

# Unscented Kalman Filters for Riemannian State-Space Systems

Henrique M. T. Menegaz, João Y. Ishihara, Hugo T. M. Kussaba

**Abstract**—Unscented Kalman Filters (UKFs) have become popular in the research community. Most UKFs work only with Euclidean systems, but in many scenarios it is advantageous to consider systems with state-variables taking values on *Riemannian manifolds*. However, we can still find some gaps in the literature's theory of UKFs for Riemannian systems: for instance, the literature has not yet i) developed Riemannian extensions of some fundamental concepts of the UKF theory (e.g., extensions of  $\sigma$ -representation, Unscented Transformation, Additive UKF, Augmented UKF, additive-noise system), ii) proofs of some steps in their UKFs for Riemannian systems (e.g., proof of sigma points parameterization by vectors, state correction equations, noise statistics inclusion), and iii) relations between their UKFs for Riemannian systems. In this work, we attempt to develop a theory capable of filling these gaps. Among other results, we propose Riemannian extensions of the main concepts in the UKF theory (including closed forms), justify all steps of the proposed UKFs, and provide a framework able to relate UKFs for particular manifolds among themselves and with UKFs for Euclidean spaces. Compared with UKFs for Riemannian manifolds of the literature, the proposed filters are more consistent, formally-principled, and general. An example of satellite attitude tracking illustrates the proposed theory.

## I. INTRODUCTION

When we want to know the value of some variables of a given system—e.g., the position and velocity of a car, the position and attitude of a satellite, the temperature of a boil, etc.—we can acquire data from the system and develop a mathematical model of it. But measurements are noisy, and models are always imperfect. Hence, to estimate the desired variables, we often must use filters, such as Unscented Kalman Filters (UKFs). Researchers have been applying UKFs in applications of diverse fields: for example, in power electronic [1], aerospace [2], and automotive [3] systems. These filters' success is partially explained by their good trade-off between estimation quality and computational complexity compared with similar techniques such as the Extended Kalman Filter (EKF) [4].

Most UKFs work only with Euclidean models (the so-called state-space systems; cf. Section IV), but sometimes modeling with *Riemannian manifolds* is better. These manifolds can i) model more systems (cf. Section I-A), ii) provide better mathematical properties than Euclidean subspaces (e.g., better metrics), and iii) be the set where measurements take value from (cf. [5]–[7] and Section I-A).

H. M. T. Menegaz ([henriquemenegaz@unb.br](mailto:henriquemenegaz@unb.br)) is with the Faculdade Gama at the Universidade de Brasília (UnB), Brazil.

J. Y. Ishihara ([ishihara@lara.unb.br](mailto:ishihara@lara.unb.br)) and Hugo T. M. Kussaba ([kussaba@lara.unb.br](mailto:kussaba@lara.unb.br)) are with the Automation and Robotics Laboratory (LARA) at the UnB. Web-page: [www.lara.unb.br](http://www.lara.unb.br).

Although some works have introduced UKFs for Riemannian systems (e.g., [5], [8]–[10]; cf. Section I-A), we can still find some gaps in the literature's theory for these UKFs. First, fundamental UKF concepts still miss for Riemannian manifolds, such as  $\sigma$ -representation ( $\sigma$ R), Unscented Transformation (UT), Additive UKF, Augmented UKF, additive-noise system, among others (cf. [11]). Second, some steps in UKFs for Riemannian manifolds are not formally justified, such as when a UKF parameterize sigma points by vectors, or correct the predicted state estimate, or consider noise statistics (cf. [5], [12], [13]; see Sections I-A and VII-D). Third, we do not know how the literature's consistent UKFs for Riemannian manifolds relates among themselves—do they follow from a same general Riemannian UKF?—or with UKFs for Euclidean Spaces—are these particular cases of those?

In this work, by continuing the research of [5], we aim to develop a formalized and systematized theory for UKFs on Riemannian manifolds. Among other results, this theory introduces Riemannian extensions of the main concepts in the UKF theory (including closed forms), justifies all steps of the proposed UKFs, and provides a framework able to relate UKFs for particular manifolds among themselves and with UKFs for Euclidean spaces.

### A. Kalman filtering in Riemannian manifolds

Riemannian manifolds can model many applications; far more than Euclidean spaces. For instance, we find i) special orthogonal groups, special Euclidean groups, unit spheres (including the set of unit quaternions), and the study quadric (the set of unit dual-quaternions) applied to many robotics applications [14]–[18], aerospace systems [12], [17]–[20], bio-engineering [8], [21], among others; ii) positive symmetric matrices applied to applications in image recognition, image registration, image tracking, and surgery [21]; iii) Grassmann and Stiefel manifolds applied to information theory [22], machine learning [23], visual recognition [23], [24], communication systems [25], and geology [26]; and iv) other Riemannian manifolds applied to quantum systems [27], and special and general relativity [28].

Some works in the literature have proposed KFs for *particular* Riemannian systems: the works [9], [29]–[31] and [12] (among others) introduced EKFs and UKFs for unit quaternions; and [10], [32] and [33] EKFs for special orthogonal groups. Other works have proposed KFs for *classes* of Riemannian systems: the works [34], [35] and [36] introduced EKFs for Lie groups; and [5] a UKF for geodesically-complete Riemannian manifolds.geodesically-complete

arXiv:1806.11012v1 [math.OC] 28 Jun 2018

Developing UKFs for Riemannian manifolds is difficult because, in general, Riemannian manifolds lack some mathematical tools used in most UKFs, such as multiplication and addition (cf. UKFs in [11], [37], [38]). An alternative is to use properties of an embedding Euclidean space and afterwards perform operations to return to the working manifold. For instance, an application on  $S^3$  can use derivatives, sums, multiplications, metrics of  $\mathbb{R}^4$  and afterwards perform a normalization. Many works take this embedding approach [13], [39], [40].

However, this approach may i) lose the physical identification (e.g., an addition of unit quaternions yields a non unit quaternion, which does not represent a rotation anymore), or ii) disregard the global properties of the manifold leading to instability. To retain the estimates on the working manifolds, literature UKFs use intrinsic manifold properties (cf. [34]–[36])—meaning we do not use properties of embedding Euclidean spaces.

In this work, we take this intrinsic approach; we combine the UKF theory we developed in [11] with the statistics for Riemannian manifolds of [41] and some results of [5] to develop a theory of UKFs for any geodesically-complete Riemannian manifolds.

## II. RIEMANNIAN MANIFOLDS

In this section, we provide a general description of the concepts from Riemannian Geometry used in this work and in Appendix A their formal definitions. This exposition is mainly based on [42], and partially on [7] and [6].

A *differentiable manifold* (Definition 5)  $\mathcal{N}$  (or  $\mathcal{N}^n$ ) can be viewed as a set whose subsets are identified through *charts* (injective mappings) with subsets of the  $\mathbb{R}^n$ . For every point  $\mathbf{a}$  on a differentiable manifold  $\mathcal{N}$ , we can define the vector space of tangent vectors at  $\mathbf{a}$  called *tangent space* and denoted by  $T_{\mathbf{a}}\mathcal{N}$  (Definition 7).

A *Riemannian manifold*  $\mathcal{N}$  (Definition 11) is a differentiable manifold endowed with a *Riemannian metric* (Definition 11)  $\langle \cdot, \cdot \rangle$  or  $g$ . For  $\mathbf{a} \in \mathcal{N}$  and  $v \in T_{\mathbf{a}}\mathcal{N}$ , the *norm* of  $v$  associated to  $\mathbf{a}$  is defined by  $\|v\|_{\mathbf{a}} := \langle v, v \rangle_{\mathbf{a}}^{1/2}$  [7].

For two points  $\mathbf{a}$  and  $\mathbf{b}$  in  $\mathcal{N}$  connected by a curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$ , the *distance* between  $\mathbf{a}$  and  $\mathbf{b}$  is defined by, for  $[a, b] \subset \mathbf{I}$ ,

$$\text{dist}(\mathbf{a}, \mathbf{b}) := \min_{\alpha} \mathbb{L}_{\mathbf{a}}^b(\alpha); \quad \alpha(a) = \mathbf{a}, \alpha(b) = \mathbf{b},$$

where  $\mathbb{L}_{\mathbf{a}}^b(\alpha)$  is the arc length (Definition 8) of  $\alpha$  in the interval  $[a, b]$ . A *geodesic ball* of center  $\mathbf{a}$  and radius  $r$  is the set defined as

$$\mathbb{B}(\mathbf{a}, r) := \{\mathbf{x} \in \mathcal{N} : \text{dist}(\mathbf{x}, \mathbf{a}) < r\}.$$

Given a tangent vector  $v_0 \in T_{\alpha(t_0)}\mathcal{N}$ ,  $t_0 \in \mathbf{I}$ , there exists only one parallel vector field  $\mathcal{X}$  (Definition 10) along  $\alpha$ , such that  $\mathcal{X}(t_0) = v_0$ ;  $\mathcal{X}(t)$  is called the *parallel transport* of  $\mathcal{X}(t_0)$  along  $\alpha$ .

A curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is called a *geodesic* at  $t_0 \in \mathbf{I}$  if

$$\frac{D}{dt}(\alpha'(t)) = 0$$

at  $t_0$ , where  $D/dt(\alpha'(t))$  is the covariant derivative of  $\alpha'(t)$  (Theorem 4); if  $\alpha$  is a geodesic at  $t$ , for all  $t \in \mathbf{I}$ , we say

$\alpha$  is a *geodesic* [42]. If a curve minimizes the *arc length* between two points of the manifold, then this curve is a geodesic, but the converse is *only valid locally*. If the definition domain of all geodesics of  $\mathcal{N}$  can be extended to  $\mathbb{R}$ , then  $\mathcal{N}$  is said to be *geodesically-complete*. There exists at least one geodesic connecting every two points of a geodesically-complete manifold.

Given a point  $\mathbf{a} \in \mathcal{N}$ , the *exponential mapping* (Definition 17), denoted by  $\exp_{\mathbf{a}}$ , associates a vector of  $T_{\mathbf{a}}\mathcal{N}$  to a point of  $\mathcal{N}$ . Geometrically,  $\exp_{\mathbf{a}}(v)$  is a point of  $\mathcal{N}$  obtained by going out the length equal to  $\|v\|$ , starting from  $\mathbf{a}$ , along a geodesic which passes through  $\mathbf{a}$  with velocity equal to  $v/\|v\|$ .

Assuming a geodesically-complete manifold, it is possible to follow the geodesic  $\exp_{\mathbf{a}}(tv)$  from  $t = 0$  to  $t \rightarrow \infty$ . It may happen, however, that from a particular value  $t_v$  to  $t \rightarrow \infty$ , the geodesics  $\exp_{\mathbf{a}}(tv)$  do not minimize the *arc length* between  $\mathbf{a}$  and  $\exp_{\mathbf{a}}(tv)$ . In this case, the subset  $\{\exp_{\mathbf{a}}(t_v v) : v \in T_{\mathbf{a}}\mathcal{N}\} \subset \mathcal{N}$  is called the *cut locus*  $\mathcal{C}(\mathbf{a})$  and the inverse image  $C(\mathbf{a}) := \exp_{\mathbf{a}}^{-1}[\mathcal{C}(\mathbf{a})]$  the *tangential cut locus* [7]. The *injectivity radius* of  $\mathcal{N}$  is defined as  $\text{inj}(\mathcal{N}) := \inf_{\mathbf{p} \in \mathcal{N}} \text{dist}(\mathbf{p}, C(\mathbf{p}))$ .

For every  $\mathbf{a} \in \mathcal{N}$ , we can reduce the domain of  $\exp_{\mathbf{a}}$  to some subsets such that  $\exp_{\mathbf{a}}$  is a diffeomorphism. The maximal of these subsets is called the *maximal definition domain*  $\Omega(\mathbf{a}) \subset T_{\mathbf{a}}\mathcal{N}$ ; this set is *bounded* by  $C(\mathbf{a})$  [7]. The inverse mapping of  $\exp_{\mathbf{a}}$  is the (Riemannian) *logarithm mapping* (Definition 17) and we denote it by either  $\log_{\mathbf{a}} \mathbf{b}$  or  $\overrightarrow{\mathbf{ba}}$ .

## III. INTRINSIC STATISTICS ON RIEMANNIAN MANIFOLDS

UKFs are based on information of *moments* of *random vectors* and of *sample moments* of *weighted sets*. To define UKFs on Riemannian manifolds, we need extensions of these concepts.

### A. Statistics of random points

Riemannian extensions of random vectors are called (*Riemannian*) *random points* [7]; the set of all random points taking values on a Riemannian manifold  $\mathcal{N}$  is denoted by  $\Phi_{\mathcal{N}}$ . Given a random point  $\mathbf{X} \in \Phi_{\mathcal{N}}$ , its *probability density function* (pdf) is denoted by  $p_{\mathbf{X}}$ , and for a real-valued function  $F : \mathcal{N} \rightarrow \mathbb{R}$  the *expected value of F relative to X* is defined by

$$\mathcal{E}_{\mathbf{X}}\{F(\mathbf{X})\} := \int_{\mathcal{N}} F(\mathbf{b})p_{\mathbf{X}}(\mathbf{b})d\mathcal{N}(\mathbf{b}). \quad (1)$$

For functions taking values on manifolds, we cannot define the expected value as in (1); thus, we define mean points following the Karcher expectation: they are the *local* minima of variances [7].

Given a point  $\mathbf{c} \in \mathcal{N}$ , the *variance*  $\sigma_{\mathbf{X}}^2(\mathbf{c})$  is defined by  $\sigma_{\mathbf{X}}^2(\mathbf{c}) := \mathcal{E}_{\mathbf{X}}\{\text{dist}^2(\mathbf{c}, \mathbf{X})\}$ . If  $\sigma_{\mathbf{X}}^2(\mathbf{c})$  is finite for every point  $\mathbf{c} \in \mathcal{N}$ , then a point  $\bar{\mathbf{X}} \in \mathcal{N}$  is an *expected point* or *mean* of  $\mathbf{X}$  if

$$\bar{\mathbf{X}} = \arg \min_{\mathbf{c} \in \mathcal{N}} \sigma_{\mathbf{X}}^2(\mathbf{c}). \quad (2)$$

The set of all means of  $\mathbf{X}$  is denoted by  $\mathbb{E}(\mathbf{X})$ . A random point can have more than one mean<sup>1</sup>.

Let  $\mathbf{X} \in \Phi_{\mathcal{N}}$  be a random point with a mean  $\bar{\mathbf{X}} \in \mathbb{E}(\mathbf{X})$ , and consider a point  $\mathbf{a} \in \mathcal{N}$ . If  $\bar{\mathbf{X}} \in \Omega(\mathbf{a})$ , then the  $j$ th (central) moment of  $\mathbf{X}$  with respect to  $\bar{\mathbf{X}}$  at  $\mathbf{a}$  is defined by, for even  $j$ ,

$$M_{\mathbf{X},\bar{\mathbf{X}}}^{a,j} := \mathcal{E}_{\mathbf{X}} \left\{ \left[ \left( \overrightarrow{\mathbf{aX}} - \overrightarrow{\mathbf{a}\bar{\mathbf{X}}} \right) (\diamond)^T \right]^{\otimes \frac{j}{2}} \right\}; \quad (3)$$

and for odd  $j$ ,

$$M_{\mathbf{X},\bar{\mathbf{X}}}^{a,j} := \mathcal{E}_{\mathbf{X}} \left\{ \left[ \left( \overrightarrow{\mathbf{aX}} - \overrightarrow{\mathbf{a}\bar{\mathbf{X}}} \right) (\diamond)^T \right]^{\otimes \frac{j-1}{2}} \otimes \left( \overrightarrow{\mathbf{aX}} - \overrightarrow{\mathbf{a}\bar{\mathbf{X}}} \right) \right\}.$$

We define joint pdf [denoted by  $p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$ ], joint expected moment ( $\mathcal{E}_{\mathbf{XY}} \{f(\mathbf{x}, \mathbf{y})\}$ ) and cross-covariance ( $\mathbf{P}_{\mathbf{XY},(\bar{\mathbf{X}},\bar{\mathbf{Y}})}^{ab}$ ) of two random points  $\mathbf{X}$  and  $\mathbf{Y}$  similarly (cf. [38]). The notation  $\mathbf{X} \sim (\bar{\mathbf{X}}, M_{\mathbf{X},\bar{\mathbf{X}}}^{a,2}, \dots, M_{\mathbf{X},\bar{\mathbf{X}}}^{a,l})_{\mathcal{N}}$  stands for a Riemannian random point  $\mathbf{X} \in \Phi_{\mathcal{N}}$  with mean  $\bar{\mathbf{X}} \in \mathbb{E}(\mathbf{X})$  and moments  $M_{\mathbf{X},\bar{\mathbf{X}}}^{a,2}, \dots, M_{\mathbf{X},\bar{\mathbf{X}}}^{a,l}$ . The second moment ( $j = 2$ ) is called *covariance* and denoted by  $\mathbf{P}_{\mathbf{XX},\bar{\mathbf{X}}}^a := M_{\mathbf{X},\bar{\mathbf{X}}}^{a,2}$ . If  $\mathbb{E}(\mathbf{X}) = \{\bar{\mathbf{X}}\}$ , we can write  $M_{\mathbf{X}}^{a,j} := M_{\mathbf{X},\bar{\mathbf{X}}}^{a,j}$  and  $\mathbf{P}_{\mathbf{XX}}^a := \mathbf{P}_{\mathbf{XX},\bar{\mathbf{X}}}^a$ , or even  $M_{\mathbf{X}}^j := M_{\mathbf{X},\bar{\mathbf{X}}}^{\bar{\mathbf{X}},j}$  and  $\mathbf{P}_{\mathbf{XX}} := \mathbf{P}_{\mathbf{XX},\bar{\mathbf{X}}}^{\bar{\mathbf{X}}}$ .

