On A Higher Order Method for Anonymous Feature Processing

James S. McCabe*
NASA Johnson Space Center, Houston, TX 77058

Some feature-driven navigation sources, such as cameras or lidars, often require measurement-to-feature associations between the collected data and an onboard feature catalog to be performed upstream of the filter. Standard navigation practice suggests the use of Kalman updates with measurements that have first passed residual editing tests, but this is often insufficient to prevent updates based upon incorrectly associated data, leading to filter degradation and divergence. Recent work has developed the anonymous feature processing (AFP) technique that eliminates reliance upon explicit feature associations outside of the filter entirely while maintaining desirable estimation performance. This paper continues by exploring the approximation employed by AFP, and a higher order approximation is presented to further improve the estimation performance of the AFP update.

I. Introduction

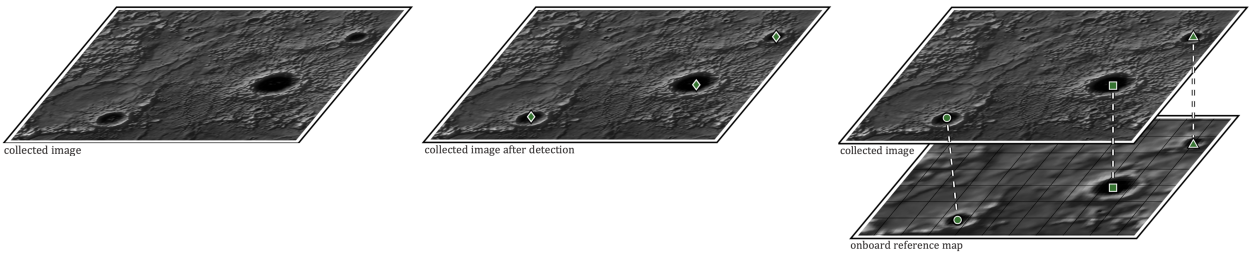
Certain sensing suites utilized in autonomous spacecraft navigation systems, such as cameras and/or lidars, present certain challenges when employed in a feature-based navigation system. Key among these challenges is the requirement that, if provided to an extended Kalman filter (EKF), measurements must be associated explicitly to features within a catalog stored onboard the spacecraft [1]. This measurement-to-feature association, herein referred to as "identification," inevitably contains errors, meaning that incorrectly identified measurements are passed to the EKF for processing. It is hoped that standard data editing techniques, such as residual editing [2], are sufficient to insulate the filter solution from these errors. However, it has been observed that, even with such editing techniques, the EKF often diverges when presented with identification errors [3].

An example pertaining to lunar terrain relative navigation (TRN) is depicted in Fig. 1. First, an image is collected, and this image is subjected to a *detection* procedure. Here, detection refers to the process by which certain coordinates in an image are located based on desired characteristics, and those coordinates are marked/saved. In this case, (b) shows that three craters have been detected, but their markers are identical because no effort has been expended to differentiate between these detections (measurements) to any specific crater on the Moon (identity) – all that has been asserted is that they are, indeed, craters. The third step is to subject these detection to the aforementioned identification process, essentially associating the detections to unique features within the onboard catalog (depicted in Fig. 1(c) as a coarse render below the image). Now, the identified craters have unique identifiers (here, square, circle, and triangle) that can be used to accomplish an EKF update. As mentioned in the previous paragraph, however, these identification procedures are bound to make identification errors at some point during flight, and the EKF may ingest misidentified measurements, resulting in filter divergence and potential loss of vehicle or mission.

To combat this, [3] derived a new measurement update that, rather than relying upon error-prone identification pre-processing, produces a navigation solution that requires only detection (and not identification) be performed (i.e., it can process what is depicted in Fig. 1(b) without needing the identification in Fig. 1(c) to be performed at all). Despite not requiring identification pre-processing, this new update, termed the anonymous feature processing (AFP) update, produces approximately the same estimation performance of the EKF [3] and can be implemented at approximately the computational complexity as the EKF in full covariance, square-root, and UDU-factorized forms [4]. The AFP update is not without its own non-ideal characteristics, however. It was observed that the AFP update self-assesses conservatively, i.e., its estimate of its own error covariance is larger than it should be. Despite establishing approximately the same *actual* performance of the EKF, the AFP update's *estimated* covariance is larger, leading to a conservative filter solution. Further, it was observed that, on a measurement-by-measurement basis, the AFP update converges somewhat slower than the EKF. These trends are depicted in terms of the filters' estimated 3σ level in Fig. 2.

Regardless, the AFP update provides desirable navigation performance without demonstrating any sensitivity whatsoever to identification pre-processing. This has some important implications for navigation system design in that, at minimum, an EKF-driven feature-based navigation application can have its robustness greatly improved by,

*GN&C Autonomous Flight Systems Engineer, EG6, Aeroscience and Flight Mechanics Division, NASA Johnson Space Center, 2101 E NASA Parkway, Houston, TX 77058



(a) Image from camera. **(b) Detected features.** **(c) Identification pre-processing.**

Fig. 1 Depiction of a standard image processing pipeline for feature-based terrain relative navigation. The collected image is depicted in (a), and a detector finds relevant features in (b). Then, (c) depicts an identification pre-processor that performs detection-to-map association between the features in (b) and some onboard map/catalog. Navigation problems induced by errors in (c), identification, are the focus of this paper.

instead, adopting the AFP update. Further, eliminating the need of identification pre-processing, which is often a very computationally burdensome procedure, provides the potential for eliminating the identification software from the system entirely, potentially freeing up a large deal of computational resources to the remainder of the vehicle's flight systems.

However, the AFP update also possesses some limitations that this present paper seeks to address. While the robustness provided by removing upstream identification entirely from the navigation solution has its advantages, as is depicted in Fig. 2, a delayed convergence and conservative state estimate is produced (when compared to the EKF). This paper explores a new, higher order AFP update that, rather than discarding feature identity and identification entirely, exploits identification information when it is available and reliable but also demonstrates insensitivity to errors in said identification. The solution that is sought is the "best of both worlds" between the EKF and AFP measurement updates. This is termed in this paper as the "higher order AFP update."

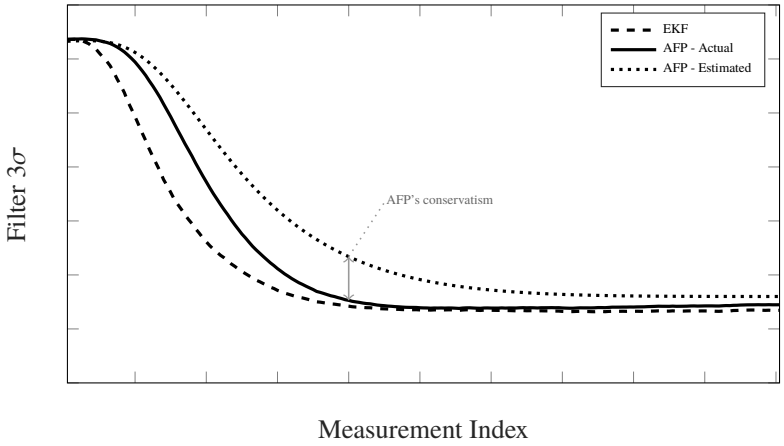


Fig. 2 Illustration of AFP's conservative self-assessment despite converging (approximately) to the solution provided by the EKF.

This paper is organized as follows. Section II develops the mathematical preliminaries for the derivations within this paper. Section III provides an improved derivation of the AFP update (compared to [3]), and Section IV leverages its intermediate results to produce a new update. Section V assesses performance of the new, higher order AFP update in a lunar TRN simulation, and, finally, Section VI provides concluding remarks.

II. Problem Formulation

Let the vehicle state be $\boldsymbol{x}_k$, where the state contains any quantities of interest, such as the vehicle's position, velocity, attitude, and estimable parameters (such as inertial measurement unit biases/misalignments/scale factor errors, sensor biases, or uncertain dynamical parameters). This paper will assume that the *a priori* vehicle state is distributed according to the Gaussian

$$\boldsymbol{x}_k \sim p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^-_{x,k}, \boldsymbol{P}^-_{xx,k}),$$

where $p_g(\boldsymbol{x}\,;\,\boldsymbol{a},\boldsymbol{A})$ denotes a Gaussian probability density in $\boldsymbol{x}$ with mean $\boldsymbol{a}$ and covariance $\boldsymbol{A}$, $\boldsymbol{m}^-_{x,k}$ is the *a priori* mean of the vehicle state, and $\boldsymbol{P}^-_{xx,k}$ is the *a priori* error covariance of the vehicle state.

Additionally, let the state of each feature be distributed according to the Gaussian density

$$\boldsymbol{\zeta}_{i,k} \sim p_g(\boldsymbol{\zeta}_{i,k}\,;\,\boldsymbol{m}_{\zeta,i,k}, \boldsymbol{P}_{\zeta\zeta,i,k}), \tag{1}$$

$\boldsymbol{m}_{\zeta,i,k}$ is the i^{th} feature's mean and $\boldsymbol{P}_{\zeta\zeta,i,k}$ is the i^{th} feature's covariance. Note that, at the expense of additional computational complexity, all of the following developments hold if Eq. (1) is a weighted sum of Gaussians (also known as Gaussian mixture (GM)). Note, too, that the state $\zeta_{i,k}$ is quite general and can contain any observable quantities of interest, such as position, velocity, size, shape, etc. In the context of lunar TRN, it is often sufficient to simply consider the 3-dimensional position coordinates of the feature in Moon-fixed coordinates.

Let the states of all the features within the known environment be elements of the random finite set (RFS) $\boldsymbol{E}_k$, and say that the environment can be partitioned into two sets $\boldsymbol{M}_k$ and $\boldsymbol{N}_k$. The first set (denoted $\boldsymbol{M}_k$ to correspond to "map") is said to contain n_M features of interest, and the second set (denoted $\boldsymbol{N}_k$ to correspond to "nuisance") is said to contain n_N features that are *a priori* known to exist but which are not to be used for updating state estimates. Therefore, the environment set is given as

$$\boldsymbol{E}_k = \boldsymbol{M}_k \cup \boldsymbol{N}_k\,,$$

such that $\boldsymbol{M}_k \cap \boldsymbol{N}_k = \emptyset$, $|\boldsymbol{E}_k| = n_E$, $|\boldsymbol{M}_k| = n_M$, $|\boldsymbol{N}_k| = n_N$, and $n_E = n_M + n_N$. Let $\boldsymbol{M}_k$ contain all of the previously discussed (uncertain) map features, $\boldsymbol{\zeta}_{i,k}$, such that

$$\boldsymbol{M}_k = \{\boldsymbol{\zeta}_{1,k}, \ldots, \boldsymbol{\zeta}_{n_M,k}\}\,.$$

Note that, since $\boldsymbol{M}_k$ is a set, it and its associated functions are all invariant to the ordering of $\boldsymbol{\zeta}_{i,k}$, i.e. their labels are insignificant as posed.

A. Observation Modeling

Let the random observations be generated according to the RFS $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k(\boldsymbol{x}_k, \boldsymbol{M}_k)$ belonging to the measurement space $\mathbb{Z}$. Based upon the vehicle state, $\boldsymbol{x}_k$, and the map, $\boldsymbol{M}_k$, the observation RFS is given as the union of the observations generated from all of the elements of $\boldsymbol{M}_k$ as

$$\boldsymbol{\Sigma}_k(\boldsymbol{x}_k, \boldsymbol{M}_k) = \boldsymbol{\Sigma}_{1,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{1,k}) \cup \cdots \cup \boldsymbol{\Sigma}_{n_M,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{n_M,k}) \cup \boldsymbol{\Theta}_k(\boldsymbol{x}_k, \boldsymbol{N}_k), \tag{2}$$

where $\boldsymbol{\Theta}_k(\boldsymbol{x}_k, \boldsymbol{N}_k)$ is the RFS of clutter returns generated by the sensor process as well as by the nuisance features in $\boldsymbol{N}_k$ (recall that the entire feature environment is the set $\boldsymbol{M}_k \cup \boldsymbol{N}_k$ with $\boldsymbol{M}_k \cap \boldsymbol{N}_k = \emptyset$). Let the sensor return m point measurements at t_k, and let these measurements be collected as the set $\boldsymbol{Z}_k \subset \boldsymbol{\Sigma}_k$ such that it is of the form

$$\boldsymbol{Z}_k = \{\boldsymbol{z}_{1,k}, \ldots, \boldsymbol{z}_{m,k}\}\,.$$

The set $\boldsymbol{Z}_k$ is an instantiation of its associated RFS $\boldsymbol{\Sigma}_k$, which is analogous to, for example, the vehicle state $\boldsymbol{x}_k$ being an instantiation of its associated random variable. The set measurement, $\boldsymbol{Z}_k$, is then an instance of the RFS $\boldsymbol{\Sigma}_k(\boldsymbol{x}_k, \boldsymbol{M}_k)$, and each $\boldsymbol{\Sigma}_{i,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \subset \boldsymbol{\Sigma}_k(\boldsymbol{x}_k, \boldsymbol{M}_k)$ is the detection produced by the i^{th} feature in $\boldsymbol{M}_k$. Let the individual observations of these features be generated according to the arbitrary nonlinear function

$$\boldsymbol{z}_{i,k} = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) + \boldsymbol{v}_k\,, \tag{3}$$

where $\boldsymbol{v}_k$ is a zero-mean white noise sequence with known covariance $\boldsymbol{P}_{vv,k}$. For example, for lunar TRN with a terrain camera, this measurement could be taken as pixel coordinates of a feature or, equivalently, bearing angles to that feature.

Within the context of this conversation, the noise sequence $\boldsymbol{v}_k$ will be constrained to being Gaussian distributed, though that is not necessary in general and, as with the feature location distributions (Eq. (1)), GMs offer another candidate tool for modeling non-Gaussian noise processes.

