2.4, where

$$g(\boldsymbol{x}) = p_{D,\text{max}} - \frac{p_{D,\text{max}} - p_{D,\text{min}}}{V_{\text{range}}} h(\boldsymbol{x}) \tag{4.155}$$

is the nonlinear function of the mean. Note that values for $p_{D,\text{max}}$, $p_{D,\text{min}}$, and R_D are specified for each analysis.

4.8.3.1 Filter Configurations

Three variants of filters are evaluated in this analyses, the configurations of which are specified here. They correspond to the low-, medium-, and high-fidelity $p_D(1, \boldsymbol{x})$ models described in Sections 4.4.2.2–4.4.2.4.

4.8.3.1.1 GSF with Residual Editing:

This filter, outlined in Section 4.7.2, is considered a framework with the lowest amount of $p_D(\boldsymbol{x})$ modeling, where the only mechanism being applied is a residual check. Not only does it not necessitate a p_{D_z} return, it also does not need any function of $p_D(\boldsymbol{x})$ to be specified. The same residual editing threshold as Section 4.8.2 of $\Upsilon = 16.45$ is used, which corresponds to a χ^2 value of 99.995%.

4.8.3.1.2 FCU-2 Filter with Zeroth-Order Approximated p_D:

This version of the proposed filter is constructed in Eqs. (4.110) and is representative of a moderate amount of $p_D(\boldsymbol{x})$ modeling as it also does not require a value of p_{D_z} to be reported. Instead, it calculates a zeroth-order approximation of the probability of detection as

$$p_D(\boldsymbol{m}) = g(\boldsymbol{m})\,.$$

4.8.3.1.3 FCU-2 Filter with State-Dependent p_D:

This filter, described in Eqs. (4.113), possesses the highest level of $p_D(\boldsymbol{x})$ modeling and requires a reported probability of detection return, p_{D_z}, from some external sensing operation. In this simulation, it is randomly sampled according to the distribution of Eq. (4.70) with mean $g(\boldsymbol{x})$ defined by Eq. (4.155).

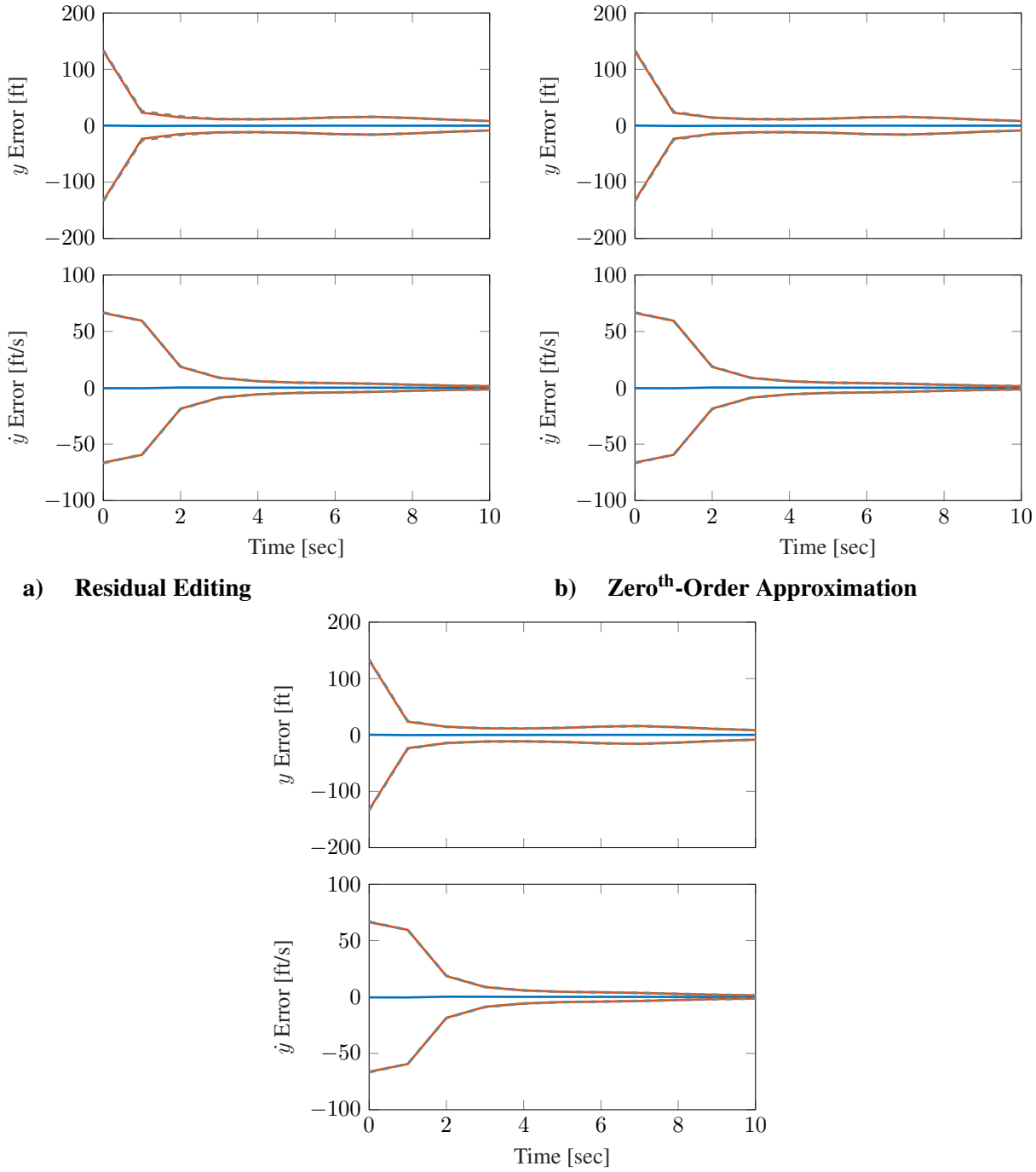
4.8.3.2 Analysis 1: Ideal Sensing Conditions

Before analyzing the filters under extreme stress, it is useful to first prove that they operate well in prototypical conditions. Therefore, this analysis is intended to mimic more ideal sensing conditions, where the probability of detection remains relatively constant and near 1 during the entirety of the simulation ($p_{D,\text{min}} = 0.95$, $p_{D,\text{max}} = 0.99$, $R_D = 0.0025$), and clutter is generated at a very low rate of $\lambda = 0.001$. The results of all three filters are presented in Fig. (4.18), where performance of each is nearly identical, and all are observed to be well-behaved as the average filter and MC standard deviations match almost exactly.

As none of the filters show signs of degeneracy, it can be said that any one of the filters is an appropriate selection in similar systems with nearly constant/unity probability of detection and essentially non-existent clutter. If minimal modeling and good computational efficiency is desired, the GSF with residual editing is advisable, as it performs as well as the other filters in this instance. This outcome is by design, of course, and this analysis provides a good baseline for the remaining analyses.

4.8.3.3 Analysis 2: Cluttered Environment with Highly Variable p_D

To evaluate filtering performance under less ideal conditions, clutter returns are generated at a rate of $\lambda = 0.25$, and the probability of detection mean of Eq. (4.155) is given parameters of $p_{D,\text{min}} = 0.10$, $p_{D,\text{max}} = 0.99$, and $R_D = 0.0025$, such that $p_D(\boldsymbol{x})$ varies much throughout the simulation. The corresponding MC results are found in Fig. (4.19), where it is immediately clear that all filters exhibit higher levels of uncertainty than in Fig. (4.18), as the quality of incoming measurements has significantly decreased. By the uncertainty profile of Fig. (4.19a), the GSF



a) Residual Editing

b) Zero$^{\text{th}}$-Order Approximation

c) Gaussian Model

Figure 4.18 Filter results in ideal sensing environment plotted as $\bar{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

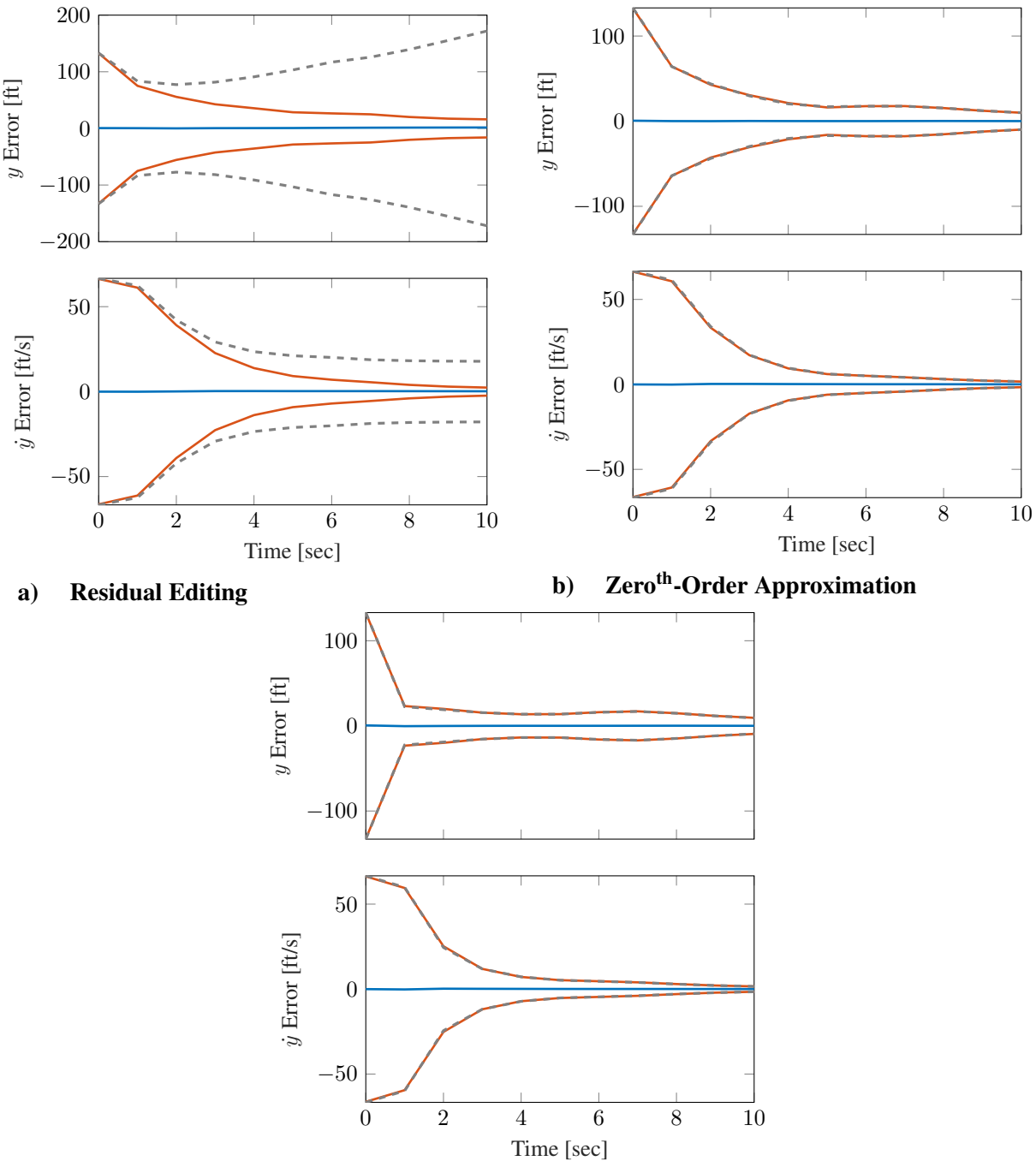
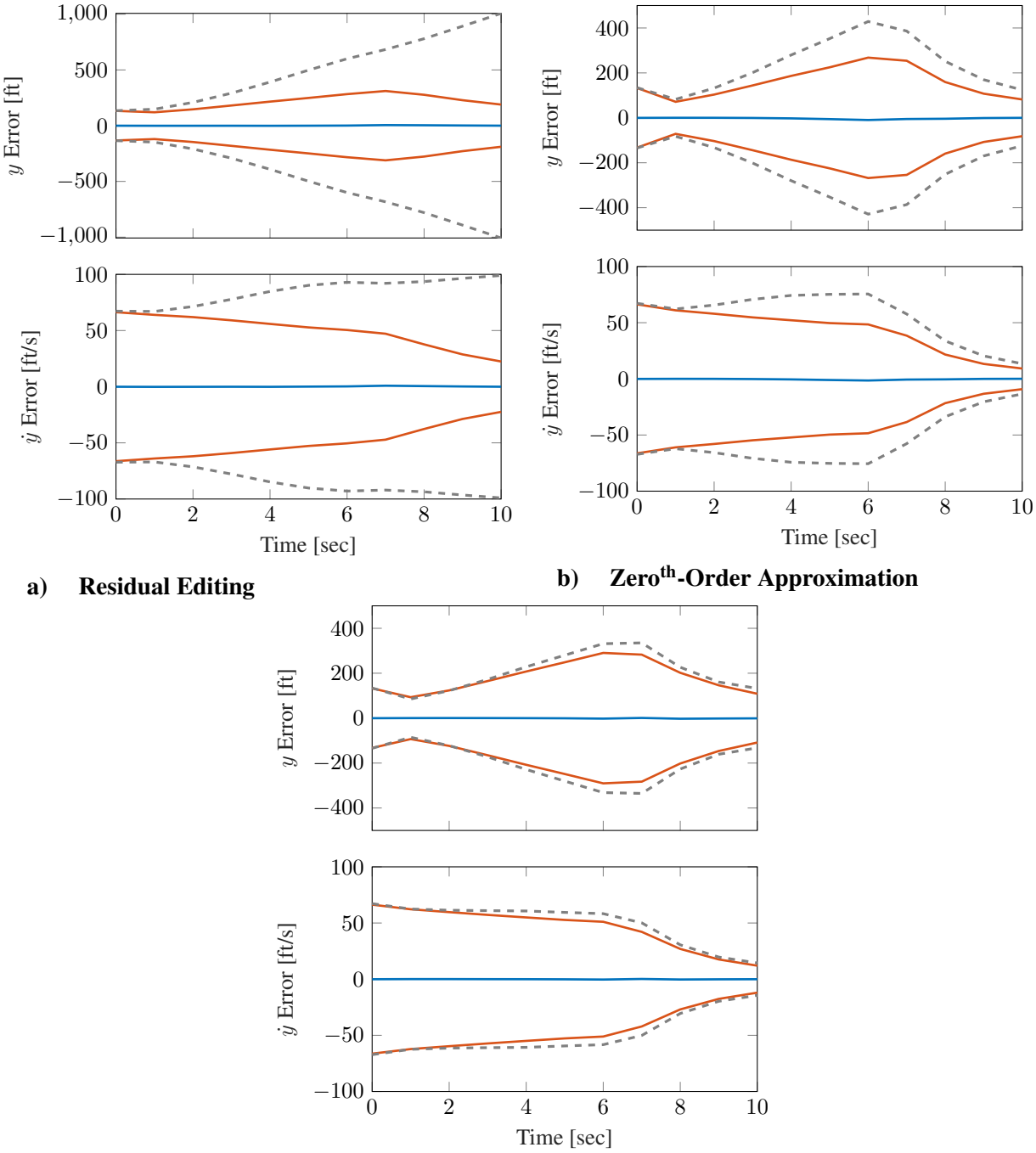


Figure 4.19 Filter results in cluttered sensing environment plotted as $\bar{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)



a) Residual Editing

b) Zeroth-Order Approximation

c) Gaussian Model

Figure 4.20 Filter results for failing sensor plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

outfitted with residual editing is no longer capable of producing a valid filtering solution and shows divergent behavior. Meanwhile, the other filters show good consistency between MC and average filtering standard deviations, indicating healthy estimation solutions. However, the filter with a Gaussian model p_D of Fig. (4.19c) produces estimates with lower uncertainty than that of the filter with zeroth-order approximation of Fig. (4.19b). This is attributed to two factors. First, as p_D varies more, there is a greater amount of information gain from the Gaussian model. Secondly, near the beginning of the simulation, the probability of detection is generally lower, such that fewer valid measurements are being generated. Therefore, information gained from p_{D_z} returns is more influential at this time. As p_D nears 1 towards the end of the simulation, valid measurements become the dominant contributor of incoming information, and both FCU-2 filters achieve similar performance.