We represent statistics of *Euclidean* manifolds without bold notation. For  $X \in \Phi_{\mathbb{R}^n}$ ,  $X$  is symmetric if  $p_X(\bar{X} + x) = p_X(\bar{X} - x)$  for every  $x \in \mathbb{R}^n$ . If  $X$  has a mean, then

$$\bar{X} = \arg \min_{c \in \mathbb{R}^n} \sigma_X^2(c) = \mathcal{E}_X \{X\};$$

and, for  $j$  even,

$$M_X^j := \mathcal{E}_X \left\{ \left[ \left( \overrightarrow{\bar{X}X} - \overrightarrow{\bar{X}\bar{X}} \right) (\diamond)^T \right]^{\otimes \frac{j}{2}} \right\} = \mathcal{E}_X \left\{ \left[ (X - \bar{X}) (\diamond)^T \right]^{\otimes \frac{j}{2}} \right\}$$

(similarly for  $j$  odd and for sample cross-covariances).

### B. Statistics of weighted sets

For a Riemannian manifold  $\mathcal{N}$  and the natural numbers  $l \geq 2$  and  $N \geq 1$ , consider the weighted set

$$\chi := \left\{ \chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \chi_i \in \mathcal{N}; \right. \\ \left. j = 1, \dots, l; w_i^m, w_i^{c,j}, w_i^{cc,j} \in \mathbb{R} \right\}_{i=1}^N.$$

The weights  $w_i^m$  are associated (below) with the definition of sample mean,  $w_i^{c,j}$  with the  $j$ th sample moment, and  $w_i^{cc}$  with the  $j$ th sample cross-moment of  $\chi$ .

The *sample variance* of  $\chi$  with respect to a point  $\mathbf{c} \in \mathcal{N}$  is defined by  $s_{\chi}^2(\mathbf{c}) := \sum_{i=1}^N w_i^m \text{dist}^2(\mathbf{c}, \chi_i)$ . If the variance  $s_{\chi}^2(\mathbf{c})$  is finite for every point  $\mathbf{c} \in \mathcal{N}$ , then a *sample expected point* or *sample mean* of  $\chi$  is defined by

$$\mu_{\chi} := \arg \min_{c \in \mathcal{N}} s_{\chi}^2(c). \quad (4)$$

The set of all sample means of  $\chi$  is represented by  $\mathcal{E}(\chi)$ . A weighted set in the form of  $\chi$  can have more than one sample mean.

<sup>1</sup>For a discussion about the existence and uniqueness of this expectation, cf. Section 4.2 of [7].

For a point  $\mathbf{a} \in \mathcal{N}$ , if  $\mu_{\chi}, \chi_1, \chi_2, \dots, \chi_N \in \mathcal{N} - \mathcal{C}(\mathbf{a})$ , then the  $j$ th *sample moment of  $\chi$  with respect to  $\bar{\mathbf{X}}$  at  $\mathbf{a}$*  is defined by, for  $j$  even,

$$\mathcal{M}_{\chi, \mu_{\chi}}^{a,j} := \sum_{i=1}^N w_i^{c,j} \left[ \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right) (\diamond)^T \right]^{\otimes \frac{j}{2}}; \quad (5)$$

and for  $j$  odd,

$$\mathcal{M}_{\chi, \mu_{\chi}}^{a,j} := \sum_{i=1}^N w_i^{c,j} \left[ \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right) (\diamond)^T \right]^{\otimes \frac{j-1}{2}} \otimes \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right).$$

The sample moment ( $j = 2$ ) is called *sample covariance* and denoted by  $\Sigma_{\chi\chi, \mu_{\chi}}^a := \mathcal{M}_{\chi, \mu_{\chi}}^{a,2}$ . If  $\mathcal{E}(\chi) = \{\mu_{\chi}\}$ , we can write  $\mathcal{M}_{\chi}^{a,j} := \mathcal{M}_{\chi, \mu_{\chi}}^{a,j}$  and  $\Sigma_{\chi\chi}^a := \Sigma_{\chi\chi, \mu_{\chi}}^a$ ; or even,  $\mathcal{M}_{\chi}^j := \mathcal{M}_{\chi, \mu_{\chi}}^{\mu_{\chi},j}$  and  $\Sigma_{\chi\chi} := \Sigma_{\chi\chi, \mu_{\chi}}^{\mu_{\chi}}$ .

In addition, for i) the Riemannian manifold  $\mathcal{R}$ , ii) a function  $f : \mathcal{N} \rightarrow \mathcal{R}$ , iii) the weighted set

$$\gamma := \left\{ \gamma_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \gamma_i = f(\chi_i); j = 1, \dots, l \right\}_{i=1}^N,$$

with a mean  $\mu_{\gamma}$ , and iv) the point  $\mathbf{b} \in \mathcal{R}$ . If  $\mu_{\gamma}, \gamma_1, \gamma_2, \dots, \gamma_N \in \mathcal{R} - \mathcal{C}(\mathbf{b})$ , then the  $j$ th *cross-moment of  $\chi$  and  $\gamma$  with respect to  $(\mu_{\chi}, \mu_{\gamma})$  at  $(\mathbf{a}, \mathbf{b})$*  is defined by, for  $j$  even,

$$\mathcal{M}_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{j,ab} := \sum_{i=1}^N w_i^{cc,j} \left[ \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right) \left( \overrightarrow{\mathbf{a}\gamma_i} - \overrightarrow{\mathbf{a}\mu_{\gamma}} \right)^T \right]^{\otimes \frac{j}{2}};$$

and for  $j$  odd,

$$\mathcal{M}_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{j,ab} := \sum_{i=1}^N w_i^{cc,j} \left[ \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right) \right. \\ \left. \times \left( \overrightarrow{\mathbf{a}\gamma_i} - \overrightarrow{\mathbf{a}\mu_{\gamma}} \right)^T \right]^{\otimes \frac{j-1}{2}} \otimes \left( \overrightarrow{\mathbf{a}\chi_i} - \overrightarrow{\mathbf{a}\mu_{\chi}} \right).$$

The second sample cross-moment ( $j = 2$ ) is called *sample cross-covariance* and denoted by  $\Sigma_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{ab} := \mathcal{M}_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{2,ab}$ . If  $\mathcal{E}(\chi) = \{\mu_{\chi}\}$  and  $\mathcal{E}(\gamma) = \{\mu_{\gamma}\}$ , we can write  $\mathcal{M}_{\chi\gamma}^{j,ab} := \mathcal{M}_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{j,ab}$  and  $\Sigma_{\chi\gamma}^{ab} := \Sigma_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{ab}$ ; or even, if  $\mathcal{M}_{\chi\gamma}^j := \mathcal{M}_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{\mu_{\chi}\mu_{\gamma},j}$  and  $\Sigma_{\chi\gamma} := \Sigma_{\chi\gamma, \mu_{\chi}\mu_{\gamma}}^{\mu_{\chi}\mu_{\gamma}}$ .

We represent *Euclidean* sets sample statistics without bold notation. For a set  $\chi$  with points  $\chi_i \in \mathbb{R}^n$ , we have

$$\mu_{\chi} = \arg \min_{c \in \mathbb{R}^n} s_{\chi}^2(c) = \sum_{i=1}^n w_i^m \chi_i;$$

and, for  $j$  even,  $\mathcal{M}_{\chi}^j = \sum_{i=1}^N w_i^{c,j} [(\chi_i - \mu_{\chi}) (\diamond)^T]^{\otimes \frac{j}{2}}$  (similarly for  $j$  odd and for sample cross-moments).

## IV. UNSCENTED KALMAN FILTERS

There are two main concepts required to define UKFs, namely:  $\sigma$ Rs and UTs [11]. Broadly, i) a  $\sigma$ R is a set of weighted points (the sigma points) approximating a random vector, and ii) a UT is a function mapping two functionally related random vectors to two sets that approximate their joint pdf.

For the natural numbers  $l \geq 2$  and  $N \geq 1$ , consider i) a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ; ii) the random vectors  $X \sim$

$(\bar{X}, M_{\bar{X}}^2, \dots, M_{\bar{X}}^l)_{\mathbb{R}^n}$  and  $Y := f(X) \sim (\bar{Y}, M_{\bar{Y}}^2, \dots, M_{\bar{Y}}^l)_{\mathbb{R}^n}$ ; and iii) the sets<sup>2</sup>

$$\chi := \left\{ \chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \chi_i \in \mathbb{R}^n; \right. \\ \left. j = 1, \dots, l; w_i^m, w_i^{c,j}, w_i^{cc,j} \in \mathbb{R} \right\}_{i=1}^N; \text{ and}$$

$$\gamma := \left\{ \gamma_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \gamma_i = f(\chi_i); j = 1, \dots, l \right\}_{i=1}^N.$$

**Definition 1** ( $\sigma$ R. Definition 1 of [11]). The set  $\chi$  is an  $l$ th order  $N$  points  $\sigma$ R ( $l$ th $N$  $\sigma$ R) of  $X$  if, for every  $j = 1, \dots, l$ :

$$w_i^m \neq 0, w_i^{c,j} \neq 0, w_i^{cc,j} \neq 0, \quad i = 1, \dots, N; \quad (6)$$

$$\mu_\chi = \bar{X}; \quad (7)$$

$$M_\chi^j = M_{\bar{X}}^j. \quad (8)$$

**Definition 2** (UT. Definition 2 of [11]). If  $\mu_\chi = \bar{X}$  and  $M_\chi^j = M_{\bar{X}}^j$  for every  $j = 2, \dots, l$ ; then the  $l$ th order UT ( $l$ UT) is defined by

$$lUT : (f, \bar{X}, M_{\bar{X}}^2, \dots, M_{\bar{X}}^l) \mapsto \\ (\mu_\gamma, M_\gamma^2, \dots, M_\gamma^l, M_{\chi\gamma}^2, \dots, M_{\chi\gamma}^l).$$

$\chi$  is called the *independent set* of an  $l$ UT, and  $\gamma$  its *dependent set*.

Every  $l$ th $N$  $\sigma$ R is an independent set of an  $l$ UT. When calling an  $l$ th $N$  $\sigma$ R of  $X$  or an  $l$ UT, the reference to the  $l$ th order can be omitted if  $l = 2$ . Also, the reference to  $N$  point and/or to  $X$  can be omitted in case they are obvious from the context or irrelevant to a discussion.

We can apply UTs in KF prediction-correction frameworks to form UKFs. UKFs estimate the state of systems described either in the additive form

$$x_k = f_k(x_{k-1}) + \varpi_k, y_k = h_k(x_k) + \vartheta_k; \quad (9)$$

or, more generally, in the form

$$x_k = f_k(x_{k-1}, \varpi_k), y_k = h_k(x_k, \vartheta_k), \quad (10)$$

where  $k$  is the time step;  $x_k \in \Phi^{n_x}$  is the internal state;  $y_k \in \Phi^{n_y}$  is the measured output; and  $\varpi_k \in \Phi^{n_\varpi}$  and  $\vartheta_k \in \Phi^{n_\vartheta}$  are the process and measurement noises respectively; the noise terms  $\varpi_k$  and  $\vartheta_k$  are assumed to be uncorrelated.

In [11], we developed consistent UKFs for these systems: the the *Additive UKF* (*AdUKF*, Algorithm 6 of [38]; see also [11]) for (9); and the *Augmented UKF* (*AuUKF*, Algorithm 7 of [38]; see also [11]) for (10). But how could we develop similar UKFs when  $x_k, y_k, \varpi_k$  and  $\vartheta_k$  are Riemannian random points? In the next section, we begin a theory towards this goal.

<sup>2</sup>Compared with [11], here we consider *simpler* sets. With this consideration, we have a clearer text and *do not lose generality* for the results relative to the UKFs.

## V. RIEMANNIAN $\sigma$ -REPRESENTATIONS

In this section, first, we define *Riemannian  $\sigma$ -representations* ( $\text{Ri}\sigma\text{R}$ ). They extend  $\sigma$ Rs to Riemannian manifolds:  $\sigma$ Rs approximate random vectors, and  $\text{Ri}\sigma$ Rs approximate Riemannian random points. Then, we show a way of extending closed forms of  $\sigma$ Rs to  $\text{Ri}\sigma$ Rs. Afterwards, we introduce results relative to the minimum number of sigma points of an  $\text{Ri}\sigma$ R. At last, we introduce some particular forms of  $\text{Ri}\sigma$ Rs.

For now on, we make the following assumptions—we explain their implications in Section VII-C—:

- 1) all Riemannian manifolds are *geodesically-complete*;
- 2) all Riemannian exponential mappings are defined with their domain allowing them to *realize diffeomorphisms*;
- 3) every set of weighted points belonging to a Riemannian manifold admits *one, and only one, Riemannian sample mean*.

For the point  $\mathbf{a} \in \mathcal{N}$  and the natural numbers  $l \geq 2$  and  $N \geq 1$ , consider i) a random point  $\mathbf{X} \sim (\bar{\mathbf{X}}, M_{\bar{\mathbf{X}}, \bar{\mathbf{X}}}^{\mathbf{a}, 2}, \dots, M_{\bar{\mathbf{X}}, \bar{\mathbf{X}}}^{\mathbf{a}, l})_{\mathcal{N}^n}$  and ii) a weighted set  $\chi := \{\chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} | \chi_i \in \mathcal{N}\}_{i=1}^N$  with sample mean  $\mu_\chi$  and sample moments  $M_\chi^j, j = 2, \dots, l$ .

**Definition 3** ( $\text{Ri}\sigma\text{R}$ . Definition 9.1 of [38]). The set  $\chi$  is a *Riemannian  $l$ th order  $N$  points  $\sigma$ -representation* ( $\text{Ri}l\text{th}N\sigma\text{R}$ ) of  $\mathbf{X}$  if, for every  $j = 1, \dots, l$ :

$$w_i^m \neq 0, w_i^{c,j} \neq 0, w_i^{cc,j} \neq 0, \quad i = 1, \dots, N; \quad (11)$$

$$\mu_\chi = \bar{\mathbf{X}}; \quad (12)$$

$$M_\chi^j = M_{\bar{\mathbf{X}}}^j, \quad j = 2, 3, \dots, l; \quad (13)$$

Moreover, assume  $\chi$  is an  $\text{Ri}l\text{th}N\sigma\text{R}$  of  $\mathbf{X}$ , then:

- $\chi$  is *normalized* if, for every  $j = 1, 2, \dots, l$ :

$$\sum_{i=1}^N w_i^m = \sum_{i=1}^N w_i^{c,j} = \sum_{i=1}^N w_i^{cc,j} = 1.$$

- $\chi$  is *homogeneous* if, for every  $j = 1, 2, \dots, l$ , the following equations are satisfied: for  $N$  odd and every  $i = 1, \dots, N-1$ :

$$w_1^m = w_i^m, w_1^{c,j} = w_i^{c,j}, w_1^{cc,j} = w_i^{cc,j}; \quad (14)$$

or, for  $N$  even and every  $i = 1, \dots, N$ :

$$w_1^m = w_i^m, w_1^{c,j} = w_i^{c,j}, w_1^{cc,j} = w_i^{cc,j}. \quad (15)$$

- $\chi$  is *symmetric* (with respect to  $\chi_N$ , without loss of generality) if

$$\overrightarrow{\mu_\chi \chi_i} - \overrightarrow{\mu_\chi \chi_N} = - \left( \overrightarrow{\mu_\chi \chi_{i+\text{int}(\frac{N}{2})}} - \overrightarrow{\mu_\chi \chi_N} \right), \\ w_i^m = w_{i+\text{int}(\frac{N}{2})}^m, w_i^{c,j} = w_{i+\text{int}(\frac{N}{2})}^{c,j}, w_i^{cc,j} = w_{i+\text{int}(\frac{N}{2})}^{cc,j}, \quad (16)$$

for every  $j = 1, 2, \dots, l$  and  $i = 1, \dots, \text{int}(N/2)$ , where  $\text{int}(N/2)$  stands for greatest integer less than or equal to  $N/2$ .

When calling an  $\text{Ri}l\text{th}N\sigma\text{R}$  of  $\mathbf{X}$ , the reference to the  $l$ th order can be omitted if  $l = 2$ . Also, the reference to  $N$  points

or to  $\mathbf{X}$  can be omitted if they are obvious from the context or irrelevant to a discussion.

Rilth $N\sigma$ R are generalizations of lth $N\sigma$ R; every lth $N\sigma$ R with an Rilth $N\sigma$ R, and every Rilth $N\sigma$ R with Euclidean points is an lth $N\sigma$ R. This follows directly from the last paragraph of Sections III-A and of III-B.

Finding closed forms for Ri $\sigma$ R may be troublesome, but the next theorem provides a way of obtaining them from closed forms of  $\sigma$ R—the reader will find several closed forms of  $\sigma$ R in [11], [37], [38].

**Theorem 1** (Theorem 9.1 of [38]). *Suppose that, for every  $i = 1, \dots, N$ ,*

- 1)  $w_i^m > 0$ ,
- 2)  $\Omega(\bar{\mathbf{X}})$  is convex, and
- 3)  $\chi_i \in \mathbb{B}(\bar{\mathbf{X}}, r) \cap \mathcal{C}(\bar{\mathbf{X}})$

where  $0 < r \leq \frac{1}{2} \min\{\text{inj}(\mathcal{N}), \pi/\sqrt{\kappa}\}$  and  $\kappa$  is an upper bound of the sectional curvatures of  $\mathcal{N}$ . Then  $\chi$  is a normalized Rilth $N\sigma$ R of  $\mathbf{X}$  if, and only if,

$$\chi := \left( \log_{\bar{\mathbf{X}}} \chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} \right)_{i=1}^N$$

is a normalized lth $N\sigma$ R of the random vector

$$X \sim ([0]_{n \times 1}, \mathbf{M}_{\mathbf{X}}^2, \dots, \mathbf{M}_{\mathbf{X}}^l)_{T_{\bar{\mathbf{X}}}\mathcal{N}}.$$

Moreover, the following statements are true:

- 1)  $\chi$  is homogeneous if, and only if,  $\chi$  is homogeneous;
- 2)  $\chi$  is symmetric if, and only if,  $\chi$  is symmetric.

The proof of Theorem 1 is given in Appendix B; for conditions to assure the convexity of  $\Omega(\bar{\mathbf{X}})$ , see [43] and references therein.

With this theorem, we can extend some results from lth $N\sigma$ R to Rilth $N\sigma$ R, such as the minimum number of sigma points of an Rilth $N\sigma$ R.