Let the RFS corresponding to the observation of the i^{th} element in $\boldsymbol{M}_k$ be either a single point measurement, $z_{i,k}$, or empty such that its corresponding RFS is

$$\boldsymbol{\Sigma}_{i,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) = \{z_{i,k}\} \cup \emptyset(p_D(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})), \tag{4}$$

where $z_{i,k}$ is a single target point measurement generated according to the likelihood function

$$z_{i,k} \sim g(z_{i,k}|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \tag{5}$$

and $\emptyset(p_D(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}))$ is a set such that

$$\Pr(\emptyset(p) = \boldsymbol{Y}) = \begin{cases} p & \text{if } \boldsymbol{Y} = \mathbb{Z} \\ 1-p & \text{if } \boldsymbol{Y} = \emptyset \\ 0 & \text{o.w.} \end{cases}. \tag{6}$$

In other words, the RFS $\boldsymbol{\Sigma}_{i,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})$ produces a point detection $z_{i,k}$ with probability $p_D(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})$ and is the empty set with probability $1 - p_D(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})$. Thus, $p_D(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})$ is referred to as the probability of detection.

It has become standard practice to utilize Taylor series expansions of the full nonlinear measurement function $\boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})$ to approximate nonlinear estimation error propagation via linearization. To that end, define the Jacobian

$$\boldsymbol{H}_{\alpha,i,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k}) \triangleq \left.\frac{\partial \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})}{\partial \boldsymbol{\alpha}_k}\right|_{(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) = (\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k})} \overset{\text{abbr.}}{=} \boldsymbol{H}_{\alpha,i,k}\,,$$

where $\boldsymbol{\alpha}_k$ is some vector that the derivative is taken with respect to and the bar notation $(\cdot)|_{(\boldsymbol{a}_1, \boldsymbol{b}_1) = (\boldsymbol{a}_2, \boldsymbol{b}_2)}$ denotes that $(\boldsymbol{a}_1, \boldsymbol{b}_1) = (\boldsymbol{a}_2, \boldsymbol{b}_2)$ should be evaluated after taking the derivative. This will be used to compute the Jacobians $\boldsymbol{H}_{x,i,k}$ and $\boldsymbol{H}_{\zeta,i,k}$, or the partials of the nonlinear measurement function with respect to $\boldsymbol{x}_k$ and $\boldsymbol{\zeta}_{i,k}$, respectively. This definition will be used throughout this paper.

B. The General Likelihood Function

When working with random vectors, it is fairly standard to work only with the probability density function (pdf) due to the nice algebraic properties useful pdfs, such as the Gaussian pdf, can offer, but other fundamental statistical descriptors are generally available for use, such as the cumulative density function (CDF) or the moment generating function (MGF). In this case, it is usually the case that the pdf is a sufficient tool for obtaining the desired mathematical results. The mathematics underpinning RFSs, however, tend to frustrate efforts to produce useful solutions and, therefore, working with pdfs directly is often impossible. Just as a random vector can be described using a pdf, a CDF, or an MGF, the statistics of an RFS can be described using any of three equivalent statistical descriptors: multitarget pdfs, belief-mass functions (BMFs), or probability generating functionals (PGFLs). Due to its useful relationship to likelihood functions for measurement models, the BMF is of particular relevance to this paper, and its useful properties are summarily described presently.

For some RFS $\boldsymbol{\Sigma}$ belonging to some space $\mathbb{Z}$, its BMF is defined by [5]

$$\beta_\Sigma(\boldsymbol{Y}) = \Pr(\boldsymbol{\Sigma} \subseteq \boldsymbol{Y})\,.$$

Furthermore, for a collection of n RFSs $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_n \subseteq \mathbb{Z}$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \cup \cdots \cup \boldsymbol{\Sigma}_n$, the BMF of all n RFSs is

$$\beta_\Sigma(\boldsymbol{Y}) = \Pr(\boldsymbol{\Sigma}_1 \subseteq \boldsymbol{Y}, \ldots, \boldsymbol{\Sigma}_n \subseteq \boldsymbol{Y}),$$

and if the $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_n$ are independent, then

$$\beta_\Sigma(\boldsymbol{Y}) = \Pr(\boldsymbol{\Sigma}_1 \subseteq \boldsymbol{Y}) \cdots \Pr(\boldsymbol{\Sigma}_n \subseteq \boldsymbol{Y})\,.$$

As noted on p. 83 in [5], BMFs are generalizations of probability-mass functions and, unsurprisingly, they behave similarly. For example, as noted in the same reference, if $\boldsymbol{\Sigma} = \{z\}$, where $z \in \mathbb{Z}$ is a standard random variable, then

$$\beta_\Sigma(\boldsymbol{Y}) = \Pr(\{z\} \subseteq \boldsymbol{Y}) = \Pr(z \in \boldsymbol{Y}),$$

which is precisely the probability mass function of z evaluated at $\boldsymbol{Y}$.

Within the context of this paper, BMFs find their use in deriving challenging likelihood functions, In particular, for some RFS $\boldsymbol{\Sigma}$, it holds that

$$f_{\Sigma}(\boldsymbol{Y}) = \frac{\delta \beta_{\Sigma}}{\delta \boldsymbol{Y}}(\emptyset), \tag{7}$$

where $\emptyset$ denotes the empty set, $\frac{\delta(\cdot)}{\delta \boldsymbol{Y}}$ denotes a set derivative with respect to the set $\boldsymbol{Y}$ [5], and $f_{\Sigma}(\boldsymbol{Y})$ is the multitarget pdf of $\boldsymbol{\Sigma}$ at some $\boldsymbol{Y} \subset \mathbb{Z}$. That is, when armed with the BMF corresponding to Eq. (2), Eq. (7) can be used to derive the corresponding likelihood function which can then subsequently be paired with a prior distribution to compute a posterior via Bayes' rule.

Another interesting property of BMFs and set derivatives is that the set derivative of a BMF with respect to the empty set recovers the same BMF, i.e., for the BMF of some RFS $\boldsymbol{\Sigma}$, it holds that

$$\frac{\delta \beta_{\Sigma}}{\delta \{\emptyset\}}(\boldsymbol{Y}) = \beta_{\Sigma}(\boldsymbol{Y}) \,. \tag{8}$$

The BMF of $\boldsymbol{\Sigma}_k$ is, by definition,

$$\beta_{\Sigma_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = \Pr(\boldsymbol{\Sigma}_k(\boldsymbol{x}_k) \subseteq \boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k),$$

which, using the definition of $\boldsymbol{\Sigma}_k$ given in Eq. (2), can be written as

$$\beta_{\Sigma_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = \Pr(\boldsymbol{\Sigma}_{1,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{1,k}) \subseteq \boldsymbol{Y}, \ldots, \boldsymbol{\Sigma}_{n_M,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{n_M,k}) \subseteq \boldsymbol{Y}, \boldsymbol{\Theta}_k(\boldsymbol{x}_k, \boldsymbol{N}_k) \subseteq \boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \,.$$

Taking all of the observations to be generated independently (i.e. measurement noise is independent and all features are detected independently), including the clutter returns, allows the BMF of $\boldsymbol{\Sigma}_k$ to be written as the product

$$\begin{aligned}\beta_{\Sigma_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) &= \Pr(\boldsymbol{\Sigma}_{1,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{1,k}) \subseteq \boldsymbol{Y}) \cdots \Pr(\boldsymbol{\Sigma}_{n_M,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{n_M,k}) \subseteq \boldsymbol{Y}) \Pr(\boldsymbol{\Theta}_k(\boldsymbol{x}_k, \boldsymbol{N}_k) \subseteq \boldsymbol{Y}|\boldsymbol{x}_k) \\ &= \beta_{\Sigma_{1,k}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \cdots \beta_{\Sigma_{n_M,k}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \beta_{\Theta_k}(\boldsymbol{Y}|\boldsymbol{x}_k) \,.\end{aligned}$$

As demonstrated in [3], the BMF of $\boldsymbol{\Sigma}_{i,k}$ can be written as

$$\beta_{\Sigma_{i,k}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = 1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) + p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \int_{\boldsymbol{Y}} g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \mathrm{d}z \,. \tag{9}$$

Additionally, taking the clutter to be Poisson, the BMF of $\boldsymbol{\Theta}_k$ is given by [3]

$$\beta_{\Theta_k}(\boldsymbol{Y}|\boldsymbol{x}_k) = \exp\left\{ \lambda \int_{\boldsymbol{Y}} c(z) \mathrm{d}z - \lambda \right\} , \tag{10}$$

where $c(z)$ is the spatial density of clutter (potentially belonging both to standard clutter returns as well as clutter returns corresponding to the nuisance elements of $\boldsymbol{N}_k$) and λ is the expected number of clutter returns per scan (which, generally, can be expected to be the number of sensor-generated clutter returns per scan plus the number of expected elements of $\boldsymbol{N}_k$ seen in a given scan).

Then, using the property of the BMF illustrated in Eq. (7), the likelihood function is given by the set derivative

$$g(\boldsymbol{Z}_k|\boldsymbol{x}_k, \boldsymbol{M}_k) = \frac{\delta \beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\emptyset|\boldsymbol{x}_k, \boldsymbol{M}_k),$$

which, on the face of it, does not appear insurmountable. However, by employing the general product rule as described in [5], this derivative takes the form*

$$\frac{\delta \beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = \sum_{\boldsymbol{W}_1 \uplus \cdots \uplus \boldsymbol{W}_{n_M+1} = \boldsymbol{Z}_k} \frac{\delta \beta_{\Sigma_{1,k}}}{\delta \boldsymbol{W}_1}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \cdots \frac{\delta \beta_{\Sigma_{n_M,k}}}{\delta \boldsymbol{W}_{n_M}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \frac{\delta \beta_{\Theta_k}}{\delta \boldsymbol{W}_{n_M+1}}(\boldsymbol{Y}|\boldsymbol{x}_k), \tag{11}$$

where the sum is taken over all (potentially empty) disjoint subsets of $\boldsymbol{Z}_k$. Equation (11) is combinatoric in that it requires the sum over all disjoint subsets of $\boldsymbol{Z}_k$ (i.e., we must consider all possible combinations – including potentially empty subsets – of the elements of $\boldsymbol{Z}_k$). This combinatorial result is undesired, and approximation of Eq. (11) is necessary to produce a practical algorithm. One such approximation produces the AFP update (Section III), and another produces the new higher order AFP update (Section IV).

*Note that this differs from [3] due to a typographical error – the final term concerning $\boldsymbol{\Theta}_k$ was omitted by mistake.

III. Anonymous Feature Processing: An Improved Derivation

A. AFP's Guiding Approximation and Deriving Its Likelihood

The guiding approximation of the AFP update in [3] seeks to interrogate each feature in $\boldsymbol{M}_k$ one at a time, each time treating all other returns as if they were clutter, regardless of whether or not they were actually generated by one of the other features. The update then mixes over all of these hypotheses to, ideally, obtain meaningful updates using the known features. Note that this is performed without any *a priori* knowledge of observation-to-feature associations (i.e., it requires no upstream identification pre-processing to accomplish an update). As an example, consider the terms that are summed over in Eq. (11) (i.e. how each of the $\boldsymbol{W}_i \subset \boldsymbol{Z}_k$ are assigned for each sum). For $\boldsymbol{Z}_k = \{z_1, \ldots, z_m\}$, allowing each target to generate at most one measurement, the subsets take the form

$$
\begin{aligned}
&\boldsymbol{Z}_k = \boldsymbol{W}_1 \uplus \cdots \uplus \boldsymbol{W}_{n_M} \uplus \boldsymbol{W}_{n_M+1} \\
&\quad \Rightarrow \underbrace{\{z_1\} \cup \{z_2\} \cup \cdots \cup \{z_m\}}_{n_M \text{ sets}} \cup \{\emptyset\} && \text{(assign all } m \text{ meas. to } n_M \text{ features)} \\
&\quad \Rightarrow \underbrace{\{z_1\} \cup \{z_2\} \cup \cdots \cup \{z_{m-1}\} \cup \{\emptyset\}}_{n_M \text{ sets}} \cup \{z_m\} && \text{(one meas. assigned as clutter)} \\
&\quad \Rightarrow \underbrace{\{z_1\} \cup \{z_2\} \cup \cdots \cup \{z_{m-2}\} \cup \{\emptyset\} \cup \{\emptyset\}}_{n_M \text{ sets}} \cup \{z_{m-1}, z_m\} && \text{(two meas. assigned as clutter)} \\
&\quad \vdots && \vdots \\
&\quad \Rightarrow \underbrace{\{z_1\} \cup \{\emptyset\} \cup \cdots \cup \{\emptyset\}}_{n_M \text{ sets}} \cup \{z_2, \ldots, z_m\} && (m-1 \text{ meas. assigned as clutter}) \\
&\vdots && \vdots
\end{aligned}
$$

where the case highlighted in blue is an example of the type of "one at a time" terms that the AFP update considers. Said another way, the AFP update of [3] is a "one at a time" approximation of Eq. (11) because it only considers the terms which assign measurements one at a time while treating all others as clutter. Note that only a few examples of partitioning are shown here to illustrate the desired case and that many more partitions exist for $\boldsymbol{Z}_k$.

