frontiers in Signal Processing

TYPE Original Research
PUBLISHED 30 June 2026
DOI 10.3389/frsip.2026.1829279

Check for updates

OPEN ACCESS

EDITED BY
Congzheng Han,
University of Chinese Academy of Sciences, China

REVIEWED BY
Mahesh Kumar Pal,
Indian Institute of Technology Roorkee, India
Junpeng Zhang,
Shenzhen Institute of Information Technology, China

*CORRESPONDENCE
Aron Persson,
aron.persson@umu.se

RECEIVED 12 March 2026
REVISED 15 May 2026
ACCEPTED 29 May 2026
PUBLISHED 30 June 2026

CITATION
Figueras J-L, Persson A and Viitasaari L (2026) Extended Kalman filtering on Stiefel manifolds.
Front. Signal Process. 6:1829279.
doi: 10.3389/frsip.2026.1829279

COPYRIGHT
© 2026 Figueras, Persson and Viitasaari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Extended Kalman filtering on Stiefel manifolds

Jordi-Lluís Figueras[1], Aron Persson[1,2]* and Lauri Viitasaari[3]

[1]Department of Mathematics, Uppsala University, Uppsala, Sweden, [2]Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden, [3]Department of Information and Service Management, Aalto University, Espoo, Finland

A generalization of the extended Kalman filter for Stiefel manifold-valued measurements is presented. Our approach is based on translating between manifold-valued measurements and the ambient space through an injective mapping between scalar true variance on the manifold and the variance of the noise. This mapping inherently takes into account the curvature of the manifold corresponding to the higher order derivatives. To illustrate our approach, we provide simulations on the 2-sphere and the space of orthogonal 4-by-2 matrices which show significant improvement of the Extended Kalman Filter compared to only relying on raw measurements. We also compare our approach to classical Extended Kalman Filter and second order Extended Kalman Filter, and complement our results by physically informed application related to spacecraft measurements. Our results indicate that our algorithm has similar performance as the second order Extended Kalman Filter without requiring the costly computations of the second order derivatives.

KEYWORDS
Kalman filter, non-linear filtering, projected normal distribution, signal processing on manifolds, Stiefel manifolds

1 Introduction

There are multiple situations where projected dynamics is inherently tied to gaining information. We shall begin by giving simple examples of two such situations in order to broaden the perspective before we go into the specifics relating to applications of Stiefel manifold-valued statistics in radiology.

Consider a radioactive dust particle travelling through an open space. Suppose also there is a moderate wind going through the area and we can only measure the direction of the incoming radiation. In physics one often models the movement of dust particles as if they travel like Brownian motion with a drift (the drift here comes from the wind). Therefore the dust particle does not travel deterministically and there is an uncertainty inherent to the position of the dust particle. Moreover, the measured radiation incoming into our sensor might have a somewhat perturbed inclination error as it collides with air particles in between the dust particle and the sensor. Hence, there is an uncertainty of the incoming direction of the radiated particles as well. In this situation the dust particle is travelling across $\mathbb{R}^3$ and our measurements are taken on the sphere $\mathbb{S}^2$. The information obtained is (non-linearly) projected onto the sphere.

Another plausible scenario would be a circular particle accelerator. As a charged particle is centripetally accelerated through electro-magnetic forces we measure its position in the accelerator using sensors throughout the walls of the accelerator. As the particle travelling through the accelerator is a quantum particle, the position of the particle is inherently probabilistic. At the same time the sensors are not perfect and there is uncertainty in the measurements. Here the particle moves inside a tubular neighbourhood of a circle and the measurements happen on the torus $\mathbb{T}^2$.

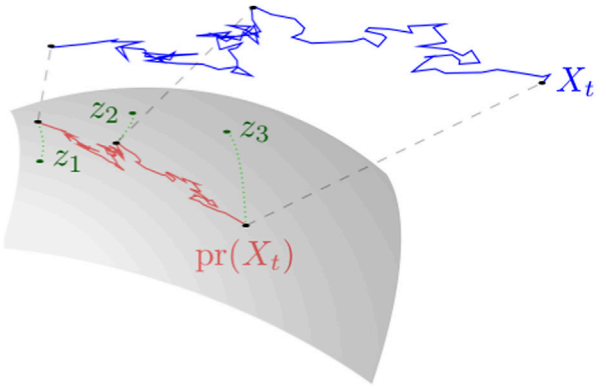

FIGURE 1
A picture of a stochastic process X_t that is (non-linearly) projected onto a manifold together with some measurements z_1, z_2, z_3.