**Corollary 1** (Corollary 9.1 of [38]). *Let i)  $\chi$  be a normalized Rilth $N\sigma$ R of  $\mathbf{X}$  with  $w_i^m > 0$  for every  $i = 1, \dots, N$ ; and ii) the rank of the covariance  $\mathbf{P}_{\mathbf{X}\mathbf{X}}$  be  $r \leq n$ . Then the following statements are true:*

- 1)  $N \geq r + 1$ . If  $N = r + 1$ , then  $\chi$  is called a minimum Rilth $N\sigma$ R of  $\mathbf{X}$ .
- 2) If  $\chi$  is symmetric, then  $N \geq 2r$ . If  $\chi$  is symmetric and  $N = 2r$ , then  $\chi$  is called a minimum symmetric Rilth $N\sigma$ R of  $\mathbf{X}$ .

Moreover, consider the set  $\chi := \{\overrightarrow{\bar{\mathbf{X}}} \chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j}\}_{i=1}^N$  and the random vector  $X \sim ([0]_{n \times 1}, \mathbf{P}_{\mathbf{X}\mathbf{X}})_{T_{\bar{\mathbf{X}}}\mathcal{N}}$ . Then the following statements are true:

- If  $\chi$  is a (normalized) homogeneous minimum symmetric  $\sigma$ R of  $X$  (HoMiSy $\sigma$ R, Corollary 3 of [11]), then  $\chi$  is also minimum and symmetric and is called a Riemannian (normalized) homogeneous minimum symmetric  $\sigma$ -representation of  $\mathbf{X}$ .
- If  $\chi$  is a Rho Minimum  $\sigma$ R of  $X$  (“it is described in the 6th row of Table I of [11] and refereed there as the “Minimum set of [12]”), then  $\chi$  is also minimum, and is called a Riemannian Rho Minimum  $\sigma$ -representation (RiRhoMi $\sigma$ R) of  $\mathbf{X}$ .

- If  $\chi$  is a Minimum  $\sigma$ R of  $X$  (Theorem 3 of [11]), then  $\chi$  is also minimum, and is called a Riemannian Minimum  $\sigma$ -representation (RiMi $\sigma$ R) of  $\mathbf{X}$ .

The proof of Corollary 1 is given in Appendix C.

With Theorem 1 and Corollary 1, we can find an Ri $\sigma$ R ( $w_i^m > 0$  for every  $i = 1, \dots, N$ ) by first finding a normalized  $\sigma$ R in the tangent space of the considered manifold; each normalized  $\sigma$ R (cf. [11] and [38]) have their associated Ri $\sigma$ R (cf. Corollary 1). For instance, suppose we want to calculate the normalized RiMi $\sigma$ R of  $\mathbf{X} \in \Phi_{\mathcal{N}}$  (Corollary 1); that is, we want<sup>3</sup>

$$\chi = \{\chi_i, w_i\}_{i=1}^{n_x+1} = \text{RiMi}\sigma\text{R}\left(\hat{\mathbf{x}}_{k-1|k-1}, \hat{\mathbf{P}}_{\mathbf{x}\mathbf{x}}^{k-1|k-1}\right).$$

We can compute the Mi $\sigma$ R (Theorem 3 of [11])

$$\chi = \{\chi_i, w_i\}_{i=1}^{n_x+1} := \text{Mi}\sigma\text{R}\left([0]_{n_x \times 1}, \hat{\mathbf{P}}_{\mathbf{x}\mathbf{x}}^{k-1|k-1}\right),$$

and then, from Theorem 1, we would have

$$\chi = \left\{ \exp_{\hat{\mathbf{x}}_{k-1|k-1}} \chi_i, w_i \right\}_{i=1}^{n_x+1}.$$

The work [5] introduced this technique [cf. (11) to (17) therein], and here, with Theorem 1 and Corollary 1, we provide its formal justification and required assumptions.

## VI. RIEMANNIAN UNSCENTED TRANSFORMATIONS

Essentially, a UT is an approximation of the joint pdf of two functionally-related random vectors by two weighted sets. For a Riemannian extension of the UT, we develop likewise.

For the natural numbers  $l \geq 2$  and  $N \geq 1$ , consider i) a function  $f : \mathcal{N} \rightarrow \mathcal{R}$ , ii) the random points  $\mathbf{X} \sim (\bar{\mathbf{X}}, \mathbf{M}_{\mathbf{X}}^2, \dots, \mathbf{M}_{\mathbf{X}}^l)_{\mathcal{N}^n}$  and  $\mathbf{Y} := f(\mathbf{X}) \sim (\bar{\mathbf{Y}}, \mathbf{M}_{\mathbf{Y}}^2, \dots, \mathbf{M}_{\mathbf{Y}}^l)_{\mathcal{R}^n}$ , and iii) the sets

$$\begin{aligned} \chi &:= \left\{ \chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \chi_i \in \mathcal{N}; \right. \\ &\quad \left. j = 1, \dots, l; w_i^m, w_i^{c,j}, w_i^{cc,j} \neq 0 \right\}_{i=1}^N \text{ and} \\ \gamma &:= \left\{ \gamma_i, w_i^m, w_i^{c,j}, w_i^{cc,j} : \gamma_i = f(\chi_i); j = 1, \dots, l \right\}_{i=1}^N. \end{aligned}$$

**Definition 4** (RilUT; Definition of 9.2 [38]). If  $\mu_{\mathbf{X}} = \bar{\mathbf{X}}$  and  $\mathcal{M}_{\mathbf{X}}^j = \mathbf{M}_{\mathbf{X}}^j$  for every  $j = 2, \dots, l$ ; then the  $l$ th order Riemannian Unscented Transformation (RilUT) is defined by

$$\begin{aligned} \text{RilUT} : (f, \bar{\mathbf{X}}, \mathbf{M}_{\mathbf{X}}^2, \dots, \mathbf{M}_{\mathbf{X}}^l) &\mapsto \\ &(\mu_{\gamma}, \mathcal{M}_{\gamma}^2, \dots, \mathcal{M}_{\gamma}^l, \mathcal{M}_{\chi\gamma}^2, \dots, \mathcal{M}_{\chi\gamma}^l). \end{aligned}$$

$\chi$  is called the independent set of RilUT, and  $\gamma$  its dependent set.

Every Rilth $N\sigma$ R is an independent set of an RilUT. If  $l = 2$  or  $l$  is irrelevant for a given discussion, we can omit the reference to  $l$  and write RiUT := Ri2UT.

RilUTs are generalizations of lUTs; every lUT is an RiUT, and every RiUT with Euclidean points is an lUT. This follows directly from the last paragraph of Sections III-A and of III-B.

<sup>3</sup>For a set  $\xi := \{\xi_i, w_i^{m,j}, w_i^{c,j}, w_i^{cc,j}\}$ , if  $w_i^{m,j} = w_i^{c,j} = w_i^{cc,j}$  for every  $j = 1, \dots, l$ ; then we write  $w_i := w_i^{m,j}$  and  $\{\xi_i, w_i\} = \xi$ .

An RiUT can be viewed as a mapping from 2 random points  $\mathbf{X} \in \Phi_{\mathcal{N}}$  and  $\mathbf{Y} := \mathbf{f}(\mathbf{X})$  to two Riemannian sets  $\chi$  and  $\gamma$  acting as a *discrete approximation* of the joint pdf of  $(\mathbf{X}, \mathbf{Y})$ . For instance, an Ri2UT can be viewed as the following approximation (this interpretation is inspired on [44])

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \approx \begin{pmatrix} \tilde{\mathbf{X}} \\ \tilde{\mathbf{Y}} \end{pmatrix} \sim \left( \begin{pmatrix} \mu_{\chi} \\ \mu_{\gamma} \end{pmatrix}, \begin{pmatrix} \Sigma_{\chi\chi} & \Sigma_{\chi\gamma} \\ \Sigma_{\chi\gamma}^T & \Sigma_{\gamma\gamma} \end{pmatrix} \right).$$

## VII. RIEMANNIAN UNSCENTED KALMAN FILTERS

At this point, we still need to develop i) Riemannian systems; and ii) state correction equations. First, UKFs estimate systems with random vectors [cf. (9) and (10)]; thus, for Riemannian UKFs (RiUKFs), we define *systems with Riemannian random points* (Section VII-A). Second, three steps compose UKFs: 1) state prediction, 2) measurement prediction, and 3) state correction (cf. [11] and [38]). The Riemannian extensions of steps 1) and 2) are trivial: since UTs compose steps 1 and 2, we extend them with *RiUTs*. But we still must extend step 3 (Section VII-B). In possession of these two results, we define RiUKFs and provide a list of some particular forms (Section VII-C).

### A. Riemannian Dynamics Systems

Up to this point, we have focused on results regarding points on manifolds. In this section, we focus on results for dynamic state-space systems on Riemannian manifolds.

The *Riemannian (stochastic discrete-time dynamic) system* in its *general* form is given by the following pair of equations:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \varpi_k), \quad \mathbf{y}_k = h_k(\mathbf{x}_k, \vartheta_k) \quad (17)$$

where  $k$  is the time step;  $\mathbf{x}_k \in \Phi_{\mathcal{N}_{\mathbf{x}}^{n_x}}$  the internal state;  $\mathbf{y}_k \in \Phi_{\mathcal{N}_{\mathbf{y}}^{n_y}}$  is the measured output;  $\varpi_k \in \Phi_{\mathcal{N}_{\varpi}^{n_{\varpi}}}$  the process noise; and  $\vartheta_k \in \Phi_{\mathcal{N}_{\vartheta}^{n_{\vartheta}}}$  the measurement noise. The noises  $\varpi_k$  and  $\vartheta_k$  are uncorrelated,  $\varpi_k$  has mean  $\bar{\varpi}_k$  and covariance  $\mathbf{Q}_k$ , and  $\vartheta_k$  mean  $\bar{\vartheta}_k$  and covariance  $\mathbf{R}_k$ .

We also want to consider an *additive variant* of (17) because filters for this class of systems are computationally cheaper. This additive variant of (17) would have i)  $\varpi_k$  acting on  $f_k(\mathbf{x}_{k-1})$  by “adding” its mean to the mean of  $f_k(\mathbf{x}_{k-1})$  and its covariance to the covariance of  $f_k(\mathbf{x}_{k-1})$ , and ii)  $\vartheta_k$  acting similarly on  $h_k(\mathbf{x}_k)$ . We can work with sums in tangent spaces using the following proposition.

**Proposition 1** (Proposition 8.2 of [38]). *Consider a Riemannian point  $\mathbf{X} \sim (\bar{\mathbf{X}}, \mathbf{P}_{\mathbf{X}\mathbf{X}})_{\mathcal{N}^n}$  and a random vector  $p \sim (\bar{p}, P_{pp})_{T_{\bar{\mathbf{X}}}\mathcal{N}^n}$ . If  $\Omega(\bar{\mathbf{X}})$  is convex, and  $\bar{p} \in \mathbb{B}(\bar{\mathbf{X}}, r) \cap \mathcal{C}(\bar{\mathbf{X}})$  where  $0 < r \leq \frac{1}{2} \min\{\text{inj}(\mathcal{N}), \pi/\sqrt{\kappa}\}$  and  $\kappa$  is an upper bound of the sectional curvatures of  $\mathcal{N}$ ; then*

$$\exp_{\bar{\mathbf{X}}}[\bar{\mathbf{X}}\bar{\mathbf{X}} + p] \sim (\exp_{\bar{\mathbf{X}}}\bar{p}, \mathbf{P}_{\mathbf{X}\mathbf{X}} + P_{pp})_{\mathcal{N}_{\mathbf{x}}}. \quad (18)$$

The proof of Proposition 1 is in Appendix D.

Consider this proposition twice: one for the process function with  $\mathbf{a} = f_k(\mathbf{x}_{k-1})$  and  $p = \varpi_k$ , and the other for the measurement function with  $\mathbf{a} = h_k(\mathbf{x}_k)$  and  $p = \vartheta_k$ . Using this reasoning, we define the *additive Riemannian (stochastic*

*discrete-time dynamic) system* as follows {equation (9.20) of [38]}:

$$\begin{aligned} \mathbf{x}_k &= \exp_{f_k(\mathbf{x}_{k-1})} \left[ \log_{f_k(\mathbf{x}_{k-1})} f_k(\mathbf{x}_{k-1}) + \varpi_k \right] \\ \mathbf{y}_k &= \exp_{h_k(\mathbf{x}_k)} \left[ \log_{h_k(\mathbf{x}_k)} h_k(\mathbf{x}_k) + \vartheta_k \right]; \end{aligned} \quad (19)$$

where  $\mathbf{x}_k \in \Phi_{\mathcal{N}_{\mathbf{x}}^{n_x}}$ ,  $\mathbf{y}_k \in \Phi_{\mathcal{N}_{\mathbf{y}}^{n_y}}$ ,  $\varpi_k \in T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$ , and  $\vartheta_k \in T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$ . The noise  $\varpi_k$  has mean  $\bar{\varpi}_k \in T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$  and covariance  $\mathbf{Q}_k \in T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x} \times T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$ , and  $\vartheta_k$  mean  $\bar{\vartheta}_k \in T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$  and covariance  $\mathbf{R}_k \in T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y} \times T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$ . Note that  $\varpi_k$  is defined in the tangent space  $T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$  and  $\vartheta_k$  in  $T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$ . In Remark 1, we discuss an alternative definition in which these noises belong to Riemannian manifolds. An example with the unit sphere manifold of dimension 3,  $S^3$ , is provided in Section VIII.

To the best of our knowledge, (19) is the *first consistent additive-noise Riemannian system*. Although the literature has introduced additive-noise discrete-time UKFs for some Riemannian manifolds, we could not find any additive-noise system retaining the random point in the working manifolds; even for simple manifolds such as  $S^3$  (cf. [12], [39], [45]).

If  $\mathcal{N}_{\mathbf{x}}^{n_x} = \mathbb{R}^{n_x}$  and  $\mathcal{N}_{\mathbf{y}}^{n_y} = \mathbb{R}^{n_y}$  then (19) is the *additive system* (9). This is a direct consequence of the following results: for  $a, b \in \mathbb{R}^n$   $\log_a b = b - a$  and  $\exp_a b = b + a$ .

Sometimes, only one of the two equations in (17) can be written with additive-noise as in (19). In this case, we define the following two partially-additive Riemannian systems:

$$\begin{aligned} \mathbf{x}_k &= f_k(\mathbf{x}_{k-1}, \varpi_k) \\ \mathbf{y}_k &= \exp_{h_k(\mathbf{x}_k)} \left[ \log_{h_k(\mathbf{x}_k)} h_k(\mathbf{x}_k) + \vartheta_k \right]; \end{aligned} \quad (20)$$

and

$$\begin{aligned} \mathbf{x}_k &= \exp_{f_k(\mathbf{x}_{k-1})} \left[ \log_{f_k(\mathbf{x}_{k-1})} f_k(\mathbf{x}_{k-1}) + \varpi_k \right] \\ \mathbf{y}_k &= h_k(\mathbf{x}_k, \vartheta_k). \end{aligned} \quad (21)$$

*Remark 1.* System (19) is defined with tangent space process and measurement noises. An alternative definition in which these noises belong to Riemannian manifolds is the following:

$$\begin{aligned} \mathbf{x}_k &= \exp_{f_k(\mathbf{x}_{k-1})} \left[ \log_{f_k(\mathbf{x}_{k-1})} f_k(\mathbf{x}_{k-1}) + \log_{f_k(\mathbf{x}_{k-1})} \varpi_k \right] \\ \mathbf{y}_k &= \exp_{h_k(\mathbf{x}_k)} \left[ \log_{h_k(\mathbf{x}_k)} h_k(\mathbf{x}_k) + \log_{h_k(\mathbf{x}_k)} \vartheta_k \right]; \end{aligned}$$

where  $\mathbf{x}_k \in \Phi_{\mathcal{N}_{\mathbf{x}}^{n_x}}$ ,  $\mathbf{y}_k \in \Phi_{\mathcal{N}_{\mathbf{y}}^{n_y}}$ ,  $\varpi_k \in \Phi_{\mathcal{N}_{\varpi}^{n_{\varpi}}}$ , and  $\vartheta_k \in \Phi_{\mathcal{N}_{\vartheta}^{n_{\vartheta}}}$ . In this case, it would be interesting to assume one of the following two cases:

- 1) That are known i) the means of  $\varpi_k$  and  $\vartheta_k$ —e.g.,  $\bar{\varpi}_k \in \mathcal{N}_{\mathbf{x}}^{n_x} - \mathcal{C}(f_k(\mathbf{x}_{k-1}))$  and  $\bar{\vartheta}_k \in \mathcal{N}_{\mathbf{y}}^{n_y} - \mathcal{C}(h_k(\mathbf{x}_k))$ —, b) the covariance of  $\varpi_k$  with respect to  $\bar{\varpi}_k$  at  $f_k(\mathbf{x}_{k-1})$ , and iii) the covariance of  $\vartheta_k$  with respect to  $\bar{\vartheta}_k$  at  $h_k(\mathbf{x}_k)$ .
- 2) That the means and covariances of  $\log_{f_k(\mathbf{x}_{k-1})} \varpi_k$  and  $\log_{h_k(\mathbf{x}_k)} \vartheta_k$  are known—e.g., the means  $\bar{\varpi}_k \in T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$  and  $\bar{\vartheta}_k \in T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$ ; and the covariances  $\mathbf{Q}_k \in T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x} \times T_{f_k(\mathbf{x}_{k-1})}\mathcal{N}_{\mathbf{x}}^{n_x}$  and  $\mathbf{R}_k \in T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y} \times T_{h_k(\mathbf{x}_k)}\mathcal{N}_{\mathbf{y}}^{n_y}$ .

## B. Correction equations

In this section, we introduce Riemannian extensions of the UKFs correction equations. Finding these extensions is not trivial because their Euclidean versions include vector operations (cf. [11]), which are not defined for all Riemannian manifolds. Thus, we proceed by first considering the simpler case  $\Phi_{\mathcal{N}_x^{n_x}} = \Phi_{\mathcal{N}_y^{n_y}}$ .