As expected, residual editing is not robust enough for cluttered sensing conditions with low detection probability, and the GSF of Section 4.7.2 is not capable of acceptable performance. Both FCU-2 filters, on the other hand, are considered well-behaved, although modeling p_D explicitly shows a significant advantage over the simple zero$^{\text{th}}$-order approximation here.

4.8.3.4 Analysis 3: Sensor Failure/Model Mismatch

As the proposed filters of this work are intended to be fault resistant, this analysis mimics a possible scenario with a failed sensor, where the filter is informed of an incorrect model for $p_D(\boldsymbol{x})$ in the time interval between two and six seconds. While the FCU-2 filters expect a relatively high probability of detection with $p_{D,\text{min,filt}} = 0.95$, $p_{D,\text{max,filt}} = 0.99$, and $R_{D,\text{filt}} = 1$, starting at two seconds, the actual probability of detection is generated according to $p_{D,\text{min,true}} = 0.05$, $p_{D,\text{max,true}} = 0.15$, and $R_{D,\text{true}} = 0.0025$. Note that the covariance $R_{D,\text{filt}}$ is inflated here to reflect a lack of trust in the model, and it is ill-advised to ever set $R_{D,\text{filt}} < R_{D,\text{true}}$. To further stress the system, the rate of faulty measurements is increased tenfold to $\lambda = 2.5$. The results of this analysis are found in Fig. (4.20).

Upon inspection, it is clear that the GSF with residual editing of Fig. (4.20a) still performs the worst by far and that the FCU-2 filters operate well, by comparison. However, the zeroth-

order approximate FCU-2 filter of Fig. (4.20b) becomes significantly more overconfident than the Gaussian model FCU-2 filter of Fig. (4.20c), while also producing higher overall errors. This implies that the FCU-2 filter with a Gaussian model for $p_D(\boldsymbol{x})$ is the most robust to extreme model mismatch, but it is noted that it can also be the most sensitive if the filter covariance is not increased enough; i.e., if $R_{D,\text{filt}}$ is too small in situations with extreme model mismatch, then the filter of Fig. (4.20c) performs much worse. It is found that in cases where the model of $p_D(\boldsymbol{x})$ is very poorly known and model mismatch of the probability of detection occurs, that setting $R_{D,\text{filt}} \geq 1$ is sufficient, as this not only expands the conventional 3σ confidence interval to include any return $0 \leq p_{D_z} \leq 1$, but also almost completely removes the information gain from the probability of detection.

4.8.4 Orbit Determination Simulation

To further test the FCU-2 update with varying models for probability of detection, a more complex simulation is introduced in this section. This orbit determination (OD) simulation involves the tracking of the satellite O3B FM7 (Catalog ID 40080), maintained in Earth centered inertial (ECI) coordinates, and initialized on August 15, 2021 at 8:30AM UTC using a two-line element set from space-track.org.

To sufficiently define the dynamics of the system, satellite motion is governed by simple Keplerian two body dynamics, and no process noise is imparted to the system. Tracking is done by a ground-based observer located at the Haleakalā observatories on Maui, where measurement scans of right-ascension and declination are simulated every minute for 2.5 days, a duration that allows for multiple satellite passes and three "night" periods for the observer. It is assumed that the optical sensor has valid noise standard deviations of 3 arcseconds with a field-of-view spanning 3° by 3°, and clutter is uniformly generated across it at a rate of $\lambda = 50$.

Following procedures in [53], the admissible region of Fig. (4.21) is formed and subsequently approximated by a GM of 165 components, which provides a distribution to initialize the filter. To limit the increasing number of components, GM pruning and merging are applied according to Sections 2.2.3.1 and 2.2.3.2, where $w_{\text{thresh}} = 10^{-10}$ and $d_{\text{merge}} = 0.05$. Additionally, the splitting

procedure of Section 2.2.3.3 is used to split each existing GM component into 729 smaller components (applied across 6 dimensions using a 3 component splitting library), which is performed any time the number of components falls below 3.

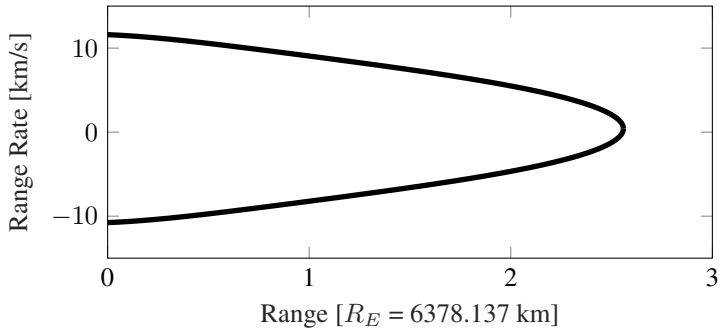


Figure 4.21 Admissible region generated for O3B FM7

Results are generated across 350 MC trials according to Eqs. (2.25) and (2.26), where both the zeroth-order approximation and Gaussian model versions of the FCU-2 filter from Eqs. (4.110)–(4.113) are tested. In this simulation, the GSF with residual editing is found to fail so often that it is infeasible to generate any meaningful MC results and is thus left out of this comparison entirely. Clearly, this indicates that a more nuanced approach to fault resistance than simply screening measurement residuals is needed for a simulation of this nature.

4.8.4.1 Probability of Satellite Detection

A four-part model of the probability of detection is built as

$$g(\boldsymbol{x}) = p_{D,1}\, p_{D,2}(\boldsymbol{x})\, p_{D,3}(\boldsymbol{x})\, p_{D,4}\,, \tag{4.156}$$

where each term $p_{D,i}$ corresponds to a different factor affecting detectability. While this model may not cover all aspects of satellite observability, it is intended to be more complete than a simple visibility check and provide a higher level of complexity with which to test the filters. Again, the simulated probability of detection is randomly sampled according to the model of Section 4.4.2.4,

where Eq. (4.156) is the state-dependent mean and $R_D = 0.03^2$ is the variance. An example of the simulated probability of detection over the duration of the simulation is presented in Fig. (4.22), the profile of which illustrates nine distinct periods of higher detectability.

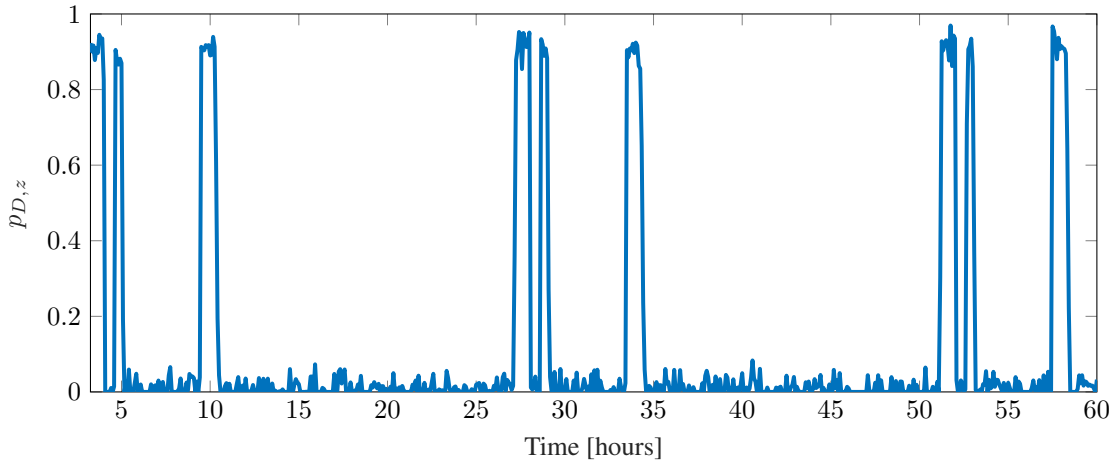


Figure 4.22 Profile of randomly sampled probabilities of detection of satellite

4.8.4.1.1 Observer Sky Brightness:

It is assumed that, for valid sensor returns to be possible, the observer's sky must be sufficiently dark, a factor that is dictated by the elevation of the Sun α_{sun} with respect to observer horizon [109]. This model assumes that the sky begins to darken at nautical twilight ($\alpha_{\text{NT}} = -6^\circ$) and reaches maximal darkness at astronomical twilight ($\alpha_{\text{AT}} = -18^\circ$), the transition between which is modeled via a sigmoid function as

$$p_{D,1} = \frac{1}{1 + \exp\{c_1(c_2 + \alpha_{\text{sun}})\}} , \tag{4.157}$$

where choosing $c_1 = 2\frac{1}{\text{deg}}$ and $c_2 = 9^\circ$ in Eq. (4.157) yields a smooth function of $p_{D,1}$ that behaves as desired, which is plotted in Fig. (4.23a). Note that this specific model assumes that the observer cannot detect satellites under daylight conditions, which neglects the recent advancements in day-

time observing [110].

4.8.4.1.2 Satellite Elevation:

For valid satellite observations to be possible, the satellite must, of course, be above the observer horizon. This is simulated via $p_{D,2}(\boldsymbol{x})$, which is also modeled as a sigmoid function as

$$p_{D,2}(\boldsymbol{x}) = \frac{1}{1 + \exp\{c_1(c_2 - \alpha_{\text{sat}})\}} \,, \tag{4.158}$$

where α_{sat} is the satellite's elevation and, in this case, $c_1 = 5\frac{1}{\text{deg}}$ and $c_2 = -0.5^\circ$ produce behavior desirable for Eq. (4.158), the profile of which is shown in Fig. (4.23b).

4.8.4.1.3 Satellite Illumination:

In order for the satellite to be visible to the observer, it must be sufficiently illuminated by the Sun. Put simply, a near-Earth satellite may be 1) not illuminated at all during an umbral eclipse, 2) partially illuminated during a penumbral eclipse, or 3) fully illuminated if no eclipse occurs [109]. Again, a sigmoid function is used to model a smooth transition between $p_{D,3} = 0$ and $p_{D,3} = 1$, such that

$$p_{D,3}(\boldsymbol{x}) = \frac{1}{1 + \exp\{c_1(c_2 - \theta)\}} \,, \tag{4.159}$$

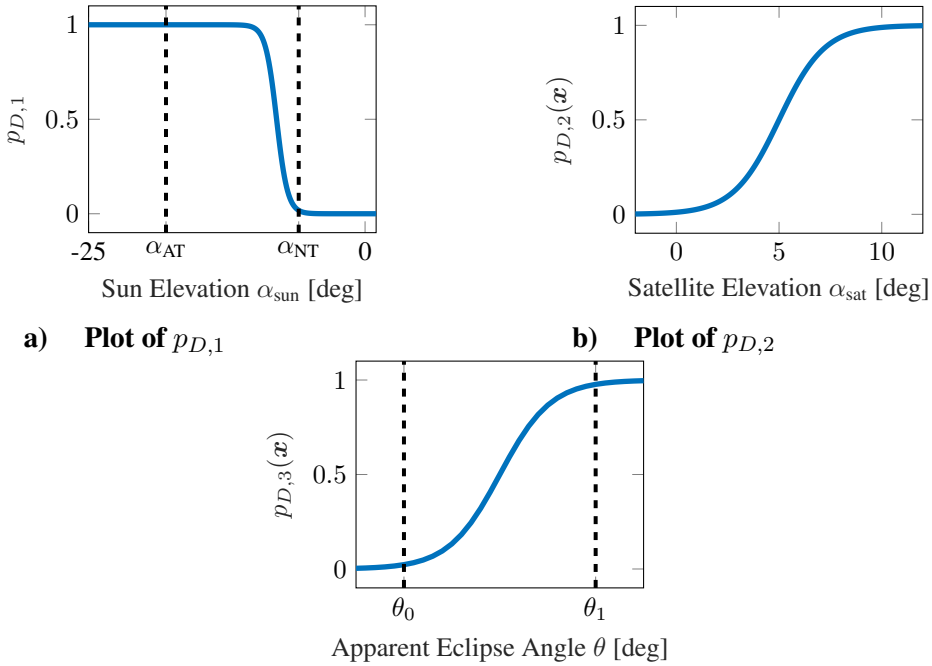
where θ is the apparent angle between the centers of the Earth and Sun, calculated as

$$\begin{aligned}
\theta &= \cos^{-1}\left(\frac{\boldsymbol{r}_E \cdot \boldsymbol{r}_S}{\left\|\boldsymbol{r}_E\right\| \cdot \left\|\boldsymbol{r}_S\right\|}\right) \\
c_1 &= \frac{7.5}{\theta_1 - \theta_0} \\
c_2 &= \theta_0 + \frac{\theta_1 - \theta_0}{2} \\
\theta_0 &= |\theta_E - \theta_S| \\
\theta_1 &= \theta_E + \theta_S \,,
\end{aligned}$$

and where θ_E and θ_S are the semi-diameters of the earth and sun, respectively, given by

$$\theta_E = \sin^{-1}\left(\frac{R_E}{\|\boldsymbol{r}_E\|}\right) \quad \text{and} \quad \theta_S = \sin^{-1}\left(\frac{R_S}{\|\boldsymbol{r}_S\|}\right).$$

Note that $\boldsymbol{r}_S$ is the position vector from the satellite to Sun, and $\boldsymbol{r}_E$ is the position vector from the satellite to Earth, whereas R_E and R_S are the radii of the Earth and Sun, respectively. More importantly, $\theta < \theta_0$ corresponds to an umbral eclipse, $\theta_0 \leq \theta \leq \theta_1$ corresponds to a penumbral eclipse, and $\theta > \theta_1$ corresponds to no eclipse at all. Figure (4.23c) contains a representative profile of the detectability function in Eq. (4.159).



a) **Plot of** $p_{D,1}$ **b)** **Plot of** $p_{D,2}$

c) **Plot of** $p_{D,3}$

Figure 4.23 Profiles of detectability factors contributing to the overall probability of detection

4.8.4.1.4 Pointing Direction:

Given the small size of the observer's FOV relative to the horizon, it is unrealistic to assume that the observer is always able to capture the satellite within the image frame. Instead, this model

proposes that the observer is equipped with a fairly successful sensor tasking scheme, such that the satellite remains in view of the sensor approximately 91% of the time. If the vector to the satellite is considered the nominal pointing direction, then perturbations of right ascension δ_{ra} and declination δ_{d} are randomly sampled from a normal distribution as

$$\delta_{\text{ra}}, \delta_{\text{d}} \sim p_g\left(\delta \middle| 0, \left(\frac{3}{4}\right)^2 \text{deg}^2\right),$$

which are then used to offset the pointing direction from its nominal state. Thus, if $\max[|\delta_{\text{ra}}|, |\delta_{\text{d}}|] > 1.5^\circ$, then the satellite is outside the FOV, and a detection does not occur. Accordingly, $p_{D,4}$ is given by

$$\begin{aligned} p_{D,4} &= \Pr\{|\delta_{\text{ra}}| \leq 1.5^\circ\} \Pr\{|\delta_{\text{d}}| \leq 1.5^\circ\} \\ &= 0.91107\,. \end{aligned} \tag{4.160}$$

4.8.4.2 Orbit Determination Simulation Discussion

The MC results for both FCU-2 filters are found in Figs. (4.24) and (4.25), which, at a glance, contain remarkably similar results. Neither of the filters fail during the simulation, and, towards the end of the simulation, both filters acheive nearly identical levels of uncertainty. However, earlier in the simulation, especially between hours three and six, the zeroth-order approximated filter of Fig. (4.24b) shows a slight overconfidence that the FCU-2 filter with Gaussian model of Fig. (4.24a) does not have. As this simulation is relatively complex, it is difficult to attribute this overconfidence to one specific factor, yet it can be stated that the proposed filter with Gaussian modeled $p_D(\boldsymbol{x})$ does exhibit better behavior than the zeroth-order approximation variant. Additionally, near the end of the simulation, the average filter uncertainty $3\sigma_{\text{filt},\iota}$ (——) of Fig. (4.25a) is slightly lower than that of Fig. (4.25b). This indicates that the high-fidelity FCU-2 filter with Gaussian model produces a more certain estimate than that of the medium-fidelity FCU-2 filter with zeroth-order approximation, which is most likely due to the additional information gain from

explicitly modeling the state-dependent probability of detection. Note that this difference in uncertainty is slight enough to the point that it is imperceptible in Fig. (4.25), which is simply because the optical measurements taken by the observer are much more accurate than the probability of detection returns supplied to the filter.

