UKF on Manifold - Notes

Luca Di Giammarino

May 2021

1 Introduction

These notes have been produced using multiple works presented in literature about Bayesian estimation and, more, in particular, Kalman filters with Unscented Transform on manifolds. The one described in detail is one of the more recent ones that seems to perform better in multiple scenarios. We start with an introduction to manifolds. Then, we present a more practical way to represent uncertainty in this space. We present in detail the current "best" UKF present in literature. Finally, we provide multiple examples and give some final remarks.

2 Manifolds as State Representation

Using a manifold has been preferred to solve the issues mentioned above. In this section, we regard a manifold as a black box encapsulating a certain (topological) structure. In practice, it will be a subset of $\mathbb{R}^s$, subject to constraints such as the *orthonormality* of a 3×3 rotation matrix that can be used as a representation of $SO(3)$. As for the representation of states, in the simple case where a state consists of a single component only, e.g. a three-dimensional orientation ($SO(3)$), we can represent a state as a single manifold. If a state consists of multiple different components, we can also represent it as a manifold since the Cartesian product [6] of the manifolds representing each component yields another compound manifold. Essentially, we can build sophisticated compound manifolds starting with a set of manifold primitives.

3 $\boxplus$-Manifold

The second important property of manifolds is that they are locally *homeomorphic* to $\mathbb{R}^n$, i.e., informally speaking, we can establish a bidirectional mapping from a local neighborhood in an n-manifold to $\mathbb{R}^n$. The $\boxplus$-method uses two encapsulation operators $\boxplus$ ("boxplus") and $\boxminus$ ("boxminus") to implement this mapping:

$$\boxplus_{\mathcal{S}} : \mathcal{S} \times \mathbb{R}^n \rightarrow \mathcal{S} \tag{1}$$

$$\boxminus_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^n \tag{2}$$

The subscript $\mathcal{S}$ is omitted when clear from the context. The operation $y = x \boxplus \delta$ adds a small perturbation as a vector $\delta \in \mathbb{R}^n$ to the state $x \in \mathcal{S}$. Conversely, $\delta = y \boxminus x$ determines the perturbation vector δ which yields y when $\boxplus$ added to x. Axiomatically, this is captured by the definition below.

3.1 Definition $\boxplus$-Manifold

A $\boxplus$-manifold is a quadruple $(\mathcal{S}, \boxplus, \boxminus, V)$ (usually referred to as just $\mathcal{S}$), consisting of a subset $\mathcal{S} \subset \mathbb{R}^s$

$$\boxplus : \mathcal{S} \times \mathbb{R}^n \to \mathcal{S} \tag{3}$$

$$\boxminus : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^n \tag{4}$$

and an open neighborhood $V \subset \mathbb{R}^n$ of 0. These data are subject to the following requirements. To begin, $\delta \to x \boxplus \delta$ must be smooth on $\mathbb{R}^n$, and for all $x \in \mathcal{S}$, $y \to y \boxminus x$ must be smooth on U_x, where $U_x = x \boxplus V$. Moreover, we impose the following axioms to hold for every $x \in \mathcal{S}$:

$$x \boxplus 0 = x \tag{5a}$$

$$\forall y \in \mathcal{S} : x \boxplus (y \boxminus x) = y \tag{5b}$$

$$\forall y \in V : (x \boxplus \delta) \boxminus x = \delta \tag{5c}$$

$$\forall \delta_1, \delta_2 \in \mathbb{R}^n : ||(x \boxplus \delta_1) \boxminus (x \boxplus \delta_2)|| \leq ||\delta_1 - \delta_2|| \tag{5d}$$

One can show that a $\boxplus$-manifold is indeed a manifold, with additional structure useful for sensor fusion algorithms. The operators $\boxplus$ and $\boxminus$ allow a generic algorithm to modify and compare manifold states as if they were flat vectors without knowing the internal structure of the manifold, which thus appears as a black box to the algorithm.