Remark 1 *A few things become apparent. First, if it was unclear before, it is now obvious why this sum is combinatorially complex in how many terms arise from the sum when enumerating by hand! Second, it is clear how many terms the "one at a time" approximation discards from the entire sum, only accepting the summands that treat a single measurement as valid and the other $m-1$ measurements as if they were clutter. In this sense, this approximation is a "first order" method. Third, it seems that the natural inclination for these sorts of things when searching for higher-order methods is to, for good reason, seek the second order method. However, looking at all of the terms that are discarded, it does not appear that the "all but* ***two*** *measurements assigned as clutter" is particularly useful, so different approximations will be pursued in Section IV.*

With this in mind, the only terms of Eq. (11) which are retained (again, only the "one at a time" terms) are

$$
\frac{\delta \beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = \sum_{\substack{z \in \boldsymbol{Z}_k \\ \boldsymbol{W} = \boldsymbol{Z} \setminus z}} \sum_{i=1}^{n_M} \frac{\delta \beta_{\Sigma_{i,k}}}{\delta \{z\}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \frac{\delta \beta_{\Theta_k}}{\delta \boldsymbol{W}}(\boldsymbol{Y}|\boldsymbol{x}_k), \tag{12}
$$

where the necessary set derivatives (corresponding to Eqs. (9) and (10), see [3] for derivation) are given as

$$
\frac{\delta \beta_{\Sigma_{i,k}}}{\delta \boldsymbol{W}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) = \begin{cases} 1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) + p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \int_{\boldsymbol{Y}} g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \mathrm{d}z & \text{if } \boldsymbol{W} = \emptyset \\ p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) g(\boldsymbol{w}|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) & \text{if } |\boldsymbol{W}| = 1, \boldsymbol{W} = \{\boldsymbol{w}\} \\ 0 & \text{if } |\boldsymbol{W}| > 1 \end{cases} \tag{13}
$$

and

$$
\frac{\delta \beta_{\Theta_k}}{\delta \boldsymbol{W}}(\boldsymbol{Y}|\boldsymbol{x}_k) = \begin{cases} \exp\{-\lambda\} & \text{if } \boldsymbol{W} = \emptyset \\ \exp\left\{\lambda \int_{\boldsymbol{Y}} c(z) \mathrm{d}z - \lambda\right\} \prod_{\boldsymbol{w} \in \boldsymbol{W}} \lambda c(\boldsymbol{w}) & \text{if } |\boldsymbol{W}| \geq 1 \end{cases}. \tag{14}
$$

Remark 2 (A Sanity Check.) *Does Eq. (12) make sense? To ensure it does, consider its terms:*

- *The first summation selects a single measurement, $\boldsymbol{z}$, from the set of all measurements, $\boldsymbol{Z}_k$, and assigns all others as clutter within the set $\boldsymbol{W} = \boldsymbol{Z}_k \setminus \boldsymbol{z}$.*
- *The second summation applies each selected measurement to a single feature of the catalog/map, $\boldsymbol{M}_k$.*
- *With both summations, a single measurement is associated to a single feature in each term with the remaining measurements being treated as if they were clutter.*

This satisfies the guiding principles of the AFP update, and thus the function in Eq. (12) passes inspection.

With these derivatives, Eq. (12) becomes

$$\frac{\delta\beta_{\Sigma_k}}{\delta\boldsymbol{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k) = \sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M} p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})\exp\left\{\lambda\int_{\boldsymbol{Y}} c(\boldsymbol{y})\mathrm{d}\boldsymbol{y}-\lambda\right\}\prod_{\boldsymbol{w}\in\boldsymbol{Z}\setminus\boldsymbol{z}}\lambda c(\boldsymbol{w}),$$

and, thanks to Eq. (7), the corresponding likelihood can be computed by setting $\boldsymbol{Y} = \emptyset$, yielding

$$\begin{aligned} g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k) &= \frac{\delta\beta_{\Sigma_k}}{\delta\boldsymbol{Z}_k}(\emptyset|\boldsymbol{x}_k,\boldsymbol{M}_k) \\ &= \sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M} p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})\exp\{-\lambda\}\prod_{\boldsymbol{w}\in\boldsymbol{Z}\setminus\boldsymbol{z}}\lambda c(\boldsymbol{w})\,. \end{aligned}$$

Note that, by pulling out terms, this can be rewritten as

$$g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k) = \lambda^m\exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k}\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]}\,, \tag{15}$$

where $m = |\boldsymbol{Z}_k|$ and we have employed the shorthand notation

$$h^{\boldsymbol{X}} = \prod_{x\in\boldsymbol{X}} h(\boldsymbol{x})\,.$$

B. Deriving the AFP Update

Equation (15) is then the likelihood function corresponding to the "one at a time" approximation of Eq. (11). Given this likelihood function, we can now propose a prior and utilize Bayes' rule to compute a posterior. To that end, recall that Bayes' rule (with set arguments) states that

$$p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_k) = \frac{g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_{k-1})}{\iint g(\boldsymbol{Z}_k|\boldsymbol{\gamma}_k,\boldsymbol{\Gamma}_k)p(\boldsymbol{\gamma}_k,\boldsymbol{\Gamma}_k|\boldsymbol{Z}_{k-1})\mathrm{d}\boldsymbol{\gamma}_k\delta\boldsymbol{\Gamma}_k}\,. \tag{16}$$

For convenience, define the normalizer

$$\eta_k = \iint g(\boldsymbol{Z}_k|\boldsymbol{\gamma}_k,\boldsymbol{\Gamma}_k)p(\boldsymbol{\gamma}_k,\boldsymbol{\Gamma}_k|\boldsymbol{Z}_{k-1})\mathrm{d}\boldsymbol{\gamma}_k\delta\boldsymbol{\Gamma}_k \tag{17}$$

such that Eq. (16) can be written as

$$p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_k) = \frac{1}{\eta_k}g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_{k-1})\,. \tag{18}$$

The likelihood, $g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)$, has already been obtained in Eq. (15), so now a prior, $p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_{k-1})$, must be specified. Take the prior vehicle density to be independent of the map density and let the map density be independent of the vehicle's collected data,[†] i.e., that $p(\boldsymbol{x}_k,\boldsymbol{M}_k|\boldsymbol{Z}_{k-1}) = p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1})p(\boldsymbol{M}_k)$. Then, substituting the likelihood function

[†]This, of course, does not mean that the map density doesn't depend on some collected data. Of course, any practical catalog, such as a lunar crater catalog, is based on *some* data. The present assertion is that the map set $\boldsymbol{M}_k$ does not depend on any of the scans collected by the vehicle up to this point (i.e., that $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ did not influence $p(\boldsymbol{M}_k)$). Note that removing this assumption transforms the problem into the SLAM problem, which is sought to be avoided here.

(Eq. (15)) and this prior into Eq. (18) yields

$$p(\boldsymbol{x}_k, \boldsymbol{M}_k|\boldsymbol{Z}_k)$$
$$= \frac{1}{\eta_k}\left\{\lambda^m \exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k}\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]}\right\}p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1})p(\boldsymbol{M}_k)\,. \quad (19)$$

Noting that we have chosen to forgo estimating $\boldsymbol{M}_k$ in favor of solely estimating $\boldsymbol{x}_k$ and that we can compute the marginal density

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x},\boldsymbol{M})\delta\boldsymbol{M}\,,$$

marginalize the map set from Eq. (19) to obtain

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k)$$
$$= \frac{1}{\eta_k}\int\left\{\lambda^m \exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k}\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]}\right\}p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1})p(\boldsymbol{M}_k)\delta\boldsymbol{M}_k\,.$$

Employing the approximation[‡]

$$p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k}) \approx p_{D,k}(\boldsymbol{m}^-_{x,k},\boldsymbol{m}_{\zeta,i,k})$$
$$[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k} \approx [1-p_{D,k}(\boldsymbol{m}^-_{x,k},\cdot)]^{\hat{\boldsymbol{M}}_k}\,,$$

where $\hat{\boldsymbol{M}}_k = \{\boldsymbol{m}_{\zeta,1,k},\ldots,\boldsymbol{m}_{\zeta,n_M,k}\}$ is taken to be a non-stochastic representation of $\boldsymbol{M}_k$ (i.e., some nominal onboard catalog or survey, for example), we can write

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \frac{1}{\eta_k}\left\{\lambda^m \exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{m}^-_{x,k},\cdot)]^{\hat{\boldsymbol{M}}_k}\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{m}^-_{x,k},\boldsymbol{m}_{\zeta,i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{m}^-_{x,k},\boldsymbol{m}_{\zeta,i,k})]}\right\}p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) \quad (20)$$

because

$$\int p(\boldsymbol{M}_k)\delta\boldsymbol{M}_k = 1\,.$$

Take the single-target measurement likelihood, $g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})$, to be Gaussian of the form

$$g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k}) = p_g(\boldsymbol{x}_k\,;\,\boldsymbol{h}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k}),\boldsymbol{P}_{vv,k})$$

(resulting from Eq. (3)). Additionally, take the vehicle state's prior conditional density to be Gaussian distributed such that

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) = p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^-_{x,k},\boldsymbol{P}^-_{xx,k}),$$

where $\boldsymbol{m}^-_{x,k}$ and $\boldsymbol{P}^-_{xx,k}$ are the vehicle state's prior mean and covariance, respectively. Substituting these into Eq. (20) yields the AFP posterior

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k)$$
$$= \frac{\lambda^m \exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{m}^-_{x,k},\cdot)]^{\hat{\boldsymbol{M}}_k}}{\eta_k}\sum_{i=1}^{n_M}\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\frac{p_{D,k}(\boldsymbol{m}^-_{x,k},\boldsymbol{m}_{\zeta,i,k})p_g(\boldsymbol{x}_k\,;\,\boldsymbol{h}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k}),\boldsymbol{P}_{vv,k})p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^-_{x,k},\boldsymbol{P}^-_{xx,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{m}^-_{x,k},\boldsymbol{m}_{\zeta,i,k})]}\,.$$

[‡]This is equivalent to assuming that the probability of detection does not depend on the stochasticity of $\boldsymbol{x}_k$ or $\boldsymbol{M}_k$.

Utilizing Lemma 1 (see Appendix A) and Eqs. (17) and (19), the posterior density produced by the AFP update is given by the GM

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \sum_{i=1}^{n_M} \sum_{\boldsymbol{z}\in\boldsymbol{Z}_k} w_k(\boldsymbol{z},i) p_g(\boldsymbol{x}_k\,;\,\tilde{\boldsymbol{m}}_{x,k}(\boldsymbol{z},i), \tilde{\boldsymbol{P}}_{xx,k}(\boldsymbol{z},i)), \tag{21}$$

where[§]

$$w_k(\boldsymbol{z},i) = \left(\sum_{j=1}^{n_M} \sum_{\boldsymbol{y}\in\boldsymbol{Z}_k} \frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,j,k}) q(\boldsymbol{y},j)}{\lambda c(\boldsymbol{y})[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,j,k})]} \right)^{-1} \frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k}) q(\boldsymbol{z},i)}{\lambda c(\boldsymbol{z})[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k})]} \tag{22}$$

$$\tilde{\boldsymbol{m}}_{x,k}(\boldsymbol{z},i) = \boldsymbol{m}^-_{x,k} + \boldsymbol{K}_{x,k}(\boldsymbol{z},i)[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k})] \tag{23}$$

$$\tilde{\boldsymbol{P}}_{xx,k}(\boldsymbol{z},i) = [\boldsymbol{I}_{n_x\times n_x} - \boldsymbol{K}_{x,k}(\boldsymbol{z},i)\boldsymbol{H}_{x,i,k}]\boldsymbol{P}^-_{xx,k}[\boldsymbol{I}_{n_x\times n_x} - \boldsymbol{K}_{x,k}(\boldsymbol{z},i)\boldsymbol{H}_{x,i,k}]^T + \boldsymbol{K}_{x,k}(\boldsymbol{z},i)\boldsymbol{P}_{vv,k}\boldsymbol{K}^T_{x,k}(\boldsymbol{z},i) \tag{24}$$

and

$$q(\boldsymbol{z},i) = p_g(\boldsymbol{z}\,;\,\boldsymbol{h}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k}), \boldsymbol{P}_{zz,k}(\boldsymbol{z},i)) \tag{25}$$

$$\boldsymbol{K}_{x,k}(\boldsymbol{z},i) = \boldsymbol{P}^-_{xx,k}\boldsymbol{H}^T_{x,i,k}[\boldsymbol{P}_{zz,k}(\boldsymbol{z},i)]^{-1} \tag{26}$$

$$\boldsymbol{P}_{zz,k}(\boldsymbol{z},i) = \boldsymbol{H}_{x,i,k}\boldsymbol{P}^-_{xx,k}\boldsymbol{H}^T_{x,i,k} + \boldsymbol{H}_{\zeta,i,k}\boldsymbol{P}_{\zeta\zeta,i,k}\boldsymbol{H}^T_{\zeta,i,k} + \boldsymbol{P}_{vv,k}\,. \tag{27}$$

The $w_k(\boldsymbol{z},i)$, $\tilde{\boldsymbol{m}}_{x,k}(\boldsymbol{z},i)$, and $\tilde{\boldsymbol{P}}_{xx,k}(\boldsymbol{z},i)$ are the weights, means, and covariances, respectively, of the GM in Eq. (21). The double sum over all $\boldsymbol{z}\in\boldsymbol{Z}_k$, $i\in\{1,\ldots,n_M\}$ effectively associates each collected measurement to each catalog feature and the weights $w_k(\boldsymbol{z},i)$ defined by Eq. (22) establish the relative likelihood over one $(\boldsymbol{z},i)$ pair versus another. With this in mind, it is clear that Eq. (21) is an $(|\boldsymbol{Z}_k|\cdot n_M)$-component GM, meaning that, since the *a priori* distribution was Gaussian but the posterior in Eq. (21) is a GM, the measurement update is analytically closed but does not behave as desired. In the interest of a practical solution, Eq. (21) is approximated as a single mean and covariance via the method of moments (a technique that finds the mean and covariance that best agrees with the mean and covariance of the entire GM), yielding

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) \approx p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^+_{x,k}, \boldsymbol{P}^+_{xx,k}), \tag{28}$$

where $\boldsymbol{m}^+_{x,k}$ and $\boldsymbol{P}^+_{xx,k}$ are the method-of-moments-merged mean and covariance, respectively. For details regarding the method of moments for GMs, see the appendices of [6] and [7].