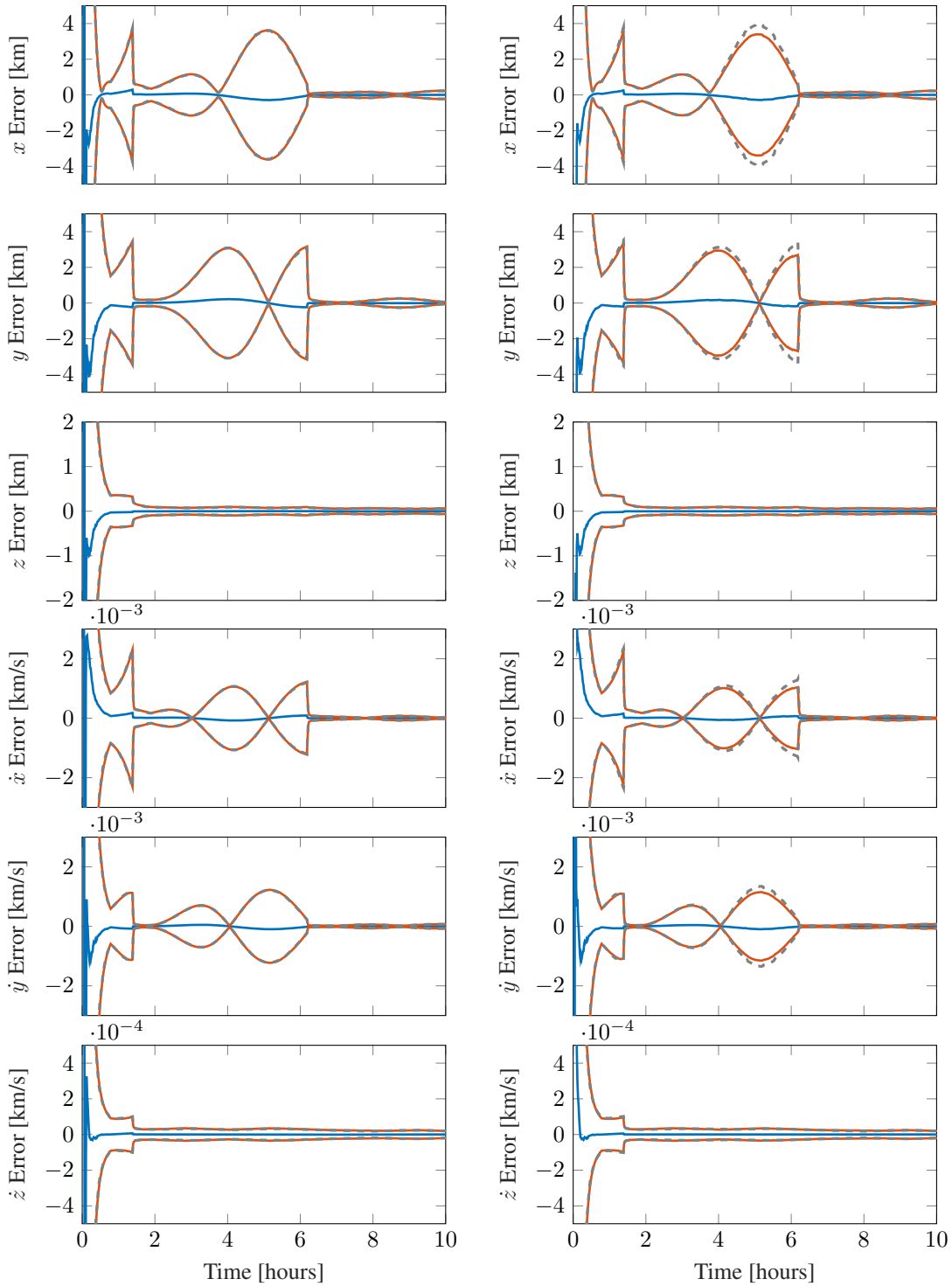


Figure 4.24 Zoomed MC results of first 10 hours of OD simulation plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

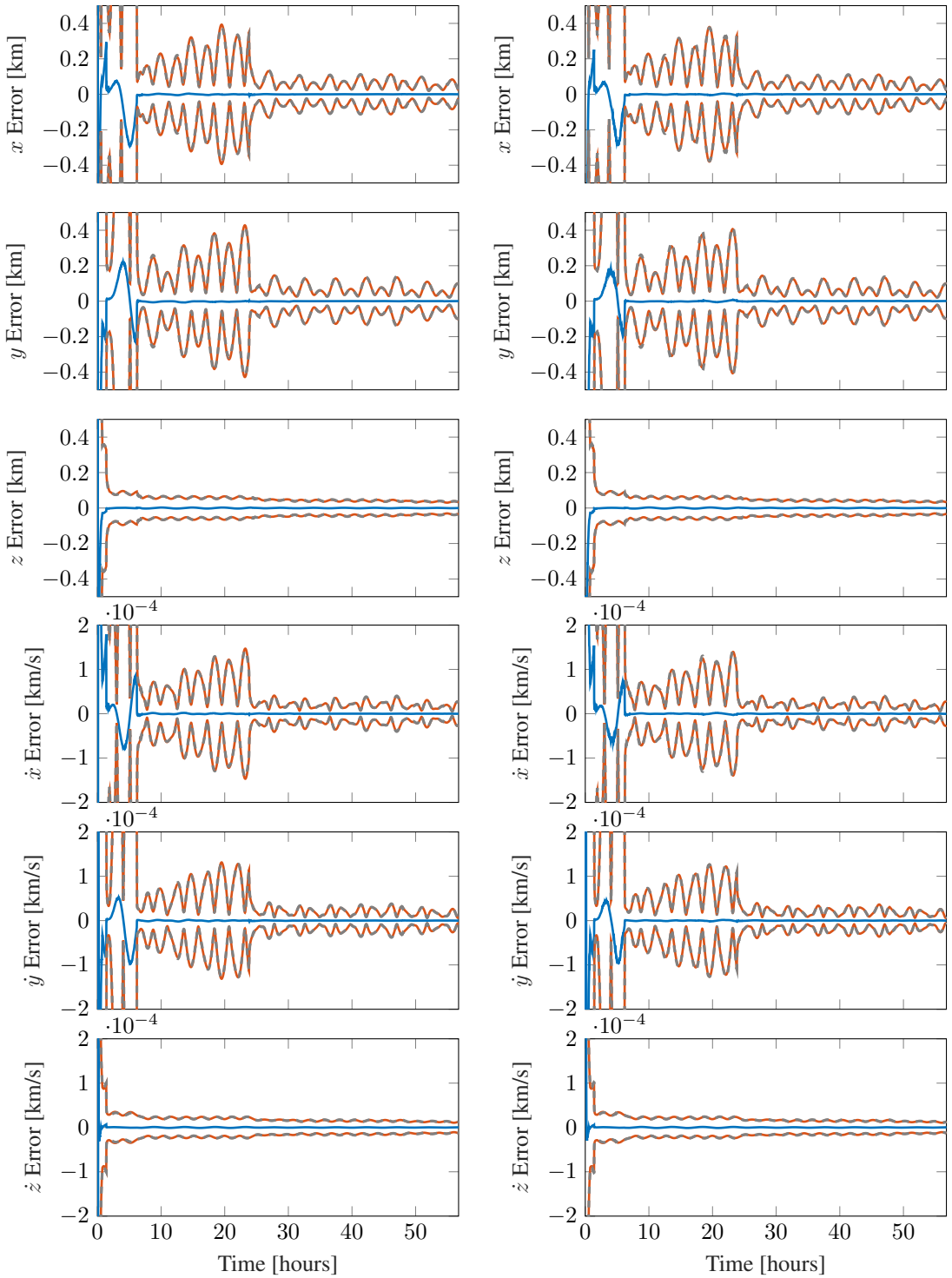


Figure 4.25 Zoomed MC results of entire OD simulation plotted as $\overline{e}_\iota$ (——), $3\sigma_{\text{filt},\iota}$ (——), and $3\sigma_{\text{MC},\iota}$ (- - -)

5. CONCLUSION

To conclude this dissertation, this chapter summarizes the research of the preceding chapters, highlights key findings from the collection of analyses, and posits potential future work.

5.1 Summary of Research

Preceded by a general introduction to robustness and filtering, the main body of this research is presented in three chapters, each focusing on a specific paradigm important to fulfilling this dissertation's purpose. As it is implemented frequently throughout this work, Chapter 2 is dedicated to nonlinear filtering. The general nonlinear framework is described by differentiating it from linear estimation, and the scope of nonlinear filtering is narrowed to Gaussian mixture (GM) filters. To illustrate the benefits of nonlinear filtering, the traditional Gaussian sum filter (GSF) is compared to its linear counterpart, the extended Kalman filter (EKF), by way of a Monte Carlo simulation.

With the utility of nonlinear filters established, Chapter 3 focuses on achieving robustness by way of non-Bayesian filtering. As Bayes' rule inherently assumes that all models are precisely known, generalized variational inference (GVI) provides an alternate foundation for updates more robust to model misspecification. Several different GVI updates are investigated via numerical optimization before a closed-form, confidence-based GVI update is derived. Both linear and nonlinear (GM) realizations of the confidence-based update are presented, which are then tested against underweighting to showcase its relevance to navigation.

While Chapter 3 seeks robustness by adjusting the confidence of a model's validity, Chapter 4 attains robustness by increasing model fidelity to include faulty measurements. Where traditional measurement models assume all incoming measurements are "valid", the fault-cognizant measurement models (FCMMs) permit the existence of undesirable or erroneous measurements. By establishing the spatial and temporal distributions of both valid and faulty measurements, the FCMMs can be processed via Bayes' rule, producing nonlinear closed-form fault-cognizant updates (FCUs). These FCUs are tested in several simulations, where residual editing is provided as

a performance baseline, as it is common to navigation.

5.2 Future Work

When conducting research, it is almost inevitable for the rate at which questions arise to outpace the time and resources required to answer them. This section is the result of this, describing several possible paths of research that solicit further investigation.

5.2.1 Alternate Closed-Form GVI Updates

While the confidence-based update of Section 3.3 is certainly useful when deriving closed-form non-Bayesian updates, other closed-form posteriors derived from generalized variational inference (GVI) are possible. Based on the promising results of Section 3.2.3.3, when the Kullback-Leibler divergence of Eq. (3.5) and the log-form of the expected γ-loss function of Eq. (3.11) are selected as the GVI cost function, then it can be shown that

$$p^{+}(\boldsymbol{x}) \propto p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\gamma-1}\Big[\int \ell(\boldsymbol{s}|\boldsymbol{x})^{\gamma}\mathrm{d}\boldsymbol{s}\Big]^{\frac{\gamma-1}{\gamma}},$$

which is derived similarly to Eq. (3.36). This update is particularly suitable for Gaussian measurement noise models, as the integral can be computed in closed-form. While this yields a rather similar result to the confidence-based filter for linear-Gaussian systems of Section 3.3.1.1, this is just one example of a possible posterior. It is the opinion of the author that a great deal more can be learned by using GVI as a genesis and testbed for the derivation of robust updates.

5.2.2 Combining Non-Bayesian and Fault-Cognizant Filtering

While stemming from different motivations, the non-Bayesian updates of Chapter 3 and fault-cognizant updates of Chapter 4 are not necessarily incohesive frameworks; the non-Bayesian updates focuses on replacing Bayesian fusion with GVI-based updates, whereas fault-cognizance is achieved by defining more complex measurement models. While the combination of the two seems justified, as it allows for a robust fusion of prior and incoming information while simultaneously modeling faults, it is unfortunately not straightforward to simply merge the two. For example,

pairing the confidence-based update of Eq. (3.38) with FCMM-1 of Eq. (4.14) yields a new update of the form

$$p^{+}(\boldsymbol{x}) = \frac{\left\{v(\boldsymbol{x})\ell_v(\boldsymbol{z}|\boldsymbol{x}) + \big[1 - v(\boldsymbol{x})\big]\ell_f(\boldsymbol{z}|\boldsymbol{x})\right\}^{\frac{1-\phi}{\phi}} p^{-}(\boldsymbol{x})}{\int \left\{v(\boldsymbol{s})\ell_v(\boldsymbol{z}|\boldsymbol{s}) + \big[1 - v(\boldsymbol{s})\big]\ell_f(\boldsymbol{z}|\boldsymbol{s})\right\}^{\frac{1-\phi}{\phi}} p^{-}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}} \,. \tag{5.1}$$

While it is possible to reach closed-form solutions for the corresponding FCU of Eq. (4.91) and the confidence-based update of Eq. (3.38) separately, this does not imply that the combined update of Eq. (5.1) has a closed-form as well. Already, it is clear that the exponential term of Eq. (5.1) is problematic, as the exponent does not distribute well among the terms that make up the fault-cognizant likelihood. Therefore, a closed-form solution to such an update requires significant additional effort to find, if one exists at all. Furthermore, if a closed-form update does exist, the resulting behavior of the combined update, if undesirable, may not warrant its existence. Therefore, further investigation is worthwhile.

If such an update is investigated in the future, it is the opinion of the author that analytically solving for the power of a GM, either exactly or approximately, is a practical first step. Specifically, if one can find a general expression for

$$\Big[\sum_{\xi=1}^{L} w_\xi p_g\big(\boldsymbol{x}|\boldsymbol{m}_\xi, \boldsymbol{P}_\xi\big)\Big]^{\alpha} \Rightarrow c(\alpha)\sum_{\xi=1}^{L} \omega_\xi(\alpha) p_g\big(\boldsymbol{x}|\boldsymbol{\mu}_\xi(\alpha), \boldsymbol{\Pi}_\xi(\alpha)\big)\,, \tag{5.2}$$

which moves the α term from the exponential to either a leading coefficient or inside the GM components, it is then possible to directly multiply Eq. (5.2) with other Gaussians, the product of which will also be Gaussian. Not only would this be a fairly useful operation for various GM applications, but it would help in calculating closed-form solutions to fault-cognizant non-Bayesian updates like Eq. (5.1) as well.

5.2.3 Practical Applications of Fault-Cognizance

Many different variations of fault-cognizant updates are presented in Section 4.6, each designed for a specific system or desired estimator behavior. The simulations of Section 4.8 analyze most of these FCUs, sufficiently illustrating the benefits of fault-cognizance in measurement modeling, but it would be impossible to exhaustively derive and test all possible FCUs. Instead, the author believes that a high-fidelity simulation, such as a digital emulation of a realistic GNC system, is a more practical way to further test fault-cognizance. Unfortunately, the resources and time required to enact such a simulation are beyond the scope of this dissertation, but should be investigated in the future.

5.3 Final Remarks

Without robust implementations, many practical applications of filters are not possible. An effective robustness approach must not only protect the filter against agents known to debase its algorithms, but also against any unknown elements within the system. To accomplish this, current practice is mostly comprised of procedure-first methods, which affix *ad hoc* mechanisms to pre-existing filter frameworks to prevent failure. For example, spacecraft navigation commonly uses residual editing, underweighted information gains, redundant sensors, and filter resets to avoid many causes of corruption to the navigation filter. While some of these causes are simply unmodeled effects of the system, other causes are long-known within the estimation community, such as the linearization error of the EKF. These procedure-first methods, while effective, are in some ways only relevant because of a continued reliance on the EKF, preferred mainly for its familiarity and computational efficiency.