1) *State and measurement in the same manifold:* Suppose that  $\mathcal{N}_x^{n_x} = \mathcal{N}_y^{n_y}$  and the measurements  $\check{y}_1, \dots, \check{y}_k$  have been acquired. Define the following random points<sup>4</sup>

$$\mathbf{x}_{k|k-1} := \mathbf{x}_k | \check{y}_{1:k-1}$$

$$\mathbf{x}_{k|k} := \mathbf{x}_k | \check{y}_{1:k}$$

$$\mathbf{y}_{k|k-1} := \mathbf{y}_k | \check{y}_{1:k-1},$$

and the following projections on the tangent space of  $\mathbf{x}_{k|k-1}$

$$x_{k|k-1}^{TM} := \log_{\bar{\mathbf{x}}_{k|k-1}} \mathbf{x}_{k|k-1} \quad (22)$$

$$x_{k|k}^{TM} := \log_{\bar{\mathbf{x}}_{k|k}} \mathbf{x}_{k|k}$$

$$y_{k|k-1}^{TM} := \log_{\bar{\mathbf{x}}_{k|k-1}} \mathbf{y}_{k|k-1} \quad (23)$$

$$\check{y}_k^{TM} := \log_{\bar{\mathbf{x}}_{k|k-1}} \check{y}_k. \quad (24)$$

Let i)  $\mathbf{x}_{k|k-1}$  and  $\mathbf{y}_{k|k-1}$  be characterized by their projection on the tangent space of  $\mathbf{x}_{k|k-1}$  according to the following equation:

$$\begin{bmatrix} x_{k|k-1}^{TM} & y_{k|k-1}^{TM} \end{bmatrix}^T \sim N \left( \begin{bmatrix} [0]_{n_x,1} \\ y_{k|k-1}^{TM} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1} & \mathbf{P}_{\mathbf{x}\mathbf{y}}^{k|k-1} \\ \left( \mathbf{P}_{\mathbf{x}\mathbf{y}}^{k|k-1} \right)^T & \mathbf{P}_{\mathbf{y}\mathbf{y}}^{k|k-1} \end{bmatrix} \right); \quad (25)$$

and ii) the projection  $x_{k|k}^{TM}$  be given by the following linear correction of  $x_{k|k-1}^{TM}$

$$x_{k|k}^{TM} = x_{k|k-1}^{TM} + \mathbf{G}_k \left( \check{y}_k^{TM} - y_{k|k-1}^{TM} \right), \quad (26)$$

where  $\mathbf{G}_k \in \mathbb{R}^{n_x \times n_x}$  is a gain matrix. From known results of the Kalman filtering theory (cf. [46]), we have

$$\mathbf{G}_k := \mathbf{P}_{\mathbf{x}\mathbf{y}}^{k|k-1} \left( \mathbf{P}_{\mathbf{y}\mathbf{y}}^{k|k-1} \right)^{-1}, \quad (27)$$

and  $x_{k|k}^{TM} \sim N(\bar{x}_{k|k}^{TM}, \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k})$  where

$$\bar{x}_{k|k}^{TM} := \mathbf{G}_k \left( \check{y}_k^{TM} - \bar{y}_{k|k-1}^{TM} \right) \quad (28)$$

$$\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k} := \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1} - (\mathbf{G}_k) \mathbf{P}_{\mathbf{y}\mathbf{y}}^{k|k-1} (\mathbf{G}_k)^T. \quad (29)$$

From (22), we have

$$\bar{\mathbf{x}}_{k|k} = \exp_{\bar{\mathbf{x}}_{k|k-1}} x_{k|k}^{TM}. \quad (30)$$

The matrix  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1, \bar{\mathbf{x}}_{k|k-1}}$  is the covariance of  $\mathbf{x}_{k|k}$  relative to  $\bar{\mathbf{x}}_{k|k}$  at  $\bar{\mathbf{x}}_{k|k-1}$ . We want the covariance  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k} := \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k, \bar{\mathbf{x}}_{k|k}}$  of  $\mathbf{x}_{k|k}$  at  $\bar{\mathbf{x}}_{k|k}$ , and the following theorem from [5] provides the mechanism to obtain  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k, \bar{\mathbf{x}}_{k|k}}$  from  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1, \bar{\mathbf{x}}_{k|k-1}}$ .

<sup>4</sup>For the random points  $\mathbf{X}$  and  $\mathbf{Y}$  and the outcomes  $\check{Y}_1, \dots, \check{Y}_l$  of  $\mathbf{Y}$ ; the random point  $\mathbf{X} | \check{Y}_{1:k-1}$  stands for  $\mathbf{X}$  conditioned to  $\mathbf{Y}_i = \check{Y}_i$  for every  $i = 1, \dots, l$ .

**Theorem 2** (Parallel Transport of a Bilinear Mapping [5]). *Let  $P$  be a symmetric bilinear mapping on the tangent space  $T_{\mathbf{a}}\mathcal{N}$  of the Riemannian manifold  $\mathcal{N}$  at  $\mathbf{a} \in \mathcal{N}$ , and  $\alpha : [0, 1] \rightarrow \mathcal{N}$  a differentiable curve on  $\mathcal{N}$  with  $\alpha(0) = \mathbf{a}$ . Since  $P$  is symmetric, it can be written as*

$$P = \sum_{i=1}^n \lambda_i v_i v_i^T$$

where  $(v_1, \dots, v_n)$  is an orthonormal basis of  $T_{\mathbf{a}}\mathcal{N}$ , and each  $\lambda_i$  is the eigenvalue of  $P$  associated with the eigenvector  $v_i$ . Let  $v_i(t)$  be the parallel transport of  $v_i$  along  $\alpha(t)$  (Definition 15). With this,

$$P_t := \sum_{i=1}^n \lambda_i v_i(t) v_i(t)^T \quad (31)$$

is the parallel transport of  $P$  along  $\alpha(t)$ .

When we do not know the closed form of a tangent vector parallel transport, we can use a numerical approach such as the *Schild's Ladder* (cf. [5]; see [47] for other implementations and algorithms of parallel transports).

We obtain  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k}$  by performing the parallel transport of  $\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1, \bar{\mathbf{x}}_{k|k-1}}$  from  $\bar{\mathbf{x}}_{k|k-1}$  to  $\bar{\mathbf{x}}_{k|k}$  as follows:

$$\mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k} = \text{PT} \left( \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1, \bar{\mathbf{x}}_{k|k-1}}, \bar{\mathbf{x}}_{k|k-1}, \bar{\mathbf{x}}_{k|k} \right), \quad (32)$$

where

$$\begin{aligned} \text{PT} : \text{Sym}(T_{\mathbf{a}}\mathcal{N}) \times \mathcal{N} \times \mathcal{N} &\rightarrow \text{Sym}(T_{\mathbf{b}}\mathcal{N}) \\ (P^{\mathbf{a}}, \mathbf{a}, \mathbf{b}) &\mapsto P^{\mathbf{b}} \end{aligned}$$

is the function mapping  $\text{Sym}(T_{\mathbf{a}}\mathcal{N}) \times \mathcal{N} \times \mathcal{N}$  to  $\text{Sym}(T_{\mathbf{b}}\mathcal{N})$  according to (31), and  $\text{Sym}(T_{\mathbf{a}}\mathcal{N})$  denotes the space of symmetric matrices of  $T_{\mathbf{a}}\mathcal{N}$ .

With this, we can define a UKF for Riemannian systems when  $\mathcal{N}_x^{n_x} = \mathcal{N}_y^{n_y}$ . Let us now consider the original more general case.

2) *State and measurement in different manifolds:* If  $\mathbf{x}_k$  belongs to a manifold  $\Phi_{\mathcal{N}_x^{n_x}}$  and  $\mathbf{y}_k$  to another manifold  $\Phi_{\mathcal{N}_y^{n_y}}$ , then we can not define  $y_{k|k-1}^{TM}$  as in (23) and  $\check{y}_k^{TM}$  as in (24); consequently, neither  $x_{k|k}^{TM}$  as in (26).

Since we know the correction equations when  $\mathcal{N}_x^{n_x} = \mathcal{N}_y^{n_y}$ , we can look for a manifold of which both  $\mathcal{N}_x$  and  $\mathcal{N}_y$  are submanifolds. The simplest of such a class is  $\mathcal{N}_x \times \mathcal{N}_y$ —the Cartesian product of two Riemannian manifolds is a Riemannian manifold [42].

Suppose  $x_{k|k-1}^{TM}$  and  $y_{k|k-1}^{TM}$  are jointly Gaussian random vectors according to (25). Define i) the Riemannian Manifold  $\mathcal{N}_{\mathbf{x}, \mathbf{y}} := \mathcal{N}_x \times \mathcal{N}_y$ ; ii) the points  $\mathbf{c} := (\mathbf{c}_x, \mathbf{c}_y) \in \mathcal{N}_{\mathbf{x}, \mathbf{y}}$ ,  $\mathbf{b}_x \in \mathcal{N}_x$ , and  $\mathbf{b}_y \in \mathcal{N}_y$  (these points are chosen); and the following random vector belonging to  $T_{\mathbf{c}}\mathcal{N}_{\mathbf{x}, \mathbf{y}}$ :

$$\begin{aligned} x_{k|k, **}^{T_{\mathbf{c}}\mathcal{N}_{\mathbf{x}, \mathbf{y}}} &:= \log_{\mathbf{c}} [\mathbf{x}_{k|k-1}, \mathbf{b}_y]^T \\ &+ \mathbf{G}_{k, **} \left( \log_{\mathbf{c}} [\mathbf{b}_x, \check{y}_k]^T - \log_{\mathbf{c}} [\mathbf{b}_x, \mathbf{y}_{k|k-1}]^T \right) \end{aligned}$$

where  $\mathbf{G}_{k, **} \in \mathbb{R}^{(n_x+n_y) \times (n_x+n_y)}$  is a gain matrix. The tangent vector  $x_{k|k, **}^{T_{\mathbf{c}}\mathcal{N}_{\mathbf{x}, \mathbf{y}}}$  is clearly related with  $x_{k|k}^{TM}$  by

$$x_{k|k}^{TM} := \left[ \hat{x}_{k|k, **}^{T_{\mathbf{c}}\mathcal{N}_{\mathbf{x}, \mathbf{y}}} \right]_{1:n_x, 1}. \quad (33)$$

By finding the mean and covariance of  $x_{k|k}^{T_c \mathcal{N}_{x,y}}$ , we find the mean and covariance of  $x_{k|k}^{TM}$ .

Since  $x_{k|k-1}^{TM}$  and  $y_{k|k-1}^{TM}$  are jointly Gaussian random vectors, it follows that—we use the same reasoning used to obtain (27), (28), (29), (30), and (32)—

$$\begin{aligned} P_{xx,**}^{k|k-1} &:= \mathcal{E} \left\{ \left( \log_c \begin{bmatrix} \mathbf{x} \\ \mathbf{b}_y \end{bmatrix} - \log_c \begin{bmatrix} \bar{\mathbf{x}}_{k|k-1} \\ \mathbf{b}_y \end{bmatrix} \right) (\diamond)^T \right\} \\ P_{yy,**}^{k|k-1} &:= \mathcal{E} \left\{ \left( \log_c \begin{bmatrix} \mathbf{b}_x \\ \mathbf{y} \end{bmatrix} - \log_c \begin{bmatrix} \mathbf{b}_x \\ \bar{\mathbf{y}}_{k|k-1} \end{bmatrix} \right) (\diamond)^T \right\} \end{aligned}$$

$$\begin{aligned} P_{xy,**}^{k|k-1} &:= \mathcal{E} \left\{ \left( \log_c \begin{bmatrix} \mathbf{x} \\ \mathbf{b}_y \end{bmatrix} - \log_c \begin{bmatrix} \bar{\mathbf{x}}_{k|k-1} \\ \mathbf{b}_y \end{bmatrix} \right) \right. \\ &\quad \times \left. \left( \log_c \begin{bmatrix} \mathbf{b}_x \\ \mathbf{y} \end{bmatrix} - \log_c \begin{bmatrix} \mathbf{b}_x \\ \bar{\mathbf{y}}_{k|k-1} \end{bmatrix} \right)^T \right\}; \end{aligned}$$

thus, the mean and covariance of  $x_{k|k}^{T_c \mathcal{N}_{x,y}}$  are given by

$$\begin{aligned} \mathbf{G}_{k,**} &:= P_{xy,**}^{k|k-1} \text{diag} \left( [0]_{n_x \times n_x}, \right. \\ &\quad \left. \mathcal{E} \{ (\log_{c_y} \mathbf{y} - \log_{c_y} \bar{\mathbf{y}}_{k|k-1}) (\diamond)^T \}^{-1} \right) \quad (34) \end{aligned}$$

$$\bar{x}_{k|k}^{T_c \mathcal{N}_{x,y}} := \log_c \begin{bmatrix} \bar{\mathbf{x}}_{k|k-1} \\ \mathbf{b}_y \end{bmatrix} + \mathbf{G}_{k,**} \log_c \begin{bmatrix} \mathbf{b}_x \\ \bar{\mathbf{y}}_{k|k-1} \end{bmatrix} \quad (35)$$

$$P_{xx,**}^{k|k, T_c M} := P_{xx,**}^{k|k-1} - (\mathbf{G}_{k,**}) P_{yy,**}^{k|k-1} (\mathbf{G}_{k,**})^T. \quad (36)$$

We can choose  $\mathbf{c}$ ,  $\mathbf{b}_x$  and  $\mathbf{b}_y$  arbitrarily, and a particular choice yields the desired correction equations.

**Theorem 3** (Theorem 9.3 of [38]). *Given (33), (34), (35), and (36); if  $\mathbf{c}_x = \mathbf{b}_x = \hat{\mathbf{x}}_{k|k-1}$  and  $\mathbf{c}_y = \mathbf{b}_y = \hat{\mathbf{y}}_{k|k-1}$ , then*

$$x_{k|k}^{TM} = \mathbf{G}_k \log_{\hat{\mathbf{y}}_{k|k-1}}(\mathbf{y}_k) \quad (37)$$

and

$$P_{xx}^{k|k, \bar{\mathbf{x}}_{k|k-1}} = \hat{P}_{xx}^{k|k-1} - \mathbf{G}_k \left( \hat{P}_{yy}^{k|k-1} \right)^{-1} (\mathbf{G}_k)^T, \quad (38)$$

where

$$\mathbf{G}_k := P_{xy}^{k|k-1} \left( P_{yy}^{k|k-1} \right)^{-1}.$$

The proof of Theorem 3 is in Appendix E.

According to this theorem, the correction equations—(27), (28), (29), (30), and (32)—are true even when the state and the measurement belong to different manifolds. Therefore, we do not have to perform calculations on the bigger manifold  $\mathcal{N}_{x,y}$  to calculate  $x_{k|k}^{TM}$  and  $P_{xx}^{k|k, \bar{\mathbf{x}}_{k|k-1}}$ . Instead, they can be calculated by (37) and (38) even when  $\mathcal{N}_x \neq \mathcal{N}_y$ .

### C. New Riemannian Unscented Kalman Filters

At this point, we are endowed with the necessary results to provide Riemannian extensions of UKFs. At every step time, the final estimates  $\hat{\mathbf{x}}_{k|k}$  and  $\hat{P}_{xx}^{k|k}$  can be calculated by (32) and Theorem 3. From (27), (32), (37), and (38) these final estimates require  $\hat{\mathbf{x}}_{k|k-1}$ ,  $\hat{P}_{xx}^{k|k-1}$ ,  $\hat{\mathbf{y}}_{k|k-1}$ ,  $\hat{P}_{yy}^{k|k-1}$ , and  $\hat{P}_{xy}^{k|k-1}$ . These last estimates can be calculated by realizing RiUTs in systems (17) and (19). For instance, from (19),

Definition 4 and Proposition 1, the estimates  $\hat{\mathbf{x}}_{k|k-1}$ ,  $\hat{P}_{xx}^{k|k-1}$  can be calculated by

$$\begin{aligned} \left( \hat{\mathbf{y}}_{k|k-1}^*, \hat{P}_{yy,**}^{k|k-1}, \hat{P}_{xy}^{k|k-1} \right) &:= \\ &\text{RiUT}_2 \left( h_k, \hat{\mathbf{x}}_{k|k-1}, \hat{P}_{xx}^{k|k-1} \right) \\ \hat{\mathbf{y}}_{k|k-1} &:= \exp_{\hat{\mathbf{y}}_{k|k-1}^*} \bar{\vartheta}_k \\ \hat{P}_{yy}^{k|k-1} &:= \hat{P}_{yy,**}^{k|k-1} + R_k. \end{aligned}$$

By similar formulas, we can obtain  $\hat{\mathbf{x}}_{k|k-1}$ ,  $\hat{P}_{xx}^{k|k-1}$ ,  $\hat{\mathbf{y}}_{k|k-1}$ ,  $\hat{P}_{yy}^{k|k-1}$ , and  $\hat{P}_{xy}^{k|k-1}$  for both (17) and (19).