It has been demonstrated that the AFP update can produce approximately the same results as the Kalman filter but with none of the divergence risks associated with misidentification errors [3]. Note that, in general, implementing the AFP update as written above introduces a significant computational burden when compared to the standard Kalman filter, but [4] discusses implementation-specific details that produce a practical algorithm with nearly the same computational complexity as the standard Kalman filter.

C. Discussion on Differences With Ref. [3] and Their Implications

The AFP update derived here is (up to notation) identical to the AFP update presented in [3] despite utilizing a wholly new derivation. Recall that the likelihood function, Eq. (15), is the "one at a time" approximation of Eq. (11) and is the likelihood function corresponding to the AFP update in [3]. However, in contrast, the derivation presented in this paper, while requiring a reader to "squint at" a combinatorial sum and carefully consider which terms are kept or discarded, does not impose any heuristic conditions as in the derivation of [3].

To summarize, [3] sought to overcome the combinatorial nature of Eq. (11) by instead subdividing the multi-feature estimation problem into a collection of single-feature estimation problems by imposing an arbitrary label to each measurement-to-feature association and the marginalizing out the labels to remove them from consideration. The advantage of this approach is that the AFP likelihood naturally "falls out" once one takes the involved set derivatives required by the FISST framework. The remaining "ugly" terms in Eq. (11) simply vanish, no longer a problem to be dealt with. Additionally, this approach seems intuitive: propose all pairwise measurement-to-feature associations via a

[§] In [3], $w_k(\boldsymbol{z},i)$, $q(\boldsymbol{z},i)$, and $\boldsymbol{P}_{zz,k}(\boldsymbol{z},i)$ were denoted as $w_{i,k}(\boldsymbol{z})$, $q_i(\boldsymbol{z})$, and $\boldsymbol{P}_{zz,i,k}$ respectively, but they have been slightly altered here for notational consistency.

measurement and label pair, and then marginalize over the labels such that all such associations are accounted for. The disadvantage of this approach is in that it leaves no room to grow or explore additional approximations to Eq. (11).

The derivation of this paper, on the other hand, directly assesses the impacts of *specific* terms of the combinatorial sum Eq. (11), permitting insight into which are useful and which can be discarded. As alluded to in Remark 1, the AFP update is, in some sense, a first-order approximation of the full combinatorial likelihood. When seeking a higher order method, it comes down to carefully inspecting the sum and determining which quantities are most useful. The heuristic-free derivation of this paper is, perhaps, harder to follow and less intuitive than that of [3] but affords us the ability to expand and produce the higher-order method in the next section.

IV. A Higher Order Approximation

In Section III, the "one at a time" approximation was employed to develop the AFP update. In other words, the update was derived by only selecting the terms of the full combinatoric likelihood Eq. (11) that assign a single measurement to a single map feature. Originally, this derivation was inspired by the desire to develop an update that was not only immune to identification errors but also permits entirely forgoing the identification process upstream of the measurement update. In essence, the identification process to moved to *within* the measurement update itself, and, thus, is immune to any upstream identification errors.

However, as was seen in the previous discussion, Eq. (11) contains a tremendous variety of terms, and AFP only considers *one* type of term: the "one at a time" terms. This section considers a system equipped with an identification algorithm upstream of the measurement update that tags each feature detection with a (potentially erroneous) identity defining its association to the onboard feature catalog/map. We seek to derive an update which is immune to errors in this identification procedure (like the AFP update) yet remains capable of utilizing this identification information to produce a stronger update. Said another way, we seek an update which is able to utilize the information provided by the upstream identification processor but is completely safe from divergence when said processor makes identification errors. It is reiterated here that this claim cannot be made for the Kalman filter – even with standard navigation insulators like residual editing, it has been demonstrated that these techniques alone are insufficient for guaranteeing stable filter performance when faced with identification errors [3].

This section will consider an approximation to the full combinatoric liklihood Eq. (11) that retains three types of terms:

1) "one at a time" — the terms which associate a single measurement to a single feature. This is synonymous with the AFP update of the previous section.
2) "identified" — the singular term which corresponds to those detections which were identified by an identification pre-processor. It will be shown that this corresponds to a modified Kalman update according to the labeled detections.
3) "null scan" — the singular term corresponding to the case where all collected returns are clutter. This is included as an additional robustness quantity.

The measurement update presented at the end of this section is capable of utilizing identified detections but, as will be demonstrated, is immune to identification errors. In the event that identification is accurate and successful, a "strong" update is performed (and converges toward the optimal Kalman update). When identification errors are present, a "weaker" update is performed (and converges toward the AFP update). To that end, consider the following approximation of the likelihood in Eq. (11):

$$g(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k) \approx \underbrace{g_1(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)}_{\text{"one at a time"}} + \underbrace{g_2(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)}_{\text{"identified"}} + \underbrace{g_3(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)}_{\text{"null scan"}} . \tag{29}$$

Each term has been labeled above, but each correspond to the "one at a time," "pre-labeled," and "null scan" variety of terms described above, respectively. All three are presented in the following discussion.

A. The "One At A Time" Term

Fortunately, the "one at a time" term of the likelihood has already been derived in Section III: it is precisely the AFP likelihood function! Therefore, utilizing Eq. (15),

$$g_1(\boldsymbol{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k) = \lambda^m \exp\{-\lambda\}\,[c(\cdot)]^{\boldsymbol{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k} \sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(\boldsymbol{z}|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(\boldsymbol{z})[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]} . \tag{30}$$

B. The "Identified" Term

Consider the scan collected at time t_k, denoted $\boldsymbol{Z}_k$. Let these detections be subjected to the aforementioned upstream identification process such that the *identified* measurement set contains measurement-to-feature labels, yielding

$$\boldsymbol{Z}_k^{(\mathrm{id})} = \{(z_{1,k}^{(\mathrm{id})}, \ell_1), \ldots, (z_{m_{\mathrm{id}}}^{(\mathrm{id})}, \ell_{m_{\mathrm{id}}})\} \subseteq \boldsymbol{Z}_k\,,$$

where ℓ_i is the label of the i^{th} identified detection $z_{i,k}^{(\mathrm{id})}$. Note that the labels define a mapping between measurements and catalog features, but there is no guarantee that the labels are correct (i.e., misidentifications are possible). Note, too, that in general $\boldsymbol{Z}_k^{(\mathrm{id})}$ is a subset of $\boldsymbol{Z}_k$, meaning that not all measurements may be identified/associated with labels.

There exists exactly *one* term of Eq. (11) which corresponds to the *correct* observation-to-feature association. This discussion will assume that this term corresponds to the observations and labels $(z_{i,k}^{(\mathrm{id})}, \ell_i) \in \boldsymbol{Z}_k^{(\mathrm{id})}$. All other observations (i.e., $z \in \boldsymbol{Z}_k \setminus \boldsymbol{Z}_k^{(\mathrm{id})}$) are treated as misdetected and/or clutter terms. This labeling effectively performs all the partitioning in Eq. (11) "for free," resulting in a single remaining term. Let the set of labels $\{o_1, \ldots, o_{n_O}\}$, where $n_O = n_M - m_{\mathrm{id}}$, be the labels of the n_O features in $\boldsymbol{M}_k$ which did *not* receive a label, ℓ_i, and let the set of these map features be contained by the set $\bar{\boldsymbol{M}}_k = \boldsymbol{M}_k \setminus \boldsymbol{M}_k^{(\mathrm{id})} \subset \boldsymbol{M}_k$, where $|\bar{\boldsymbol{M}}_k| = n_O$. Then, the corresponding term of Eq. (11) is given as

$$\begin{aligned}
&\frac{\delta\beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \\
&= \left[\frac{\delta\beta_{\Sigma_{\ell_1,k}}}{\delta\{z_{1,k}^{(\mathrm{id})}\}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \cdots \frac{\delta\beta_{\Sigma_{\ell_{m_{\mathrm{id}}},k}}}{\delta\{z_{m_{\mathrm{id}},k}^{(\mathrm{id})}\}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k)\right] \left[\frac{\delta\beta_{\Sigma_{o_1,k}}}{\delta\{\emptyset\}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \cdots \frac{\delta\beta_{\Sigma_{o_{n_O},k}}}{\delta\{\emptyset\}}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k)\right] \frac{\delta\beta_{\Theta_k}}{\delta(\boldsymbol{Z}_k \setminus \boldsymbol{Z}_k^{(\mathrm{id})})}(\boldsymbol{Y}|\boldsymbol{x}_k),
\end{aligned}$$

where the first term corresponds to all identified features, the second term corresponds to all unidentified features (assumed to be misdetected), and the third term accounts for clutter.

Substituting Eqs. (13) and (14) allows us to write

$$\begin{aligned}
&\frac{\delta\beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k, \boldsymbol{M}_k) \\
&= \left[p_{D,k}(\boldsymbol{x}_k, \cdot) g(\cdot|\boldsymbol{x}_k, \cdot)\right]^{\boldsymbol{Z}_k^{(\mathrm{id})}} \left[1 - p_{D,k}(\boldsymbol{x}_k, \cdot) + p_{D,k}(\boldsymbol{x}_k, \cdot) \int_{\boldsymbol{Y}} g(z|\boldsymbol{x}_k, \cdot)\mathrm{d}z\right]^{\bar{\boldsymbol{M}}_k} \exp\left\{\lambda \int_{\boldsymbol{Y}} c(z)\mathrm{d}z - \lambda\right\} \prod_{z \in \boldsymbol{Z}_k \setminus \boldsymbol{Z}_k^{(\mathrm{id})}} \lambda c(z),
\end{aligned}$$

where the $h^{\boldsymbol{X}}$ shorthand has been overloaded slightly to imply

$$h^{\boldsymbol{Z}} = \prod_{(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{Z}} h(\boldsymbol{x}, \boldsymbol{y})\,.$$

Finally, evaluating $\frac{\delta\beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\emptyset|\boldsymbol{x}_k, \boldsymbol{M}_k)$ yields the likelihood function

$$\begin{aligned}
g_2(\boldsymbol{Z}_k|\boldsymbol{x}_k, \boldsymbol{M}_k) &= \frac{\delta\beta_{\Sigma_k}}{\delta \boldsymbol{Z}_k}(\emptyset|\boldsymbol{x}_k, \boldsymbol{M}_k) \\
&= \exp\{-\lambda\} \left[\lambda c(\cdot)\right]^{\boldsymbol{Z}_k \setminus \boldsymbol{Z}_k^{(\mathrm{id})}} \left[1 - p_{D,k}(\boldsymbol{x}_k, \cdot)\right]^{\boldsymbol{M}_k \setminus \boldsymbol{M}_k^{(\mathrm{id})}} \prod_{(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}} p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}),
\end{aligned}$$

and pulling out terms from the summation provides the final form

$$g_2(\boldsymbol{Z}_k|\boldsymbol{x}_k, \boldsymbol{M}_k) = \lambda^m \exp\{-\lambda\} \left[c(\cdot)\right]^{\boldsymbol{Z}_k} \left[1 - p_{D,k}(\boldsymbol{x}_k, \cdot)\right]^{\boldsymbol{M}_k} \prod_{(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})]}\,. \tag{31}$$

Note that, in contrast to the summmations present in the "on at a time" likelihood in Eq. (30), the "identified" likelihood in Eq. (31) consists principally of *products*. This lends some intuition into the differences between the two: the "one at a time" (aka AFP) likelihood operations merge (unidentified) associations using OR conditions via summation, whereas the "identified" likelihood operations merge (identified) associations using AND conditions via products. The interpretation lends theoretical basis to the observations in [3] that the AFP update is conservative when compared to the idealized Kalman filter (and, in fact, it will be shown later this section that the "identified" likelihood will be used to produce to a slightly modified Kalman update).

C. The "Null Scan" Term

The third and final term in the likelihood we must consider is that which treats all measurements in the collected scan as clutter. For example, consider an image processing anomaly in a crater-based TRN system that, perhaps due to lighting or exposure changes, falsely detects craters randomly throughout a collected image. The reason for considering this term is that neither the "one at a time" or "identified" terms consider this possibility, and the Kalman filter has no mechanism for formally accounting for it either. Additionally, this highlights another advantage to the present approach: in contrast to both the standard Kalman filter and the original AFP developments of [3], the present approach allows a designer to tailor their measurement update according to the specific problem at hand, lending a tremendous amount of design flexibility.