These procedure-first/EKF-centric tendencies within the field of navigation motivate the majority of this dissertation, where the focus is achieving model-first robustness through nonlinear updates. Instead of amending existing frameworks to make them robust, the many updates of this research are robustly derived from first principles and with nonlinear GM realizations that circumvent both the linearization errors and Gaussian-reliant models afflicting linear estimators. This

model-first mindset is approached in two ways.

First, since models only approximate the reality of physical systems, Bayes' rule is theoretically inappropriate, as it inherently assumes that the models for prior and incoming information are exactly known. This solicits the investigation of non-Bayesian updates, the result of which is a GVI-based update that allows for variable confidence in the validity of system models. Whereas a naive Bayesian filter forms more efficient estimates, assuming nothing will disagree with its models, the confidence-based filter produces more conservative estimates, reflecting the possibility that its models are misspecified. While conservative estimation can functionally be achieved in a variety of manners (filter tuning, fictitious process noise, underweighting, etc.), this confidence-based update is a more nuanced approach that is theoretically well-founded.

The second approach to model-first robustness in this work challenges traditional measurement modeling, where measurements are assumed to be generated according to a single, valid distribution. It goes without saying that filters are functionally "aware" of a system only to the extent of their models; if a (Bayesian) filter is constructed with a Gaussian measurement model, it assumes that all measurements, Gaussian or not, must abide by the prescribed model and treats them as such. In this way, any measurement that does obey the valid model is faulty and may cause significant damage to the filter. Traditionally, measurements are screened such that any existing outside of the filters expectations are rejected. Alternatively, this work suggests fault-cognizance be imparted to the filter by attempting to model these faulty measurements instead of simply screening for them. This requires extra modeling and assumptions that are highly system-specific, which results in many different fault-cognizant updates. Analyzing these updates, it is shown that fault-cognizance, while computationally more expensive, is more robust to faulty measurements (model mismatch) than residual editing.

Regardless of the precise approach selected, the culmination of this dissertation is a collection of updates, each designed with a slightly different system or behavior in mind. The functional differences between these updates may be large or small, but all differences can be attributed to the assumptions unique to each derivation. While the author would certainly be pleased if any future

applications are able to directly apply the algorithms of this work, that is not the main goal of this dissertation; it is unreasonable to exhaustively construct a filter designed for every possible system. Instead, it is the author's intent to convey the following: *every filter can be wholly derived with assumptions suitable for its specific system.* While many reuse established estimators like the EKF, there is often a more suitable framework that is much better suited for the conditions at hand, the only caveat being it has yet to be created. Of course, deriving a customized filter is a challenging feat, which may require statistical expertise and in-depth system knowledge. If the contents of this dissertation—wherein robust updates are progressively derived from statistical theory to practical closed-forms—provides instruction (or, even better, encouragement) that makes the tailored derivation of filters seem more tenable, then this dissertation has fulfilled its intended purpose. After all, when building a filter, one is not limited solely to the catalog of existing architectures, and this work is a single step towards a future where more individuals are comfortable with designing estimators from the ground-up.

REFERENCES

[1] Abdelaziz, A. M., Tealib, S. K., and Molotov, I., "Analytical study of Egyptian TIBA-1 satellite orbit from Optical Satellite Tracking Station (OSTS), NRIAG-Egypt," *Astrophysics and Space Sciences*, Vol. 366, No. 81, 2021. https://doi.org/10.1007/s10509-021-03987-7.

[2] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, Vol. 82, No. 1, 1960, pp. 35–45. https://doi.org/10.1115/1.3662552.

[3] Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *Transactions of the ASME–Journal of Basic Engineering*, Vol. 83, No. 1, 1961, pp. 95–108. https://doi.org/10.1115/1.3658902.

[4] Schmidt, S. F., "State Space Techniques Applied to the Design of a Space Navigation System," *Joint Automatic Control Conference*, 1962.

[5] Schmidt, S. F., "Applications of State Space Methods to Navigation Problems," *Advances in Control Systems*, Vol. 3, 1966, pp. 293–340. https://doi.org/10.1016/B978-1-4831-6716-9.50011-4.

[6] Lear, W. M., "Multi-Phase Navigation Program for the Space Shuttle Orbiter," Internal Memo No. 73-FM-132, NASA Johnson Space Center, Houston, TX, 1973.

[7] Zanetti, R., Holt, G., Gay, R., D'Souza, C., Sud, J., Mamich, H., Begley, M., King, E., and Clark, F. D., "Absolute Navigation Performance of the Orion Exploration Flight Test 1," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 5, 2017, pp. 1106–1116. https://doi.org/10.2514/1.G002371.

[8] van de Geer, S. A., *Least Squares Estimation*, John Wiley & Sons, Ltd, Chichester, England, 2005, Vol. 2, pp. 1041–1045. https://doi.org/10.1002/0470013192.bsa199.

[9] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Dover Publications, Mineola, NY, 2007.

[10] Särkkä, S., *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, UK, 2013. https://doi.org/10.1017/CBO9781139344203.

[11] Ross, S. M., *Introduction to Probability Models*, 11th ed., Academic Press, Cambridge, MA, 2014.

[12] Huber, P. J., *Robust Statistics*, Springer, Berlin, Heidelberg, 2011, pp. 1248–1251. https://doi.org/10.1007/978-3-642-04898-2_594.

[13] Box, G. E. P., "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society: Series A (General)*, Vol. 143, No. 4, 1980, pp. 383–404. https://doi.org/10.2307/2982063.

[14] Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., et al., "An Overview of Robust Bayesian Analysis," *Test*, Vol. 3, No. 1, 1994, pp. 5–124. https://doi.org/10.1007/BF02562676.

[15] Huber, P. J., and Ronchetti, E. M., *Robust Statistics*, 2nd ed., John Wiley & Sons, Inc., 2009. https://doi.org/10.1002/9780470434697.

[16] Masreliez, C. I., and Martin, R., "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Transactions on Automatic Control*, Vol. 22, No. 3, 1977, pp. 361–371. https://doi.org/10.1109/TAC.1977.1101538.

[17] Calvet, L. E., Czellar, V., and Ronchetti, E. M., "Robust Filtering," *Journal of the American Statistical Association*, Vol. 110, No. 512, 2015, pp. 1591–1606. https://doi.org/10.1080/01621459.2014.983520.

[18] Schick, I. C., and Mitter, S. K., "Robust Recursive Estimation in the Presence of Heavy-Tailed Observation Noise," *The Annals of Statistics*, Vol. 22, No. 2, 1994, pp. 1045–1080.

[19] McCabe, J. S., "Multitarget Tracking and Terrain-Aided Navigation Using Square-Root Consider Filters," Ph.D. thesis, Missouri University of Science and Technology, 2018.

[20] Schmidt, S. F., "The Kalman Filter: Its Recognition and Development for Aerospace Applications," *Journal of Guidance and Control*, Vol. 4, No. 1, 1981, pp. 4–7. https://doi.org/10.2514/3.19713.

[21] Carpenter, J. R., and D'Souza, C. N., "Navigation Filter Best Practices," Technical Report TP–2018–219822, NASA Goddard Space Flight Center, April 2018.

[22] Lefferts, E., Markley, F., and Shuster, M., "Kalman Filtering for Spacecraft Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 5, No. 5, 1982, pp. 417–429. https://doi.org/10.2514/3.56190.

[23] Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Dover Publications, Mineola, NY, 2006.

[24] Mejías, L., Saripalli, S., Campoy, P., and Sukhatme, G. S., "Visual Servoing of an Autonomous Helicopter in Urban Areas using Feature Tracking," *Journal of Field Robotics*, Vol. 23, No. 3-4, 2006, pp. 185–199. https://doi.org/10.1002/rob.20115.

[25] Tapley, B., Schutz, B., and Born, G. H., *Statistical Orbit Determination*, Elsevier, 2004.

[26] Adurthi, N., Majji, M., and Singla, P., "Quadrature-Based Nonlinear Joint Probabilistic Data Association Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 42, No. 11, 2019, pp. 2369–2381. https://doi.org/10.2514/1.G004313.

[27] Särkkä, S., Tolvanen, V., Kannala, J., and Rahtu, E., "Adaptive Kalman Filtering and Smoothing for Gravitation Tracking in Mobile Systems," *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2015, pp. 1–7. https://doi.org/10.1109/IPIN.2015.7346762.

[28] Lin, C.-L., Chu, T.-C., Wu, C.-E., Chang, Y.-M., Lin, T.-C., Chen, J.-F., Chuang, C.-Y., and Chiu, W.-C., "Tracking Touched Trajectory on Capacitive Touch Panels Using an Adjustable Weighted Prediction Covariance Matrix," *IEEE Transactions on Industrial Electronics*, Vol. 64, No. 6, 2017, pp. 4910–4916. https://doi.org/10.1109/TIE.2017.2669887.

[29] Cutolo, F., Mamone, V., Carbonaro, N., Ferrari, V., and Tognetti, A., "Ambiguity-Free Optical–Inertial Tracking for Augmented Reality Headsets," *Sensors*, Vol. 20, No. 5, 2020, p. 1444. https://doi.org/10.3390/s20051444.

[30] Kaplan, E. D., and Hegarty, C., *Understanding GPS/GNSS: Principles and Applications*, Artech House, 2017.

[31] Yun, S., "Sequential Monte Carlo Filtering with Gaussian Mixture Models for Highly Nonlinear Systems," Ph.D. thesis, The University of Texas at Austin, 2021. https://doi.org/10.26153/tsw/13639.

[32] Gelb, A., *Applied Optimal Estimation*, MIT press, Cambridge, MA, 1974.

[33] Maybeck, P. S., *Stochastic Models, Estimation, and Control*, Vol. 141-2, Academic Press, New York, NY, 1982.

[34] Ito, K., and Xiong, K., "Gaussian Filters for Nonlinear Filtering Problems," *IEEE Transactions on Automatic Control*, Vol. 45, No. 5, 2000, pp. 910–927. https://doi.org/10.1109/9.855552.

[35] Wan, E. A., and van der Merwe, R., "The Unscented Kalman Filter for Nonlinear Estimation," *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, IEEE, 2000, pp. 153–158. https://doi.org/10.1109/ASSPCC.2000.882463.

[36] Arasaratnam, I., Haykin, S., and Elliott, R. J., "Discrete-Time Nonlinear Filtering Algo-

rithms Using Gauss-Hermite Quadrature," *Proceedings of the IEEE*, Vol. 95, No. 5, 2007, pp. 953–977. https://doi.org/10.1109/JPROC.2007.894705.

[37] Arasaratnam, I., and Haykin, S., "Cubature Kalman Filters," *IEEE Transactions on Automatic Control*, Vol. 54, No. 6, 2009, pp. 1254–1269. https://doi.org/10.1109/TAC.2009.2019800.

[38] Kumar, M., and Chakravorty, S., "Nonlinear Filter Based on the Fokker-Planck Equation," *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 1, 2012, pp. 68–79. https://doi.org/10.2514/1.54070.

[39] Bazik, M., Flewelling, B., Majji, M., and Mundy, J., "Bayesian Inference of Spacecraft Pose using Particle Filtering," *19th AMOS Technical Conference*, 2018, pp. 757–762.

[40] Raihan, D., and Chakravorty, S., "Particle Gaussian Mixture Filters-I," *Automatica*, Vol. 98, 2018, pp. 331–340. https://doi.org/10.1016/j.automatica.2018.07.024.

[41] Vo, B.-N., and Ma, W.-K., "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, Vol. 54, No. 11, 2006, pp. 4091–4104. https://doi.org/10.1109/TSP.2006.881190.

[42] De Moivre, A., "A Method of approximating the Sum of the Terms of the Binomial $(a+b)^n$ expanded into a Series, from whence are deduced some practical Rules to eliminate the Degree of Assent which is to be given to Experiements," *The Doctrine of Chances*, 1738, pp. 235–243.

[43] Lyon, A., "Why are Normal Distributions Normal?" *The British Journal for the Philosophy of Science*, Vol. 65, No. 3, 2014, pp. 621–649. https://doi.org/10.1093/bjps/axs046.

[44] Alspach, D. L., and Sorenson, H. W., "Nonlinear Bayesian Estimation using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, Vol. AC-17, No. 4, 1972, pp. 439–448. https://doi.org/10.1109/TAC.1972.1100034.

[45] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.

[46] Sorenson, H. W., and Alspach, D. L., "Recursive Bayesian Estimation using Gaussian Sums," *Automatica*, Vol. 7, No. 4, 1971, pp. 465–479. https://doi.org/10.1016/0005-1098(71)90097-5.

[47] van der Merwe, R., "Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models," Ph.D. thesis, Oregon Health & Science University, Portland, OR, April 2004.

[48] Mahler, R. P. S., *Advances in Statistical Multisource-Multitarget Information Fusion*, Artech House, Norwood, MA, 2014.

[49] Crouse, D. F., Willett, P., Pattipati, K., and Svensson, L., "A Look at Gaussian Mixture Reduction Algorithms," *14th International Conference on Information Fusion*, IEEE, Chicago, IL, 2011, pp. 1–8.

[50] Kristan, M., Skočaj, D., and Leonardis, A., "Incremental Learning with Gaussian Mixture Models," *Computer Vision Winter Workshop*, Moravske Toplice, Slovenia, 2008, pp. 25–32.

[51] DeMars, K. J., Bishop, R. H., and Jah, M. K., "Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 4, 2013, pp. 1047–1057. https://doi.org/10.2514/1.58987.

[52] Tuggle, K., and Zanetti, R., "Automated Splitting Gaussian Mixture Nonlinear Measurement Update," *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 3, 2018, pp. 725–734. https://doi.org/10.2514/1.G003109.

[53] DeMars, K. J., and Jah, M. K., "Probabilistic Initial Orbit Determination Using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 5, 2013, pp. 1324–1335. https://doi.org/10.2514/1.59844.

[54] Dizikes, P., "Explained: Monte Carlo simulations," *MIT News*, 2010. URL https://news.mit.edu/2010/exp-monte-carlo-0517.

[55] Bar-Shalom, Y., Li, X.-R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, John Wiley & Sons, Inc., New York, NY, 2001. https://doi.org/10.1002/0471221279.

[56] Sorenson, H. W., and Sacks, J. E., "Recursive Fading Memory Filtering," *Information Sciences*, Vol. 3, No. 2, 1971, pp. 101–119. https://doi.org/10.1016/S0020-0255(71)80001-4.

[57] Wishner, R. P., Tabaczynski, J. A., and Athans, M., "A Comparison of Three Non-Linear Filters," *Automatica*, Vol. 5, 1969, pp. 487–496. https://doi.org/10.1016/0005-1098(69)90110-1.

[58] Fritsch, G. S., and DeMars, K. J., "Adaptive Confidence Filter Update for High Uncertainty Environments," *Proceedings of the AIAA SciTech Forum*, Orlando, FL, 2020. https://doi.org/10.2514/6.2020-1696.

[59] Bayes, T., "An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S." *Philosophical Transactions of the Royal Society of London*, 1763, pp. 370–418. https://doi.org/10.1098/rstl.1763.0053.

[60] Blei, D. M., Kucukelbir, A., and McAuliffe, J. D., "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, Vol. 112, No. 518, 2017, pp. 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[61] Knoblauch, J., Jewson, J., and Damoulas, T., "Generalized Variational Inference: Three Arguments for Deriving New Posteriors," , December 2019. https://doi.org/10.48550/arXiv.1904.02063.

[62] DeMars, K. J., and Jah, M. K., "Evaluation of the Information Content of Observations with Application to Sensor Management for Orbit Determination," *Advances in the Astronautical Sciences*, Vol. 142, Univelt, Inc., 2011, pp. 3169–3188.

[63] Cichocki, A., and Amari, S., "Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities," *Entropy*, Vol. 12, No. 6, 2010, pp. 1532–1568. https://doi.org/10.3390/e12061532.

[64] Kullback, S., and Leibler, R. A., "On Information and Sufficiency," *The Annals of Mathematical Statistics*, Vol. 22, No. 1, 1951, pp. 79–86. URL https://www.jstor.org/stable/2236703.