In MIMO radio-systems for cellular networks a configuration of n antennas interfere constructively and a directed radio signal is produced with high energy efficiency. This signal is then received at and responded by k receivers. These receivers move around, in part, in a predictive way (a person moving in a certain direction will probably move in that direction for a while) and, in part, with some uncertainty. Hence, these receivers may be modelled as stochastic processes with some dynamical component. Moreover, the antenna configurations can be modelled as being a vector in $\mathbb{R}^n$ for each receiver. In total we consider the state as an n by k matrix. However, as these configurations do not depend on the total strength of the currents, only their relative strength and phase, so measurements are realised as n by k orthogonal matrices, see Hussien et al. (2014); Pitaval and Tirkkonen (2013); Seddik et al. (2017). These realisations may be computed by only retaining the orthogonal part, Q, in the polar decomposition of the n by k matrix. This Q can be computed so that it is a projection onto the space of n by k orthogonal matrices, i.e. the Stiefel manifold $\mathrm{St}_{n,k}$. These measurements are assumed to be noisy and the measured optimal configuration of the antennae has some uncertainty.

All these three situations illuminates a mathematical problem: an object moves around with a drift in some (possibly an open subset of) vector space, see Figure 1 for a picture of these situations. If one performs measurements on a non-linear space, i.e. if $x, y \in X$, then it is generally not true that $x + y \in X$. How would one filter out this information using both the known drift of the system and the measurements? In all the above examples the non-linear spaces are manifolds which have sufficient geometrical structure which allows for measurement of distances between points. Using distances and curves on these manifolds it is then possible to weight the prediction together with the measurement. In a nutshell, working on the original filtering equations developed in Kálmán (1960a); Kálmán (1960b); Kálmán and Bucy (1961) we will filter the predicted point and the measurement by using a curve that connects these two points. Then one can obtain the filtered mean by taking a weighted average along the curve.

In recent time, quite a few manifold-valued filters has been developed. Some Snoussi and Mohammad-Djafari (2006); Chahbazian et al. (2022); Boulanger et al. (2016); de Figueredo et al. (2022) apply some type of particle filters in order to solve the manifold-valued filtering problem. However, particle-filters are accurate but computationally expensive, especially as the dimension of the manifold goes up. For example, problems in radiology where real time estimations of (ideally ≥ 100 dimensional) manifolds the particle filters are non-tractable. In Boulanger et al. (2016) the authors explicitly explain the curse of dimensionality of these methods by saying

> "For higher dimensions ($n > 10$), the use of a particle filter should not be considered. Indeed, as n increases, the dimension of the estimation space increases like $O(n^2)$ and a very large number of particles is required to perform a proper estimation".

Any Manifold-valued Extended Kalman filters that has been previously considered Brossard et al. (2021); Barrau and Bonnabel (2018); Barrau and Bonnabel (2016) only consider the case when the manifold itself is a Lie group or by lifting the problem to a Lie group Mahony et al. (2022). Therefore, there is a need to develop some kind of Extended Kalman filter with projected normal observations on the Stiefel manifolds that can also deal with high dimensions.

More broadly, this paper can be viewed as part of the vast field of directional statistics. The area of directional statistics has found applications in medicine, see Craig and Oppe (2010); Demir and Bilgin (2019); Karaibrahimoglu et al. (2021), in meteorology, see Li et al. (2018); Nuñez-Antonio et al. (2015); Nuñez-Antonio et al. (2011), and in robotics, see Lang and Feiten (2012); Sveier and Egeland (2018), to name a few. Applications of these manifold-valued filtering methods are found in diverse fields. In aerospace and navigation, they are utilized for spacecraft attitude stabilization and control Bloch and Marsden (1990); Astolfi and Lovera (2002) as well as terrain-based positioning Murangira et al. (2016). Robotics and sensor networks employ these techniques for rigid body localization Chepuri et al. (2014) and kinematic state estimation Park et al. (2008). In computer vision and medical imaging, they facilitate action and object recognition Turaga et al. (2011); Lui (2012); Liu et al. (2004) and the statistical analysis of diffusion tensor MRI Lenglet et al. (2006). Furthermore, specific applications are found in biomechanics for assessing climbing fluency Seifert et al. (2014), in wireless communications for predictive quantization Schwarz and Rupp (2014), and in deep learning for enforcing orthogonality constraints in neural network training Bansal et al. (2018); Wang et al. (2020); Fiori (2001); Fiori (2008).

This paper is mainly aimed towards the engineering/application viewpoint of the Extended Kalman filter with Stiefel manifold-valued observations. While detailed convergence proofs are provided in a sister paper Figueras et al. (2025a), we present the mathematical derivation and Bayesian motivation for the filter design in Section 3. However, in order to make any technical conclusions one has to make a few approximations of some of the maps used in the filter. These approximations are not necessary in order to apply our proposed filter and the approximations give rise to additional error propagation. Therefore, we keep the more exact formulation and application relevant version of the filter in this paper with emphasis on application and proof of concept.