Below, we introduce the Riemannian UKFs (RiUKFs): UKFs for the Riemannian systems (17) and (19). For the filter of (17), define the augmented functions

$$\begin{aligned} f_k^{aug}([\mathbf{x}_{k-1}, \boldsymbol{\varpi}_k]^T) &:= f_k(\mathbf{x}_{k-1}, \boldsymbol{\varpi}_k) \quad (39) \\ h_k^{aug}([\mathbf{x}_k, \boldsymbol{\vartheta}_k]^T) &:= h_k(\mathbf{x}_k, \boldsymbol{\vartheta}_k). \end{aligned}$$

Consider system (17) and suppose that i) the initial state is  $\mathbf{x}_0 \sim (\bar{\mathbf{x}}_0, P_{xx}^0)_{\mathcal{N}_x}$ , and ii) the measurements  $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_{k_f}$  are given. Then the *Riemannian Augmented Unscented Kalman Filter* (RiAuUKF) is given by the following algorithm:

**Algorithm 1** (RiAuUKF; Algorithm 19 of [38]). *Set the initial estimates  $\hat{\mathbf{x}}_{0|0} := \bar{\mathbf{x}}_0$  and  $\hat{P}_{xx}^{0|0} := P_{xx}^0$ . For  $k = 1, \dots, k_f$ , perform the following steps:*

1) *State prediction.*

$$\begin{aligned} \hat{\mathbf{x}}_{k-1|k-1}^{aug} &:= \left[ \hat{\mathbf{x}}_{k-1|k-1}^T, \bar{\boldsymbol{\varpi}}_k^T \right]^T \\ \hat{P}_{xx,aug}^{k-1|k-1} &:= \text{diag} \left( \hat{P}_{xx}^{k-1|k-1}, \mathbf{Q}_k \right) \end{aligned}$$

$$\begin{aligned} \left( \hat{\mathbf{x}}_{k|k-1}, \hat{P}_{xx}^{k|k-1} \right) &:= \\ &\text{RiUT}_1 \left( J_k^{aug}, \hat{\mathbf{x}}_{k-1|k-1}^{aug}, \hat{P}_{xx,aug}^{k-1|k-1} \right). \quad (40) \end{aligned}$$

2) *Measurement prediction.*

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1}^{aug} &:= \left[ \hat{\mathbf{x}}_{k|k-1}^T, \bar{\boldsymbol{\vartheta}}_k^T \right]^T \\ \hat{P}_{xx,aug}^{k|k-1} &:= \text{diag} \left( \hat{P}_{xx}^{k|k-1}, \mathbf{R}_k \right). \end{aligned}$$

$$\begin{aligned} \left( \hat{\mathbf{y}}_{k|k-1}, \hat{P}_{yy}^{k|k-1}, \hat{P}_{xy,a}^{k|k-1} \right) &:= \\ &\text{RiUT}_2 \left( h_k^{aug}, \hat{\mathbf{x}}_{k|k-1}^{aug}, \hat{P}_{xx,aug}^{k|k-1} \right) \quad (41) \end{aligned}$$

$$\hat{P}_{xy}^{k|k-1} := \left[ \hat{P}_{xy,aug}^{k|k-1} \right]_{(1:n_x), (1:n_y)}.$$

3) *State correction.*

$$\begin{aligned} \mathbf{G}_k &:= \hat{P}_{xy}^{k|k-1} \left( \hat{P}_{yy}^{k|k-1} \right)^{-1} \quad (42) \\ \hat{\mathbf{x}}_{k|k}^{TM} &:= \mathbf{G}_k \log_{\hat{\mathbf{y}}_{k|k-1}}(\hat{\mathbf{y}}_k) \\ \hat{\mathbf{x}}_{k|k} &:= \exp_{\hat{\mathbf{x}}_{k|k-1}}(\hat{\mathbf{x}}_{k|k}^{TM}) \\ \hat{P}_{xx}^{k|k, \bar{\mathbf{x}}_{k|k-1}} &:= \hat{P}_{xx}^{k|k-1} - \mathbf{G}_k \hat{P}_{yy}^{k|k-1} \mathbf{G}_k^T \end{aligned}$$



$$\hat{\mathbf{P}}_{xx}^{k|k} := \text{PT}\left(\hat{\mathbf{P}}_{xx}^{k|k, \hat{\mathbf{x}}_{k|k-1}}, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{x}}_{k|k}\right).$$

Consider the system (19) and suppose that i) the initial state is  $\mathbf{x}_0 \sim (\bar{\mathbf{x}}_0, \mathbf{P}_{xx}^0)_{\mathcal{N}_x}$ , and ii) the measurements  $\check{\mathbf{y}}_1, \check{\mathbf{y}}_2, \dots, \check{\mathbf{y}}_{k_f}$  are given. Then the *Riemannian Additive Unscented Kalman Filter* (RiAdUKF) is given by the following algorithm:

**Algorithm 2** (RiAdUKF; Algorithm 21 of [38]). *Set the initial estimates  $\hat{\mathbf{x}}_{0|0} := \bar{\mathbf{x}}_0$  and  $\hat{\mathbf{P}}_{xx}^{0|0} := \mathbf{P}_{xx}^0$ . For  $k = 1, \dots, k_f$ , perform the following steps:*

1) *State prediction.*

$$\left(\hat{\mathbf{x}}_{k|k-1}^*, \hat{\mathbf{P}}_{xx,*}^{k|k-1}\right) := \text{RiUT}_1\left(f_k, \hat{\mathbf{x}}_{k-1|k-1}, \hat{\mathbf{P}}_{xx}^{k-1|k-1}\right) \quad (43)$$

$$\hat{\mathbf{x}}_{k|k-1} := \exp_{\hat{\mathbf{x}}_{k|k-1}^*} \bar{\omega}_k \quad (44)$$

$$\hat{\mathbf{P}}_{xx}^{k|k-1} := \hat{\mathbf{P}}_{xx,*}^{k|k-1} + Q_k. \quad (45)$$

2) *Measurement prediction.*

$$\left(\hat{\mathbf{y}}_{k|k-1}^*, \hat{\mathbf{P}}_{yy,*}^{k|k-1}, \hat{\mathbf{P}}_{xy}^{k|k-1}\right) := \text{RiUT}_2\left(h_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{xx}^{k|k-1}\right) \quad (46)$$

$$\hat{\mathbf{y}}_{k|k-1} := \exp_{\hat{\mathbf{y}}_{k|k-1}^*} \bar{\vartheta}_k \quad (47)$$

$$\hat{\mathbf{P}}_{yy}^{k|k-1} := \hat{\mathbf{P}}_{yy,*}^{k|k-1} + R_k. \quad (48)$$

3) *State correction.*

$$\mathbf{G}_k := \hat{\mathbf{P}}_{xy}^{k|k-1} \left(\hat{\mathbf{P}}_{yy}^{k|k-1}\right)^{-1} \quad (49)$$

$$\hat{\mathbf{x}}_{k|k}^{TM} := \mathbf{G}_k \log_{\hat{\mathbf{y}}_{k|k-1}}(\check{\mathbf{y}}_k) \quad (50)$$

$$\hat{\mathbf{x}}_{k|k} := \exp_{\hat{\mathbf{x}}_{k|k-1}}(\hat{\mathbf{x}}_{k|k}^{TM}) \quad (51)$$

$$\hat{\mathbf{P}}_{xx}^{k|k, \hat{\mathbf{x}}_{k|k-1}} := \hat{\mathbf{P}}_{xx}^{k|k-1} - \mathbf{G}_k \hat{\mathbf{P}}_{yy}^{k|k-1} \mathbf{G}_k^T \quad (52)$$

$$\hat{\mathbf{P}}_{xx}^{k|k} := \text{PT}\left(\hat{\mathbf{P}}_{xx}^{k|k, \hat{\mathbf{x}}_{k|k-1}}, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{x}}_{k|k}\right).$$

All steps of the RiUKFs are justified by and coherent with the other results of this work. Among these, the most important are  $\sigma$ R, RiUT and Riemannian systems.

The notations RiUT<sub>1</sub> and RiUT<sub>2</sub> [in (40), (41), (43), and (46)] indicate these RiUTs can have different forms. The output of RiUT<sub>1</sub> has only two terms—which is different from the number of mapped variables in Definition 4—meaning that only the first two variables of the output of Definition 4 are needed.

We can consider *not regenerating the independent set of RiUT<sub>2</sub>* when RiUT<sub>1</sub> = RiUT<sub>2</sub>. Let  $\chi_*^{k|k-1}$  be the *dependent set of RiUT<sub>1</sub>* and  $\chi_{i,*}^{k|k-1}$  the *dependent set of RiUT<sub>2</sub>*. Because, from (41) and (46),  $\chi_{i,*}^{k|k-1}$  and  $\chi_*^{k|k-1}$  are different objects, we say  $\chi_*^{k|k-1}$  is regenerated. Nonetheless, we could set  $\chi_{i,*}^{k|k-1} = \chi_*^{k|k-1}$ ; consequently, the computational effort of the filter would decrease—calculating a new  $\chi_*^{k|k-1}$  can be computationally because it includes calculating a square-root matrix of  $\hat{\mathbf{P}}_{xx, aug}^{k|k-1}$  or  $\hat{\mathbf{P}}_{xx}^{k|k-1}$ . But in this case, i) the estimation quality of the RiAdUKF would possibly deteriorate—it has been shown for the Euclidean case (cf. Section 5.1 of [38])—and ii) the reasoning behind the RiUKFs explained

in the second paragraph of this section would not be true anymore.

After choosing the manifolds' atlases, *all* expressions for the Riemannian exponentials, logarithms, etc., *must be coherent* with the chosen parameterizations. These transformations, as well as other elements in these filters such as covariances, have different expressions *depending on the parameterizations* defining the manifolds.

We can find Ri $\sigma$ Rs (with  $w_i^m > 0$  for every  $i = 1, \dots, N$ ) by first finding  $\sigma$ Rs in tangent spaces (see the last paragraph of Section V). The independent sets of RiUT<sub>1</sub> and RiUT<sub>2</sub> can be difficult to find. Fortunately, closed forms of Ri $\sigma$ Rs (which can be independent sets of RiUTs) can be found from closed forms of normalized  $\sigma$ Rs by using Theorem 1.

The method for obtaining the sample means of RiUT<sub>1</sub> and RiUT<sub>2</sub> affects the computation efforts of the RiUKFs because, following [7], we define these sample means as optimization problems (Section III-B). Sometimes there exist closed forms, but more often it requires optimization algorithms. The reader will find efficient options in [6], [7], [48], [49] and in the MATLAB and Python toolbox ManOpt [50]<sup>5</sup>.

Computational efforts of the RiUKFs also varies with the underlying manifolds and their atlases because the expressions for exponentials, logarithms and parallel transports change with them. The reader can also refer to the ManOpt toolbox for many efficient implementations of these operations.

Apart from these three factors, computational efforts majorly depends on the square-rooting involved in the Ri $\sigma$ R calculations and the  $\hat{\mathbf{P}}_{yy}^{k|k-1}$  inversion in the Kalman gain calculations. Since we can find Ri $\sigma$ Rs by finding  $\sigma$ Rs in tangent spaces and, to the best of our knowledge, all known  $\sigma$ Rs require square-rooting a covariance matrix (cf. [11]), the computational complexity of these operations in (40) is  $\mathcal{O}([n_x + n_\varpi]^3)$ , in (41)  $\mathcal{O}([n_y + n_\vartheta]^3)$ , in (43) is  $\mathcal{O}(n_x^3)$ , and in (46)  $\mathcal{O}(n_y^3)$ . The computational complexity of the  $\hat{\mathbf{P}}_{yy}^{k|k-1}$  inversion is  $\mathcal{O}(n_y^3)$  in both (42) and (49).

RiUKFs are *generalizations of UKFs*. Every UKF is a RiUKF, and every RiUKF for Euclidean state-variables is a UKF. It is easy to see that, if  $\mathcal{N}_x$  and  $\mathcal{N}_y$  are Euclidean spaces, then RiAuUKF is equivalent to AuUKF (Algorithm 7 of [38]), and RiAdUKF to AdUKF (Algorithm 6 of [38]).

Since Cartesian products of Riemannian manifolds are also Riemannian manifolds (e.g.,  $S^3 \times \mathbb{R}^n$ ) [42], the proposed RiUKF also estimates systems with state variables belonging to *Cartesian products* of Riemannian manifolds.

The Kalman gain  $\mathbf{G}_k$  in (42) and (49) could be defined in a more general way, as done in (34). However, it would imply more computational effort—the dimension of the sigma points and matrices would be higher—at the exchange of no advantage, at least at present; perhaps benefits can be obtained from (34) in future works.

The three assumptions cited at the beginning of Section V impose some limitations on the RiUKFs. Assumption 1 limits the RiUKFs to the case of geodesically-complete Riemannian manifolds: still there are many of these manifolds useful for practical applications, such as unit spheres, special orthogonal

<sup>5</sup>Available for download at <https://www.manopt.org/>.

Table I  
RIUKF VARIANTS FOR SOME RI $\sigma$ RS.

$\sigma$ R <sup>1</sup>	AuUKF <sup>1</sup>	AdUKF <sup>1</sup>
RiMi $\sigma$ R	RiMiAuUKF	RiMiAdUKF
RiRhoMi $\sigma$ R	RiRhoMiAuUKF	RiRhoMiAdUKF
RiMiSy $\sigma$ R	RiMiSyAuUKF	RiMiSyAdUKF
RiHoMiSy $\sigma$ R	RiHoMiSyAuUKF	RiHoMiSyAdUKF

<sup>1</sup>Ad for Additive, Au for Augmented, Ho for Homogeneous, Mi for Minimum, Ri stands for Riemannian,  $\sigma$ R for  $\sigma$ -Representation, Sy for Symmetric, UKF for Unscented Kalman Filter. Rho stand for Rho itself; see also the acronyms list in Appendix F

groups, special Euclidean groups, real projective spaces, special unitary groups, Grassmann manifolds, among others (cf. [6] and Section I-A). Assumption 2 imposes careful choice of  $P_{xx}^0$ ,  $Q_k$ ,  $R_k$  (or  $Q_k$  and  $R_k$  for the RiAdUKF): their values should be consistent with the logarithms in their definitions [or in (19) in the case of the RiAdUKF]; since these covariances are tuning parameters and are often set based on intuition, an user could chose inconsistent (too great) values; this would probably result on either inconsistent sigma points—because the tangent sigma points would be outside the tangent cut locus—or on some divergence in the algorithm, such as non-positive state covariance matrix. Assumption 3 will not, in most cases, impose other limitations if the user model the system equations and parameters consistently.

We can find *particular cases* of RiUKFs by choosing particular forms of Ri $\sigma$ Rs; Table I shows some cases for RiUT<sub>1</sub> = RiUT<sub>2</sub>—the second and third columns contain the filters. Each filter is the resulting variant of using i) the corresponding RiUKF in the *heading row* of its column (RiAuUKF or RiAdUKF), and ii) the corresponding Ri $\sigma$ R written in the first column of its row. For instance, the Riemannian Minimum AuUKF (RiMiAuUKF in the first row and second column), is the result of the RiAuUKF with the RiMi $\sigma$ R (Corollary 1). All filters in Table I are *new*.

An RiUKF for the partially-additive system (20) is given by step 1 of the RiAuUKF with steps 2 and 3 of the RiAdUKF, and for (21) is given by step 1 of the RiAdUKF with steps 2 and 3 of the RiAuUKF.

For (19), (20) and (21) when either  $f_k$  or  $h_k$  are the identity function, we can simplify their filters by skipping sigma points calculations; hence saving computation effort. If, for example,  $f_k(x) = x$ , then the following two equations can replace the state prediction (e.g., the step 1 of the RiAdUKF):

$$\begin{aligned} \hat{x}_{k|k-1} &:= \exp_{\hat{x}_{k-1|k-1}} \bar{\varpi}_k \\ \hat{P}_{xx}^{k|k-1} &:= \hat{P}_{xx}^{k-1|k-1} + Q_k. \end{aligned} \quad \left( \log_{\hat{y}_{k|k-1}} (\gamma_i^{k|k-1}) \right)^T.$$

The case  $h_k(x) = x$  is similar.

#### D. Relation with the literature

To the best of our knowledge, the *UKF for Riemannian manifolds (UKFRM)* of [5] is the only UKF for any geodesically-complete Riemannian manifold in the literature. Consider system (17) and define the following functions—cf. (1) and (2) of [5]—:

$$f_k^*(x_{k-1}) := f_k(x_{k-1}, \varpi_{k-1}), h_k^*(x_k) := h_k(x_k, \vartheta_k). \quad (53)$$

Suppose that i) the initial state  $x_0$  is characterized by  $x_0 \sim (\bar{x}_0, P_{xx}^0)_{\mathcal{N}_x}$ , and ii) the measurements  $\check{y}_1, \check{y}_2, \dots, \check{y}_{k_f}$  are given. Let

$$\text{HoMiSy}\sigma\text{R} : (\bar{X}, P_{XX}) \mapsto \{\chi_i, w_i\}_{i=1}^N$$

be a function mapping the mean  $\bar{X}$  and covariance  $P_{XX}$  of a given random vector  $X$  to a HoMiSy $\sigma$ R (Corollary 3 of [11]). Then the UKFRM of [5] is given by the following algorithm:

**Algorithm 3** (UKFRM of [5]). *Set  $N := 2n_x + 1$  and the initial estimates  $\hat{x}_{0|0} := \bar{x}_0$  and  $\hat{P}_{xx}^{0|0} := P_{xx}^0$ . For  $k = 1, \dots, k_f$ , perform the following steps:*

1) State prediction.

$$\{\chi_{i,k-1|k-1}^{TM}, w_i\}_{i=1}^N := \text{HoMiSy}\sigma\text{R} \left( [0]_{n_x}, \hat{P}_{xx}^{k-1|k-1} \right) \quad (54)$$

$$\chi_i^{k-1|k-1} := \exp_{\hat{x}_{k-1|k-1}} (\chi_{i,k-1|k-1}^{TM}), i = 1, \dots, N \quad (55)$$

$$\chi_{i,*}^{k|k-1} := f_k^*(\chi_i^{k-1|k-1}), i = 1, \dots, N$$

$$\hat{x}_{k|k-1} := \arg \min_{a \in \mathcal{N}_x} \sum_{i=1}^N w_i \text{dist}^2(\chi_{i,*}^{k|k-1}, a)$$

$$\hat{P}_{xx}^{k|k-1} := \sum_{i=1}^N w_i \left( \log_{\hat{x}_{k|k-1}} (\chi_{i,*}^{k|k-1}) \right) (\diamond)^T. \quad (56)$$

2) Measurement prediction.

$$\{\chi_{i,k|k-1}^{TM}, w_i\}_{i=1}^N := f_k \text{HoMiSy}\sigma\text{R} \left( [0]_{n_x}, \hat{P}_{xx}^{k|k-1} \right) \quad (57)$$

$$\chi_i^{k|k-1} := \exp_{\hat{x}_{k|k-1}} (\chi_{i,k|k-1}^{TM}), i = 1, \dots, N \quad (58)$$

$$\gamma_i^{k|k-1} := h_k^*(\chi_i^{k|k-1}), i = 1, \dots, N$$

$$\hat{y}_{k|k-1} := \arg \min_{b \in \mathcal{N}_y} \sum_{i=1}^N w_i \text{dist}^2(\gamma_i^{k|k-1}, b)$$

$$\hat{P}_{yy}^{k|k-1} := \sum_{i=1}^N w_i \left( \log_{\hat{y}_{k|k-1}} (\gamma_i^{k|k-1}) \right) (\diamond)^T \quad (59)$$

$$\hat{P}_{xy}^{k|k-1} := \sum_{i=1}^N w_i \left( \log_{\hat{x}_{k|k-1}} (\chi_i^{k|k-1}) \right)$$

3) State correction.

$$\begin{aligned} G_k &:= \hat{P}_{xy}^{k|k-1} \left( \hat{P}_{yy}^{k|k-1} \right)^{-1} \\ \hat{x}_{k|k}^{TM} &:= \hat{x}_{k|k-1}^{TM} + G_k \log_{\hat{y}_{k|k-1}} (\check{y}_k) \end{aligned} \quad (60)$$

$$\hat{x}_{k|k} := \exp_{\hat{x}_{k|k-1}} (\hat{x}_{k|k}^{TM})$$

$$\hat{P}_{xx}^{k|k, \hat{x}_{k|k-1}} := \hat{P}_{xx}^{k|k-1} - G_k \hat{P}_{yy}^{k|k-1} G_k^T$$

$$\hat{P}_{xx}^{k|k} := \text{PT} \left( \hat{P}_{xx}^{k|k, \hat{x}_{k|k-1}}, \hat{x}_{k|k-1}, \hat{x}_{k|k} \right).$$

Compared with the UKFRM of [5], we can point out the following five improvements of the RiUKFs:

1) The noises are incorporated into the RiUKFs, but in the UKFRM they are not. In the RiAuUKF, the noises are incorporated by realizing the augmented sigma points in the process and measurement functions [equations (40) and (41)]; and in the RiAdUKF, by “adding” (in the tangent space) their means and covariances [equations (44), (45), (47), (48)].