[65] Shannon, C. E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, No. 3, 1948, pp. 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[66] Caticha, A., "Relative Entropy and Inductive Inference," American Institute of Physics, 2004, pp. 75–96. https://doi.org/10.1063/1.1751358.

[67] van Erven, T., and Harremoës, P., "Rényi Divergence and Kullback-Leibler Divergence," *IEEE Transactions on Information Theory*, Vol. 60, No. 7, 2014, pp. 3797–3820. https://doi.org/10.1109/TIT.2014.2320500.

[68] Rényi, A., "On Measures of Entropy and Information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 1, University of California Press, 1961, pp. 547–562.

[69] Fujisawa, H., and Eguchi, S., "Robust parameter estimation with a small bias against heavy contamination," *Journal of Multivariate Analysis*, Vol. 99, No. 9, 2008, pp. 2053–2081. https://doi.org/10.1016/j.jmva.2008.02.004.

[70] Jewson, J., Smith, J. Q., and Holmes, C., "Principles of Bayesian Inference Using General Divergence Criteria," *Entropy*, Vol. 20, No. 6, 2018, p. 442. https://doi.org/10.3390/e20060442.

[71] Boyd, S. P., and Vandenberghe, L., *Convex Optimization*, Cambridge University Press, New York, NY, 2004.

[72] Eliason, S. R., *Maximum Likelihood Estimation*, SAGE Publications, Inc., Newbury Park, CA, 1993. https://doi.org/10.4135/9781412984928.

[73] Hung, H., Jou, Z.-Y., and Huang, S.-Y., "Robust Mislabel Logistic Regression without Modeling Mislabel Probabilities," *Biometrics*, Vol. 74, No. 1, 2018, pp. 145–154. https://doi.org/10.1111/biom.12726.

[74] Pardo, L., *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC, New York, NY, 2005. https://doi.org/10.1201/9781420034813.

[75] Gil, M., Alajaji, F., and Linder, T., "Rényi Divergence Measures for Commonly used Univariate Continuous Distributions," *Information Sciences*, Vol. 249, 2013, pp. 124–131. https://doi.org/10.1016/j.ins.2013.06.018.

[76] Ho, Y.-C., and Lee, R. C. K., "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Transactions on Automatic Control*, Vol. 9, No. 4, 1964, pp. 333–339. https://doi.org/10.1109/TAC.1964.1105763.

[77] Zanetti, R., DeMars, K. J., and Bishop, R. H., "Underweighting Nonlinear Measurements," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 5, 2010, pp. 1670–1675. https://doi.org/10.2514/1.50596.

[78] Jenkins, S. C., and Geller, D. K., "State Estimation and Targeting for Autonomous Rendezvous and Proximity Operations," *Proceedings of the AAS/AIAA Astrodynamics Specialists Conference*, Mackinac Island, MI, 2007.

[79] Petersen, K. B., and Pedersen, M. S., "The Matrix Cookbook," , November 2012. URL https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf.

[80] Carpenter, J. R., and Bishop, R. H., "Flight Data Results of Estimate Fusion for Spacecraft Rendezvous Navigation from Shuttle Mission STS-69," Technical Report TM–111666, NASA Johnson Space Center, August 1996.

[81] Kriegsman, B. A., and Tao, Y.-C., "Shuttle Navigation System for Entry and Landing Mission Phases," *Journal of Spacecraft and Rockets*, Vol. 12, No. 4, 1975, pp. 213–219. https://doi.org/10.2514/3.56966.

[82] Wiesel, W. E., *Spaceflight Dynamics*, 2nd ed., McGraw-Hill Science Engineering, New York, NY, 1997.

[83] Fritsch, G. S., and DeMars, K. J., "Nonlinear Gaussian Mixture Filtering with Intrinsic Fault Resistance," *Journal of Guidance, Control, and Dynamics*, Vol. 44, No. 12, 2021, pp. 2172–2185. https://doi.org/10.2514/1.G005965.

[84] Vo, B.-N., Mallick, M., Bar-Shalom, Y., Coraluppi, S., Osborne III, R., Mahler, R. P. S., and Vo, B.-T., "Multitarget Tracking," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015, pp. 1–15. https://doi.org/10.1002/047134608X.W8275.

[85] Dennehy, C. J., and Carpenter, J. R., "A Summary of the Rendezvous, Proximity Operations, Docking, and Undocking (RPODU) Lessons Learned from the Defense Advanced Research Project Agency (DARPA) Orbital Express (OE) Demonstration System Mission," Technical Report TM–2011–217088, NASA Langley Research Center, April 2011.

[86] Mahler, R. P. S., "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39, No. 4, 2003, pp. 1152–1178. https://doi.org/10.1109/TAES.2003.1261119.

[87] Akca, A., and Efe, M. O., "Multiple Model Kalman and Particle Filters and Applications: A Survey," *IFAC-PapersOnLine*, Vol. 52, No. 3, 2019, pp. 73–78. https://doi.org/10.1016/j.ifacol.2019.06.013, 15th IFAC Symposium on Large Scale Complex Systems.

[88] Granström, K., Baum, M., and Reuter, S., "Extended Object Tracking: Introduction, Overview and Applications," *Journal of Advances in Information Fusion*, Vol. 12, No. 2, 2016, pp. 139–174. https://doi.org/10.48550/arXiv.1604.00970.

[89] Winternitz, L. M. B., Hassouneh, M. A., Mitchell, J. W., Valdez, J. E., Price, S. R., Semper, S. R., Wayne, H. Y., Ray, P. S., Wood, K. S., Arzoumanian, Z., et al., "X-Ray Pulsar Navigation Algorithms and Testbed for SEXTANT," *2015 IEEE Aerospace Conference*, IEEE, Big Sky, MT, 2015, pp. 1–14. https://doi.org/10.1109/AERO.2015.7118936.

[90] McCabe, J. S., and DeMars, K. J., "Terrain Relative Navigation With Anonymous Features," *AIAA SciTech 2019 Forum*, San Diego, CA, 2019. https://doi.org/10.2514/6.2019-0923.

[91] Hendeby, G., and Karlsson, R., "Gaussian Mixture PHD Filtering with Variable Probability of Detection," *17th International Conference on Information Fusion*, IEEE, 2014, pp. 1–7.

[92] Huntington, A. S., Williams, G. M., and Lee, A. O., "Modeling False Alarm Rate and Related Characteristics of Lidar Avalanche Photodiode Photoreceivers," Tech report, Allegro MicroSystems, Manchester, NH, 2021. URL https://www.allegromicro.com/en/insights-and-innovations/technical-documents/p0175-apdlidar-modeling-far.

[93] Varshney, P. K., *Distributed Detection and Data Fusion*, Springer Science & Business Media, New York, NY, 1997. https://doi.org/10.1007/978-1-4612-1904-0.

[94] Williams, G. M., "Relationship Between False-Alarm Rate and Probability of False Alarm," Tech report, Allegro MicroSystems, Manchester, NH, 2021. URL https://www.allegromicro.com/en/insights-and-innovations/technical-documents/p0176-far-pfa-relationship.

[95] Koch, W., *Tracking and Sensor Data Fusion*, Springer, Heidelberg, DE, 2014. https://doi.org/10.1007/978-3-642-39271-9.

[96] Williams, G. M., and Huntington, A., "Laser Rangefinder Effective Range," Tech report, Voxtel, Beaverton, OR, 2018.

[97] Mahler, R. P. S., *Statistical Multisource-Multitarget Information Fusion*, Artech House, Norwood, MA, 2007.

[98] Bar-Shalom, Y., Willet, P. K., and Tian, X., *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, 2011.

[99] Blackman, S. S., "Multiple Hypothesis Tracking for Multiple Target Tracking," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 19, No. 1, 2004, pp. 5–18. https://doi.org/10.1109/MAES.2004.1263228.

[100] Murty, K. G., "Letter to the Editor—An Algorithm for Ranking all the Assignments in Order of Increasing Cost," *Operations Research*, Vol. 16, No. 3, 1968, pp. 682–687. https://doi.org/10.1287/opre.16.3.682.

[101] Vo, B.-N., Vo, B.-T., and Phung, D., "Labeled Random Finite Sets and the Bayes Multi-Target Tracking Filter," *IEEE Transactions on Signal Processing*, Vol. 62, No. 24, 2014, pp. 6554–6567. https://doi.org/10.1109/TSP.2014.2364014.

[102] Lear, W. M., "RTCC Requirements for Mission G: MSFN Tracking Data Processor for Powered Flight Lunar Ascent/Descent Navigation," MSC Internal Memo No. 69-FM-36, NASA Manned Spacecraft Center, Houston, TX, 1969.

[103] DART Mishap Investigation Board, "Overview of the DART Mishap Investigation Results," Tech report, NASA. Available at http://www.nasa.gov/pdf/148072main_DART_mishap_overview.pdf, 2006.

[104] Doyle, R., Some, R., Powell, W., Mounce, G., Goforth, M., Horan, S., and Lowry, M., "High Performance Spaceflight Computing (HPSC) Next-Generation Space Processor (NGSP): a Joint Investment of NASA and AFRL," *Proceedings of the Workshop on Spacecraft Flight Software*, 2013.

[105] Powell, W. (ed.), *High-Performance Spaceflight Computing (HPSC) Program Overview*, Space Computing & Connected Enterprise Resiliency Conference, NASA Goddard Space Flight Center Electrical Engineering Division, Bedford, MA, 2018.

[106] Townes, S., Israel, D., Chandler, F., Neil Dennehy, C. J., D'Souza, C. N., Miranda, F., Podolski, D., Vitalpur, S., and Bryan, T., "TA 5: Communications, Navigation, and Orbital Debris Tracking and Characterization Systems," *2015 NASA Technology Roadmaps*, NASA, 2015.

[107] Lear, W. M., "Kalman Filtering Techniques," Tech Report JSC-20688, NASA Johnson Space Center, Houston, TX, 1985.

[108] Wang, J.-G., "Test Statistics in Kalman Filtering," *Journal of Global Positioning Systems*, Vol. 7, No. 1, 2008, pp. 81–90. https://doi.org/10.5081/jgps.7.1.81.

[109] Kelso, T. S., "Visually Observing Earth Satellites," *Satellite Times*, Vol. 3, No. 1, 1996, pp. 80–82.

[110] Shaddix, J., Brannum, J., Ferris, A., Hariri, A., Larson, A., Mancini, T., and Aristoff, J., "Daytime GEO Tracking with "Aquila": Approach and Results from a New Ground-Based SWIR Small Telescope System," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, 2019, p. 82.

[111] DeMars, K. J., and McCabe, J. S., "Multi-sensor Data Fusion in Non-Gaussian Orbit Determination," *AIAA/AAS Astrodynamics Specialist Conference*, San Diego, CA, 2014, p. 4310. https://doi.org/10.2514/6.2014-4310.

[112] Golub, G. H., and Van Loan, C. F., *Matrix Computations*, 4th ed., The Johns Hopkins University Press, Baltimore, MD, 2013.

[113] Cover, T. M., and Thomas, J. A., *Elements of Information Theory*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2006.

[114] Kataria, K. K., "A Probabilistic Proof of the Multinomial Theorem," *The American Mathematical Monthly*, Vol. 123, No. 1, 2016, pp. 94–96. https://doi.org/10.4169/amer.math.monthly.123.1.94.

[115] DeMars, K. J., and Darling, J. E., "Proof of Equation 32 in A Bayesian Approach to Problems in Stochastic Estimation and Control by Yu-Chi Ho and R. C. K. Lee," Technical Report AREUS 1501, Missouri S&T, 2015.

APPENDIX A

SELECT IDENTITIES

A.1 Gaussian Identities

As the Gaussian distribution is frequently encountered and invoked throughout this dissertation, this section defines a few important Gaussian identities.

A.1.1 The Guassian Distribution

Let $p_g(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P})$ represent a multivariate Gaussian distribution of $\boldsymbol{x}$ with mean $\boldsymbol{m}$ and covariance $\boldsymbol{P} = \boldsymbol{P}^T > 0$, such that

$$p_g(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P}) = \left|2\pi\boldsymbol{P}\right|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}) \right\}. \tag{A.1}$$

A.1.2 Ho's Rule for Nonlinear Measurement Models

Given $\boldsymbol{h}(\cdot)$, $\boldsymbol{R}$, $\boldsymbol{m}$, and $\boldsymbol{P}$ are of appropriate dimensions and $\boldsymbol{R}$ and $\boldsymbol{P}$ are symmetric, positive-definite [44, 76]

$$p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{x}), \boldsymbol{R}) p_g(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P}) = p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}), \boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m}) + \boldsymbol{R}) p_g(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Pi}), \tag{A.2a}$$

where

$$\boldsymbol{\mu} = \boldsymbol{m} + \boldsymbol{K}[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m})] \tag{A.2b}$$

$$\boldsymbol{\Pi} = \boldsymbol{P} - \boldsymbol{K}\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P} \tag{A.2c}$$

$$\boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m})[\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m}) + \boldsymbol{R}]^{-1}. \tag{A.2d}$$

Note that $\boldsymbol{H}(\boldsymbol{m})$ is the Jacobian of $\boldsymbol{h}(\boldsymbol{x})$ evaluated at $\boldsymbol{x} = \boldsymbol{m}$. The relationship in Eq. (A.2a) is found using a first-order Taylor series expansion of the nonlinear function $\boldsymbol{h}(\boldsymbol{x})$ about $\boldsymbol{m}$.

A.1.3 Generalized Ho's Rule

A more generalized version of Eqs. (A.2) can be shown to be

$$p_g(\boldsymbol{z}|\boldsymbol{H}(\boldsymbol{m})\boldsymbol{x}+\boldsymbol{b},\boldsymbol{R})p_g(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P}) = p_g(\boldsymbol{z}|\boldsymbol{H}(\boldsymbol{m})\boldsymbol{m}+\boldsymbol{b},\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m})+\boldsymbol{R})p_g(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Pi})\,, \tag{A.3}$$

where

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{m}+\boldsymbol{K}[\boldsymbol{z}-\boldsymbol{H}(\boldsymbol{m})\boldsymbol{m}-\boldsymbol{b}]\\ \boldsymbol{\Pi} &= \boldsymbol{P}-\boldsymbol{K}\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\\ \boldsymbol{K} &= \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m})[\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m})+\boldsymbol{R}]^{-1}\,,\end{aligned}$$

and where $\boldsymbol{b}$ must be deterministic.

A.1.4 Integral Form of Ho's Rule

Given $\boldsymbol{h}(\cdot)$, $\boldsymbol{R}$, $\boldsymbol{m}$, and $\boldsymbol{P}$ are of appropriate dimensions and $\boldsymbol{R}$ and $\boldsymbol{P}$ are symmetric, positive-definite, then

$$\int p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{R})p_g(\boldsymbol{x};\boldsymbol{m},\boldsymbol{P})\mathrm{d}\boldsymbol{x} = p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{m}),\boldsymbol{H}(\boldsymbol{m})\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{m})+\boldsymbol{R})\,, \tag{A.4a}$$

which is derived by integrating Eq. (A.2a) with respect to the variable $\boldsymbol{x}$, since $p_g(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Pi})$, is a pdf.