The paper is organized as follows. Section 2 explains the foundational concepts for the problem at hand. In Section 3 we give an exposition on how to use the extended Kalman filter when the measurements are points on the Stiefel manifold $\mathrm{St}_{n,k}$. The reader with little interest of the geometrical meaning behind the Extended Kalman filter on $\mathrm{St}_{n,k}$ may jump directly to Algorithm 2 at the end of this section. In Sections 4 and 5 we give results from simulations for the 2-sphere and $\mathrm{St}_{4,2}$, respectively. In Section 7 we discuss some

limitations and we present some potential drawbacks of the extended Kalman filter on Stiefel manifolds.

2 Setting

Suppose an object X_t is represented as an n by k matrix, that is $X_t \in \mathbb{R}^{n\times k}$. As t increases we suppose it moves around with known dynamics and with non-zero noise (therefore the movement is not fully deterministic). This may be mathematically represented as an SDE (Stochastic Differential Equation)

$$\mathrm{d}X_t = AX_t\mathrm{d}t + \nu\mathrm{d}B_t, \qquad X_0 \overset{\mathrm{d}}{=} \mathrm{N}(\mu_0, \nu_0^2 \mathrm{id}_{\mathbb{R}^{n\times k}}). \tag{1}$$

Here $\mathrm{id}_{\mathbb{R}^{n\times k}}$ denotes the identity map from the space of n by k matrices to itself, i.e. for given $Y \in \mathbb{R}^{n\times k}$, $\mathrm{id}_{\mathbb{R}^{n\times k}}(Y) = Y$. More explicitly, if $Y_{\mathrm{vec}} \in \mathbb{R}^{nk}$ is the vectorization of an n times k matrix Y, then if $\mathrm{id}^{\mathrm{vec}}_{\mathbb{R}^{n\times k}}$ is the corresponding vectorized linear map, it is the $nk \times nk$ identity matrix. The initial value X_0 is the starting (normal) distribution of X_t and $A \in \mathbb{R}^{n\times n}$ describes the dynamics which is the infinitesimal displacement over time. The term $\mathrm{d}B_t$ is a formal stochastic differential of the Brownian motion which we scale by a factor ν. After a time step $\delta t > 0$ and if $A = 0$, then solving Equation 1 one would obtain $X_t = X_0 + \nu\int_0^{\delta t} \mathrm{d}B_t$ which has variance matrix $(\nu_0^2 + \nu^2\delta t)\mathrm{id}_{\mathbb{R}^{n\times k}}$ and constant mean μ_0. More generally, if A is anti-symmetric, i.e. $A^T = -A$, then one obtains the full solution to Equation 1 as

$$X_t \overset{\mathrm{d}}{=} \mathrm{N}(\exp_{\mathrm{M}}(tA)\mu_0, (\nu_0^2 + t\nu^2)\mathrm{id}_{\mathbb{R}^{n\times k}}).$$

Here $\exp_{\mathrm{M}}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ denotes the matrix exponential defined as

$$\exp_{\mathrm{M}}(A) = \sum_{j=0}^{\infty} \frac{A^j}{j!},$$

where A^j denotes j-times matrix multiplication and $A^0 = I_{n\times n}$, the $n \times n$ identity matrix. (Recall that this series is absolutely convergent for any A).

Remark 2.1. Note that this solution is only true when the factor ν is a scalar (for which we then call the noise isotropic). If the factor ν is not a scalar then there will appear more time dependent terms in the covariance part.

Suppose now the process X_t from Equation 1 is observed, but we only observe the orthogonal part in the polar decomposition of X_t. This observation is a projection onto the space of $n \times k$ matrices with pairwise orthonormal columns, the Stiefel manifold

$$\mathrm{St}_{n,k} \coloneqq \{X \in \mathbb{R}^{n\times k}\colon X^T X = I_k\}.$$

We write $\mathrm{pr}\colon \mathbb{R}^{n\times k} \to \mathrm{St}_{n,k}$, and its computation can be done by a polar decomposition using SVD algorithms. If $X_t = Q_t D_t S_t^*$ is the thin SVD-decomposition, see Bai et al. (2000), where D_t and S_t are $k \times k$ matrices and Q_t is $n \times k$, then $\mathrm{pr}(X_t) = Q_t S_t^*$. Note that this map is heavily non-linear and should not be confused with (linear) projections seen in linear algebra. A conceptual picture on how this projection of the stochastic process X_t might look like is shown in Figure 2.