However, the UKFRM exclude the noises. Even though the UKFRM of [5] considers a system with process and measurement noises [cf. (53)], they do *not* influence any estimate within the UKFRM; these noises’ statistics *do not appear at any step* of the UKFRM—commonly, filters consider these statistics when calculating the predicted covariances, but this is also not the case for the UKFRM [cf. (56) and (59)].

We can point out at least two consequences of this absence of the noise elements:

- a) the Euclidean case of the UKFRM is *not equivalent to any* (Euclidean) UKF. This can be seen by considering Euclidean manifolds in Algorithm 3 (cf. the last paragraph of Sections III-A and of III-B). Besides, to the best of our knowledge, there is no UKF without process and measurement noises covariance (cf. [11], [38]).
- b) the UKFRM might *diverge* in situations in which the RiUKFs do not. This behavior can be seen in the following simple example: consider (19) and (53) with  $\mathcal{N}_{\mathbf{x}^x}^{n_x} = \mathcal{N}_{\mathbf{x}^y}^{n_y} = \mathcal{N}_{\mathbf{x}^z}^{n_z} = \mathcal{N}_{\mathbf{x}^w}^{n_w} = \mathbb{R}$ . Suppose that i) the initial state is  $x_0 \sim (1, 1)_{\mathbb{R}}$ , ii) the noise covariances are  $Q_k = R_k = 1$ , iii) the system functions are  $f_k(x_{k-1}) = f_k^*(x_{k-1}) = x_{k-1}$  and  $h_k(x_k) = h_k^*(x_k) = 1 - x_k$ , and iv) the measurements are  $\check{y}_1 = \dots = \check{y}_{k_f} = 1$ . For this example, we ran the (linear) KF (cf. [51]), the RiAdUKF, and the UKFRM. Both the KF and the RiAdUKF provided the same estimates, but the UKFRM did not provide consistent results; the simulation was halted because the corrected covariance ( $\hat{P}_{\mathbf{x}\mathbf{x}}^{2|2}$ ) lost its positiveness. Similar results occurred in the simulations of Section VIII.
- 2) We introduced a consistent definition [equation (19)] for the system associated with the RiAdUKF. To the best of our knowledge, (19) is the first consistent additive-noise Riemannian stochastic discrete-time dynamic system.
- 3) To the best of our knowledge, the RiUKFs are the *first* UKFs for Riemannian state-space systems considering noises with *non-zero means*. Even for simple manifolds such as the unit sphere, we could not find a UKF considering this case.
- 4) All the equations of our RiUKFs are *formally justified*. These justifications are the following ones:
  - a) *The equations of steps 1 and 2 of the RiUKFs are justified by Definition 3, Theorem 1 and Corollary VI.*
  - b) *Equations (42) and (49) (the Kalman Gains) are justified in Section VII-B2.* This form of the Kalman gain  $G_k$  in (42) and (49) follows as a particular case of the Kalman gain of a more general system ( $G_{k,**}$ ) where the state and the measurement belong to the product  $\mathcal{N}_{\mathbf{x}} \times \mathcal{N}_{\mathbf{y}}$ .
  - c) *The equations of step 3 of the RiUKFs are justified in*

*Section VII-B.* We showed that they follow from considering i)  $x_{k|k-1}^{TM}$  and  $y_{k|k-1}^{TM}$  normally-joint distributed [equation (25)], and ii)  $x_{k|k}^{TM}$  given by a linear correction of  $x_{k|k-1}^{TM}$  by  $(\check{y}_k^{TM} - y_{k|k-1}^{TM})$  [equation (26)].

5) (Euclidean) UKFs are particular cases of the RiUKFs (cf. Section VII-C).

Altogether, we can say the RiUKFs have novelties compared with the UKF for Riemannian state-space systems of the literature.

## VIII. EXAMPLE: SATELLITE ATTITUDE TRACKING

In this section, we apply the developed theory to estimate the attitude of a satellite in a realistic scenario (cf. [52]).

The set of possible attitudes of a rotating body is not a Euclidean space, but a three dimensional smooth manifold known as  $SO(3)$ . This manifold has many different topological properties from a Euclidean space: for instance, it is compact whilst Euclidean spaces are not. Due to this difference, Euclidean UKFs designed over Euclidean spaces may not work properly: its estimates may not stay within the state-space manifold, resulting in poor performance and poor accuracy [12].

Although we could apply an RiUKFs for  $SO(3)$  in this example, we prefer to apply an RiUKF for the set of unit quaternions  $S^3$  because they represent, without singularities [53], attitudes using the minimal set of parameters. Let  $\mathbf{q}_i = [\eta_i \ \epsilon_i^T]^T \in \mathbb{R}^4$ , where  $\eta_i \in \mathbb{R}$  and  $\epsilon_i \in \mathbb{R}^3$ . It is possible to prove that the three dimensional sphere

$$S^3 = \{(q_1, q_2, q_3, q_4) \in \mathbb{R}^4 : q_1^2 + q_2^2 + q_3^2 + q_4^2 = 1\} \quad (61)$$

is a Riemannian manifold and the product

$$\mathbf{q}_1 \otimes \mathbf{q}_2 = \begin{bmatrix} \eta_1 \eta_2 - \epsilon_1^T \epsilon_2 \\ \eta_1 \epsilon_2 + \eta_2 \epsilon_1 + \epsilon_1 \times \epsilon_2 \end{bmatrix}.$$

is closed. For a rotation of an angle  $\theta$  around an unit vector  $\mathbf{n}$ , there are two associated unit quaternions  $\mathbf{q}$  and  $\mathbf{q}'$  such that

$$\mathbf{q} = \cos\left(\frac{\theta}{2}\right) + \mathbf{v}_m \mathbf{n} \sin\left(\frac{\theta}{2}\right), \quad \mathbf{q}' = -\mathbf{q}.$$

Let  $\mathbf{q}(t) \in S^3$  be the attitude of the satellite at the time instant  $t$ , and  $\boldsymbol{\omega}(t) \in \mathbb{R}^3$  its the angular velocity. The evolution of  $\mathbf{q}(t)$  over time can be described by the following differential equation [54]:

$$\dot{\mathbf{q}}(t) = \frac{1}{2} \boldsymbol{\omega}(t) \otimes \mathbf{q}(t), \quad (62)$$

where  $\boldsymbol{\omega} \in \mathbb{R}^4$  is given by  $\boldsymbol{\omega} = [0 \ \boldsymbol{\omega}^T]^T$ .

We generate synthetic data by a fourth order Runge-Kutta integration of (62) over the interval [0s, 20s] with angular velocity

$$\boldsymbol{\omega}(t) = \begin{bmatrix} 0.03 \sin([\pi t/600]^\circ) \\ 0.03 \sin([\pi t/600]^\circ - 300^\circ) \\ 0.03 \sin([\pi t/600]^\circ - 600^\circ) \end{bmatrix}$$

and initial state  $\mathbf{q}(0) = 0.96 + \mathbf{v}_m [0.13, 0.19, \sqrt{1 - 0.96^2 - 0.13^2 - 0.19^2}]^T$ .

Table II  
ROOT MEAN SQUARE ERROR ( $\times 10^{-6}$ ) OF EACH RIUKF IN TABLE I  
CONSIDERING 1,000 SIMULATIONS OF A SATELLITE ATTITUDE TRACKING  
EXAMPLE.

RiMiAuUKF	RiRhoMiAuUKF	RiMiSyAuUKF	RiHoMiSyAuUKF
2,612	2,614	2,614	2,614
RiMiAdUKF	RiRhoMiAdUKF	RiMiSyAdUKF	RiHoMiSyAdUKF
2,612	2,613	2,613	2,613

For filtering, we consider (19) with  $\mathbf{x}_k = \mathbf{q}(k\delta t)$

$$\theta(t) := \|\omega(t)\| \frac{\delta t}{2}$$

$$f_k(\mathbf{x}_{k-1}) = \begin{bmatrix} \cos \theta(t) & \frac{\omega^T(t)}{\|\omega(t)\|} \sin \theta(t) \end{bmatrix}^T \otimes \mathbf{x}_{k-1}$$

$$h_k(\mathbf{x}_k) = \mathbf{x}_k,$$

$\bar{\omega}_k = \bar{\vartheta}_k = [0]_{3 \times 1}$ ,  $\mathbf{Q}_k = (0.31236 \times 10^{-6})^2 I_3$ , and  $\mathbf{R}_k = (0.5\pi/180 \times 10^{-6})^2 I_3$ . These values for  $\mathbf{Q}_k$  and  $\mathbf{R}_k$  were chosen according to [12].

We performed 1,000 simulations with the RiUKFs of Table I and the UKFRM of [5]. To calculate Riemannian means, we used the gradient descent method of [49] with a threshold of  $10^{-6}$ ; and for Riemannian exponentials, Riemannian logarithms, and parallel transport, we used the MATLAB toolbox ManOpt [50].

For all simulations, the RiUKFs of Table I provided good estimates, with a Root Mean Square Error in the order of  $10^{-6}$  (Table II). The RiMiAdUKF or the RiRhoMiAdUKF are the best alternatives for this example because i) it demands less computational effort than the other filters—it is additive and is composed of the least number of sigma points (cf. Corollary 1)—and ii) all RiUKFs performed almost equally.

The UKFRM failed in all the 1,000 simulations; in every simulation, the state covariance estimate lost its positiveness. Nonexistence of noise terms in the UKFRM might explain this problematic behavior (cf. Section VII-D).

## IX. CONCLUSIONS

In this work, we extend the systematization of the Unscented Kalman Filtering theory we developed in [11] towards estimating the state of Riemannian systems. In this systematization, we introduce the following results<sup>6</sup> (all results are mathematically justified):

- 1) A Riemannian extension of the  $\sigma$ -representation ( $\sigma\mathbf{R}$ ): the Riemannian  $\sigma$ -representation (Ri $\sigma\mathbf{R}$ , Section V).
- 2) A technique to obtain closed forms of the Ri $\sigma\mathbf{R}$  by closed forms of the  $\sigma\mathbf{R}$  (Theorem 1). Using this result, we discover (Corollary 1)
  - a) the minimum number of sigma points of an Ri $\sigma\mathbf{R}$ ,
  - b) the minimum number of sigma points of a symmetric Ri $\sigma\mathbf{R}$ ,
  - c) closed forms for the minimum Ri $\sigma\mathbf{R}$ , and
  - d) closed forms for the minimum symmetric Ri $\sigma\mathbf{R}$ .
- 3) An additive-noise Riemannian system definition (Section VII-A). We require this definition to introduce additive-noise Riemannian UKFs.

<sup>6</sup>These results were first presented in Menegaz's PhD thesis [38].

- 4) Kalman correction equations on Riemannian manifolds (Section VII-B).
- 5) New discrete-time Riemannian UKFs (RiUKFs), namely the Riemannian Additive UKF and the Riemannian Augmented UKF (Section VII-C). Besides, we
  - a) provide a list of particular variants of these filters (Table I); all these variants are new. Compared with the literature's UKF for Riemannian manifolds (in [5]), our RiUKFs are more consistent, formally-principled, and general.
  - b) numerically compare all these particular variants with the literature's UKF on Riemannian manifolds in a satellite attitude tracking scenario. For all 1,000 simulations, the new variants provided good estimates, but the literature's filter diverged; in every simulation, the state covariance estimate lost its positiveness.

With this work, we hope to have expanded the literature's knowledge on Kalman filtering and provided a tool for the research community to improve the performance and stability of many UKFs.

Following this study, we recommend the research community searching for computationally-implementable variants of RiUKFs. Since concepts of the Riemannian manifold theory can be very abstract, depending on the underlying manifold, developing RiUKFs variants is not trivial.

This task is even harder without a generalizing base theory: that is one of the reasons why, in this work, we develop a general consistent systematized theory of Unscented Kalman Filters for Riemannian State-Space Systems.

## REFERENCES

- [1] J. Meng, G. Luo, and F. Gao, "Lithium Polymer Battery State-of-Charge Estimation Based on Adaptive Unscented Kalman Filter and Support Vector Machine," *IEEE Transactions on Power Electronics*, vol. 31, no. 3, pp. 2226–2238, 2016.
- [2] A. Rahimi, K. D. Kumar, and H. Alighanbari, "Enhanced adaptive unscented Kalman filter for reaction wheels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 2, pp. 1568–1575, 2015.
- [3] A. N. Vargas, H. M. T. Menegaz, J. Y. Ishihara, and L. Acho, "Unscented Kalman Filters for Estimating the Position of an Automotive Electronic Throttle Valve," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4627–4632, jun 2016.
- [4] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [5] S. Hauberg, F. Lauze, and K. S. Pedersen, "Unscented Kalman Filtering on Riemannian Manifolds," *J. Math. Imaging Vis.*, vol. 46, no. 1, pp. 103–120, jan 2013.
- [6] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.
- [7] X. Pennec, "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," *J. Math. Imaging Vision*, vol. 25, no. 1, pp. 127–154, jul 2006.
- [8] N. Enayati, E. D. Momi, and G. Ferrigno, "A Quaternion-Based Unscented Kalman Filter for Robust Optical/Inertial Motion Tracking in Computer-Assisted Surgery," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 8, pp. 2291–2301, jan 2015.
- [9] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Trans. Autom. Control*, vol. 61, no. 1, pp. 172–177, jan 2016.
- [10] T. Lee, "Global unscented attitude estimation via the matrix Fisher distributions on SO(3)," in *2016 Americ. Control Conf.*, Boston, MA, jul 2016, pp. 4942–4947.
- [11] H. M. T. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, "A Systematization of the Unscented Kalman Filter Theory," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2583–2598, oct 2015.