Axiom (5a) makes 0 the neutral element of $\boxplus$. Axiom (5b) ensures that from an element x, every other element $y \in \mathcal{S}$ can be reached via $\boxplus$, thus making $\delta \to x \boxplus \delta$ *surjective*. Axiom (5c) makes $\delta \to x \boxplus \delta$ *injective* on V, which defines the range of perturbations for which the parametrization by $\boxplus$ is unique. Obviously, this axiom cannot hold globally in general since otherwise, we could have used $\mathbb{R}^n$ as a universal state representation in the first place. Instead, $\boxplus$ and $\boxminus$ create a local vectorized view of the state space. Intuitively, x is a reference point that defines the "center" of a local neighborhood in the manifold and, thus, also the coordinate system of δ in the part of $\mathbb{R}^n$ onto which the local neighborhood in the manifold is mapped.

Additionally, we demand that the operators are **smooth** (i.e., sufficiently often differentiable) in δ and y (for $y \in U_x$). This makes limits and derivatives of δ correspond to limits and derivatives of $x \boxplus \delta$, essential for any estimation algorithm (formally, $\delta \to x \boxplus \delta$ is a *diffeomorphism* from V to U_x).

4 Uncertainity representation on $\boxplus$-Manifold

From this section on we start to go technically inside the problem of Bayesian estimation on manifolds. Here we present a more schematic nomenclature that will be kept for the rest of the manuscript.

Our goal is to estimate the state $\mathrm{X} \in \mathcal{S}$ given all the sensor measurements. As sensors are flawed, it is impossible to exactly reconstruct X. Instead, a filter maintains a "belief" about the state: its statistical distribution given past sensors' readings. The Kalman filter in $\mathbb{R}^n$ typically maintains a Gaussian belief such that $\mathrm{X} \sim \mathcal{N}(\hat{\mathrm{X}}, \mathbf{\Sigma})$, which may be re-written in the form

$$\mathrm{X} = \hat{\mathrm{X}} + \xi, \quad \xi \sim \mathcal{N}(0, \mathbf{\Sigma}) \tag{6}$$

We see that the belief is encoded using only a mean estimate $\hat{\mathrm{X}}$, and a covariance matrix $\mathbf{\Sigma}$ that encodes the extent of belief dispersion around the estimate.

Consider a manifold $\mathcal{S}$, and let $\{V_1, V_2, , V_n\}$ denote the associated vector fields. To devise a similar belief on $\mathcal{S}$, one needs, of course, local coordinates to write the mean $\hat{\mathrm{X}} \in \mathcal{S}$. This poses no problem, though. The harder part is to find a way to encode dispersion around the estimate $\hat{\mathrm{X}}$. It is now commonly admitted that the tangent space at $\hat{\mathrm{X}}$ should encode such dispersion and that covariance $\mathbf{\Sigma}$ should hence reflect dispersion in the tangent space. As additive noise (6) makes no sense for $\mathrm{X} \in \mathcal{S}$, we define a probability distribution $\mathrm{X} \sim \mathcal{N}_{\boxplus}(\hat{\mathrm{X}}, \mathbf{\Sigma})$ for the random variable $\mathrm{X} \in \mathcal{S}$ as

$$\mathrm{X} = \hat{\mathrm{X}} \boxplus \xi, \quad \xi \sim \mathcal{N}(0, \mathbf{\Sigma}) \tag{7}$$

where $\boxplus : \mathcal{S} \times \mathbb{R}^n$ is the smooth function chosen by the user and satisfying axioms (5a), (5b), (5c), (5d), ξ is a random Gaussian vector that encodes directions of the tangent space at $\hat{\mathrm{X}}$, $\mathcal{N}(\cdot)$ is the classical Gaussian distribution in Euclidean space, and $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ the associated covariance matrix.