A.1.5 Powers of Gaussians

Given a vector $\boldsymbol{x}$ of n random variables that is distributed as an n-dimensional multivariate Gaussian of mean $\boldsymbol{m}$ and covariance $\boldsymbol{P}$, the a^{th} power of said Gaussian can be shown to be [111]

$$\left[p_g(\boldsymbol{x}|\boldsymbol{m},\,\boldsymbol{P})\right]^a = |a|^{-n/2}|2\pi\boldsymbol{P}|^{(1-a)/2}p_g(\boldsymbol{x}|\boldsymbol{m},\,{}^1\!/_a\boldsymbol{P})\,. \tag{A.5}$$

A.1.6 Product of N Multivariate Gaussians

Given N Gaussian pdfs of the form $p_g(\boldsymbol{x}|\boldsymbol{m}_i, \boldsymbol{P}_i)$, the product of N Gaussian pdfs yields

$$\prod_{i=1}^{N} p_g(\boldsymbol{x}|\boldsymbol{m}_i, \boldsymbol{P}_i) = \gamma p_g(\boldsymbol{x}|\tilde{\boldsymbol{m}}, \tilde{\boldsymbol{P}}) \,, \tag{A.6}$$

where

$$\begin{aligned}
\tilde{\boldsymbol{P}} &= \Big(\sum_{i=1}^{N} \boldsymbol{P}_i^{-1} \Big)^{-1} \\
\tilde{\boldsymbol{m}} &= \tilde{\boldsymbol{P}} \Big(\sum_{i=1}^{N} \boldsymbol{P}_i^{-1} \boldsymbol{m}_i \Big) \\
\gamma &= \frac{|2\pi\tilde{\boldsymbol{P}}|^{\frac{1}{2}}}{\prod_{i=1}^{N} |2\pi\boldsymbol{P}_i|^{\frac{1}{2}}} \prod_{i<j} \exp\Big\{ -\frac{1}{2} (\boldsymbol{m}_i - \boldsymbol{m}_j)^T \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{P}} \boldsymbol{P}_j^{-1} (\boldsymbol{m}_i - \boldsymbol{m}_j) \Big\} \,.
\end{aligned}$$

A.2 Linear Algebra Identities

Two identities useful for matrix manipulation are presented here.

A.2.1 Matrix Inversion Lemma

Given matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$, if $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible, then [112]

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1} \,. \tag{A.7}$$

A.2.2 Sylvester's Determinant Theorem

Given matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of sizes $n \times m$ and $m \times n$, respectively, [79]

$$\big|\boldsymbol{I}_n + \boldsymbol{BA}\big| = \big|\boldsymbol{I}_m + \boldsymbol{AB}\big| \,. \tag{A.8}$$

A.3 Optimization Identities

As the topic of optimization is discussed fairly often in this work, mostly when creating GVI updates, several identities useful to optimization are provided here.

A.3.1 Convexity of Functions

A function f is considered to be convex if [71]

$$f\big(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\big) \leq \theta f\big(\boldsymbol{x}\big) + (1-\theta) f\big(\boldsymbol{y}\big)\,, \tag{A.9}$$

where $\boldsymbol{x}, \boldsymbol{y} \in \text{dom}(f)$ and $0 \leq \theta \leq 1$. Furthermore, if

$$f\big(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\big) < \theta f\big(\boldsymbol{x}\big) + (1-\theta) f\big(\boldsymbol{y}\big)\,, \tag{A.10}$$

then the function f is considered to be strictly convex. Additionally, if

$$f\big(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\big) = \theta f\big(\boldsymbol{x}\big) + (1-\theta) f\big(\boldsymbol{y}\big)\,, \tag{A.11}$$

then the function f is considered to be affine.

A.3.2 Properties of Convex Functions

A.3.2.1 Weighted Sum of Convex Functions

Given a set of N convex functions denoted by f_i, the non-negative weighted sum of these functions is also convex, such that [71]

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} w_i f_i(\boldsymbol{x})\,, \tag{A.12}$$

is convex, where w_i are non-negative weights. Furthermore, it can be shown that the weighted sum of strictly convex functions is strictly convex. It is also of note that Eq. (A.12) applies to weighted sums of a single term, meaning that any (strictly) convex function multiplied by a non-negative

scalar is also (strictly) convex.

A.3.2.2 Sum of Convex and Strictly Convex Functions

Consider f to be a convex function defined by Eq. (A.9) and g to be a strictly convex function defined by Eq. (A.10). When subjected to the definition of convexity, it can be shown that

$$(f+g)\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big) = f\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big) + g\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big)$$

and

$$\theta(f+g)\big(\boldsymbol{x}\big) + (1-\theta)(f+g)\big(\boldsymbol{y}\big) = \theta f\big(\boldsymbol{x}\big) + (1-\theta) f\big(\boldsymbol{y}\big) + \theta g\big(\boldsymbol{x}\big) + (1-\theta) g\big(\boldsymbol{y}\big)\,.$$

Noting again that f is convex and g is strictly convex, it can be shown that

$$\begin{aligned} f\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big) + g\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big) &< \theta f\big(\boldsymbol{x}\big) + (1-\theta) f\big(\boldsymbol{y}\big) + \theta g\big(\boldsymbol{x}\big) + (1-\theta) g\big(\boldsymbol{y}\big) \\ (f+g)\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big) &< \theta(f+g)\big(\boldsymbol{x}\big) + (1-\theta)(f+g)\big(\boldsymbol{y}\big)\,, \end{aligned} \tag{A.13}$$

demonstrating that the sum of a convex and strictly convex function is strictly convex.

A.3.3 First Variation

The first variation of some functional $F(x)$ in the direction of y is defined as

$$\delta F(x,y) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big[F(x+\epsilon y)\Big]_{\epsilon=0}\,, \tag{A.14}$$

where x and y are functions and ϵ is a scalar.

A.3.4 Log Sum Inequality

Consider the following sums of N terms

$$a = \sum_{i=1}^{N} a_i \text{ and } b = \sum_{i=1}^{N} b_i\,,$$

where $a_i,\ b_i \geq 0 \forall i = 1,\ 2,\ \ldots,\ N$. The log sum inequality states that [113]

$$\sum_{i=1}^{N} a_i \log\left\{\frac{a_i}{b_i}\right\} \geq a \log\left\{\frac{a}{b}\right\}, \tag{A.15}$$

where the equality holds if and only if

$$\frac{a_i}{b_i} = \frac{a_j}{b_j} \qquad \forall i,\ j = 1,\ 2,\ \ldots,\ N\,.$$

A.4 Other Identities

A.4.1 Total Probability Theorem

If events B_i for $i = 1, \ldots, n$ are mutually exclusive and exhaustive such that

$$\Pr\{B_i, B_j\} = 0\,, \quad \forall i \neq j \tag{A.16a}$$

$$\sum_{i=1}^{n} \Pr\{B_i\} = 1\,, \tag{A.16b}$$

where $\Pr\{\cdot\}$ denotes probability. Then, for any event A,

$$\Pr\{A\} = \sum_{i=1}^{n} \Pr\{A|B_i\} \Pr\{B_i\}\,. \tag{A.17}$$

Similarly, for continuous random variables this can be represented as

$$p(x) = \int_{-\infty}^{\infty} p(x|y)p(y)\ \mathrm{d}y\,, \tag{A.18}$$

where $p(\cdot)$ denotes a probability density function. For situations with both random variables and events, this theorem produces

$$p(x) = \sum_{i=1}^{n} p(x|B_i) \Pr\{B_i\}\,. \tag{A.19}$$

A.4.2 Multinomial Theorem

For a positive integer k and some non-negative integer n, the multinomial theorem states that [114]

$$\Big(\sum_{i=1}^{k} x_i\Big)^n = \sum_{b_1+b_2+\cdots+b_k=n} \binom{n}{b_1, b_2, \ldots, b_k} \prod_{i=1}^{k} x_i^{b_i} \,, \tag{A.20}$$

where

$$\binom{n}{b_1, b_2, \ldots, b_k} = \frac{n!}{b_1! b_2! \ldots b_k!} \,.$$

Note that the number of terms on the right hand side of Eq. (A.20) is $\binom{n+k-1}{n}$.

APPENDIX B

SELECT DERIVATIONS

B.1 Minimum Mean Square Error Derivation

As a supplement to Section 1.1.3, the minimum mean square error (MMSE) estimate $\hat{\boldsymbol{x}}$ is derived by minimizing the cost function of Eq. (1.7) with the risk function defined in Eq. (1.8), which is given by

$$\hat{\boldsymbol{x}} = \min_{\tilde{\boldsymbol{x}}} \left\{ \mathbb{E}_{p(\boldsymbol{x})} \left[(\boldsymbol{x} - \tilde{\boldsymbol{x}})^T (\boldsymbol{x} - \tilde{\boldsymbol{x}}) \right] \right\}. \tag{B.1}$$

Making use of the linearity of the expectation operator, the cost function of Eq. (B.1) can be manipulated to yield

$$\begin{aligned} J &= \mathbb{E}_{p(\boldsymbol{x})} \left[\tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}} \right] - \mathbb{E}_{p(\boldsymbol{x})} \left[\tilde{\boldsymbol{x}}^T \boldsymbol{x} \right] - \mathbb{E}_{p(\boldsymbol{x})} \left[\boldsymbol{x}^T \tilde{\boldsymbol{x}} \right] + \mathbb{E}_{p(\boldsymbol{x})} \left[\boldsymbol{x}^T \boldsymbol{x} \right] \\ &= \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^T \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] - \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]^T \tilde{\boldsymbol{x}} + \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}^T \boldsymbol{x}] \,. \end{aligned} \tag{B.2}$$

To find possible $\hat{\boldsymbol{x}}$ that minimize Eq. (B.1), the partial derivative of Eq. (B.2) is taken with respect to $\tilde{\boldsymbol{x}}$ as

$$\begin{aligned} \frac{\partial J}{\partial \tilde{\boldsymbol{x}}} &= \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}} \right\} - \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ \tilde{\boldsymbol{x}}^T \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] \right\} - \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]^T \tilde{\boldsymbol{x}} \right\} + \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}^T \boldsymbol{x}] \right\} \\ &= 2 \big(\tilde{\boldsymbol{x}} - \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] \big)^T \,. \end{aligned} \tag{B.3}$$

When Eq. (B.3) is set equal to zero the solution for $\tilde{\boldsymbol{x}}$ is

$$\tilde{\boldsymbol{x}} = \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}] \,, \tag{B.4}$$

which is the mean of the random variable $\boldsymbol{x}$. At this point, Eq. (B.4) has only been shown to be an extremum of the cost function of Eq. (B.2) and has yet to be proven a minimum satisfying Eq. (B.1). To do this, a sufficient condition for minimality is convexity of the cost function of Eq. (B.2) at the extremum of Eq. (B.4), which can be accomplished by confirming that the second derivative of Eq. (B.2) is non-negative (or, in the case of vector calculus, a positive semi-definite matrix). Therefore, starting with Eq. (B.3), an additional derivative is taken as

$$\begin{aligned} \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ \frac{\partial J}{\partial \tilde{\boldsymbol{x}}} \right\}^T &= \frac{\partial}{\partial \tilde{\boldsymbol{x}}} \left\{ 2\big(\tilde{\boldsymbol{x}} - \mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]\big) \right\} \\ &= 2\boldsymbol{I} \;\; \forall\, \tilde{\boldsymbol{x}}\,, \end{aligned} \tag{B.5}$$

which indicates that Eq. (B.2) is not only convex with respect to $\tilde{\boldsymbol{x}}$, but strictly convex as well, as $2\boldsymbol{I}$ is positive definite [71]. Furthermore, since Eq. (B.5) has no dependency on $\tilde{\boldsymbol{x}}$ this strict convexity holds across all values of $\tilde{\boldsymbol{x}}$, such that Eq. (B.2) is, in fact, a strictly convex function of $\tilde{\boldsymbol{x}}$. Since any strictly convex cost function will take on a unique minimum, Eq. (B.4) is the unique, globally optimal MMSE estimate of $p(\boldsymbol{x})$.

B.2 Confidence-based Ho's Equation for a Linear-Gaussian Update

To supplement the confidence-based updates of Section 3.3.1, a more in-depth derivation of the confidence-based Ho's equation is presented here, and is similar to the approach taken in [115]. Note that an exceeding amount of detail is contained in this derivations as it is intended to be a step-by-step guide through Ho's rule. To start, it is assumed that $p^+(\boldsymbol{x}_{k-1})$ is a Gaussian distribution such that, after propagating according to linear-Gaussian dynamics, $p^-(\boldsymbol{x}_k)$ is also a Gaussian distribution of the form

$$p^-(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m},\, \boldsymbol{P})\,, \tag{B.6}$$

where the time index k is omitted for brevity, since the remainder of this update will exist within the scope of a single time step. Furthermore, assume that the likihood function $\ell(\boldsymbol{z}|\boldsymbol{x})$ exists in its

linear-Gaussian form of

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\ \boldsymbol{R})\,. \tag{B.7}$$

Now, taking the numerator of the confidence-based update from Eq. (3.38), as well as the expressions in Eqs. (B.6) and (B.7), it can be shown that

$$\begin{aligned}
p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{1-\phi}{\phi}} &= p_g(\boldsymbol{x}|\boldsymbol{m},\ \boldsymbol{P})p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{x},\ \boldsymbol{R})^{\frac{1-\phi}{\phi}} \\
&= \left|2\pi\boldsymbol{P}\right|^{-1/2}\exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m})\Big\} \\
&\quad\times\left[\left|2\pi\boldsymbol{R}\right|^{-1/2}\exp\Big\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})\Big\}\right]^{\frac{1-\phi}{\phi}} \\
&= \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m})\Big\} \\
&\quad\times\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\frac{1-\phi}{\phi}}\exp\Big\{-\frac{1}{2}\Big(\frac{1-\phi}{\phi}\Big)(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})\Big\} \\
&= \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\frac{1-\phi}{\phi}} \\
&\quad\times\exp\Big\{-\frac{1}{2}\Big[(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}) + \Big(\frac{1-\phi}{\phi}\Big)(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})\Big]\Big\}\,.
\end{aligned} \tag{B.8}$$

For brevity, the scalar α is defined to be

$$\alpha = \frac{1-\phi}{\phi}\,. \tag{B.9}$$

Taking the expression found inside $[\cdots]$ and expanding the variables yields

$$\begin{aligned}
&(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}) + \alpha(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^T\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x}) = \\
&\quad= \boldsymbol{x}^T\boldsymbol{P}^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{P}^{-1}\boldsymbol{m} - \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{x} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} \\
&\qquad+ \alpha[\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} - \boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}]
\end{aligned}$$

$$= \boldsymbol{x}^T(\alpha \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H} + \boldsymbol{P}^{-1})\boldsymbol{x} - 2\boldsymbol{x}^T(\alpha \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{m}) + \boldsymbol{m}^T \boldsymbol{P}^{-1}\boldsymbol{m} + \alpha \boldsymbol{z}^T \boldsymbol{R}^{-1}\boldsymbol{z}\,. \tag{B.10}$$

The argument of the quadratic term in $\boldsymbol{x}$ is defined as

$$\boldsymbol{\Pi}^{-1} \triangleq \alpha \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H} + \boldsymbol{P}^{-1}\,, \tag{B.11}$$

which can be solved for using the matrix inversion lemma of Eq. (A.7) to yield

$$\boldsymbol{\Pi} = \boldsymbol{P} - \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{H}\boldsymbol{P}\,. \tag{B.12}$$

If Eq. (B.11) is used in Eq. (B.10), then Eq. (B.8) can be reformed into

$$p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}\exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\Pi}(\alpha\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{m}) + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z}\Bigg]\Bigg\}\,. \tag{B.13}$$

Notice that an additional $\boldsymbol{\Pi}^{-1}\boldsymbol{\Pi}$ has been introduced. This is to support the use of $\boldsymbol{\mu}$ defined by

$$\boldsymbol{\mu} \triangleq \boldsymbol{\Pi}(\alpha\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{m})\,. \tag{B.14}$$

Substituting Eq. (B.14) into Eq. (B.13) yields

$$p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}\exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z}\Bigg]\Bigg\}\,. \tag{B.15}$$

Once again, taking the terms contained within $[\cdots]$ of Eq. (B.15), the following manipulations can be made:

$$\begin{aligned}
&\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} = \\
&\quad = \boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} \\
&\quad = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Pi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z}\,,
\end{aligned}$$

which, when substituted back into Eq. (B.15) yields

$$\begin{aligned}
p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}\exp\Bigg\{ -\frac{1}{2}\Bigg[&(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Pi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
&- \boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} + \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z}\Bigg]\Bigg\}\,.
\end{aligned} \tag{B.16}$$

At this point, new variables will need to be introduced in order to help manipulate the $\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}$ term. First, $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} \triangleq \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1}\,. \tag{B.17}$$