- [12] J. L. Crassidis and F. L. Markley, "Unscented Filtering for Spacecraft Attitude Estimation," *J. Guid. Control Dynam.*, vol. 26, no. 4, pp. 536–542, 2003.
- [13] M. S. Challa, J. G. Moore, and D. J. Rogers, "A Simple Attitude Unscented Kalman Filter: Theory and Evaluation in a Magnetometer-Only Spacecraft Scenario," *IEEE Access*, vol. 4, pp. 1845–1858, apr 2016.
- [14] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems*, ser. Texts in Applied Mathematics. Springer Verlag, 2004.
- [15] B. V. Adorno, "Two-arm Manipulation: From Manipulators to Enhanced Human-Robot Collaboration," Ph.D. dissertation, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) Université Montpellier 2, 2011.
- [16] J. Selig, *Geometric Fundamentals of Robotics*, 2nd ed., ser. Monographs in Computer Science. Springer, 2005.
- [17] A. Barrau and S. Bonnabel, "The Invariant Extended Kalman Filter as a Stable Observer," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1797–1812, apr 2017.
- [18] S. Bonnabel, P. Martin, and P. Rouchon, "Symmetry-Preserving Observers," *IEEE Trans. Autom. Control*, vol. 53, no. 11, pp. 2514–2526, dec 2008.
- [19] H. D. Curtis, *Orbital Mechanics for Engineering Students*, 3rd ed. Oxford: Butterworth-Heinemann, 2014.
- [20] B. Wie, *Space Vehicle Dynamics and Control*, 2nd ed., ser. AIAA Education Series. Reston, VA: AIAA, jan 2008.
- [21] X. Pennec, "Statistical Computing on Manifolds for Computational Anatomy," L'habilitation à diriger des recherches, Université de Nice Sophia-Antipolis, dec 2006.
- [22] R.-A. Pitaval, L. Wei, O. Tirkkonen, and C. Hollanti, "Density of Spherically Embedded Stiefel and Grassmann Codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 1–1, jan 2017.
- [23] M. Harandi, M. Salzmann, and R. Hartley, "Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 48–62, jan 2018.
- [24] F. Hajati, M. Tavakolian, S. Gheisari, Y. Gao, and A. S. Mian, "Dynamic Texture Comparison Using Derivative Sparse Representation: Application to Video-Based Face Recognition," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 6, pp. 970–982, dec 2017.
- [25] K. G. Seddik, R. H. Gohary, M. T. Hussien, M. Shaqfeh, H. Alnuweiri, and H. Yanikomeroglu, "Multi-Resolution Multicasting Over the Grassmann and Stiefel Manifolds," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5296–5310, aug 2017.
- [26] S. Chepushtanova and M. Kirby, "Sparse Grassmannian Embeddings for Hyperspectral Data Representation and Classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 434–438, mar 2017.
- [27] C. Altafini and F. Ticozzi, "Modeling and control of quantum systems: an introduction," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 1898–1917, 2012.
- [28] L. Godinho and J. Natário, *An Introduction to Riemannian Geometry: With Applications to Mechanics and Relativity*, ser. Universitext. Springer International Publishing, 2014.
- [29] J.-P. Condomines, C. Seren, and G. Hattenberger, "Nonlinear state estimation using an invariant unscented Kalman filter," in *AIAA Guid. Navig. Control Conf.* Boston, MA: AIAA, aug 2013, pp. 1–15.
- [30] —, "Pi-Invariant Unscented Kalman Filter for Sensor Fusion," in *53rd IEEE Conf. Decis. Control*. Los Angeles, CA: IEEE, dec 2014, pp. 1035–1040.
- [31] S.-G. Kim, J. L. Crassidis, Y. Cheng, A. M. Fosbury, and J. L. Junkins, "Kalman Filtering for Relative Spacecraft Attitude and Position Estimation," *J. Guid. Control Dynam.*, vol. 30, no. 1, pp. 133–143, 2007.
- [32] A. H. J. de Ruiter, "SO(3)-constrained Kalman filtering with application to attitude estimation," in *2014 Americ. Control Conf.*, Portland, OR, jun 2014, pp. 4937–4942.
- [33] I. Marković, J. Česić, and I. Petrović, "On wrapping the Kalman filter and estimating with the SO(2) group," in *2016 19th Int. Conf. Inform. Fusion*, Heidelberg, BW, jul 2016, pp. 2245 – 2250.
- [34] M. Barczyk and A. F. Lynch, "Invariant Extended Kalman Filter Design for a Magnetometer-plus-GPS Aided Inertial Navigation System," in *50th IEEE Conf. Decis. Control and Europ. Control Conf.*, Orlando, FL, dec 2011, pp. 5389–5394.
- [35] S. Bonnabel, P. Martin, and E. Salaün, "Invariant Extended Kalman Filter: theory and application to a velocity-aided attitude estimation problem," in *Proc. 48th IEEE Conf. Decis. Control and 28th Chin. Control Conf.*, Shanghai, China, dec 2009, pp. 1297–1304.
- [36] S. Bonnabel, "Left-invariant extended Kalman filter and attitude estimation," in *46th IEEE Conf. on Decis. Control*, New Orleans, LA, dec 2007, pp. 1027–1032.
- [37] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, UK: Cambridge University Press, 2013.
- [38] H. M. T. Menegaz, "Unscented Kalman Filtering on Euclidean and Riemannian Manifolds," Ph. D., Universidade de Brasília, 2016. [Online]. Available: <http://repositorio.unb.br/handle/10482/21617>
- [39] P. Vartiainen, T. Bragge, J. P. Arokoski, and P. a. Karjalainen, "Nonlinear State-Space Modeling of Human Motion Using 2-D Marker Observations," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2167–2178, jul 2014.
- [40] B. O. Teixeira, J. Chandrasekar, L. A. Tôres, L. A. Aguirre, and D. S. Bernstein, "State estimation for linear and non-linear equality-constrained systems," *Int. J. Control*, vol. 82, no. 5, pp. 918–936, may 2009.
- [41] X. Pennec, "L'Incertain dans les problemes de reconnaissance et de recalage: application en imagerie medicale et biologie moleculaire," PhD Thesis, Ecole Polytechnique, Palaiseau, dec 1996.
- [42] M. P. Do Carmo, *Riemannian Geometry*, ser. Mathematics: Theory & Applications. Woodbine, NJ: Birkhäuser Boston, 1992.
- [43] A. Figalli, T. O. Gallouët, and L. Rifford, "On the convexity of injectivity domains on nonfocal manifolds," *SIAM Journal on Mathematical Analysis*, vol. 47, no. 2, pp. 969–1000, 2015.
- [44] S. Särkkä, "On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1631–1641, 2007.
- [45] L. Chang, F. Qin, and F. Zha, "Pseudo Open-Loop Unscented Quaternion Estimator for Attitude Estimation," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4460–4469, jun 2016.
- [46] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, ser. Information and System Science Series, T. Kailath, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1979, no. 2.
- [47] M. Lorenzi and X. Pennec, "Efficient Parallel Transport of Deformations in Time Series of Images: From Schild's to Pole Ladder," *J. Math. Imaging Vision*, vol. 50, no. 1-2, pp. 5–17, sep 2014.
- [48] M. Moakher, "Means and Averaging in the Group of Rotations," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 1–16, 2002.
- [49] X. Pennec, "Computing the Mean of Geometric Features Application to the Mean Rotation," INRIA, Tech. Rep., mar 1998. [Online]. Available: <https://hal.inria.fr/inria-00073318>
- [50] N. Boumal, B. Mishra, P. -A. Absil, and R. Sepulchre, "Manopt, a Matlab Toolbox for Optimization on Manifolds," *J. Mach. Learn. Res.*, vol. 15, pp. 1455–1459, 2014.
- [51] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [52] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of Nonlinear Attitude Estimation Methods," *J. Guid. Control Dynam.*, vol. 30, no. 1, pp. 12–28, jan 2007.
- [53] J. Stuelpnagel, "On the Parametrization of the Three-Dimensional Rotation Group John Stuelpnagel," *SIAM Review*, vol. 6, no. 4, pp. 422–430, 1964.
- [54] P. H. Zipfel, *Modeling and Simulation of Aerospace Vehicle Dynamics*, 2nd ed., ser. AIAA Education Series. Reston, VA: AIAA, jan 2007.
- [55] R. S. Palais and C.-L. Terng, *Critical point theory and submanifold geometry*. Springer, 2006, vol. 1353.
- [56] F. Nielsen and R. Bhatia, *Matrix information geometry*. Springer, 2013.

## APPENDIX

### A. Results relative to Riemannian manifolds

In this appendix, we provide some results relative to the theory of Riemannian manifolds. These definitions are mainly based on [42].

**Definition 5** (Differentiable manifold [42]). A *differentiable manifold* of dimension  $n$  is a pair  $(\mathcal{N}, \mathcal{A})$  where  $\mathcal{N}$  is a set, and  $\mathcal{A} = \{(U_a, \varphi_a)\}$ , called atlas, a family of injective mappings (*charts*)  $\varphi_a : U_a \subset \mathbb{R}^n \rightarrow \mathcal{N}$  of open sets  $U_a$  of  $\mathbb{R}^n$  into  $\mathcal{N}$  such that:

- 1)  $\bigcup_a \varphi_a(U_a) = \mathcal{N}$ .
- 2) for any pair  $a, b$ , with  $\varphi_a(U_a) \cap \varphi_b(U_b) =: W \neq \emptyset$ , the sets  $\varphi_a^{-1}(W)$  and  $\varphi_b^{-1}(W)$  are open sets in  $\mathbb{R}^n$ , and the mappings  $\varphi_b^{-1} \circ \varphi_a$  and  $\varphi_a^{-1} \circ \varphi_b$  are differentiable.

3) The family  $\mathcal{A} = \{(U_a, \varphi_a)\}$  is *maximal* relative to the conditions 1) and 2).

A pair  $(U_a, \varphi_a)$  (or the mapping  $\varphi_a$ ) with  $\mathbf{a} \in \varphi_a(U_a)$  is called a *parameterization* of  $\mathcal{N}$  at  $\mathbf{a}$ . For simplicity, we can denote a differentiable manifold  $(\mathcal{N}, \mathcal{A})$  of dimension  $n$  by  $\mathcal{N}$  or  $\mathcal{N}^n$ .

**Definition 6** (Differentiable function [42]). Let  $\mathcal{N}_1^n$  and  $\mathcal{N}_2^m$  be differentiable manifolds. A mapping  $f : \mathcal{N}_1 \rightarrow \mathcal{N}_2$  is *differentiable at*  $\mathbf{a} \in \mathcal{N}_1$  if, given a parameterization  $\varphi_2 : V \subset \mathbb{R}^m \rightarrow \mathcal{N}_2$  at  $f(\mathbf{a})$ , there exists a parameterization  $\varphi_1 : U \subset \mathbb{R}^n \rightarrow \mathcal{N}_1$  at  $\mathbf{a}$  such that  $f(\varphi_1(U)) \subset \varphi_2(V)$  and the mapping

$$\tilde{f} := \varphi_2^{-1} \circ f \circ \varphi_1 : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (63)$$

is differentiable at  $\varphi_1^{-1}(\mathbf{a})$ . We say  $f$  is differentiable on an open set of  $\mathcal{N}_1$  if it is differentiable at all of the points of this open set.

In this work, we suppose that all functions are differentiable unless otherwise stated.

**Definition 7** (Tangent space [42]). Let  $\mathcal{N}$  be a differentiable manifold. A differentiable function  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is called a (differentiable) *curve* in  $\mathcal{N}$ . Suppose  $\alpha(0) = \mathbf{a} \in \mathcal{N}$ , and let  $\mathcal{D}_a(\mathcal{N})$  be the set of all functions  $f : \mathcal{N} \rightarrow \mathbb{R}$  that are differentiable at  $\mathbf{a}$ . The *tangent vector to the curve*  $\alpha$  at  $t = 0$  is a function  $\alpha'(0) : \mathcal{D}_a(\mathcal{N}) \rightarrow \mathbb{R}$  given by

$$\alpha'(0)f = d(f \circ \alpha)/dt|_{t=0}, \quad f \in \mathcal{D}_a(\mathcal{N}).$$

Note that  $\alpha'(0)$  is an operator taking  $f \in \mathcal{D}_a(\mathcal{N})$  to a scalar  $d(f \circ \alpha)/dt|_{t=0}$ . A *tangent vector at*  $\mathbf{a}$  is a tangent vector of some curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  with  $\alpha(0) = \mathbf{a}$  at  $t = 0$ . The set of all tangent vectors to  $\mathcal{N}$  at  $\mathbf{a}$  will be indicated by  $T_a\mathcal{N}$ .

The set  $T_a\mathcal{N}$  forms a vector space of dimension  $n$  and is called the *tangent space of*  $\mathcal{N}$  at  $\mathbf{a}$ .

**Definition 8** (Arc length [7]). Given an open interval  $\mathbf{I} \subset \mathbb{R}$ , a differentiable function (Definition 6)  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is called a (differentiable) *curve* in  $\mathcal{N}$ . Given a curve  $\alpha$  on  $\mathcal{N}$ , the arc length of  $\alpha$  in the interval  $[a, b] \subset \mathbf{I}$  is defined by

$$\mathbb{L}_a^b(\alpha) := \int_a^b \|\alpha'(t)\|_{\alpha(t)} dt.$$

**Definition 9** (Differential of a function). Let  $\mathcal{N}_1$  and  $\mathcal{N}_2$  be differentiable manifolds and  $f : \mathcal{N}_1 \rightarrow \mathcal{N}_2$  a differentiable mapping. For every  $\mathbf{a} \in \mathcal{N}_1$  and for each  $v \in T_a\mathcal{N}_1$ , choose a differentiable curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}_1$  with  $\alpha(0) = \mathbf{a}$ ,  $\alpha'(0) = v$ . Take  $\beta = f \circ \alpha$ . Then it can be shown that the operator  $df_a(v)$  defined by

$$df_a(v) := \beta'(0)$$

is a tangent vector of  $T_{f(\mathbf{a})}\mathcal{N}_2$ . Moreover the mapping the

$$df_a : T_a\mathcal{N}_1 \rightarrow T_{f(\mathbf{a})}\mathcal{N}_2 : v \mapsto \beta'(0)$$

is linear and does not depend on the choice of  $\alpha$  [42]. This linear mapping  $df_a$  is called the *differential* of  $f$  at  $\mathbf{a}$ .

**Definition 10** (Vector field [6], [42]). A *vector field*  $\mathcal{X}$  on a differentiable manifold  $\mathcal{N}$  is a correspondence that associates

to each point  $\mathbf{a} \in \mathcal{N}$  a vector  $\mathcal{X}(\mathbf{a}) \in T_a\mathcal{N}$ . Given a vector field  $\mathcal{X}$  on  $\mathcal{N}$  and a differentiable real-valued function  $f : \mathcal{N} \rightarrow \mathbb{R}$ , we let  $\mathcal{X}f$  denote the real-valued function on  $\mathcal{N}$  defined by

$$\begin{aligned} (\mathcal{X}f) : \mathcal{N} &\rightarrow \mathbb{R} \\ \mathbf{a} &\mapsto vf, \quad v \in T_a\mathcal{N}. \end{aligned}$$

The set of all vector fields of  $\mathcal{N}$  is denote by  $\mathcal{X}(\mathcal{N})$ .

The multiplication of a vector field  $\mathcal{X}$  by a function  $f : \mathcal{N} \rightarrow \mathbb{R}$  is defined by  $f\mathcal{X} : \mathcal{N} \rightarrow T_a\mathcal{N} : \mathbf{a} \mapsto f(\mathbf{a})v$ ,  $v \in T_a\mathcal{N}$ ; and the addition of two vector fields  $\mathcal{X}$  and  $\mathcal{Y}$  by  $\mathcal{X} + \mathcal{Y} : \mathcal{N} \rightarrow T_a\mathcal{N} : \mathbf{a} \mapsto \mathcal{X}(\mathbf{a}) + \mathcal{Y}(\mathbf{a})$ . The *Lie bracket of vector fields* is defined as the unique vector field  $[\mathcal{X}, \mathcal{Y}]$  satisfying  $([\mathcal{X}, \mathcal{Y}]f) := (\mathcal{X}(\mathcal{Y}f)) - (\mathcal{Y}(\mathcal{X}f))$  for all real valued smooth functions  $f$  defined on  $\mathcal{N}$ . A *vector field*  $\mathcal{X}$  along a curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is a differentiable mapping that associates to every  $t \in \mathbf{I}$  a tangent vector  $\mathcal{X}(t) \in T_{\alpha(t)}\mathcal{N}$ .

**Definition 11** (Riemannian manifold). A *Riemannian metric*  $\langle \cdot, \cdot \rangle$  or  $g$  on a differentiable manifold  $\mathcal{N}$  is a correspondence which associates to each point  $\mathbf{a}$  of  $\mathcal{N}$  an inner product  $g_a := \langle \cdot, \cdot \rangle_a$  on a tangent space  $T_a\mathcal{N}$ , with  $\langle \cdot, \cdot \rangle_a$  varying differentially in the following sense: if  $\varphi : U \subset \mathbb{R}^n \rightarrow \mathcal{N}$  is a system of coordinates (or chart) around  $\mathbf{a}$ , with  $\varphi(u_1, u_2, \dots, u_n) = \mathbf{a} \in \varphi(U)$  and  $\partial/\partial u_i(\mathbf{a}) = d\varphi_a(0, \dots, 0, 1, 0, \dots, 0)$ , then

$$g_{i,j}(u_1, u_2, \dots, u_n) = \left\langle \frac{\partial}{\partial u_i}(\mathbf{a}), \frac{\partial}{\partial u_j}(\mathbf{a}) \right\rangle_a$$

is a differentiable function on  $U$  [42]. We delete the index  $\mathbf{a}$  in the functions  $g_a$  and  $\langle \cdot, \cdot \rangle_a$  whenever there is no possibility of confusion.

The pair  $(\mathcal{N}, g)$  is called a *Riemannian manifold* [6]. For simplicity, we can also denote the Riemannian manifold  $(\mathcal{N}, g)$  by the set  $\mathcal{N}$ .

**Definition 12** (Riemannian gradient [42]). Let  $\mathcal{N}$  be a Riemannian manifold. Given a smooth function  $f : \mathcal{N} \rightarrow \mathbb{R}$ , the *Riemannian gradient* of  $f$  at  $\mathbf{x}$ , denoted by  $\text{grad}f(\mathbf{x})$  is defined as the unique element of  $T_x\mathcal{N}$  that satisfies

$$\langle \text{grad}f(\mathbf{x}), v \rangle_x = df_x(v), \quad \forall v \in T_x\mathcal{N}.$$

**Definition 13** (Critical point [55]). Let  $\mathcal{N}$  and  $\mathcal{R}$  be smooth manifolds. If  $f : \mathcal{N} \rightarrow \mathcal{R}$  is a smooth map, then a point  $\mathbf{x} \in \mathcal{N}$  is a *critical point* of  $f$  if  $df_x : T_x\mathcal{N} \rightarrow T_{f(\mathbf{x})}\mathcal{R}$  is not surjective. In the particular case that  $\mathcal{R} = \mathbb{R}$ , then the critical points of  $f$  are exactly the points  $\mathbf{x}$  which  $df_x = 0$ . Moreover, if  $\mathcal{N}$  is a Riemannian manifold, the critical points are the points  $\mathbf{x} \in \mathcal{N}$  such that  $\text{grad}f(\mathbf{x}) = 0$ .

**Definition 14** (Affine connection [42]). An *affine connection*  $\nabla$  on a differentiable manifold  $\mathcal{N}$  is a mapping  $\nabla : \mathcal{X}(\mathcal{N}) \times \mathcal{X}(\mathcal{N}) \rightarrow \mathcal{X}(\mathcal{N})$  which is denoted by  $(\mathcal{X}, \mathcal{Y}) \mapsto \nabla_{\mathcal{X}}\mathcal{Y}$  and which satisfies the following properties, for  $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{X}(\mathcal{N})$  and  $f, g \in \mathcal{D}\mathcal{N}$ :

- 1)  $\nabla_{f\mathcal{X}+g\mathcal{Y}}\mathcal{Z} = f\nabla_{\mathcal{X}}\mathcal{Z} + g\nabla_{\mathcal{Y}}\mathcal{Z}$ ,
- 2)  $\nabla_{\mathcal{X}}(\mathcal{Y} + \mathcal{Z}) = \nabla_{\mathcal{X}}\mathcal{Y} + \nabla_{\mathcal{X}}\mathcal{Z}$ ,
- 3)  $\nabla_{\mathcal{X}}(f\mathcal{Y}) = f\nabla_{\mathcal{X}}\mathcal{Y} + (\mathcal{X}f)\mathcal{Y}$ .

If  $\nabla$  satisfies the following additional properties:

- 1)  $\mathcal{X}\langle\mathcal{Y}, \mathcal{Z}\rangle = \langle\nabla_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}\rangle + \langle\mathcal{Y}, \nabla_{\mathcal{X}}\mathcal{Z}\rangle$ , for all  $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{X}(\mathcal{N})$ ,
- 2)  $\nabla_{\mathcal{X}}\mathcal{Y} - \nabla_{\mathcal{Y}}\mathcal{X} = [\mathcal{X}, \mathcal{Y}]$ , for all  $\mathcal{X}, \mathcal{Y} \in \mathcal{X}(\mathcal{N})$ ,

then  $\nabla$  is known as the *Riemannian connection* of  $\mathcal{N}$ . The Levi-Cevita theorem [42] says that any Riemannian manifold has a Riemannian connection and it is unique.