So, presume that *everything* that was collected was clutter. That is, all m measurements in $\mathbf{Z}_k$ received and to be processed by the filter were generated according to the clutter RFS $\boldsymbol{\Theta}_k$. This corresponds to a single term in Eq. (11) and is of the form

$$\frac{\delta\beta_{\Sigma_k}}{\delta\mathbf{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)=\frac{\delta\beta_{\Sigma_{1,k}}}{\delta\{\emptyset\}}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)\cdots\frac{\delta\beta_{\Sigma_{n_M,k}}}{\delta\{\emptyset\}}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)\frac{\delta\beta_{\Theta_k}}{\delta\mathbf{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k)\,.$$

Noting the set derivative property in Eq. (8) and utilizing the shorthand $\beta_{\Sigma_{i,k}}=\beta_{\Sigma_{i,k}}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)$, this can be written as

$$\frac{\delta\beta_{\Sigma_k}}{\delta\mathbf{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)=\beta_{\Sigma_{1,k}}\cdots\beta_{\Sigma_{n_M,k}}\frac{\delta\beta_{\Theta_k}}{\delta\mathbf{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k)\,.$$

Substituting Eqs. (9) and (14) into the above expression yields

$$\frac{\delta\beta_{\Sigma_k}}{\delta\mathbf{Z}_k}(\boldsymbol{Y}|\boldsymbol{x}_k,\boldsymbol{M}_k)=\left[1-p_{D,k}(\boldsymbol{x}_k,\cdot)+p_{D,k}(\boldsymbol{x}_k,\cdot)\int_{\boldsymbol{Y}}g(\boldsymbol{y}|\boldsymbol{x}_k,\cdot)\mathrm{d}\boldsymbol{y}\right]^{\boldsymbol{M}_k}\exp\left\{\lambda\int_{\boldsymbol{Y}}c(\boldsymbol{y})\mathrm{d}\boldsymbol{y}-\lambda\right\}\prod_{z\in\mathbf{Z}_k}\lambda c(z),$$

and setting $\boldsymbol{Y}=\emptyset$ (recall Eq. (7)) yields the likelihood function

$$\begin{aligned}g_3(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)&=\frac{\delta\beta_{\Sigma_k}}{\delta\mathbf{Z}_k}(\emptyset|\boldsymbol{x}_k,\boldsymbol{M}_k)\\&=\left[1-p_{D,k}(\boldsymbol{x}_k,\cdot)\right]^{\boldsymbol{M}_k}\exp\{-\lambda\}\prod_{z\in\mathbf{Z}_k}\lambda c(z)\,.\end{aligned}$$

Finally, rearranging terms gives

$$g_3(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)=\lambda^m\exp\{-\lambda\}\left[1-p_{D,k}(\boldsymbol{x}_k,\cdot)\right]^{\boldsymbol{M}_k}[c(\cdot)]^{\mathbf{Z}_k}\,,\tag{32}$$

the likelihood term corresponding to the "null scan" case.

D. Deriving the Higher Order AFP Update

With all the likelihood terms found, an update can now be derived using Bayes' rule. First, start with the higher order likelihood approximation Eq. (29), where $g_1(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)$, $g_2(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)$, and $g_3(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)$ are defined in Equations (30), (31), and (32), respectively, such that

$$\begin{aligned}g(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)=\;&\lambda^m\exp\{-\lambda\}\,[c(\cdot)]^{\mathbf{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k}\sum_{z\in\mathbf{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(z|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]}\\&+\lambda^m\exp\{-\lambda\}\,[c(\cdot)]^{\mathbf{Z}_k}\left[1-p_{D,k}(\boldsymbol{x}_k,\cdot)\right]^{\boldsymbol{M}_k}\prod_{(z,\ell)\in\mathbf{Z}_k^{(\mathrm{id})}}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})g(z|\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})]}\\&+\lambda^m\exp\{-\lambda\}\left[1-p_{D,k}(\boldsymbol{x}_k,\cdot)\right]^{\boldsymbol{M}_k}[c(\cdot)]^{\mathbf{Z}_k}\,.\end{aligned}$$

Grouping and pulling out terms yields the expression

$$\begin{aligned}g(\mathbf{Z}_k|\boldsymbol{x}_k,\boldsymbol{M}_k)=\;&\lambda^m\exp\{-\lambda\}\,[c(\cdot)]^{\mathbf{Z}_k}[1-p_{D,k}(\boldsymbol{x}_k,\cdot)]^{\boldsymbol{M}_k}\\&\times\left[1+\sum_{z\in\mathbf{Z}_k}\sum_{i=1}^{n_M}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})g(z|\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{i,k})]}+\prod_{(z,\ell)\in\mathbf{Z}_k^{(\mathrm{id})}}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})g(z|\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})]}\right]\,.\end{aligned}\tag{33}$$

Equation (33) is the likelihood corresponding to the higher order AFP update, and this will be used (in similar fashion to Section III) with Bayes' rule and a prior to compute a corresponding higher order posterior.

Just as in Eq. (18), we can write Bayes' rule as

$$p(\boldsymbol{x}_k, \boldsymbol{M}_k|\boldsymbol{Z}_k) = \frac{1}{\tilde{\eta}_k} g(\boldsymbol{Z}_k|\boldsymbol{x}_k, \boldsymbol{M}_k) p(\boldsymbol{x}_k, \boldsymbol{M}_k|\boldsymbol{Z}_{k-1}), \tag{34}$$

where, again, $\tilde{\eta}_k$ is the normalizer (adorned with a tilde to differentiate it from Eq. (18)), defined as the integral of the Bayes' rule numerator. As before, assume the vehicle state and the map RFS to be uncorrelated and that the map RFS does not depend on the data collected by the vehicle such that we can write the prior as

$$p(\boldsymbol{x}_k, \boldsymbol{M}_k|\boldsymbol{Z}_{k-1}) = p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) p(\boldsymbol{M}_k)\,.$$

Then, substituting this factorized prior and Eq. (33) into Bayes' rule gives

$$\begin{aligned} p(\boldsymbol{x}_k, \boldsymbol{M}_k|\boldsymbol{Z}_k) = \frac{1}{\tilde{\eta}_k} \lambda^m \exp\{-\lambda\} [c(\cdot)]^{\boldsymbol{Z}_k} [1 - p_{D,k}(\boldsymbol{x}_k, \cdot)]^{\boldsymbol{M}_k} \\ \times \left[1 + \sum_{z \in \boldsymbol{Z}_k} \sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})]} + \prod_{(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})]} \right] p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) p(\boldsymbol{M}_k) \end{aligned} \tag{35}$$

First, compute the normalizer, $\tilde{\eta}_k$. To that end, recall that it is defined as the integral of the Bayes' rule numerator, i.e.,

$$\begin{aligned} \tilde{\eta}_k &= \iint g(\boldsymbol{Z}_k|\boldsymbol{x}_k, \boldsymbol{M}_k) p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) p(\boldsymbol{M}_k) \mathrm{d}\boldsymbol{x}_k \delta\boldsymbol{M}_k \\ &= \iint \lambda^m \exp\{-\lambda\} [c(\cdot)]^{\boldsymbol{Z}_k} [1 - p_{D,k}(\boldsymbol{x}_k, \cdot)]^{\boldsymbol{M}_k} \\ &\quad \times \left[1 + \sum_{z \in \boldsymbol{Z}_k} \sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})]} + \prod_{(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})]} \right] p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) p(\boldsymbol{M}_k) \mathrm{d}\boldsymbol{x}_k \delta\boldsymbol{M}_k\,. \end{aligned}$$

Utilizing the same probability of detection approximation as in Section III,

$$p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}) \approx p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k}),$$

and recalling $\hat{\boldsymbol{M}}_k = \{\boldsymbol{m}_{\zeta,1,k}, \ldots, \boldsymbol{m}_{\zeta,n_M,k}\}$ permits the expression

$$\begin{aligned} \tilde{\eta}_k &= \lambda^m \exp\{-\lambda\} [c(\cdot)]^{\boldsymbol{Z}_k} [1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \cdot)]^{\hat{\boldsymbol{M}}_k} \left(\int p(\boldsymbol{M}_k) \delta\boldsymbol{M}_k \right) \\ &\quad \times \int \left[1 + \sum_{z \in \boldsymbol{Z}_k} \sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k})]} + \prod_{(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}} \frac{p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k}) g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k})]} \right] p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) \mathrm{d}\boldsymbol{x}_k \end{aligned}$$

Assuming the prior state density is Gaussian of the form $p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) = p_g(\boldsymbol{x}_k\,;\, \boldsymbol{m}_{x,k}^-, \boldsymbol{P}_{xx,k}^-)$, the likelihood is Gaussian of the form $g(z|\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell,k}) = p_g(\boldsymbol{x}_k\,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{i,k}), \boldsymbol{P}_{vv,k})$, utilizing Lemma 1 (see Appendix A), and carrying out integration yields

$$\begin{aligned} \tilde{\eta}_k = \lambda^m \exp\{-\lambda\} [c(\cdot)]^{\boldsymbol{Z}_k} [1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \cdot)]^{\hat{\boldsymbol{M}}_k} \\ \times \left\{ 1 + \sum_{z \in \boldsymbol{Z}_k} \sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k}) q(z,i)}{\lambda c(z)[1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,i,k})]} + \left[\frac{p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,\cdot,k}) q(\cdot,\cdot)}{c(\cdot)[1 - p_{D,k}(\boldsymbol{m}_{x,k}^-, \boldsymbol{m}_{\zeta,\cdot,k})]} \right]^{\boldsymbol{Z}_k^{(\mathrm{id})}} \right\}, \end{aligned} \tag{36}$$

where $q(\cdot,\cdot)$ is defined in Eq. (25), which is the final form for the normalizer.

Returning to Eq. (35) and carrying out the same procedure that produced Eq. (36) (i.e., employ the probability of detection approximation, group terms, assume the vehicle state and single-feature likelihood are Gaussian, and apply Lemma 1), substituting Eq. (36) into Eq. (35), and marginalizing out $\boldsymbol{M}_k$, as in Eq. (20), lets us write[¶]

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \underbrace{\theta_k^{-1} p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^-_{x,k}, \boldsymbol{P}^-_{xx,k})}_{\text{“null scan”}} + \underbrace{\sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M} \breve{w}_k(\boldsymbol{z},i) p_g(\boldsymbol{x}_k\,;\,\tilde{\boldsymbol{m}}_{x,k}(\boldsymbol{z},i), \tilde{\boldsymbol{P}}_{xx,k}(\boldsymbol{z},i))}_{\text{“one at a time” (AFP)}} + \underbrace{w_k^{(\mathrm{id})} p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^{(\mathrm{id})}_{x,k}, \boldsymbol{P}^{(\mathrm{id})}_{xx,k})}_{\text{“identified”}}, \tag{37}$$

where[‖]

$$\begin{aligned}
\theta_k &= 1 + \sum_{\boldsymbol{z}\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M} \frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k}) q(\boldsymbol{z},i)}{\lambda c(\boldsymbol{z})[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k})]} + \left[\frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k}) q(\cdot,\cdot)}{c(\cdot)[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k})]}\right]^{\boldsymbol{Z}_k^{(\mathrm{id})}} \\
\breve{w}_k(\boldsymbol{z},i) &= \frac{1}{\theta_k} \cdot \frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k}) q(\boldsymbol{z},i)}{\lambda c(\boldsymbol{z})[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,i,k})]} \\
w_k^{(\mathrm{id})} &= \frac{1}{\theta_k}\left[\frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k}) q(\cdot,\cdot)}{c(\cdot)[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k})]}\right]^{\boldsymbol{Z}_k^{(\mathrm{id})}},
\end{aligned}$$

$\tilde{\boldsymbol{m}}_{x,k}(\boldsymbol{z},i)$ is defined in Eq. (23), $\tilde{\boldsymbol{P}}_{xx,k}(\boldsymbol{z},i)$ is defined in Eq. (24), $q(\boldsymbol{z},i)$ is defined in Eq. (25), and $\boldsymbol{m}^{(\mathrm{id})}_{x,k}$ and $\boldsymbol{P}^{(\mathrm{id})}_{xx,k}$ are obtained by performing sequential Kalman updates to $\boldsymbol{m}^-_{x,k}$ and $\boldsymbol{P}^-_{xx,k}$ using all identified detections $(\boldsymbol{z},\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}$. That is,

$$\begin{aligned}
&\textit{(I. init. mean and cov.)} \begin{cases} \boldsymbol{m}^*_{x,k} = \boldsymbol{m}^-_{x,k} \\ \boldsymbol{P}^*_{xx,k} = \boldsymbol{P}^-_{xx,k} \end{cases} \\
&\textit{(II. for all } (\boldsymbol{z},\ell) \in \boldsymbol{Z}_k^{(id)}\textit{)} \begin{cases} \boldsymbol{K}(\boldsymbol{z},\ell) = \boldsymbol{P}^*_{xx,k}\boldsymbol{H}^T_{x,\ell,k}(\boldsymbol{H}_{x,\ell,k}\boldsymbol{P}^*_{xx,k}\boldsymbol{H}^T_{x,\ell,k} + \boldsymbol{H}_{\zeta,\ell,k}\boldsymbol{P}_{\zeta\zeta,\ell,k}\boldsymbol{H}^T_{\zeta,\ell,k} + \boldsymbol{P}_{vv,k})^{-1} \\ \boldsymbol{m}^*_{x,k} \leftarrow \boldsymbol{m}^*_{x,k} + \boldsymbol{K}(\boldsymbol{z},\ell)[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{m}^*_{x,k}, \boldsymbol{m}_{\zeta,\ell,k})] \\ \boldsymbol{P}^*_{xx,k} \leftarrow [\boldsymbol{I}_{n_x\times n_x} - \boldsymbol{K}(\boldsymbol{z},\ell)\boldsymbol{H}_{x,\ell,k}]\boldsymbol{P}^*_{xx,k}[\boldsymbol{I}_{n_x\times n_x} - \boldsymbol{K}(\boldsymbol{z},\ell)\boldsymbol{H}_{x,\ell,k}]^T + \boldsymbol{K}(\boldsymbol{z},\ell)\boldsymbol{P}_{vv,k}\boldsymbol{K}^T(\boldsymbol{z},\ell) \end{cases} \\
&\textit{(III. updated mean and cov.)} \begin{cases} \boldsymbol{m}^{(\mathrm{id})}_{x,k} = \boldsymbol{m}^*_{x,k} \\ \boldsymbol{P}^{(\mathrm{id})}_{xx,k} = \boldsymbol{P}^*_{xx,k} \end{cases}
\end{aligned}$$

For proof that the rightmost term of Eq. (37) indeed produces these sequential Kalman updates, see Appendix B.