Suppose further that when this projected process is observed there is an error of magnitude ξ^2 inherent to these measurements. That is, if $\mathrm{pr}(X_t)$ is the orthogonal part of X_t using the polar decomposition, the measurement's error of X_t is assumed to have mean $\mathrm{pr}(X_t)$ and (scalar) variance ξ^2 (we shall delve into what this means in the next section). In Figure 1 there is an illustration of how these faulty measurements may look like.

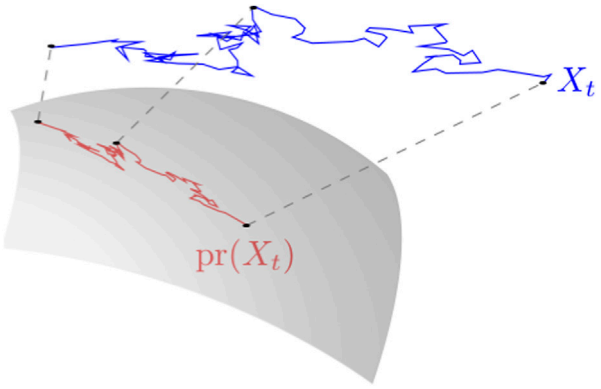

FIGURE 2
A picture of a stochastic process which is then non-linearly projected onto a non-linear space.

Now the goal is to estimate $\mathrm{pr}(X_t)$ given that we know the system parameters $A \in \mathbb{R}^{n\times n}$, $\nu \in \mathbb{R}$ and the measurement error. In order to capture parts of this non-linearity we shall use the tangent space. The tangent space $T_X\mathrm{St}_{n,k}$ contains all possible velocities (aka. tangent vectors) at the point $X \in \mathrm{St}_{n,k}$ such that any curve going through X with velocity in $T_X\mathrm{St}_{n,k}$ will (infinitesimally) stay on $\mathrm{St}_{n,k}$, see Figure 3 for an illustration. Explicitly, the tangent space is given by

$$T_X\mathrm{St}_{n,k} = \{V \in \mathbb{R}^{n\times k}\colon V^T X = -X^T V\}$$

which may be deduced from that any curve $Y(t)$ on $\mathrm{St}_{n,k}$ must satisfy $Y^T(t)Y(t) = I_k$. Differentiating this expression yields exactly the above set of tangent vectors.

Now, considering $\mathrm{St}_{n,k}$ as a subset[1] of $\mathbb{R}^{n\times k}$, one can measure distances on $\mathrm{St}_{n,k}$ by considering the length of curves on $\mathrm{St}_{n,k}$ with constant velocity by restricting a modified version of the Frobenius norm onto the tangent space at every point. Recall that the classical Frobenius norm of a matrix $B \in \mathbb{R}^{n\times k}$ is defined by $\|B\|^2 \coloneqq \|B\|_F^2 = \mathrm{tr}(B^T B)$. Through the notion of distance, one can relate points on the tangent space $T_X\mathrm{St}_{n,k}$ to points near X: by denoting $\exp_X(V)$ as the point $Y \in \mathrm{St}_{n,k}$ which is reached by taking the constant velocity curve starting at X with starting velocity direction $V/\|V\|$ and is the distance $\|V\|$ away. Note here that $\exp_X$ is not the same as the matrix exponential map $\exp_{\mathrm{M}}$ defined earlier. The map $\exp_X\colon T_X\mathrm{St}_{n,k} \to \mathrm{St}_{n,k}$ is called the (Riemannian) exponential map (we have provided a picture of this exponential map in Figure 4). We denote the inverse of $\exp_X$ by $\log_X$, and call it the logarithm map.

Remark 2.2. Explicitly, one may compute $\exp_X(V)$ by first computing the QR-decomposition

1 This is not an isometric embedding into $\mathbb{R}^{n\times k}$ when equipped with the Frobenius norm, see Remark 2.3.

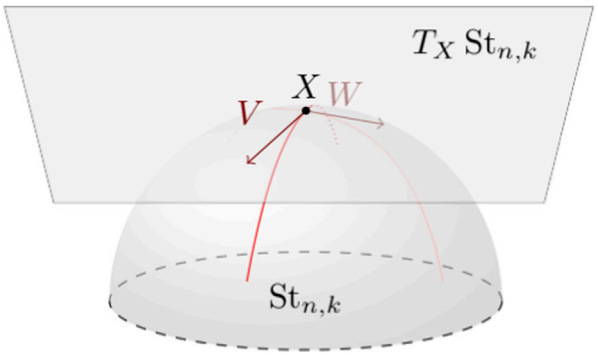

FIGURE 3
A conceptual picture of the tangent space at a point $X \in \mathrm{St}_{n,k}$, together with two tangent vectors $V, W \in T_X\mathrm{St}_{n,k}$ and their corresponding curves that has the corresponding momentous velocity at X.