From this definition of $\boldsymbol{K}$, it is clear that $\boldsymbol{\Pi}$ in Eq. (B.12) can also be written as

$$\boldsymbol{\Pi} = \boldsymbol{P} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}\,. \tag{B.18}$$

Note that we can find an alternate definition of by taking Eq. (B.11) and modifying it as follows:

$$\begin{aligned}
\boldsymbol{\Pi}^{-1} &= \alpha\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \boldsymbol{P}^{-1} \\
\alpha\boldsymbol{\Pi}^{-1} &= \alpha^2\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \alpha\boldsymbol{P}^{-1} \\
\alpha\boldsymbol{\Pi}\boldsymbol{\Pi}^{-1}\boldsymbol{P} &= \alpha^2\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P} + \alpha\boldsymbol{\Pi}\boldsymbol{P}^{-1}\boldsymbol{P} \\
\alpha\boldsymbol{P} &= \alpha^2\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P} + \alpha\boldsymbol{\Pi}
\end{aligned}$$

$$
\begin{aligned}
\alpha \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1} &= \alpha^2\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1} + \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1} \\
\alpha \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1} &= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\alpha\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1} + \boldsymbol{I}) \\
\alpha \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1} &= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\alpha\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R})\boldsymbol{R}^{-1} \\
\alpha \boldsymbol{P}\boldsymbol{H}^T &= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\alpha\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R}) \\
\alpha \boldsymbol{P}\boldsymbol{H}^T(\alpha\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R})^{-1} &= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1} \\
\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1} &= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\,. \qquad \text{(B.19)}
\end{aligned}
$$

Recalling the definition from Eq. (B.17) and substituting into Eq. (B.19), it is clear that $\boldsymbol{K}$ can also be defined as

$$
\boldsymbol{K} = \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\,. \qquad \text{(B.20)}
$$

Later on, it will be useful to note that Eq. (B.20) can be rearranged to the form

$$
\boldsymbol{K}^T\boldsymbol{\Pi}^{-1} = \alpha\boldsymbol{R}^{-1}\boldsymbol{H}\,. \qquad \text{(B.21)}
$$

Consequently, we can use Eqs. (B.18) and (B.20) to substitute in for the expression of $\boldsymbol{\mu}$ in Eq. (B.14) to yield

$$
\begin{aligned}
\boldsymbol{\mu} &= \boldsymbol{\Pi}(\alpha\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{m}) \\
&= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{\Pi}\boldsymbol{P}^{-1}\boldsymbol{m} \\
&= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + (\boldsymbol{P} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P})\boldsymbol{P}^{-1}\boldsymbol{m} \\
&= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{m} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{m} \\
&= \alpha\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{m} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{m} \\
&= \boldsymbol{K}\boldsymbol{z} + \boldsymbol{m} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{m} \\
&= \boldsymbol{m} + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})\,. \qquad \text{(B.22)}
\end{aligned}
$$

Now, $\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}$ can be manipulated by introducing the definition of $\boldsymbol{\mu}$ from Eq. (B.22) as well as Eq. (B.21), producing

$$\begin{aligned}
\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} &= (\boldsymbol{m}^T + (\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})^T\boldsymbol{K}^T)\boldsymbol{\Pi}^{-1}(\boldsymbol{m} + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})) \\
&= \boldsymbol{m}^T\boldsymbol{\Pi}^{-1}\boldsymbol{m} + 2(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})^T\boldsymbol{K}^T\boldsymbol{\Pi}^{-1}\boldsymbol{m} + (\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})^T\boldsymbol{K}^T\boldsymbol{\Pi}^{-1}\boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}) \\
&= \boldsymbol{m}^T\boldsymbol{\Pi}^{-1}\boldsymbol{m} + 2(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})^T\alpha\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} + (\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m})^T\alpha\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}) \\
&= \boldsymbol{m}^T\boldsymbol{\Pi}^{-1}\boldsymbol{m} + 2\alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} \\
&\quad - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} + \alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}\boldsymbol{m}\,. \qquad \text{(B.23)}
\end{aligned}$$

Substituting Eq. (B.11) into Eq. (B.23) yields

$$\begin{aligned}
&\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} \\
&= \boldsymbol{m}^T(\alpha\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \boldsymbol{P}^{-1})\boldsymbol{m} + 2\alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} \\
&\quad - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} + \alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}\boldsymbol{m} \\
&= \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + 2\alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} - \alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} \\
&\quad - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} + \alpha(\boldsymbol{H}\boldsymbol{m})^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}\boldsymbol{m} \\
&= \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}\boldsymbol{z} - \alpha(\boldsymbol{H}\boldsymbol{m})^T[\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}]\boldsymbol{H}\boldsymbol{m} \\
&\quad - 2\alpha(\boldsymbol{H}\boldsymbol{m})^T[\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}]\boldsymbol{z}\,. \qquad \text{(B.24)}
\end{aligned}$$

The definition of $\boldsymbol{K}$ from Eq. (B.17) can be substituted into the term $\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K}$ in Eq. (B.24) as

$$\begin{aligned}
\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K} &= \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{K} \\
&= \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1} \\
&= \boldsymbol{R}^{-1}[\boldsymbol{I} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1}] \\
&= \boldsymbol{R}^{-1}[\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T](\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1}
\end{aligned}$$

$$= \alpha^{-1}\boldsymbol{R}^{-1} - 1\boldsymbol{R}(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}$$
$$= \alpha^{-1}(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\,. \tag{B.25}$$

Substituting the result of Eq. (B.25) into Eq. (B.24) results in a form of $\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}$ expressed as

$$\begin{aligned}
&\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} \\
&\quad = \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{HKz} - (\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{Hm} \\
&\qquad - 2(\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{z}\,.
\end{aligned} \tag{B.26}$$

Equation (B.25) can be manipulated to yield

$$\boldsymbol{R}^{-1}\boldsymbol{HK} = \boldsymbol{R}^{-1} - \alpha^{-1}(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\,, \tag{B.27}$$

which can be applied directly to Eq. (B.26) to produce

$$\begin{aligned}
&\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu} \\
&\quad = \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T[\boldsymbol{R}^{-1} - \alpha^{-1}(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}]\boldsymbol{z} \\
&\qquad - (\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{Hm} - 2(\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{z} \\
&\quad = \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} - \boldsymbol{z}^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{z} \\
&\qquad - (\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{Hm} - 2(\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}\boldsymbol{z} \\
&\quad = \boldsymbol{m}^T\boldsymbol{P}^{-1}\boldsymbol{m} + \alpha\boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z} - (\boldsymbol{z} - \boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}(\boldsymbol{z} - \boldsymbol{Hm})\,.
\end{aligned} \tag{B.28}$$

With Eq. (B.28), this form of $\boldsymbol{\mu}^T\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}$ is substituted back into Eq. (B.16) to yield

$$\begin{aligned}
p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}\exp\Bigg\{ \\
-\frac{1}{2}\Big[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Pi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + (\boldsymbol{z}-\boldsymbol{Hm})^T(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})^{-1}(\boldsymbol{z}-\boldsymbol{Hm})\Big]\Bigg\}\,.
\end{aligned} \tag{B.29}$$

The exponential in Eq. (B.29) can be separated into two distinct exponentials as

$$p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha} \tag{B.30}$$
$$\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Pi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})\right\}.$$

The exponentials in Eq. (B.30) can forcibly be made Gaussian by introducing normalizing constants, resulting in

$$\begin{aligned}
&p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha \\
&= \frac{\left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}}{\left|2\pi\boldsymbol{\Pi}\right|^{-\frac{1}{2}}\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|^{-\frac{1}{2}}}\left|2\pi\boldsymbol{\Pi}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Pi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\
&\quad\times\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{m})\right\} \\
&= \frac{\left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}}{\left|2\pi\boldsymbol{\Pi}\right|^{-\frac{1}{2}}\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|^{-\frac{1}{2}}}p_g(\boldsymbol{x}|\boldsymbol{\mu},\,\boldsymbol{\Pi})p_g(\boldsymbol{z}|\boldsymbol{H}\boldsymbol{m},\,\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R}).
\end{aligned} \tag{B.31}$$

The leading term in Eq. (B.31) can be simplified as

$$\begin{aligned}
\frac{\left|2\pi\boldsymbol{P}\right|^{-\frac{1}{2}}\left|2\pi\boldsymbol{R}\right|^{-\frac{1}{2}\alpha}}{\left|2\pi\boldsymbol{\Pi}\right|^{-\frac{1}{2}}\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|^{-\frac{1}{2}}} &= \left[\frac{\left|2\pi\boldsymbol{\Pi}\right|\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|}{\left|2\pi\boldsymbol{P}\right|\left|2\pi\boldsymbol{R}\right|^{\alpha}}\right]^{\frac{1}{2}} \\
&= \left[\frac{(\alpha^{-1})^m\left|2\pi\boldsymbol{\Pi}\right|\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|}{(\alpha^{-1})^m\left|2\pi\boldsymbol{P}\right|\left|2\pi\boldsymbol{R}\right|^{\alpha-1}\left|2\pi\boldsymbol{R}\right|}\right]^{\frac{1}{2}} \\
&= \left|2\pi\boldsymbol{R}\right|^{\frac{1-\alpha}{2}}\left[\frac{\alpha^{-m}\left|2\pi\boldsymbol{\Pi}\right|\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|}{\left|2\pi\boldsymbol{P}\right|\left|2\pi\alpha^{-1}\boldsymbol{R}\right|}\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}}\left|2\pi\boldsymbol{R}\right|^{\frac{1-\alpha}{2}}\left[\frac{\left|2\pi\boldsymbol{\Pi}\right|\left|2\pi(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|}{\left|2\pi\boldsymbol{P}\right|\left|2\pi\alpha^{-1}\boldsymbol{R}\right|}\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}}\left|2\pi\boldsymbol{R}\right|^{\frac{1-\alpha}{2}}\left[\frac{\left|\boldsymbol{\Pi}\right|\left|(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\right|}{\left|\boldsymbol{P}\right|\left|\alpha^{-1}\boldsymbol{R}\right|}\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}}\left|2\pi\boldsymbol{R}\right|^{\frac{1-\alpha}{2}}\left[\left|\boldsymbol{\Pi}\boldsymbol{P}^{-1}\right|\left|(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\alpha^{-1}\boldsymbol{R})\alpha\boldsymbol{R}^{-1}\right|\right]^{\frac{1}{2}}
\end{aligned}$$

$$= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|\alpha \boldsymbol{HPH}^T \boldsymbol{R}^{-1} + \boldsymbol{I}\right|\right]^{\frac{1}{2}}, \tag{B.32}$$

where m is defined by the size of matrix $\boldsymbol{R}$ such that $\boldsymbol{R}$ is an $m \times m$ matrix. Now, using Sylvester's theorem from Eq. (A.8) and noting that Eq. (B.11) can be rearranged to

$$\alpha \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H} = \boldsymbol{\Pi}^{-1} - \boldsymbol{P}^{-1}\,, \tag{B.33}$$

Eq. (B.32) can be manipulated to yield

$$\begin{aligned}
\frac{\left|2\pi \boldsymbol{P}\right|^{-\frac{1}{2}} \left|2\pi \boldsymbol{R}\right|^{-\frac{1}{2}\alpha}}{\left|2\pi \boldsymbol{\Pi}\right|^{-\frac{1}{2}} \left|2\pi(\boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})\right|^{-\frac{1}{2}}} &= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|\alpha \boldsymbol{HPH}^T \boldsymbol{R}^{-1} + \boldsymbol{I}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|\alpha \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{HP} + \boldsymbol{I}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|(\boldsymbol{\Pi}^{-1} - \boldsymbol{P}^{-1})\boldsymbol{P} + \boldsymbol{I}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|\boldsymbol{\Pi}^{-1}\boldsymbol{P} - \boldsymbol{P}^{-1}\boldsymbol{P} + \boldsymbol{I}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\right| \left|\boldsymbol{\Pi}^{-1}\boldsymbol{P}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{P}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{\Pi P}^{-1}\boldsymbol{P}\boldsymbol{\Pi}^{-1}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \left[\left|\boldsymbol{I}\right|\right]^{\frac{1}{2}} \\
&= \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}}\,.
\end{aligned} \tag{B.34}$$

Substituting Eq. (B.34) into Eq. (B.31) produces

$$p^-(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha = \alpha^{-\frac{m}{2}} \left|2\pi \boldsymbol{R}\right|^{\frac{1-\alpha}{2}} \times p_g(\boldsymbol{x}|\boldsymbol{\mu},\, \boldsymbol{\Pi}) p_g(\boldsymbol{z}|\boldsymbol{Hm},\, \boldsymbol{HPH}^T + \alpha^{-1}\boldsymbol{R})\,. \tag{B.35}$$

Recalling the definition of α from Eq. (B.9), Eq. (B.35) assumes the form

$$p^{-}(\boldsymbol{x})\ell(\boldsymbol{z}|\boldsymbol{x})^{\frac{1-\phi}{\phi}} = \left[\frac{\phi}{1-\phi}\right]^{\frac{m}{2}} \left|2\pi\boldsymbol{R}\right|^{\frac{2\phi-1}{2\phi}} \times p_g\Big(\boldsymbol{x}\Big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\Big) p_g\left(\boldsymbol{z}\middle|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\right), \tag{B.36a}$$

where

$$\boldsymbol{\mu} = \boldsymbol{m} + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\boldsymbol{m}) \tag{B.36b}$$

$$\boldsymbol{\Pi} = \boldsymbol{P} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P} \tag{B.36c}$$

$$\boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \alpha^{-1}\boldsymbol{R})^{-1}. \tag{B.36d}$$

If the identity in Eq. (B.36a) is used in the update of Eq. (3.38), then

$$\begin{aligned} p^{+}(\boldsymbol{x}) &= \frac{\left[\frac{\phi}{1-\phi}\right]^{\frac{m}{2}} \left|2\pi\boldsymbol{R}\right|^{\frac{2\phi-1}{2\phi}} \times p_g\big(\boldsymbol{x}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big) p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)}{\int \left[\frac{\phi}{1-\phi}\right]^{\frac{m}{2}} \left|2\pi\boldsymbol{R}\right|^{\frac{2\phi-1}{2\phi}} \times p_g\big(\boldsymbol{s}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big) p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)\mathrm{d}\boldsymbol{s}} \\ &= \frac{p_g\big(\boldsymbol{x}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big) p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)}{\int p_g\big(\boldsymbol{s}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big) p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)\mathrm{d}\boldsymbol{s}}. \end{aligned} \tag{B.37}$$

Considering only the denominator of Eq. (B.37), it can be shown that

$$\begin{aligned} &\int p_g\big(\boldsymbol{s}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big) p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)\mathrm{d}\boldsymbol{s} \\ &\qquad = p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big) \int p_g\big(\boldsymbol{s}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big)\mathrm{d}\boldsymbol{s}, \end{aligned} \tag{B.38}$$

which is valid as the terms taken outside of the integral have no dependence on $\boldsymbol{s}$. Furthermore, since it is assumed that $p_g(\cdot|\cdot,\cdot)$ is a valid pdf, it is known that integrating over the support will result in unity, and therefore Eq. (B.38) can be further reduced to

$$p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big) \int p_g\big(\boldsymbol{s}\big|\boldsymbol{\mu},\, \boldsymbol{\Pi}\big)\mathrm{d}\boldsymbol{s} = p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\, \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big). \tag{B.39}$$

Using the result of Eq. (B.39) to replace the denominator of Eq. (B.37), it can be shown that

$$\begin{aligned} p^{+}(\boldsymbol{x}) &= \frac{p_g\big(\boldsymbol{x}\big|\boldsymbol{\mu},\ \boldsymbol{\Pi}\big)p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\ \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)}{p_g\big(\boldsymbol{z}\big|\boldsymbol{H}\boldsymbol{m},\ \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \frac{\phi}{1-\phi}\boldsymbol{R}\big)} \\ &= p_g\big(\boldsymbol{x}\big|\boldsymbol{\mu},\ \boldsymbol{\Pi}\big)\,. \end{aligned} \tag{B.40}$$