Just as with Eq. (28), Eq. (37) is a GM and is ultimately approximated via its mean and covariance found via the method of moments. That is, the final estimate provided by the higher order AFP update is the posterior

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) \approx p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}^+_{x,k}, \boldsymbol{P}^+_{xx,k}),$$

where $\boldsymbol{m}^+_{x,k}$ and $\boldsymbol{P}^+_{xx,k}$ are the mean and covariance resulting from applying the method of moments to the GM in Eq. (37). Similar to η_k in Eq. (20), θ_k serves as the normalizer for the combinations of the higher order update's weights, ensuring that $p(\boldsymbol{x}_k|\boldsymbol{Z}_k)$ is a proper pdf. This completes the higher order AFP update, and $\boldsymbol{m}^+_{x,k}$ can $\boldsymbol{P}^+_{xx,k}$ then be used for further sequential filtering operations (propagation/update cycles).

E. Discussion on the Higher Order AFP Update

An inspection of Eq. (37) permits the conclusion that the posterior, $p(\boldsymbol{x}_k|\boldsymbol{Z}_k)$ is a GM with components directly related to the three approximations developed for the higher order AFP update:

[¶]The steps are omitted here but they are identical to the developments of Section III and the derivation of Eq. (36).

[‖]Note that $\breve{w}_k(\boldsymbol{z},i)$ here does not equal $w_k(\boldsymbol{z},i)$ in Eq. (22). Equation (24) has been normlized by its first term, whereas $w_k(\boldsymbol{z},i)$ has to be normalized by its higher order normalizer: θ_k.

- The first component is a Gaussian with a weight of θ_k^{-1}, corresponding to the "null scan" term. Note that the prior is unmodified in this case, leaving simply the term $p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}_{x,k}^-, \boldsymbol{P}_{xx,k}^-)$, the prior Gaussian, which is intuitive in that clutter should not provide a change to the vehicle state estimate. Should $\boldsymbol{Z}_k$ truly be a null scan, this term permits the higher order AFP update to correctly discard all the measurements and leave the prior unaffected.
- The second component is an $n_M \cdot |\boldsymbol{Z}_k|$ GM resulting from standard AFP updates (the "one at a time" approximation), the same as was found in Section III and [3]. This GM is identical to the GM in Eq. (21) but that its weights, $\breve{w}_k(z,\ell)$, are now normalized by θ_k and not $\sum_{z\in\boldsymbol{Z}_k}\sum_{i=1}^{n_M}\tilde{w}_k(z,\ell)$ (i.e., the first term in Eq. (25)). This term permits the higher order AFP update to perform meaningful updates in the event that (i) no identification information is available (i.e., $\boldsymbol{Z}_k^{(\mathrm{id})} = \emptyset$) or (ii) the identification information provided by the pre-processor is poor (i.e., $\boldsymbol{Z}_k^{(\mathrm{id})}$ contains misidentfications).
- The third component is a single Gaussian resulting from sequential Kalman updates according to all *identified* measurement-label pairs, i.e., $(z,\ell) \in \boldsymbol{Z}_k^{(\mathrm{id})}$. This components weight, $w_k^{(\mathrm{id})}$, is how the higher order update assesses how valuable and reliable each (z,ℓ) is, and this is primarily performed via weighting with the term $q(z,\ell)$. Since $w_k^{(\mathrm{id})}$ consists of *products*, this term rapidly vanishes if any of the associations are very unreliable (i.e., $q(z,\ell)$ is very small), relegating all of the "strength" of the update to the "null scan" and "one at a time" (AFP) components. Conversely, if all (z,ℓ) are deemed viable (i.e., each $q(z,\ell)$ is of a non-negligible, non-zero value), their product dominates the GM Eq. (37) and this results in a very strong update (an approximately optimal update over $\boldsymbol{Z}_k^{(\mathrm{id})}$, in fact, since it is due to the Kalman update).

The intent of this higher order update is to seek the "best of both worlds": an update which is invulerable to misidentification errors, performs well in the absense of identification entirely, but is capable of safely utilizing any idenfication data should it be available. The hope at the outset was to have an update which autonomously "blends" the Kalman filter and the standard AFP update based on the quality of incoming identification data. Equation (37) suggests that the higher order AFP update should behave in a somewhat bounded fashion between the Kalman filter and the standard AFP update, with the solution approaching the Kalman filter as identification quality is high and trending toward the standard AFP solution when identification quality is poor. Section V explores this hypothesis numerically.

V. Terrain Relative Navigation Simulation

This section compares the new, higher order AFP update with an idealized (i.e., perfect identification) extended Kalman filter (EKF) and the identification-independent AFP update. Accordingly, consider a spacecraft in a circular, equatorial, 100 [km] altitude orbit about the Moon. Let this spacecraft be equipped with a camera having a field of view of 48.75°, and let the spacecraft's orientation be such that this camera is always pointed in the nadir direction (i.e., directly pointed to the surface below at all times). Each filter estimates the same state consisting of the inertial, position, velocity, and attitude of the vehicle such that

$$\boldsymbol{x}_k = \begin{bmatrix} \boldsymbol{r}_{\mathrm{icrf}\to\mathrm{veh}}^{\mathrm{icrf}} \\ \boldsymbol{v}_{\mathrm{icrf}\to\mathrm{veh}}^{\mathrm{icrf}} \\ \bar{\boldsymbol{q}}_{\mathrm{icrf}\to\mathrm{veh}} \end{bmatrix},$$

where $\boldsymbol{r}_{\mathrm{icrf}\to\mathrm{veh}}^{\mathrm{icrf}}$ is the position of the vehicle wrt the International Celestial Reference Frame (ICRF) origin coordinatized in the ICRF frame, $\boldsymbol{v}_{\mathrm{icrf}\to\mathrm{veh}}^{\mathrm{icrf}}$ is the velocity of the vehicle wrt the ICRF frame coordinatized in the ICRF frame, and $\bar{\boldsymbol{q}}_{\mathrm{icrf}\to\mathrm{veh}}$ is the attitude quaternion describing the vehicle's orientation within the ICRF frame. For simplicity, consider the position and orientation of the camera with respect to the vehicle frame to be such that

$$\begin{aligned} \boldsymbol{r}_{\mathrm{veh}\to\mathrm{cam}}^{\mathrm{veh}} &= [0 \quad 0 \quad 0]^T \\ \bar{\boldsymbol{q}}_{\mathrm{veh}\to\mathrm{cam}} &= \bar{\boldsymbol{q}}_{\boldsymbol{I}}\,, \end{aligned}$$

where $\bar{\boldsymbol{q}}_{\boldsymbol{I}}$ is the identity quaternion.

Let the spacecraft be equipped with a crater detector such that craters in the collected images are used for TRN (note that this detector is described in the next paragraph). The spacecraft considers an extensive crater catalog consisting of a pared-down collection of the 1.3 million crater catalog distributed in [8], this collection being pared down to only consider those craters which are nearly equatorial (±3° latitude). The result is a catalog consisting of 30,374 craters utilized for navigation. Figure 3 depicts the crater catalog employed by all filters in these studies, where blue markers correspond to the center of each cataloged crater.

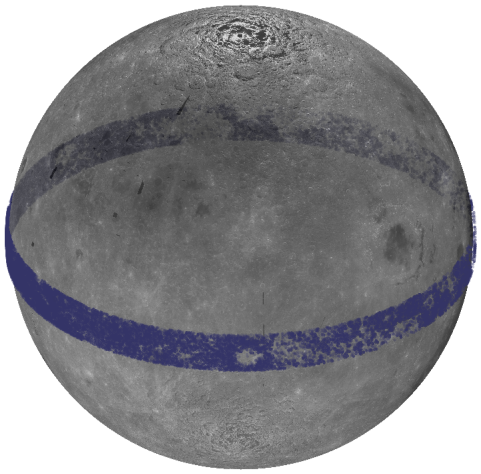


Fig. 3 Graphical depiction of the crater catalog utilized by the spacecraft in its equatorial orbit.

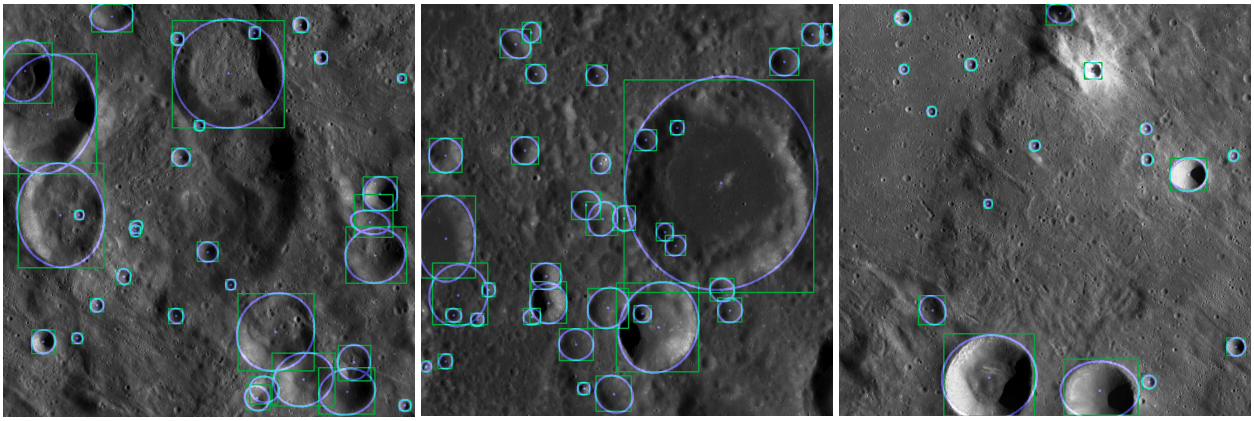


Fig. 4 Example outputs of the machine learning-based detector [10] on the simulated imagery constructed from LROC global maps [9].

The imagery considered in this simulation is simulated using cropped image sections from the Lunar Reconnaissance Orbiter Camera's Global Morphologic Maps [9], where the cropping is performed based on the varying trajectory of the vehicle/camera. Examples of these images can be seen in Fig. 4. This provides the system with images which can then be subjected to detection; that is, the images can be provided to a detection algorithm to find craters within an image. The detector considered in this paper is a machine-learning-based detector, based upon the Mask Region-based Convolutional Neural Network framework and is described in [10]. For the present discussion, it suffices to say that the detector receives one of the aforementioned images as input, and it attempts to locate craters within those images. Examples of detected output can also be seen in Fig. 4. So, given these detections, each filter considered in this study will compare its onboard catalog to the detections provided by the detector.

Detections alone are sufficient to successfully perform AFP-based measurement updates. However, recall that the EKF requires these detections to be further identified to successfully perform an update. That is, detections alone are insufficient for an EKF update, and the EKF must also be told specifically *which* crater a given detection belongs to in order to update its state. Additionally, the higher order AFP update derived in this paper seeks to utilize identification information as well. Accordingly, an identification pre-processor is applied to the detection output. In this case, knowledge of ground truth allows the true detection-to-crater associations to be computed. Then, a probability of successful labeling, p_L, is specified. Each detection is subjected to the test

$$u < p_L \,,$$

where u is a random sample from the standard uniform distribution. If this condition passes, the detection is taken to be perfectly identified, and the correct crater identity is passed along with the detection. If this condition fails, the detection is said to be *misidentified*, and its label switches with a neighboring crater. The purpose of this is to emulate a failing identification process such that its impact on the new update can be quantified. So, as examples, $p_L = 1$ suggests that all identification results passed to the filters are error-free, $p_L = 0.5$ suggests that the identification pre-processor is wrong approximately half the time, and $p_L = 0$ suggests that the identification pre-processor never successfully identifies detections.

All filters in this study accept the crater center as the measurement to be processed. This is processed in the form of horizontal and vertical bearing angles from the camera to that crater center, defined as

$$\alpha_h = \text{atan2}\left\{x^{\text{cam}}_{\text{cam}\to\zeta_{i,k}},\ z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}\right\}$$
$$\alpha_v = \text{atan2}\left\{y^{\text{cam}}_{\text{cam}\to\zeta_{i,k}},\ z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}\right\},$$

respectively, where

$$\boldsymbol{r}^{\text{cam}}_{\text{cam}\to\zeta_{i,k}} = \begin{bmatrix} x^{\text{cam}}_{\text{cam}\to\zeta_{i,k}} \\ y^{\text{cam}}_{\text{cam}\to\zeta_{i,k}} \\ z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}} \end{bmatrix} = \boldsymbol{T}_{\text{icrf}\to\text{cam}}\left(\boldsymbol{T}_{\text{mcmf}\to\text{icrf}}\boldsymbol{\zeta}^{\text{mcmf}}_{i,k} - \boldsymbol{r}^{\text{icrf}}_{\text{icrf}\to\text{cam}}\right),$$

$\boldsymbol{T}_{\text{icrf}\to\text{cam}} = \boldsymbol{T}_{\text{icrf}\to\text{cam}}(\bar{\boldsymbol{q}}_{\text{icrf}\to\text{veh}}, \bar{\boldsymbol{q}}_{\text{veh}\to\text{cam}})$ is the DCM corresponding to the orientation of the camera within the ICRF frame, $\boldsymbol{T}_{\text{mcmf}\to\text{icrf}}$ is the DCM corresponding to Moon-Centered Moon-Fixed (MCMF) coordinate frame orientation with respect to the ICRF frame, $\boldsymbol{\zeta}^{\text{mcmf}}_{i,k}$ is the position vector of the i^{th} catalog feature coordinatized in the MCMF frame (taken from the crater catalog), and $\boldsymbol{r}^{\text{icrf}}_{\text{icrf}\to\text{cam}}$ is the position of the camera with respect to the ICRF frame coordinatized in the ICRF frame. The corresponding measurement Jacobian can be found to be

$$\boldsymbol{H}_{x,i,k} = \begin{bmatrix} -\boldsymbol{H}_{r,i}\boldsymbol{T}_{\text{icrf}\to\text{cam}} & \boldsymbol{0}_{3\times3} & \boldsymbol{H}_{r,i}[\boldsymbol{r}^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}\times] \end{bmatrix}$$
$$\boldsymbol{H}_{r,i} = \begin{bmatrix} \frac{z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}}{(x^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2+(z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2} & 0 & \frac{-x^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}}{(x^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2+(z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2} \\ 0 & \frac{z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}}{(y^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2+(z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2} & \frac{-y^{\text{cam}}_{\text{cam}\to\zeta_{i,k}}}{(y^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2+(z^{\text{cam}}_{\text{cam}\to\zeta_{i,k}})^2} \end{bmatrix},$$

where $[\boldsymbol{a}\times]$ is the skew-symmetric matrix formed by the 3-vector $\boldsymbol{a}$.