**Theorem 4** (Covariant derivative [42]). *Let  $\mathcal{N}$  be a differentiable manifold with an affine connection  $\nabla$ . There exists a unique correspondence which associates to a vector field  $\mathcal{X}$  along the differentiable curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  another vector field  $D\mathcal{X}/dt$  along  $\alpha$ , called the covariant derivative of  $\mathcal{N}$  along  $\alpha$ , such that:*

- 1)  $\frac{D}{dt}(\mathcal{X} + \mathcal{Y}) = \frac{D\mathcal{X}}{dt} + \frac{D\mathcal{Y}}{dt}$ ;
- 2)  $\frac{D}{dt}(\mathcal{X}V) = \frac{df}{dt}\mathcal{X} + f\frac{D\mathcal{X}}{dt}$ , where  $f$  is a differentiable function on  $\mathbf{I}$ ;
- 3) if  $\mathcal{X}$  is induced by a vector field  $\mathcal{Z} \in \mathcal{X}(\mathcal{N})$ , i.e.,  $\mathcal{X}(t) = \mathcal{Z}(\alpha(t))$ , then  $D\mathcal{X}/dt = \nabla_{\alpha'(t)}\mathcal{Z}$ .

**Definition 15** (Parallel Transport [42]). Let  $\mathcal{N}$  be a differentiable manifold with an affine connection  $\nabla$ . A vector field  $\mathcal{X}$  along a curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is called *parallel* when

$$\frac{D\mathcal{X}}{dt}(t) = 0, \text{ for all } t \in \mathbf{I}.$$

Moreover, let  $\alpha$  be differentiable and  $v_0$  a vector tangent to  $\mathcal{N}$  at  $\alpha(t_0)$ ,  $t_0 \in \mathbf{I}$ . Then there exists a unique parallel vector field  $\mathcal{X}$  along  $\alpha$ , such that  $\mathcal{X}(t_0) = v_0$ ;  $V(t)$  is called the *parallel transport* of  $\mathcal{X}(t_0)$  along  $\alpha$ .

**Definition 16** (Geodesic [42]). A parameterized curve  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  is a *geodesic* at  $t_0 \in \mathbf{I}$  if

$$\frac{D}{dt}(\alpha'(t)) = 0$$

at the point  $t_0$ ; if  $\alpha$  is a geodesic at  $t$ , for all  $t \in \mathbf{I}$ , we say that  $\alpha$  is a *geodesic* [42]. If the definition domain of all geodesics of  $\mathcal{N}$  can be extended to  $\mathbb{R}$ , then  $\mathcal{N}$  is said to be *geodesically-complete*.

**Definition 17** (Exponential and logarithm mappings [7]). Consider a point  $\mathbf{a} \in \mathcal{N}$  and let  $V \subset T_{\mathbf{a}}\mathcal{N}$  be an open set of  $T_{\mathbf{a}}\mathcal{N}$ . For a given vector  $v \in V$  and  $1 \in \mathbf{I}$ , consider the *geodesic*  $\alpha : \mathbf{I} \rightarrow \mathcal{N}$  passing through  $\mathbf{a}$  with initial velocity  $\alpha'(0) = v$ . Then the mapping  $\exp_{\mathbf{a}} : V \rightarrow \mathcal{N}$  defined by  $v \mapsto \alpha(1)$  is well-defined [42] and is called the (Riemannian) *exponential mapping* on  $V$ .

The mapping  $\exp_{\mathbf{a}}$  is differentiable, and there is a neighborhood  $\mathcal{U}$  of  $\mathbf{a}$  such that the exponential map at  $\mathbf{a}$  is a diffeomorphism from the tangent space to the manifold. For  $\mathcal{U}$  being this neighborhood and  $\mathbf{a}, \mathbf{b} \in \mathcal{U}$ ,  $\mathbf{b} = \exp_{\mathbf{a}}(v)$ , then the inverse mapping  $\log_{\mathbf{a}} : \mathcal{U} \rightarrow T_{\mathbf{a}}\mathcal{N}$  defined by  $\mathbf{b} \mapsto v$  is called the (Riemannian) *logarithm mapping*. For brevity, we can also write  $\overrightarrow{\mathbf{a}\mathbf{b}}$  in the place of  $\log_{\mathbf{a}}(\mathbf{b})$ .

**Definition 18** (Riemannian curvature tensor and sectional curvatures [42]). Let  $\mathbb{X}(\mathcal{N})$  be the set of mappings from  $\mathcal{X}(\mathcal{N})$  to  $\mathcal{X}(\mathcal{N})$ . The *Riemannian curvature tensor*  $R$  of a differentiable manifold  $\mathcal{N}$  is the correspondence  $R : \mathcal{X}(\mathcal{N}) \times \mathcal{X}(\mathcal{N}) \rightarrow \mathbb{X}(\mathcal{N})$  that associates to each pair of vector fields

$\mathcal{X}, \mathcal{Y} \in \mathcal{X}(\mathcal{N})$  the application  $R(\mathcal{X}, \mathcal{Y}) : \mathcal{X}(\mathcal{N}) \rightarrow \mathcal{X}(\mathcal{N})$  given by

$$R(\mathcal{X}, \mathcal{Y})\mathcal{Z} := \nabla_{\mathcal{Y}}\nabla_{\mathcal{X}}\mathcal{Z} - \nabla_{\mathcal{X}}\nabla_{\mathcal{Y}}\mathcal{Z} + \nabla_{[\mathcal{X}, \mathcal{Y}]}\mathcal{Z},$$

where  $\nabla$  is the Riemannian connection of  $\mathcal{N}$ . A notion closely related to the Riemannian curvature tensor is the sectional curvatures of  $\mathcal{N}$ . Given two linearly independent tangent vectors  $u$  and  $v$  at the same point, the expression

$$K(u, v) := \frac{\langle R(u, v)u, v \rangle}{\langle u, u \rangle \langle v, v \rangle - \langle u, v \rangle^2}$$

does not depend on the choice of  $u, v$ , but only on the subspace  $\sigma$  spanned by them [42]. Given a point  $\mathbf{p} \in \mathcal{N}$  and a bidimensional subspace  $\sigma$  of  $T_{\mathbf{p}}\mathcal{N}$ , the real number  $K(u, v) = K(\sigma)$  where  $\{u, v\}$  is any basis of  $\sigma$ , is the *sectional curvature* of  $\sigma$  in  $\mathbf{p}$ .

### B. Proof of Theorem 1

Suppose  $\chi$  is a  $\text{RilthN}\sigma\text{R}$  of  $\mathbf{X}$ . Then, from (11), (6) is satisfied. Because  $\chi$  is a  $\text{RilthN}\sigma\text{R}$  of  $\mathbf{X}$ , from (12),  $\bar{\mathbf{X}}$  is a Riemannian sample mean of  $\chi$  and, therefore, from (4),  $\bar{\mathbf{X}}$  minimizes the function

$$g(\mathbf{x}) := \sum_{i=1}^N w_i^m \text{dist}^2(\mathbf{x}, \exp_{\bar{\mathbf{X}}}\chi_i).$$

The function  $g \circ \exp_{\bar{\mathbf{X}}} : \Omega(\bar{\mathbf{X}}) \subset T_{\bar{\mathbf{X}}}\mathcal{N} \rightarrow [0, \infty)$  is a real valued function defined in a subset of the vector space  $T_{\bar{\mathbf{X}}}\mathcal{N}$ . Since  $\Omega(\bar{\mathbf{X}})$  is convex by hypothesis and its second derivative is positive, then  $g \circ \exp_{\bar{\mathbf{X}}}$  is a strictly convex function. Because it is also a differentiable function,  $g \circ \exp_{\bar{\mathbf{a}}}$  has a unique minimum  $x^* \in \Omega(\bar{\mathbf{a}})$  and it is a critical point of  $g \circ \exp_{\bar{\mathbf{a}}}$ . Thus

$$[0]_{n \times 1} = \left. \frac{d(g \circ \exp_{\bar{\mathbf{X}}})(x)}{dx} \right|_{x=x^*} \Leftrightarrow x^* = \sum_{i=1}^N w_i^m \chi_i. \quad (64)$$

By Theorem 7.9 of [56],  $\bar{\mathbf{X}}$  is the unique minimum and critical point of  $g$ . Thus  $\log_{\bar{\mathbf{a}}}\bar{\mathbf{X}}$  is a critical point of  $g \circ \exp_{\bar{\mathbf{a}}}$  and

$$[0]_{n \times 1} = \overrightarrow{\bar{\mathbf{X}}\bar{\mathbf{X}}} = x^* = \sum_{i=1}^N w_i^m \chi_i =: \mu_{\chi}. \quad (65)$$

Hence, (7) is satisfied.

Now let us prove the converse for the mean. Suppose all points  $\chi_i$  belong to the domain of  $\exp_{\bar{\mathbf{X}}}$ , and that  $\chi$  is an  $l\text{thN}\sigma\text{R}$  of  $X := \overrightarrow{\bar{\mathbf{X}}\bar{\mathbf{X}}}$ . Define the set

$$\mathcal{X} := \{\exp_{\bar{\mathbf{X}}}\chi_i, w_i^m, w_i^{c,j}, w_i^{cc,j} | \chi_i \in \mathcal{N}; w_i^{cc,j}, w_i^{cc,j}, w_i^{cc,j} > 0\}_{i=1}^N \quad (66)$$

Then, from (6) and (66), (11) is satisfied. From (64) and (65), we have that  $\exp_{\bar{\mathbf{X}}}(\mu_{\chi}) = \exp_{\bar{\mathbf{X}}}(\overrightarrow{\bar{\mathbf{X}}\bar{\mathbf{X}}}) = \bar{\mathbf{X}}$  minimizes  $g$  and (12) is satisfied.

For even  $j$ , we have, from (7) and (5),

$$\mathcal{M}_{\chi}^j = \sum_{i=1}^N w_i^{cc,j} [(\chi_i - \mu_{\chi}) \diamond]^T \otimes^{\frac{j}{2}} =: \mathcal{M}_{\chi}^j;$$

and from (8), it follows  $\mathcal{M}_{\chi}^j = \mathcal{M}_{\chi}^j = M_{\chi}^j$ ; for odd  $j$ , the reasoning is similar. The remaining is straightforward.

### C. Proof of Corollary 1

From Theorem 1,  $\chi$  is a normalized  $l$ th  $N\sigma R$  of  $X \sim (\log_{\bar{X}}(\mathbf{X}), \mathbf{P}_{\mathbf{X}\mathbf{X}})^n$ . From Corollary 1 of [11], it follows that i)  $N \geq r + 1$ ; and ii), if  $\chi$  is symmetric, then  $N = 2r$ . The remaining of the proof is a direct consequence of Theorem 1.

### D. Proof of Proposition 1

A Riemannian mean  $\bar{X}$  of  $\mathbf{X}$  is such that it solves (2). Consider the following optimization problem

$$\begin{aligned} & \text{minimize } \tilde{g}(\tilde{c}) := g \circ \exp_{\bar{a}}(\vec{\bar{a}\tilde{c}}) = \sigma_{\vec{\bar{a}\tilde{c}}+p}^2(\tilde{c}) \\ & \text{subject to } \tilde{c} \in \mathcal{N}; \end{aligned} \quad (67)$$

From a reasoning similar to the proof of Theorem 1 (Appendix B), if  $\tilde{c}$  solves (67), then  $\log_{\bar{a}}^{-1} \tilde{c} = \exp_{\bar{a}} \tilde{c}$  solves (2), and  $\bar{X} = \exp_{\bar{a}} \tilde{c}$ . Since  $\sigma_{\vec{\bar{a}\tilde{c}}+p}^2(\tilde{c})$  is the variance of  $\vec{\bar{a}\tilde{c}} + p$  it follows that  $\mathcal{E}_{\vec{\bar{a}\tilde{c}}+p} \{\vec{\bar{a}\tilde{c}} + p\} = \bar{p}$  minimizes  $\tilde{g}(\tilde{c})$ ; thus  $\bar{X} := \exp_{\bar{a}} \bar{p}$ . For the covariance part, we have

$$\mathbf{P}_{\mathbf{X}\mathbf{X}} := \int_{\mathcal{N}-\mathcal{C}(\bar{X})} \bar{X} \mathbf{x} (\bar{X} \mathbf{x})^T p_{\mathbf{X}}(\mathbf{x}) d\mathcal{N}(\mathbf{x}) = \mathbf{P}_{\mathbf{a}\mathbf{a}} + \mathbf{P}_{pp}.$$

### E. Proof of Theorem 3

First, by considering  $\mathbf{c}_x = \mathbf{b}_x = \hat{x}_{k|k-1}$  and  $\mathbf{c}_y = \mathbf{b}_y = \hat{y}_{k|k-1}$  in the definitions of  $\mathbf{P}_{\mathbf{x}\mathbf{x},**}^{k|k-1}$ ,  $\mathbf{P}_{\mathbf{y}\mathbf{y},**}^{k|k-1}$ , and  $\mathbf{P}_{\mathbf{x}\mathbf{y},**}^{k|k-1}$ , (34) yields

$$\mathbf{G}_{k,**} = \begin{bmatrix} [0]_{n_x \times n_x} & \mathbf{G}_k \\ [0]_{n_y \times n_x} & [0]_{n_y \times n_y} \end{bmatrix}, \quad (68)$$

and substituting  $\mathbf{c}_x = \mathbf{b}_x = \hat{x}_{k|k-1}$ ,  $\mathbf{c}_y = \mathbf{b}_y = \hat{y}_{k|k-1}$ , and (68) into (35) gives  $\bar{x}_{k|k,**}^{T_c \mathcal{N}_{\mathbf{x},\mathbf{y}}} = [\mathbf{G}_k \log_{\hat{y}_{k|k-1}}(\hat{y}_k), [0]_{n_y \times 1}]^T$ ; consequently, from (33),  $x_{k|k}^{TM} := [x_{k|k,**}^{T_c \mathcal{N}_{\mathbf{x},\mathbf{y}}}]_{1:n_x,1} = \mathbf{G}_k \log_{\hat{y}_{k|k-1}}(\hat{y}_k)$ . Second, considering (68) into (36) yields

$$\mathbf{P}_{\mathbf{x}\mathbf{x},**}^{k|k,T_c M} = \text{diag} \left( \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1} - \mathbf{G}_k (\mathbf{P}_{\mathbf{y}\mathbf{y}}^{k|k-1})^{-1} \mathbf{G}_k^T, [0]_{n_y \times n_y} \right);$$

and, from (33), it follows that  $\mathbf{P}_{\mathbf{x}\mathbf{x},**}^{k|k,\hat{x}_{k|k-1}} = \mathbf{P}_{\mathbf{x}\mathbf{x}}^{k|k-1} - \mathbf{G}_k (\mathbf{P}_{\mathbf{y}\mathbf{y}}^{k|k-1})^{-1} \mathbf{G}_k^T$ .

### F. Notation and Acronyms

Throughout this paper, we use the following notations:

- for a matrix  $A$ ,  $(A) (\diamond)^T$  stands for  $(A) (A)^T$ , and  $\sqrt{A}$  for a square-root matrix of  $A$  such that  $A = \sqrt{A} \sqrt{A}^T$ .
- $\otimes$  stands for the Kronecker product operator, and  $A^{\otimes n} := A \otimes \dots \otimes A$ .
- $[A]_{p \times q}$  stands for a block matrix consisting of the matrix  $A$  being repeated  $p$  times in the rows and  $q$  times in the columns.
- $[A]_{i_1:i_2, j_1:j_2}$  stands for a sub-matrix of the matrix  $A$  formed by the rows  $i_1$  to  $i_2$  and the columns  $j_1$  to  $j_2$  of  $A$ .
- $\mathbf{I}$  stands for an open interval in  $\mathbb{R}$ .

Below, we provide a list of acronyms and parts of acronyms—these parts end with an ‘-’ and are followed by examples—along with their meaning. There are other

acronyms in the text that can be composed by i) concatenating some items below (e.g.,  $\text{Mi}\sigma R$  [Mi- with  $\sigma R$ ] standing for Minimum  $\sigma$ Representation), and ii) adding Ri- (standing for Riemannian; e.g.,  $\text{RiMi}\sigma R$  [Ri- with  $\text{Mi}\sigma R$ ] standing for  $\text{RiMi}\sigma R$ ):

- **AdUKF**: Additive Unscented Kalman Filter.
- **AuUKF**: Augmented Unscented Kalman Filter.
- **EKF**: Extended Kalman Filter.
- **HoMiSy-**: Homogeneous Minimum Symmetric- (e.g.,  $\text{HoMiSy}\sigma R$ ,  $\text{RiHoMiSy}\sigma R$ ,  $\text{RiHoMiSyAdUKF}$ ,  $\text{RiHoMiSyAuUKF}$ ).
- **KF**: Kalman Filter.
- **$l$ th  $N\sigma R$** :  $l$ th order  $N$  points  $\sigma$ -representation.
- **$l$ UT**:  $l$ th order UT.
- **Mi-**: Minimum- (e.g.,  $\text{RiMi}\sigma R$ ,  $\text{RiMiAdUKF}$ ,  $\text{RiMiAuUKF}$ ).
- **MiSy-**: Minimum Symmetric- (e.g.,  $\text{RiMiSy}\sigma R$ ,  $\text{RiMiSyAdUKF}$ ,  $\text{RiMiSyAuUKF}$ ).
- **RhoMi-**: Rho Minimum- (e.g.,  $\text{RiRhoMi}\sigma R$ ,  $\text{RiRhoMiAdUKF}$ ,  $\text{RiRhoMiAuUKF}$ ).
- **$\sigma R$** :  $\sigma$ -Representation.
- **UKF**: Unscented Kalman Filter.
- **UT**: Unscented Transformation.



**Henrique M. T. Menegaz** received the B.S. degree in Electrical Engineering from the Universidade de Brasília (UnB), Brazil, in 2007. He received the M.S. and Ph.D. degrees in Engineering of Electronic Systems and Automation from the UnB in 2011 and 2016 respectively. He is currently an Assistant Professor with Faculdade Gama, UnB. His major field of study is filtering of nonlinear dynamic systems and their applications.



**João Y. Ishihara** received the Ph.D. degree in Electrical Engineering from the University of São Paulo, Brazil, in 1998. He is currently an Associate Professor at the University of Brasília, Brazil. His research interests include robust filtering and control theory, singular systems, and robotics.



**Hugo T. M. Kussaba** received the B.S. degree in control engineering and the M.S. degree in engineering of electronic systems and automation from the University of Brasília (UnB), Brazil, in 2012 and 2014 respectively. Currently, he is a Ph.D. student at the same university. His research interests include control and estimation on Lie groups, robust control and linear matrix inequalities, and hybrid dynamical systems.