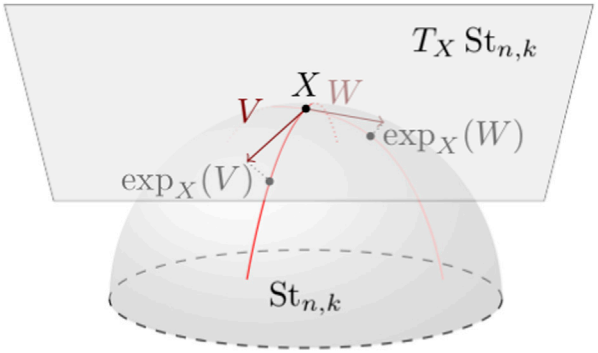

FIGURE 4
An illustration of the exponential map and how it maps tangent vectors $V, W \in T_X\mathrm{St}_{n,k}$ to points on $\mathrm{St}_{n,k}$.

$$QR = (I_{n\times n} - XX^T)V$$

(here Q is a n by k orthogonal matrix and R is a k by k upper triangular matrix), then

$$\exp_X(V) = \begin{pmatrix} X & Q \end{pmatrix} \exp_{\mathrm{M}}\left(\begin{pmatrix} X^TV & -R^T \\ R & 0_{k\times k} \end{pmatrix}\right)\begin{pmatrix} I_{k\times k} \\ 0_{k\times k} \end{pmatrix}.$$

In Theorem 2.1 and Corollary 2.2 in Edelman et al. (1998) one can find more information about this exponential map. To compute the logarithm, which is a little bit more involved to compute but still possible, we refer the reader to Zimmermann (2017).

The inner product, aka metric, on the tangent space $T_X\mathrm{St}_{n,k}$, with $X \in \mathrm{St}_{n,k}$, is computed as follows

$$\langle V, W\rangle_X = \mathrm{tr}\left(V^T\left(I_{n\times n} - \frac{1}{2}XX^T\right)W\right). \tag{2}$$

Remark 2.3. An important key point to note is that $\mathrm{St}_{n,k}$ is not considered as a subset of $\mathbb{R}^{n\times k}$ with the standard Frobenius norm. Here, the norm induced by Equation 2 is the norm inherited from the Frobenius norm on the Lie group $O(n)$ of which the Stiefel manifold is a quotient space. This actually makes $\mathrm{St}_{n,k}$ a isometrically submerged submanifold of $O(n)$ and by extension $\mathbb{R}^{n\times n}$. It is still an open question how and if one can isometrically embed $\mathrm{St}_{n,k}$ in $\mathbb{R}^{n\times k}$ with an easily expressed norm defined on $\mathbb{R}^{n\times k}$.

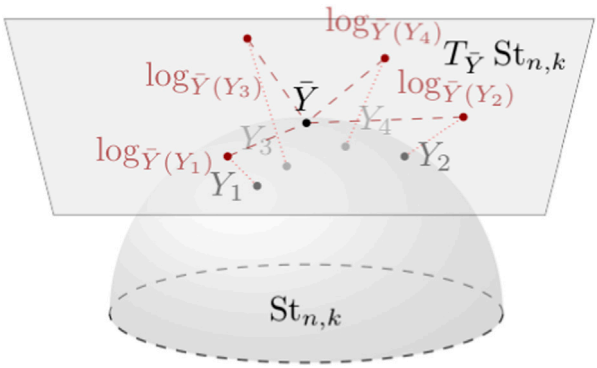

FIGURE 5
A picture on how the inherent average $\bar{Y}$ relates to a set of points $Y_1, Y_2, Y_3, Y_4 \in \mathrm{St}_{n,k}$.

Through the logarithm map one can define a notion of average which is inherent (aka. intrinsic) to $\mathrm{St}_{n,k}$. Given points $Y_1, \ldots, Y_N \in \mathrm{St}_{n,k}$, their average is defined as $\bar{Y} \in \mathrm{St}_{n,k}$ which minimises the distance squared from all the points $Y_1, \ldots, Y_N$. Locally, the average $\bar{Y} \in \mathrm{St}_{n,k}$ explicitly satisfies the barycenter equation

$$\sum_{i=1}^{N} \log_{\bar{Y}}(Y_i) = 0.$$

This average is illustrated in Figure 5. Note that this is a non-linear generalisation of the standard definition of average in a vector space: in this case the average $\bar{y} \in \mathbb{R}$ satisfies