Let the vehicle collect and process images every $\Delta t = 5$ seconds over the span of three full orbits. Three filter configurations are compared, and each are initialized and parameterized identically:

1) **Case "EKF":** an idealized implementation of the extended Kalman filter. This case is the Kalman filter provided *perfectly identified measurements* at all times (i.e., $p_L = 1$), meaning this solution serves as a lower error bound as point of comparison. Previous work (see [3]) has demonstrated the divergent behavior of the EKF in the presence of misidentification errors. This study makes no effort to reiterate this but, instead, utilizes the EKF case as an ideal, (approximately) optimal estimator to serve as a baseline for comparison.
2) **Case "AFP":** the standard AFP update derived in [3] and Section III. Note that this procedure does not rely upon identification pre-processing in any way, so no matter the probability of successful identification, given identical measurements, identical results are produced (i.e., p_L is irrelevant).
3) **Case "AFP2":** the new higher order AFP update derived in Section IV, where the probability of successful identification is varied from $p_L = 1$ down to $p_L = 0$ in increments of 0.1. The results are compared to the EKF and AFP cases to determine if the hypothesis posited at the end of Section IV holds.

Let the initial conditions of the spacecraft be such that

$$\boldsymbol{r}^{\text{icrf}}_{\text{icrf}\to\text{veh}}(t_0) = [1837400 \quad 0 \quad 0]^T \quad \text{[m]}$$
$$\boldsymbol{v}^{\text{icrf}}_{\text{icrf}\to\text{veh}}(t_0) = [0 \quad 1633.5 \quad 0]^T \quad \text{[m/s]}$$
$$\bar{\boldsymbol{q}}_{\text{icrf}\to\text{veh}}(t_0) = \begin{bmatrix}\boldsymbol{q}_v^T & q_s\end{bmatrix}^T = [0.5 \quad 0.5 \quad -0.5 \quad -0.5]^T.$$

These studies consider initial position, velocity, and attitude 3σ uncertainties of $10,000$ [m], $1,000$ [m/s], and 5 [deg], respectively. Therefore, the initial state covariance is given as

$$\boldsymbol{P}_{xx,0} = \text{blkdiag}\left\{(\tfrac{1}{3}10,000)^2\boldsymbol{I}_{3\times3},\ (\tfrac{1}{3}1,000)^2\boldsymbol{I}_{3\times3},\ (\tfrac{1}{3}5)^2\boldsymbol{I}_{3\times3}\right\}.$$

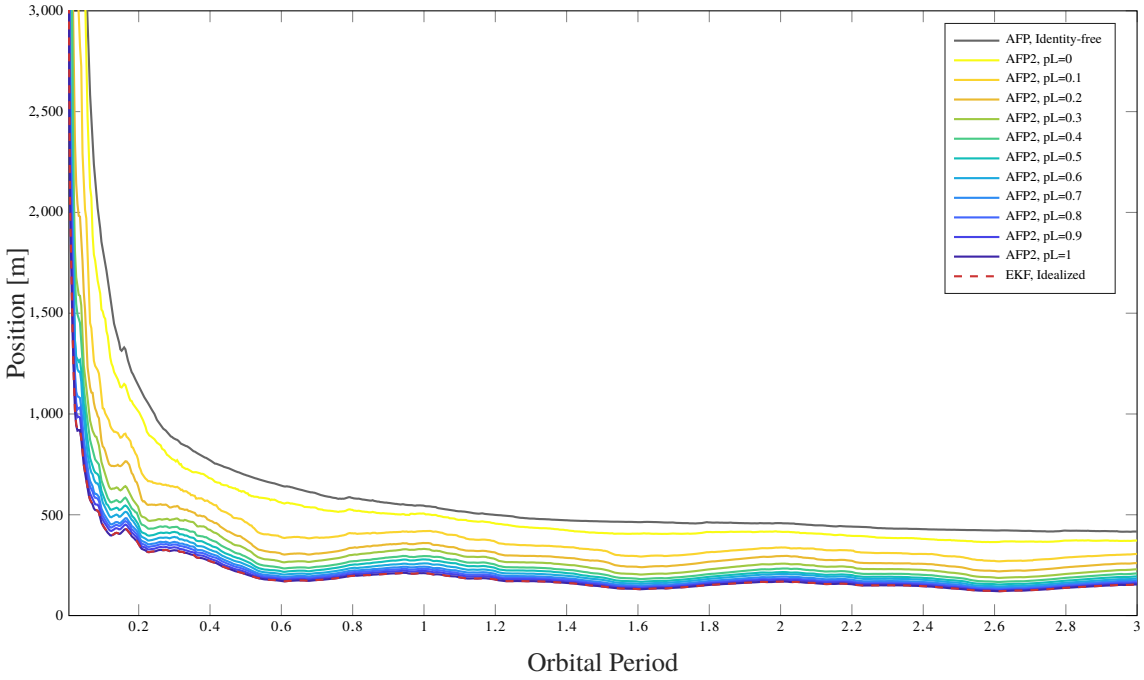


Fig. 5 Norm 3σ performance for the position states.

Letting $\boldsymbol{x}_0$ denote the true initial state, the each filter's initial state estimate, $\boldsymbol{m}_{x,0}$, is computed by drawing a random sample from a multivariate Gaussian centered at $\boldsymbol{x}_0$ with covariance $\boldsymbol{P}_{xx,0}$. Each filter's process noise covariance, $\boldsymbol{P}_{ww,k}$, is set using the state noise compensation model [11]

$$\boldsymbol{P}_{ww,k} = \boldsymbol{P}_{ww} = \begin{bmatrix} \frac{1}{3}q_{\text{trans}}\Delta t^3\boldsymbol{I}_{3\times3} & \frac{1}{2}q_{\text{trans}}\Delta t^2\boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \frac{1}{2}q_{\text{trans}}\Delta t^2\boldsymbol{I}_{3\times3} & q_{\text{trans}}\Delta t\boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & q_{\text{rot}}\Delta t\boldsymbol{I}_{3\times3} \end{bmatrix},$$

where $q_{\text{trans}} = 1 \times 10^{-9}$, $q_{\text{rot}} = 1 \times 10^{-9}$, and Δt is the length of time of the propagation interval between measurements (here, $\Delta t = 5$ [sec]). The horizontal and vertical bearing measurement noise covariance is taken to be $\boldsymbol{R} = (0.5/3)^2\boldsymbol{I}_{2\times2}$ [deg^2]. In addition, all filters apply measurement underweighting with an underweighting parameter of 0.2 [2].

Consider a trial of the previously described scenario, with the 3 orbits resulting in a total of 8,481 images to be processed. Three times the root-sum-square of the position, velocity, and attitude covariance (i.e., three times the square-root of the trace of the corresponding blocks of $\boldsymbol{P}^+_{xx,k}$) estimated by each filter can be seen in Figs. 5, 6, and 7, respectively. A reader can loosely interpret this as 3σ intervals corresponding to the norm estimation error of position, velocity, and attitude. Immediately, a striking trend becomes apparent: as hypothesized, the AFP2 solution remains "bounded" below by the EKF and above by the standard AFP solution depending on the rate of successful identification (i.e., the value of p_L). As identification is successful, $p_L \to 1$, AFP2 converges to the solution estimated by the EKF. As identification degrades, $p_L \to 0$, AFP2 converges to the standard AFP update. This suggests that, as desired, the higher order AFP update (AFP2) meets the criteria upon which this investigation started: an update which can utilize identification information should it be available and reliable, but which demonstrates stable estimation behavior when identification degrades.

Note that the adaptive behavior of AFP2 is possible with zero knowledge of the identification success rate, p_L. Said another way, AFP2 produces these results without knowledge of p_L as an input parameter. The mathematics employed to produce and implement Eq. (37) produces an estimator which is naturally adaptive to the quality of the incoming identified measurements within $\boldsymbol{Z}^{(\text{id})}_k$.

Note, too, that this is a study considering a single trial and not a comprehensive Monte Carlo study. Future work involves assessing the Monte Carlo performance of AFP2 to compare it with known trends produced by the EKF and standard AFP update. The purpose of the present study is to assess the feasibility of the new higher order AFP update and to determine if its behavior is in family with the behavior hypothesized at the outset of this effort. The numerical

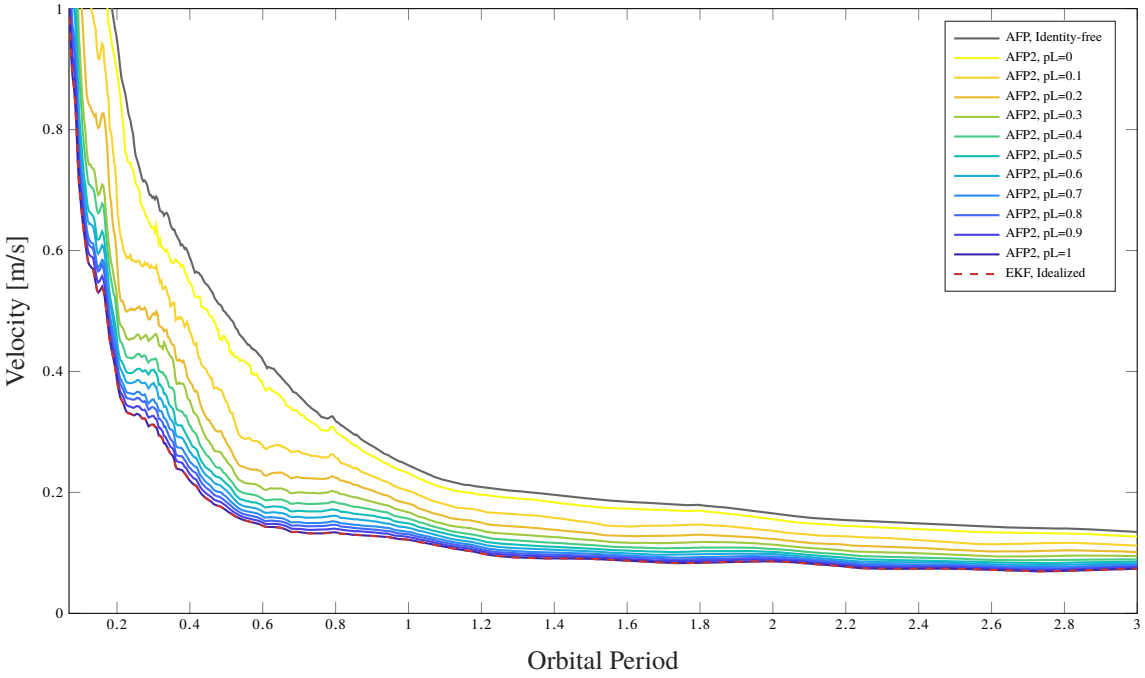


Fig. 6 Norm 3σ performance for the velocity states.

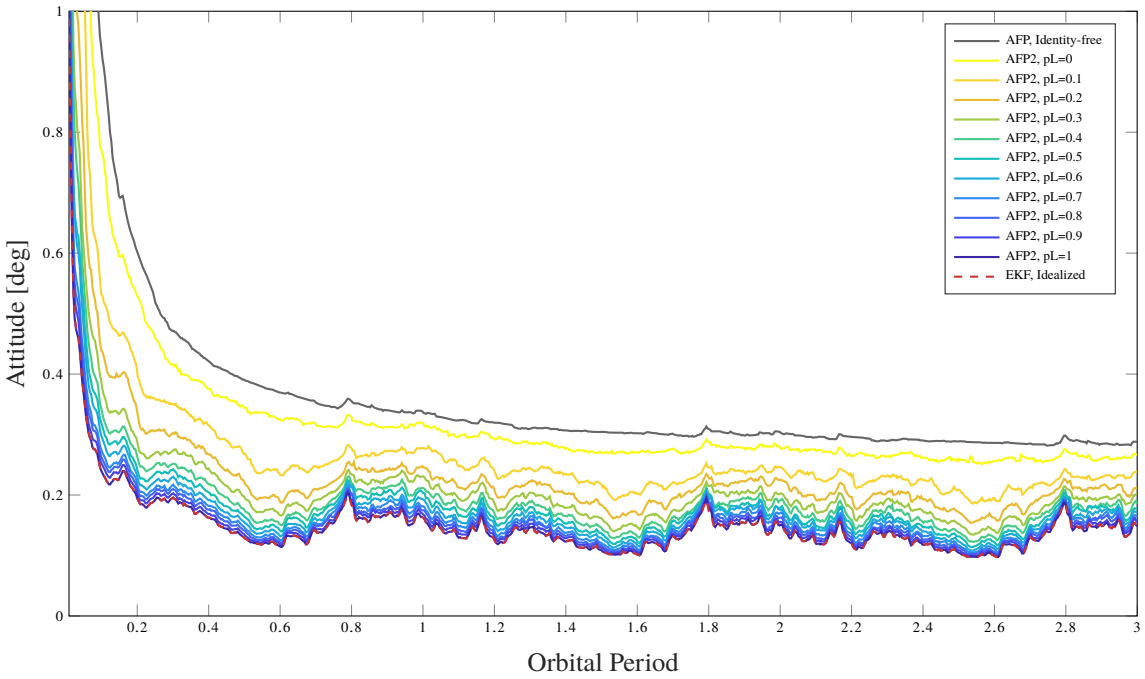


Fig. 7 Norm 3σ performance for the attitude states.

results presented here buoy the belief that AFP2 "marries" the EKF and AFP solutions for these types of data, via an adaptive continuum of AFP2 solutions, and the author believes that AFP2 deserves further study.