B.2.1 Extension for a Nonlinear System

Now, consider systems with measurement models that are nonlinear functions of $\boldsymbol{x}$. Under these types of systems, the Gaussian likelihood equivalent of the measurement model from Eq. (2.1c) is

$$\ell(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{x}),\ \boldsymbol{R})\,, \tag{B.41}$$

which has been restated here without time dependencies. As it is common practice in filtering to linearize such models using first-order Taylor series approximations (as like the EKF), such an approach is made here. Again, assuming that an *a priori* estimate of the state exists as

$$p^{-}(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^{-},\ \boldsymbol{P}^{-})\,,$$

Equation (B.41) can be linearized via a first-order Taylor series about the mean $\boldsymbol{m}^{-}$ to produce

$$\begin{aligned} \ell(\boldsymbol{z}|\boldsymbol{x}) &= p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^{-}) + \boldsymbol{H}(\boldsymbol{m}^{-})[\boldsymbol{x} - \boldsymbol{m}^{-}],\ \boldsymbol{R}) \\ &= p_g(\boldsymbol{z}|\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{x} + [\boldsymbol{h}(\boldsymbol{m}^{-}) - \boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{m}^{-}],\ \boldsymbol{R})\,. \end{aligned} \tag{B.42}$$

Now, considering the numerator of the confidence-based update in Eq. (3.38) to form

$$\ell(\boldsymbol{z}|\boldsymbol{x})^{\alpha}p^{-}(\boldsymbol{x}) = p_g(\boldsymbol{z}|\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{x} + [\boldsymbol{h}(\boldsymbol{m}^{-}) - \boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{m}^{-}], \boldsymbol{R})^{\alpha}p_g(\boldsymbol{x}|\boldsymbol{m}^{-}, \boldsymbol{P}^{-})\,.$$

Utilizing the power of a Gaussian identity from Eq. (A.5) yields

$$\begin{aligned}&\ell(\boldsymbol{z}|\boldsymbol{x})^{\alpha}p^{-}(\boldsymbol{x}) \\ &\quad= |\alpha|^{-\frac{n}{2}}|2\pi\boldsymbol{R}|^{\frac{1-\alpha}{2}}p_g\Big(\boldsymbol{z}|\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{x}+[\boldsymbol{h}(\boldsymbol{m}^{-})-\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{m}^{-}],\frac{1}{\alpha}\boldsymbol{R}\Big)p_g(\boldsymbol{x}|\boldsymbol{m}^{-},\boldsymbol{P}^{-})\,.\end{aligned} \tag{B.43}$$

Instead of rederiving Ho's generalized equation here, it is provided in the appendix as Eq. (A.3), such that the expression in Eq. (B.43) becomes

$$\ell(\boldsymbol{z}|\boldsymbol{x})^{\alpha}p^{-}(\boldsymbol{x}) = |\alpha|^{-\frac{n}{2}}|2\pi\boldsymbol{R}|^{\frac{1-\alpha}{2}}p_g\Big(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^{-}),\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})+\frac{1}{\alpha}\boldsymbol{R}\Big)p_g(\boldsymbol{x}|\boldsymbol{m}^{+},\boldsymbol{P}^{+})\,, \tag{B.44}$$

where

$$\begin{aligned}\boldsymbol{m}^{+} &= \boldsymbol{m}^{-}+\boldsymbol{K}[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{m}^{-})] \\ \boldsymbol{P}^{+} &= \boldsymbol{P}^{-}-\boldsymbol{K}\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-} \\ \boldsymbol{K} &= \boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})\Big[\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})+\frac{1}{\alpha}\boldsymbol{R}\Big]^{-1}\,.\end{aligned}$$

The denominator of the confidence-based update from Eq. (3.38) can be solved for by integrating Eq. (B.44) as

$$\begin{aligned}&\int\ell(\boldsymbol{z}|\boldsymbol{s})^{\alpha}p^{-}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &\quad= \int|\alpha|^{-\frac{n}{2}}|2\pi\boldsymbol{R}|^{\frac{1-\alpha}{2}}p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^{-}),\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})+\frac{1}{\alpha}\boldsymbol{R})p_g(\boldsymbol{x}|\boldsymbol{m}^{+},\boldsymbol{P}^{+})\mathrm{d}\boldsymbol{s} \\ &\quad= |\alpha|^{-\frac{n}{2}}|2\pi\boldsymbol{R}|^{\frac{1-\alpha}{2}}p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^{-}),\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})+\frac{1}{\alpha}\boldsymbol{R})\int p_g(\boldsymbol{x}|\boldsymbol{m}^{+},\boldsymbol{P}^{+})\mathrm{d}\boldsymbol{s} \\ &\quad= |\alpha|^{-\frac{n}{2}}|2\pi\boldsymbol{R}|^{\frac{1-\alpha}{2}}p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^{-}),\boldsymbol{H}(\boldsymbol{m}^{-})\boldsymbol{P}^{-}\boldsymbol{H}^{T}(\boldsymbol{m}^{-})+\frac{1}{\alpha}\boldsymbol{R})\,,\end{aligned} \tag{B.45}$$

where it is recalled that a Gaussian must integrate to unity. Next, substituting the results from Eqs. (B.44) and (B.45) into the update of Eq. (3.38) yields

$$\begin{aligned} p^+(\boldsymbol{x}) &= \frac{\ell(\boldsymbol{z}|\boldsymbol{x})^\alpha p^-(\boldsymbol{x})}{\int \ell(\boldsymbol{z}|\boldsymbol{s})^\alpha p^-(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}} \\ &= \frac{|\alpha|^{-\frac{n}{2}}\det\{2\pi\boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^-), \boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-) + \frac{1}{\alpha}\boldsymbol{R}) p_g(\boldsymbol{x}|\boldsymbol{m}^+, \boldsymbol{P}^+)}{|\alpha|^{-\frac{n}{2}}\det\{2\pi\boldsymbol{R}\}^{\frac{1-\alpha}{2}} p_g(\boldsymbol{z}|\boldsymbol{h}(\boldsymbol{m}^-), \boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-) + \frac{1}{\alpha}\boldsymbol{R})} \\ &= p_g(\boldsymbol{x}|\boldsymbol{m}^+, \boldsymbol{P}^+)\,, \end{aligned}$$

such that the complete equations for the confidence-based update for a nonlinear Gaussian system are

$$p^+(\boldsymbol{x}) = p_g(\boldsymbol{x}|\boldsymbol{m}^+, \boldsymbol{P}^+)\,, \tag{B.46}$$

where

$$\begin{aligned} \boldsymbol{m}^+ &= \boldsymbol{m}^- + \boldsymbol{K}[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m}^-)] \\ \boldsymbol{P}^+ &= \boldsymbol{P}^- - \boldsymbol{K}\boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^- \\ \boldsymbol{K} &= \boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-)\Big[\boldsymbol{H}(\boldsymbol{m}^-)\boldsymbol{P}^-\boldsymbol{H}^T(\boldsymbol{m}^-) + \frac{1}{\alpha}\boldsymbol{R}\Big]^{-1}. \end{aligned}$$

Due to the linearization via a first-order Taylor series, it is noted that this recursion only achieves approximate closure, as a Gaussian prior is only approximated to remain Gaussian under the update for nonlinear systems.

B.3 Gaussian KLD

Assuming that $p^-(\boldsymbol{x})$ and $p(\boldsymbol{x})$ are Gaussian distributions of the form

$$\begin{aligned} p^-(\boldsymbol{x}) &= p_g(\boldsymbol{x}|\boldsymbol{m}^-,\, \boldsymbol{P}^-) \\ p(\boldsymbol{x}) &= p_g(\boldsymbol{x}|\boldsymbol{m},\, \boldsymbol{P})\,, \end{aligned}$$

it can be shown that the KLD becomes [74]

$$\begin{aligned}\mathcal{D}_{KL}[p\|p^-] &= \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{p^-(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} \\ &= \mathbb{E}_p\big\{\ln[p(\boldsymbol{x})] - \ln[p^-(\boldsymbol{x})]\big\} \\ &= \frac{1}{2}\Bigg[\ln\left(\frac{|\boldsymbol{P}^-|}{|\boldsymbol{P}|}\right) - \mathbb{E}_p\big\{(\boldsymbol{x}-\boldsymbol{m})^T(\boldsymbol{P})^{-1}(\boldsymbol{x}-\boldsymbol{m})\big\} \\ &\qquad + \mathbb{E}_p\big\{(\boldsymbol{x}-\boldsymbol{m}^-)^T(\boldsymbol{P}^-)^{-1}(\boldsymbol{x}-\boldsymbol{m}^-)\big\}\Bigg], \end{aligned} \tag{B.47}$$

where using properties of traces, as found in [79], it can be shown that

$$\begin{aligned}\mathbb{E}_p\big\{(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m})\big\} &= \mathbb{E}_p\big\{\mathrm{tr}\big\{(\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{P}^{-1}\big\}\big\} \\ &= \mathrm{tr}\big\{\mathbb{E}_p\big[(\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m})^T\big]\boldsymbol{P}^{-1}\big\} \\ &= \mathrm{tr}\big\{\boldsymbol{P}\boldsymbol{P}^{-1}\big\} \\ &= n_{\boldsymbol{x}}\,, \end{aligned} \tag{B.48}$$

where $n_{\boldsymbol{x}}$ is the size of the state vector $\boldsymbol{x}$. Similarly, using properties from [79], it can be shown that

$$\mathbb{E}_p\big\{(\boldsymbol{x}-\boldsymbol{m}^-)^T(\boldsymbol{P}^-)^{-1}(\boldsymbol{x}-\boldsymbol{m}^-)\big\} = \mathrm{tr}\big\{(\boldsymbol{P}^-)^{-1}\boldsymbol{P}\big\} + \big\{(\boldsymbol{m}^- - \boldsymbol{m})^T(\boldsymbol{P}^-)^{-1}(\boldsymbol{m}^- - \boldsymbol{m})\big\}\,. \tag{B.49}$$

Substituting the results of Eqs. (B.48) and (B.49) into Eq. (B.47), the KLD under Gaussian assumptions is shown to be

$$\mathcal{D}_{KL}[p\|p^-] = \frac{1}{2}\left[\ln\left(\frac{|\boldsymbol{P}^-|}{|\boldsymbol{P}|}\right) - n_{\boldsymbol{x}} + \mathrm{tr}\big\{(\boldsymbol{P}^-)^{-1}\boldsymbol{P}\big\} + (\boldsymbol{m}^- - \boldsymbol{m})^T(\boldsymbol{P}^-)^{-1}(\boldsymbol{m}^- - \boldsymbol{m})\right]. \tag{B.50}$$

B.4 Convexity of Kullback-Leibler Divergence

The Kullback-Leibler divergence (KLD) has the form

$$\mathcal{D}_{KL}[p\|q] = \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}\,, \tag{B.51}$$

where $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are probability density functions (pdfs). It is well known that the KLD is a convex function with respect to both densities p and q; however, when q is fixed, it can be shown that the KLD becomes strictly convex. Considering the definition of convexity in Eq. (A.9), it can be shown that

$$\mathcal{D}_{KL}[\theta p + (1-\theta)\pi\|q] = \int \big(\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})\big) \ln\left(\frac{\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \tag{B.52}$$

and also that

$$\begin{aligned}
&\theta\mathcal{D}_{KL}[p\|q] + (1-\theta)\mathcal{D}_{KL}[\pi\|q] \\
&\qquad = \theta \int p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} + (1-\theta) \int \pi(\boldsymbol{x}) \ln\left(\frac{\pi(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \\
&\qquad = \int \left\{ \theta p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) + (1-\theta)\pi(\boldsymbol{x}) \ln\left(\frac{\pi(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \right\} \mathrm{d}\boldsymbol{x} \\
&\qquad = \int \left\{ \theta p(\boldsymbol{x}) \ln\left(\frac{\theta p(\boldsymbol{x})}{\theta q(\boldsymbol{x})}\right) + (1-\theta)\pi(\boldsymbol{x}) \ln\left(\frac{(1-\theta)\pi(\boldsymbol{x})}{(1-\theta) q(\boldsymbol{x})}\right) \right\} \mathrm{d}\boldsymbol{x}\,.
\end{aligned} \tag{B.53}$$

Relating the expressions inside of the integrals of Eqs. (B.52) and (B.53) yields

$$\begin{aligned}
&\big(\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})\big) \ln\left(\frac{\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \\
&\qquad\qquad < \theta p(\boldsymbol{x}) \ln\left(\frac{\theta p(\boldsymbol{x})}{\theta q(\boldsymbol{x})}\right) + (1-\theta)\pi(\boldsymbol{x}) \ln\left(\frac{(1-\theta)\pi(\boldsymbol{x})}{(1-\theta) q(\boldsymbol{x})}\right),
\end{aligned} \tag{B.54}$$

where the inequality follows directly from the log sum inequality identity of Eq. (A.15). Note that the inequality of Eq. (B.54) is strictly less than as

$$\frac{\theta p(\boldsymbol{x})}{\theta q(\boldsymbol{x})} \neq \frac{(1-\theta)\pi(\boldsymbol{x})}{(1-\theta)q(\boldsymbol{x})} \,.$$

Therefore, integrating Eq. (B.54) results in

$$\begin{aligned}\int \big(\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})\big) \ln\left(\frac{\theta p(\boldsymbol{x}) + (1-\theta)\pi(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \\ < \int \left\{\theta p(\boldsymbol{x}) \ln\left(\frac{\theta p(\boldsymbol{x})}{\theta q(\boldsymbol{x})}\right) + (1-\theta)\pi(\boldsymbol{x}) \ln\left(\frac{(1-\theta)\pi(\boldsymbol{x})}{(1-\theta)q(\boldsymbol{x})}\right)\right\} \mathrm{d}\boldsymbol{x} \,,\end{aligned}$$

or, equivalently,

$$\mathcal{D}_{KL}[\theta p + (1-\theta)\pi || q] < \theta \mathcal{D}_{KL}[p||q] + (1-\theta)\mathcal{D}_{KL}[\pi||q] \,, \tag{B.55}$$

such that by the definition of Eq. (A.10), the Kullback-Leibler divergence of $p(\boldsymbol{x})$ with $q(\boldsymbol{x})$ fixed is strictly convex.

B.5 Convexity of Expected Loss Function

Consider the expectation of the loss function given by

$$\mathbb{E}_p\big[\mathcal{L}(\boldsymbol{z}, \boldsymbol{x})\big] = \int_{\boldsymbol{x}\in\mathbb{X}} p(\boldsymbol{x})\mathcal{L}(\boldsymbol{z}, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \,, \tag{B.56}$$

where $p(\boldsymbol{x})$ is a probability density function of $\boldsymbol{x}$ over the domain $\mathbb{X}$. Given the definition of convexity from Eq. (A.9), it can be shown that

$$\begin{aligned}\mathbb{E}_{\left[\theta p + (1-\theta) g\right]}\big[\mathcal{L}(\boldsymbol{z}, \boldsymbol{x})\big] &= \int_{\boldsymbol{x}\in\mathbb{X}} \big[\theta p(\boldsymbol{x}) + (1-\theta) g(\boldsymbol{x})\big] \mathcal{L}(\boldsymbol{z}, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \theta \int_{\boldsymbol{x}\in\mathbb{X}} p(\boldsymbol{x})\mathcal{L}(\boldsymbol{z}, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} + (1-\theta) \int_{\boldsymbol{x}\in\mathbb{X}} g(\boldsymbol{x})\mathcal{L}(\boldsymbol{z}, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \theta \mathbb{E}_p\big[\mathcal{L}(\boldsymbol{z}, \boldsymbol{x})\big] + (1-\theta)\mathbb{E}_g\big[\mathcal{L}(\boldsymbol{z}, \boldsymbol{x}))\big] \,,\end{aligned}$$

which indicates that Eq. (B.56) is affine (as well as convex) with respect to $p(\boldsymbol{x})$.