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}) = 0$$

which is a bit more convoluted way of saying

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Let B_j be an orthonormal basis for the vector space $T_{\bar{Y}}\mathrm{St}_{n,k}$ under the inner product Equation 2. The sample covariance given points $Y_1, \ldots, Y_N \in \mathrm{St}_{n,k}$, is then the $nk - \frac{1}{2}k(k+1)$ by $nk - \frac{1}{2}k(k+1)$ matrix

$$(\mathrm{Cov}_Y)_{i,j} = \frac{1}{N-1}\sum_{\ell=1}^{N}\langle \log_{\bar{Y}}(Y_\ell), B_i\rangle_{\bar{Y}}\langle \log_{\bar{Y}}(Y_\ell), B_j\rangle_{\bar{Y}}$$

with corresponding scalar variance

$$\frac{\mathrm{tr}(\mathrm{Cov}_Y)}{nk - \frac{1}{2}k(k+1)}. \tag{3}$$

Note also that the notion of variance has a meaning of measuring the squared distance, and the mean has the meaning of being the point which minimises the least squared distance, see (Figueras et al., 2025b, Definition 2.1 and 2.2).

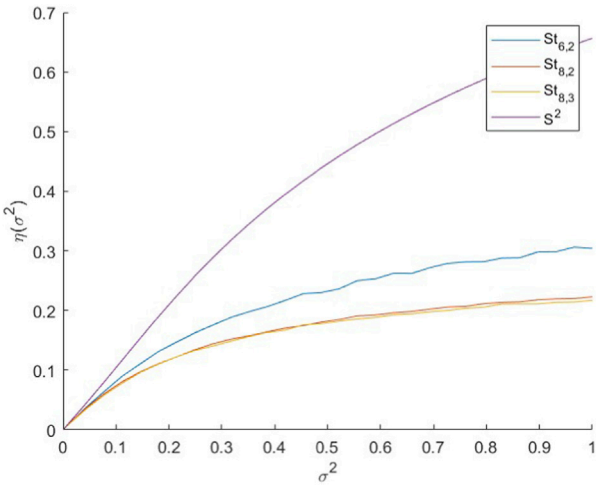

FIGURE 6
The intrinsic scalar variance of the projected normal as a function of the scalar variance of the original normal distribution.

Observe that spheres $\mathbb{S}^{n-1}$ are special cases of $\mathrm{St}_{n,k}$. Specifically note that $\mathbb{S}^{n-1}$ is the same as $\mathrm{St}_{n,1}$. On the spheres, we know that given a normal random variable X in $\mathbb{R}^{n+1}$ with mean μ and isotropic covariance $\sigma^2 I_{n+1}$ the projected random variable $\mathrm{pr}(X) = \frac{X}{\|X\|}$ has inherent average $\frac{\mu}{\|\mu\|}$. Moreover, it has been shown that there is unique, one-to-one, intrinsic scalar variance of $\mathrm{pr}(X)$ for every choice of σ for $\mathbb{S}^2$, see (Figueras et al., 2025b, Theorem 2.8). Along this paper we shall assume that the same is true for all $\mathrm{St}_{n,k}$.

Let $X \in \mathbb{R}^{n\times k}$ is a normal random variable with mean $\mathbb{E}[X] \in \mathrm{St}_{n,k}$ and covariance matrix $\mathrm{Cov}(X) = \sigma^2 \mathrm{id}_{\mathbb{R}^{n\times k}}$. Then the inherent scalar variance can be computed by

$$\eta(\sigma^2) = \frac{1}{nk - \frac{1}{2}k(k+1)} \int_{\mathrm{St}_{n,k}} \mathrm{dist}^2(\mathbb{E}[X], x) p_{\mathrm{pr}(X)}(x) \mathrm{dVol}_{\mathrm{St}_{n,k}}(x), \tag{4}$$

which is the true scalar variance of the random variable $\mathrm{pr}(X)$ corresponding to the sample scalar variance in Equation 3. Because the intrinsic covariance is inherently respecting the curvature of the manifold, the mapping η gives a (virtually) cost-free way of taking the curvature of the manifold into consideration when moment-matching the covariance. This allows to translate between true scalar variance and the corresponding variance in the ambient space, see Section 3. Due to Figueras et al. (2025b), in the case of $\mathrm{St}_{3,1} = \mathbb{S}^2$, η is known to be injective and is explicitly given by

$$\eta(\sigma^2) \coloneqq \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\right) \int_0^\pi \phi^2 \left(\frac{\cos(\phi)}{\sigma} + \left(\frac{\cos^2(\phi)}{\sigma^2} + 1\right) \frac{\Phi\left(\frac{\cos(\phi)}{\sigma}\right)}{\varphi\left(\frac{\cos(\phi)}{\sigma}\right)}\right) \sin(\phi) \mathrm{d}\phi. \tag{5}$$

As of today, there is no such explicit formula for a general Stiefel manifold since $p_{\mathrm{pr}(X)}$ is difficult to express for $\mathrm{St}_{n,k}$ when $k \geq 2$. Because of this, we shall need the following conjecture in order to perform parameter inference on the projected observations.