Using the $\boxplus$-manifold property, we implicity use coordinates in the tangent space, as $\xi = (\xi(1), \xi(2), \ldots, \xi(n))^T \in \mathbb{R}^n$ encodes the tangent vector $\xi(1)V_1(\hat{\mathrm{X}}), \ldots, \xi(n)V_n(\hat{\mathrm{X}})$. In (7), the noise-free quantity $\hat{\mathrm{X}}$ is viewed as the mean, and the dispersion arises through $\boxplus$. It is important to point out that the distribution defined at (7) it is not Gaussian. It is "only" Gaussian in coordinates related to map $\boxplus$. Practically $\xi \in \mathbb{R}^n$ can be viewed as a 'small,' uncertain perturbation (i.e., zero-mean noise).

5 Kalman Filters on $\boxplus$-Manifolds

Since its inception in the late 1950s, the Kalman filter [9] and its many variants have successfully been applied to a wide variety of state estimation and control problems. In its original form, the Kalman filter provides a framework for continuous and discrete time state estimation of linear Gaussian systems. Many real world problems, however, are intrinsically nonlinear, which gives rise

to the idea of modifying the Kalman filter algorithm to work with nonlinear process models (mapping old to new state) and measurement models (mapping state to expected measurements) as well. The two most popular extensions of this kind are the Extended Kalman Filter (EKF) [1], more recent robustified version like [3] the Invariant Extended Kalman Filter I-EKF and the Unscented Kalman Filter (UKF) [8, 6, 5, 4]. The EKF linearizes the process and measurement models through first-order Taylor series expansion. The UKF, on the other hand, is based on the *unscented transform*, which approximates the respective probability distributions through deterministically chosen samples, so-called *sigma points*, propagates these directly through the non-linear process and measurement models and recover the statistics of the transformed distribution from the transformed samples. Thus, intuitively, the EKF relates to the UKF as a tangent to a secant. We will focus on the UKF here since it is generally better at handling non-linearities and does not require (explicit or numerically approximated) Jacobians of the process and measurement models, i.e. it is a derivative-free filter.

5.1 Bayesian Estimation Using the Unscented Transform

Consider a random variable $\mathrm{X} \in \mathcal{S}$ with prior probability distribution $p(\mathrm{X})$. Suppose we obtain some additional information about X through a measurement $\mathbf{z}$. The goal is to compute the posterior distribution $p(\mathrm{X}|\mathbf{z})$. Let

$$\mathbf{z} = h(\mathrm{X}) + \delta \tag{8}$$

be a measurement, where $h(\cdot) : \mathcal{S} \to \mathbb{R}^n$ represents the observation function and $\delta \sim \mathcal{N}(0, \mathbf{R})$ is a white Gaussian noise in $\mathbb{R}^n$ with known characteristics. The problem of Bayesian estimation we consider is as follow:

1. assume the prior distribution to follow (7) with know parameters, respectively mean $\hat{\mathrm{X}}$ and covariance $\mathbf{\Sigma}$
2. assume one measurement $\mathbf{z}$ of (8) is available
3. approximate the posterior distribution as

$$p(\mathrm{X}|\mathbf{z}) \sim \hat{\mathrm{X}}^+ \boxplus \xi^+ \tag{9}$$

where $\xi^+ \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^+)$, and find parameters X^+ and $\mathbf{\Sigma}^+$

Letting $\mathrm{X} = \hat{\mathrm{X}} \boxplus \xi$ in (8), we see $\mathbf{z}$ provides an information about $\xi \sim \mathcal{N}(0, \mathbf{\Sigma})$ and we may use the unscented transform of [7] to approximate the posterior $p(\xi|\mathbf{z})$ for ξ (see Algorithm 1). We compute a finite number of samples $\xi_i, i = 1, ..., 2n$, and pass each of these so-called **sigma points** through the measurement function

$$\mathbf{z}_i = h(\hat{\mathrm{X}} \boxplus \xi_i), \quad i = 1, \ldots, 2n. \tag{10}$$