VI. Conclusions

Measurements which require identification pre-processing to perform measurement-to-feature association for feature-based navigation have been demonstrated to cause a myriad of issues for navigation filters. Previous work sought to address this by producing the anonymous feature processing (AFP) update which forgoes the identification pre-processing entirely in favor of allowing the mathematics internal to the filter to assess associations directly. However, should identified data become available, the AFP update has no mechanism by which to extract this information. This paper derived a new, higher order AFP update which is designed to operate when identified data is available but is also insensitive to errors in this identification process when they appear. Numerical results suggest an intriguing trend that the higher order AFP update produces a continuum of solutions between that provided by the extended Kalman filter and the standard AFP update, converging to the former when identification rates are high and to the latter when they are low.

VII. Acknowledgments

The author would like to thank Sofia Catalan for the development of the simulated lunar imagery with the LROC global data and for preparation and pre-processing of the images with their machine learning-based crater detector.

A. A Useful Property of Gaussians

Lemma 1 (Ho's Equation: Product Form) *Given terms of appropriate dimension and provided that* $\boldsymbol{R}$ *and* $\boldsymbol{P}$ *are positive definite, [12]*

$$p_g(z\,;\,\boldsymbol{H}\boldsymbol{x}+\boldsymbol{b},\boldsymbol{R})p_g(\boldsymbol{x}\,;\,\boldsymbol{m},\boldsymbol{P}) = q(z)p_g(\boldsymbol{x}\,;\,\boldsymbol{\mu},\boldsymbol{\Sigma}),$$

where

$$\begin{aligned} q(z) &= p_g(z\,;\,\boldsymbol{H}\boldsymbol{m},\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{R}) \\ \boldsymbol{\mu} &= \boldsymbol{m}+\boldsymbol{K}(z-\boldsymbol{H}\boldsymbol{m}-\boldsymbol{b}) \\ \boldsymbol{\Sigma} &= (\boldsymbol{I}-\boldsymbol{K}\boldsymbol{H})\boldsymbol{P} \\ \boldsymbol{K} &= \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{R})^{-1}. \end{aligned}$$

It is common to utilize a second form for $\boldsymbol{\Sigma}$*, commonly referred to as Joseph's form, that has more desirable numerical properties and is given as*

$$\boldsymbol{\Sigma} = (\boldsymbol{I}-\boldsymbol{K}\boldsymbol{H})\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{H})^T+\boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T.$$

B. The Third Term in Eq. (37)

It was claimed that the third term in Eq. (37) corresponds to a weighted result of sequential Kalman updates according to identified observations $(z,\ell)\in\boldsymbol{Z}_k^{(\mathrm{id})}$. To see this, consider only the relevant term in Eq. (35):

$$\prod_{(z,\ell)\in\boldsymbol{Z}_k^{(\mathrm{id})}}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})g(z|\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})]}p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1}) = \prod_{(z,\ell)\in\boldsymbol{Z}_k^{(\mathrm{id})}}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})p_g(z\,;\,\boldsymbol{h}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k}),\boldsymbol{P}_{vv,k})p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}_{x,k}^-,\boldsymbol{P}_{xx,k}^-)}{\lambda c(z)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell,k})]}.$$

It will be demonstrated that this product actually corresponds to a weighted sequential application of the Kalman filter according to all (z,ℓ). To accomplish this, recall that $\boldsymbol{Z}_k^{(\mathrm{id})}=\{(z_1,\ell_1),\ldots,(z_{m_{\mathrm{id}}},\ell_{m_{\mathrm{id}}})\}$ such that we can index over $j=\{1,\ldots,m_{\mathrm{id}}\}$ and write

$$\prod_{j=1}^{m_{\mathrm{id}}}\frac{p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell_j,k})p_g(z_j\,;\,\boldsymbol{h}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell_j,k}),\boldsymbol{P}_{vv,k})p_g(\boldsymbol{x}_k\,;\,\boldsymbol{m}_{x,k}^-,\boldsymbol{P}_{xx,k}^-)}{\lambda c(z_j)[1-p_{D,k}(\boldsymbol{x}_k,\boldsymbol{\zeta}_{\ell_j,k})]}. \tag{38}$$

Consider only the first term ($j = 1$), given as

$$\frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k}) p_g(z_1 \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k}), \boldsymbol{P}_{vv,k}) p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^-_{x,k}, \boldsymbol{P}^-_{xx,k})}{\lambda c(z_1)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k})]} \,.$$

Applying Lemma 1 to the product of Gaussians gives

$$\underbrace{\frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k}) q(z_1, \ell_1)}{\lambda c(z_1)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k})]}}_{\text{weighting}} \underbrace{p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{P}^{(1)}_{xx,k})}_{\text{Kalman update}}, \tag{39}$$

where

$$\begin{aligned}
q(z_1, \ell_1) &= p_g(z_1 \,;\, \boldsymbol{h}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\ell_1,k}), \boldsymbol{H}_{x,\ell_1,k} \boldsymbol{P}^-_{xx,k} \boldsymbol{H}^T_{x,\ell_1,k} + \boldsymbol{P}_{vv,k}) \\
\boldsymbol{m}^{(1)}_{x,k} &= \boldsymbol{m}^-_{x,k} + \boldsymbol{K}^{(1)}[z_1 - \boldsymbol{h}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\ell_1,k})] \\
\boldsymbol{P}^{(1)}_{xx,k} &= (\boldsymbol{I}_{n_x \times n_x} - \boldsymbol{K}^{(1)} \boldsymbol{H}_{x,\ell_1,k}) \boldsymbol{P}^-_{xx,k} (\boldsymbol{I}_{n_x \times n_x} - \boldsymbol{K}^{(1)} \boldsymbol{H}_{x,\ell_1,k})^T + \boldsymbol{K}^{(1)} \boldsymbol{P}_{vv,k} (\boldsymbol{K}^{(1)})^T \\
\boldsymbol{K}^{(1)} &= \boldsymbol{P}^-_{xx,k} \boldsymbol{H}^T_{x,\ell_1,k} (\boldsymbol{H}_{x,\ell_1,k} \boldsymbol{P}^-_{xx,k} \boldsymbol{H}^T_{x,\ell_1,k} + \boldsymbol{P}_{vv,k})^{-1} \,.
\end{aligned}$$

Note that, weighting aside, Eq. (39) is mathematically identical to a Kalman update applied according to (z_1, ℓ_1)! Substituting Eq. (39) into Eq. (38) yields

$$\prod_{j=2}^{m_{\text{id}}} \frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_j,k}) p_g(z_j \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_j,k}), \boldsymbol{P}_{vv,k})}{\lambda c(z_j)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_j,k})]} \left(\frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k}) q(z_1, \ell_1)}{\lambda c(z_1)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_1,k})]} p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{P}^{(1)}_{xx,k}) \right) \,.$$

Terms not involving Gaussians can be grouped as

$$\left[\frac{p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\cdot,k}) p_g(\cdot \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\cdot,k}), \boldsymbol{P}_{vv,k})}{\lambda c(\cdot)[1 - p_{D,k}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\cdot,k})]} \right]^{\boldsymbol{z}^{(\text{id})}_k} q(z_1, \ell_1) \left(\prod_{j=2}^{m_{\text{id}}} p_g(z_j \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_j,k}), \boldsymbol{P}_{vv,k}) \right) p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{P}^{(1)}_{xx,k}) \,.$$

Then, stripping the $j = 2$ term and evaluating the product

$$p_g(z_2 \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_2,k}), \boldsymbol{P}_{vv,k}) p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{P}^{(1)}_{xx,k})$$

yields

$$p_g(z_2 \,;\, \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\zeta}_{\ell_2,k}), \boldsymbol{P}_{vv,k}) p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{P}^{(1)}_{xx,k}) = q(z_2, \ell_2) p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(2)}_{x,k}, \boldsymbol{P}^{(2)}_{xx,k})$$

where

$$\begin{aligned}
q(z_2, \ell_2) &= p_g(z_2 \,;\, \boldsymbol{h}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\ell_2,k}), \boldsymbol{H}_{x,\ell_2,k} \boldsymbol{P}^{(1)}_{xx,k} \boldsymbol{H}^T_{x,\ell_2,k} + \boldsymbol{P}_{vv,k}) \\
\boldsymbol{m}^{(2)}_{x,k} &= \boldsymbol{m}^{(1)}_{x,k} + \boldsymbol{K}^{(2)}[z_2 - \boldsymbol{h}(\boldsymbol{m}^{(1)}_{x,k}, \boldsymbol{m}_{\zeta,\ell_2,k})] \\
\boldsymbol{P}^{(2)}_{xx,k} &= (\boldsymbol{I}_{n_x \times n_x} - \boldsymbol{K}^{(2)} \boldsymbol{H}_{x,\ell_2,k}) \boldsymbol{P}^{(1)}_{xx,k} (\boldsymbol{I}_{n_x \times n_x} - \boldsymbol{K}^{(2)} \boldsymbol{H}_{x,\ell_2,k})^T + \boldsymbol{K}^{(2)} \boldsymbol{P}_{vv,k} (\boldsymbol{K}^{(2)})^T \\
\boldsymbol{K}^{(2)} &= \boldsymbol{P}^{(1)}_{xx,k} \boldsymbol{H}^T_{x,\ell_2,k} (\boldsymbol{H}_{x,\ell_2,k} \boldsymbol{P}^{(1)}_{xx,k} \boldsymbol{H}^T_{x,\ell_2,k} + \boldsymbol{P}_{vv,k})^{-1} ,
\end{aligned}$$

which is just another weighted Kalman update, this time to $\boldsymbol{m}^{(1)}_{x,k}$ and $\boldsymbol{P}^{(1)}_{xx,k}$ according to (z_2, ℓ_2). Continuing sequentially in this fashion (that is, performing the Gaussian products and collecting any non-Gaussian terms) until $j = m_{\text{id}}$ yields (utilizing definitions from Section IV)

$$\frac{1}{\theta_k} \left[\frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k}) q(\cdot,\cdot)}{c(\cdot)[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k})]} \right]^{\boldsymbol{z}^{(\text{id})}_k} p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(m_{\text{id}})}_{x,k}, \boldsymbol{P}^{(m_{\text{id}})}_{xx,k}) = \frac{1}{\theta_k} \left[\frac{p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k}) q(\cdot,\cdot)}{c(\cdot)[1 - p_{D,k}(\boldsymbol{m}^-_{x,k}, \boldsymbol{m}_{\zeta,\cdot,k})]} \right]^{\boldsymbol{z}^{(\text{id})}_k} p_g(\boldsymbol{x}_k \,;\, \boldsymbol{m}^{(\text{id})}_{x,k}, \boldsymbol{P}^{(\text{id})}_{xx,k})$$

as claimed.

References

[1] Johnson, A., Aaron, S., Chang, J., Cheng, Y., Montgomery, J., Mohan, S., Schroeder, S., Tweddle, B., Trawny, N., and Zheng, J., "The Lander Vision System for Mars 2020 Entry Descent and Landing," 2017.

[2] Lear, W. M., "Multi-Phase Navigation Program for the Space Shuttle Orbiter," Tech. rep., 1973. NASA Internal Note 73-FM-132.

[3] McCabe, J. S., and DeMars, K. J., "Anonymous Feature-Based Terrain Relative Navigation," *Journal of Guidance, Control, and Dynamics*, 2019. doi:10.2514/1.G004423.

[4] McCabe, J. S., and DeMars, K. J., "Anonymous Feature Processing for Efficient Onboard Navigation," *AIAA SciTech Forum*, 2020. doi:10.2514/6.2020-0598.

[5] Mahler, R. P., *Advances in Statistical Multisource-Multitarget Information Fusion*, Artech House, Inc., Norwood, MA, USA, 2014.

[6] DeMars, K. J., "Nonlinear Orbit Uncertainty Prediction and Rectification for Space Situational Awareness," Ph.D. thesis, The University of Texas at Austin, Austin, Texas, 2010.

[7] McCabe, J. S., and DeMars, K. J., "Terrain Relative Navigation With Anonymous Features," *AIAA SciTech Forum*, 2019. doi:10.2514/6.2018-1332.

[8] Robbins, S. J., "A New Global Database of Lunar Impact Craters >12km: 1. Crater Locations and Sizes, Comparisons With Published Databases, and Global Analysis," *Journal of Geophysical Research: Planets*, Vol. 124, No. 4, 2019, pp. 871–892. doi:https://doi.org/10.1029/2018JE005592, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JE005592`.

[9] Speyerer, E., Robinson, M., and Denevi, B., "Lunar Reconnaissance Orbiter Camera Global Morphological Map of the Moon," 2011.

[10] Catalan, S. G., McCabe, J. S., and Jones, B. A., "Implementation of Machine Learning Methods for Crater-Based Navigation," *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference*, 2021.

[11] Tapley, B. D., Schutz, B. E., and Born, G. H., *Statistical Orbit Determination*, Elselvier Academic Press, New York, NY, 2004. Chapters 5 and 6.

[12] Ho, Y. C., and Lee, R., "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, Vol. 9, No. 4, 1964, pp. 333–339. doi:10.1109/TAC.1964.1105763.