Conjecture 2.4. *Let $X \in \mathbb{R}^{n\times k}$ be a normal random variable with $\mathbb{E}[X] \in \mathrm{St}_{n,k}$ and covariance $\mathrm{Cov}(X) = \sigma^2 \mathrm{id}_{\mathbb{R}^{n\times k}}$. Then the projected normal random variable has inherent average $\mathbb{E}[X]$. Moreover, there exists an injective function η: $[0,\infty) \to [0,\infty)$ such that the inherent scalar variance is uniquely determined by $\eta(\sigma^2)$ for each $\sigma > 0$.*

Remark 2.5. Moreover, the function η is conjectured to have derivative 1 at 0. This has been numerically verified as seen in Figure 6. This is also a reasonable conjecture intuitively. When the variance is small then the projected normal lives in a small neighborhood on the manifold and the manifold looks flat.

The pseudo-code in Algorithm 1 help us to numerically estimate the function η using the Monte Carlo method, in the case that η is difficult to explicitly express. That Algorithm 1 converges follows from (Figueras et al., 2025b, Theorem 2.5).

```
Given σ² > 0 and μ = I_{n,k}.
Draw L samples {X_i} from N(μ, σ² id_{R^{n×k}})
Compute Z_i = pr(X_i) for each i.
Compute V_i = log_μ(Z_i) for each i.
Let B_j be an orthonormal basis for T_μ St_{n,k} and estimate
```

$$\eta(\sigma^2) \approx \frac{1}{L} \frac{1}{nk - \frac{1}{2}k(k+1)} \sum_{i=1}^{L} \sum_{j=1}^{\left(nk - \frac{1}{2}k(k+1)\right)} \langle V_i, B_j \rangle^2$$

Algorithm 1. An outline of how to estimate the function η, defined by **Equation 4**, on $\mathrm{St}_{n,k}$.

Using Algorithm 1, we have numerically verified Conjecture 2.4 for the Stiefel manifolds $\mathbb{S}^2$, $\mathrm{St}_{6,2}$, $\mathrm{St}_{8,2}$ and $\mathrm{St}_{8,3}$, see Figure 6. For each $\sigma^2 \in (0,1]$ and each Stiefel manifold, 100 pseudo-random samples of $\mathrm{N}(I_{n,k}, \sigma^2 \mathrm{id}_{\mathbb{R}^{n\times k}})$ were drawn. Then, $\eta(\sigma^2)$ was estimated for each σ^2 and each Stiefel manifold. Notice that $\mathbb{S}^2$ is the only exception, in this case $\eta(\sigma^2)$ was computed by numerically approximating the integral in Equation 5. Notice also that each η is roughly monotonically increasing and does not contradict Conjecture 2.4.

Now we have sufficient amount of tools to do filtering on $\mathrm{St}_{n,k}$.

3 Extended Kalman filtering on Stiefel manifolds

Suppose we have the following filtering problem

$$\begin{cases} \mathrm{d}X_t = AX_t \mathrm{d}t + \nu \mathrm{d}B_t, & X_0 \overset{\mathrm{d}}{=} \mathrm{N}(\mu_0, \sigma_0^2 \mathrm{id}_{\mathbb{R}^{n\times k}}) \\ Z_m = \mathrm{pr}(\mathrm{pr}(X_{t_m}) + \varepsilon_m) \end{cases}, \tag{6}$$

where t_m is an increasing sequences of discrete times, $\varepsilon_m \in T_{\mathrm{pr}(X_{t_m})}\mathrm{St}_{n,k}$ are i. i.d. such that $\varepsilon_m \overset{\mathrm{d}}{=} \mathrm{N}(0, \xi^2 \mathrm{id}_{T_{\mathrm{pr}(X_{t_m})}St_{n,k}})$, $\nu \in \mathbb{R}$ and A: $\mathbb{R}^n \to \mathbb{R}^n$ is an anti-symmetric linear mapping $(A^T = -A)$.