Algorithm 1: Bayes updating, UKF correction

inputs: $\hat{\mathrm{X}}$, $\boldsymbol{\Sigma}$, $\mathbf{z}$, $\mathbf{R}$

begin

// set sigma points

$\xi_i = \mathrm{col}(\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i \quad i = 1, \dots, n$

$\xi_i = -\mathrm{col}(\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_{i-n} \quad i = n+1, \dots, 2n$

// compute measurements sigma points

$\mathbf{z}_0 = h(\hat{\mathrm{X}} \boxplus \mathbf{0})$

$\mathbf{z}_i = h(\hat{\mathrm{X}} \boxplus \xi_i) \quad i = 1, \dots, n$

// infer covariance matrices

$\bar{\mathbf{z}} = w_m \mathbf{z}_0 + \sum_{i=1}^{2n} w_i \mathbf{z}_i$

$\boldsymbol{\Sigma}_{zz} = \sum_{i=1}^{2n} w_i (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T + \mathbf{R}$

$\boldsymbol{\Sigma}_{\xi z} = \sum_{i=1}^{2n} w_i \xi_i (\mathbf{z}_i - \bar{\mathbf{z}})^T$

// update state and covariance

$\mathbf{K} = \boldsymbol{\Sigma}_{\xi z} \boldsymbol{\Sigma}_{zz}^{-1}$ // Kalman gain

$\hat{\mathrm{X}}^+ = \hat{\mathrm{X}} \boxplus \mathbf{K}(\mathbf{z} - \bar{\mathbf{z}})$

$\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma} - \mathbf{K} \boldsymbol{\Sigma}_{zz} \mathbf{K}^T$

end

output: $\hat{\mathrm{X}}^+$, $\boldsymbol{\Sigma}^+$

By noting $\mathbf{z}_0 = h(\hat{\mathrm{X}} \boxplus \mathbf{0})$ we then compute successively the measurement mean $\bar{\mathbf{z}} = w_m \mathbf{z}_0 + \sum_{i=1}^{2n} w_i \mathbf{z}_i$, the measurement covariance $\boldsymbol{\Sigma}_{zz} = \sum_{i=0}^{2n} w_i (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T + \mathbf{R}$ and the cross-covariance $\boldsymbol{\Sigma}_{\xi z} = \sum_{i=1}^{2n} w_i \xi_i (\mathbf{z}_i - \bar{\mathbf{z}})^T$, where w_m, w_j and λ are weights and parameters defined in Sec. 5.3. We then derive the conditional distribution of $\xi \in \mathbb{R}^n$ as

$$p(\xi|\mathbf{z}) \sim (\bar{\xi}, \boldsymbol{\Sigma}^+), \text{where} \tag{11}$$

$$\mathbf{K} = \boldsymbol{\Sigma}_{\xi z} \boldsymbol{\Sigma}_{zz}^{-1}, \ \bar{\xi} = \mathbf{K}(\mathbf{z} - \bar{\mathbf{z}}), \ \boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma} - \mathbf{K} \boldsymbol{\Sigma}_{zz} \mathbf{K}^T \tag{12}$$

This may be viewed as a Kalman update on the error ξ, in the vein of error state Kalman filtering. The problem is then to convert this into a distribution on the manifold in the form (7). We first represent $p(\xi|\mathbf{z})$ as $\bar{\xi} + \xi^+$ with $\xi^+ \sim \mathcal{N}(0, \boldsymbol{\Sigma}^+)$ and $\bar{\xi}$ considered as a noise free mean. We suggest to define the posterior $p(\mathrm{X}|\mathbf{z})$ as

$$\mathrm{X} \sim \hat{\mathrm{X}}^+ \boxplus \xi^+, \quad \xi^+ \sim \mathcal{N}(0, \boldsymbol{\Sigma}^+) \tag{13}$$

where

$$\hat{\mathrm{X}}^+ = \hat{\mathrm{X}} \boxplus \bar{\xi} \tag{14}$$

Approximation written in (13)-(14) could be rewritten as

$$\hat{\mathrm{X}} \boxplus (\bar{\xi} + \xi^+) \approx (\hat{\mathrm{X}} \boxplus \bar{\xi}) \boxplus \xi^+ \tag{15}$$

When $\mathcal{S} = \mathbb{R}^n$ the latter equality holds up to the first order in the dispersions $\bar{\xi}$, ξ^+, both assumed small. In the case where $\mathcal{S}$ is not a vector space, it may

be geometrically interpreted as saying that moving from $\hat{\mathrm{X}}$ along the direction $\bar{\xi} + \xi^+$ approximately consists in moving from $\hat{\mathrm{X}}$ along $\bar{\xi}$ and then from the obtained point on $\mathcal{S}$ along ξ^+.

It is important to clarify that in Algorithm 1, $\mathbf{z}$ is the raw measurement from the sensor, not the estimated one. In fact, the error is inferred as $\mathbf{z} - \bar{\mathbf{z}}$, where the second is the measurement means coming from the state X and the observation function $h(\cdot)$.

5.2 Unscented Kalman Filter on $\boxplus$-Manifold

Consider the dynamics

$$\mathrm{X}_t = f(\mathrm{X}_{t-1}, \mathbf{u}_t, \epsilon_t) \tag{16}$$

where the state X_t lives in a manifold $\mathcal{S}$, $\mathbf{u}_t$ is a known input variable and $\epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ is a white Gaussian noise in $\mathbb{R}^q$. We consider the observation of the form

$$\mathbf{z}_t = h(\mathrm{X}_t) + \delta_t \tag{17}$$

where $\delta_t \sim \mathcal{N}(0, \mathbf{R}_t)$ is a white Gaussian noise with known covariance (we assume this addictive only for the clarity of algorithm derivation). For system (16), (17), we model the state posterior conditioned on past measurements using the uncertainty representation (7). To propagate the state we start from the prior distribution $p(\mathrm{X}_{t-1}) \sim \hat{\mathrm{X}}_{t-1} \boxplus \xi_{t-1}$ with $\xi_{t-1} \sim \mathcal{N}(0, \mathbf{\Sigma}_{t-1})$ and $\hat{\mathrm{X}}_{t-1}$, $\mathbf{\Sigma}_{t-1}$ known, and we seek to compute the state propagated distribution in the form

$$p(\mathrm{X}_t|\mathrm{X}_{t-1}) \sim \hat{\mathrm{X}}_t \boxplus \xi_t, \quad \text{with } \xi_t \sim \mathcal{N}(0, \mathbf{\Sigma}). \tag{18}$$

We define sigma points using (7) and the statistics of noise ϵ_t, and pass them through prediction function (16). Then, to find $\hat{\mathrm{X}}_t$ one is faced with the optimization problem of computing a weighted mean on $\mathcal{S}$. To keep the implementation simple and analog to the EKF, we suggest merely propagating the mean using the unnoisy state model, leading to

$$\hat{\mathrm{X}}_t = f(\hat{\mathrm{X}}_{t-1}, \mathbf{u}_t, 0) \tag{19}$$

To compute the covariance $\mathbf{\Sigma}_t$ from $\mathbf{\Sigma}_{t-1}$ of ξ_{t-1} we use the fact ϵ_t and ξ_{t-1} are uncorrelated and proceed in two steps:

1. we generate sigma points in $\mathbb{R}^n$ corresponding to $\mathbf{\Sigma}_{t-1}$ and pass them through the unnoisy model (15) for nonlinear propagation of $\mathbf{\Sigma}_{t-1}$ through f. We obtain points in X_t^i on the manifold $\mathcal{S}$, and the distribution of propagated state is described as $\hat{\mathrm{X}}_t \boxplus \xi_t$, with $\hat{\mathrm{X}}_t$ known from (19). We then use the $\boxminus$ operator to find the map such that

$$\hat{\mathrm{X}} \boxminus (\hat{\mathrm{X}} \boxplus \xi) = \xi + O(||\xi||^2), \tag{20}$$

that is, a map that allows one ot asses the discrepancy between $\hat{\mathrm{X}}$ and $\hat{\mathrm{X}} \boxplus \xi$ is ξ indeed. Then we use $\boxminus$ to map sigma points X_t^i back into $\mathbb{R}^n$ and compute their empirical covariance $\mathbf{\Sigma}_t$.

2. we then generate sigma points for process noise X_t^i back into $\mathbb{R}^n$ and compute their noise ϵ_t similarly and obtain another covariance matrix encoding dispersion in $\mathbb{R}^n$ owed to noise, that adds up to $\mathbf{\Sigma}_t$ and thus clearly distinguish the contribution of the state error dispersion ξ from noise ϵ_t.

When a new measurement arrives, belief is updated via Algorithm 1. Algorithm 2 summarizes both steps, where the weights defined through `set_weights(n,` α) depend on a scale parameter α (generally set between 10^{-3} and 1), and sigma point dimension.

In Algorithm 2, n corresponds to the X state dimension, $\mathrm{X} \in \mathbb{R}^n$ while q corresponds to the input noise covariance of dimension $\mathbf{Q} \in \mathbb{R}^{q \times q}$.

Algorithm 2: Algorithm 2: UKF on $\boxplus$-Manifold

inputs: $\hat{\mathrm{X}}_{t-1}$, $\mathbf{\Sigma}_{t-1}$, $\mathbf{u}_t$, $\mathbf{Q}_t$ $\mathbf{z}_t$, α

Prediction

begin

`// propagate mean state`

$\hat{\mathrm{X}}_t = f(\hat{\mathrm{X}}_{t-1}, \mathbf{u}_t, \mathbf{0})$

`// propagate state error covariance`

$\lambda, \{w_i\}_{i=0,\dots,2n} =$ `set_weights`(n, α)

$\xi_i = \mathrm{col}(\sqrt{(\lambda+n)\mathbf{\Sigma}_{t-1}})_i \quad i = 1, \dots, n$

$\xi_i = -\mathrm{col}(\sqrt{(\lambda+n)\mathbf{\Sigma}_{t-1}})_{i-n} \quad i = n+1, \dots, 2n$

`// use boxplus into manifold`

$\hat{\mathrm{X}}_t^i = f(\hat{\mathrm{X}}_{t-1} \boxplus \xi_i, \mathbf{u}_t, \mathbf{0})$ $i = 1, \dots, 2n$

`// use boxminus to go back in` $\mathbb{R}^n$

$\mathbf{\Sigma}_t = \sum_{i=1}^{2n} w_i (\hat{\mathrm{X}}_t \boxminus \hat{\mathrm{X}}_t^i)(\hat{\mathrm{X}}_t \boxminus \hat{\mathrm{X}}_t^i)^T$

`// proceed similarly for noise`

$\lambda, \{w_i\}_{i=0,\dots,2q} =$ `set_weights`(q, α)

$\epsilon^i = \mathrm{col}(\sqrt{(\lambda+q)\mathbf{Q}_t})_i \quad i = 1, \dots, q$

$\epsilon^i = -\mathrm{col}(\sqrt{(\lambda+q)\mathbf{Q}_t})_{i-n} \quad i = n+1, \dots, 2q$

`// get final version of covariance`

$\tilde{\mathrm{X}}_t^i = f(\hat{\mathrm{X}}_{t-1}, \mathbf{u}_t, \epsilon^i) \quad i = 1, \dots, 2q$

$\mathbf{\Sigma}_t = \mathbf{\Sigma}_t + \sum_{i=1}^{2q} w_i (\hat{\mathrm{X}}_t \boxminus \tilde{\mathrm{X}}_t^i)(\hat{\mathrm{X}}_t \boxminus \tilde{\mathrm{X}}_t^i)^T$

end

Update (when measurement $\mathbf{z}_t$ arrives), Algorithm 1

compute $\hat{\mathrm{X}}_t^+$ and $\mathbf{\Sigma}_t^+$ from $\hat{\mathrm{X}}_t$ and $\mathbf{\Sigma}_t$

5.3 Defining Unscented Transform parameters

5.3.1 Sigma Points

The set of $2n+1$ sigma points that are used to approximate an n-dimensional Gaussian distribution or random variable X with $\hat{\mathrm{X}}$ mean and covariance $\mathbf{\Sigma}$ is computed as follows:

$$\mathrm{X}^0 = \hat{\mathrm{X}} \tag{21a}$$
$$\mathrm{X}^i = \hat{\mathrm{X}} + (\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i, \quad w_i = 1/2(n+\lambda) \tag{21b}$$
$$\mathrm{X}^{i+n} = \hat{\mathrm{X}} - (\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i, \quad w_{i+n} = 1/2(n+\lambda) \tag{21c}$$

where $\lambda \in \mathbb{R}$, $(\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i$ is the i-th column of the matrix square root of $(\lambda+n)\boldsymbol{\Sigma}$ (obtained through Cholesky decomposition $(\lambda+n)\boldsymbol{\Sigma} = (\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i(\sqrt{(\lambda+n)\boldsymbol{\Sigma}})_i^T$ and w_i is the weight which is associated with the i-th point.

The weights parameters are computed as follows:

$$\lambda = (\alpha^2 - 1)n \tag{22}$$
$$w_\lambda = \lambda \tag{23}$$
$$w_{w_i} = 1/2(n+\lambda) \tag{24}$$
$$w_{w_{m0}} = \lambda/(\lambda+n) \tag{25}$$
$$w_{w_{c0}} = \lambda/(\lambda+n) + 3 - \alpha^2 \tag{26}$$

Weights w_i determine the distance of the sigma points from the original covariance. Above n is the dimension of the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$. If we are estimating the parameters for state propagation w.r.t. noise rather than covariance, then n is substituted with q, which is the noise dimension (see Sec. 5.2).

6 Pratically important $\boxplus$-Manifolds

TBA

7 Final Remarks

It has been demonstrated that this version of UKF performs particularly well in heterogeneous tasks. The authors [5, 4] compared this approach against the standard EKF, standard UKF, and the more recent Invariant-EKF (IEKF), an EKF variant on Lie groups that comes with guaranteed convergence properties [3]. Other main benefits of this approach are the fast prototyping and the derivative-free approach.

Other studies and more in-depth studies explicitly related to covariance propagation [2] show that the general outcome depends strictly on the application. Different methods have advantages, with the higher-order methods always doing better than the lower ones. The choice of whether to use sigma points to propagate uncertainty depends on the application: for instance, no for pose compounding but yes to propagate uncertainty through a nonlinear camera model.

References

[1] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[2] Timothy D Barfoot and Paul T Furgale. Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Transactions on Robotics*, 30(3):679–693, 2014.

[3] Axel Barrau and Silvère Bonnabel. The invariant extended kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4):1797–1812, 2016.

[4] Martin Brossard, Axel Barrau, and Silvere Bonnabel. A code for unscented kalman filtering on manifolds (ukf-m). In *International Conference on Robotics and Automation (ICRA)*, pages 5701–5708. IEEE, 2020.

[5] Martin Brossard, Silvere Bonnabel, and Jean-Philippe Condomines. Unscented kalman filtering on lie groups. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 2485–2491. IEEE, 2017.

[6] Christoph Hertzberg, René Wagner, Udo Frese, and Lutz Schröder. Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds. *Information Fusion*, 14(1):57–77, 2013.

[7] Simon J Julier. The scaled unscented transformation. In *American Control Conference*, volume 6, pages 4555–4559. IEEE, 2002.

[8] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193. International Society for Optics and Photonics, 1997.

[9] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.