Next we shall adopt the classical Extended Kalman filter to the setting of Stiefel manifolds. Suppose we have prior $X_0 \overset{\mathrm{d}}{=} \mathrm{N}(\mu_0, \sigma_0^2 \mathrm{id}_{\mathbb{R}^{n\times k}})$, and suppose after time t_1 we make measurement z_1. Now our predicted average is $\mu_{\mathrm{pred.}} = \exp_{\mathrm{M}}(t_1 A)\mu_0$ and the (inherent) predicted scalar variance

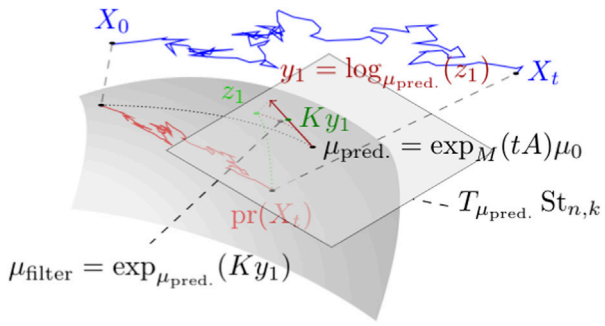

FIGURE 7
A geometrical picture of how the filtered average μ^K is computed after a timestep t as a weighted average with weight Kalman gain K between the predicted mean $\mu_{\text{pred.}}$ and the measurement z_1.

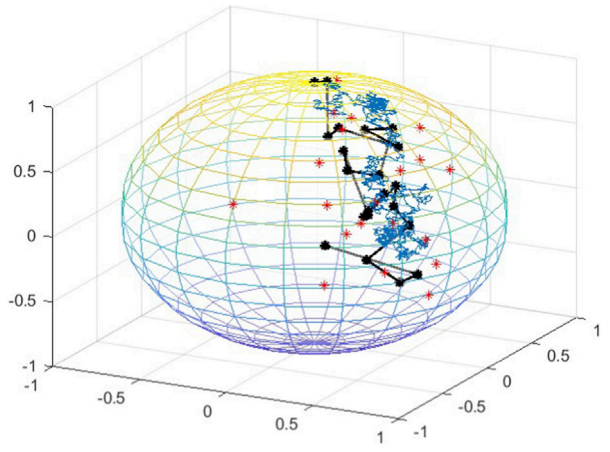

FIGURE 8
In blue the (projected) underlying stochastic process X_t, in red measurements Z_m and in black the mean filtered estimate, with parameters A given in Equation 8, $\nu^2 = 1$, and $\xi^2 = 0.1$. The initial distribution is given by $\Sigma_0 = 0.1I_3$ and μ_0 is given by Equation 9.

is, by a linear approximation, $P_{\text{pred.}} = \eta(\sigma_0^2 + t_1\nu^2)$. Since the measurements have scalar variance ξ^2 we can compute the Kalman gain by

$$K = \frac{\sigma_0^2 + t_1\nu^2}{\sigma_0^2 + t_1\nu^2 + \xi^2}.$$

We use the logarithm map to get the tangent vector $y_1 \in T_{\mu_{\text{pred.}}}\text{St}_{n,k}$ which gives the velocity of the constant speed curve starting at $\mu_{\text{pred.}}$ which reaches z_1, i.e. $\exp_{\mu_{\text{pred.}}}(y_1) = z_1$. Now we rescale this vector by the Kalman gain K to get a weighted average of the predicted point $\mu_{\text{pred.}}$ and the measurement z_1. We end up with the mean estimate $\exp_{\mu_{\text{pred.}}}(Ky_1)$ and the estimate $(1-K)P_{\text{pred.}}$ for the inherent (scalar) variance. Hence our posterior, after lifting this distribution back to the ambient space $\mathbb{R}^{n\times k}$, is

$$\text{N}\left(\exp_{\mu_{\text{pred.}}}(Ky_1), \eta^{-1}((1-K)P_{\text{pred.}})\text{id}_{\mathbb{R}^{n\times k}}\right).$$

In Figure 7 we give a picture of the geometry behind the filtering. Moreover, the entire extended Kalman filtering algorithm is given step-by-step in Algorithm 2 as pseudo-code.

3.1 Mathematical interpretation

The proposed filter can be interpreted within the framework of Bayesian estimation on manifolds. The filtering problem aims to recursively compute the posterior probability distribution of the state X_t given the sequence of measurements $z_1, \ldots, z_m$. Due to the non-linear nature of the projection map $\text{pr}(\cdot)$ and the Stiefel manifold geometry, the exact posterior is generally non-Gaussian and computationally intractable to represent exactly.

Our approach approximates the posterior distribution at each step k by a Projected Normal distribution $\text{PrN}(\mu_k, P_k)$, which is characterized by a central point $\mu_k \in \text{St}_{n,k}$ and a covariance matrix P_k defined on the tangent space $T_{\mu_k}\text{St}_{n,k}$.

3.2 Bayesian interpretation

The filter operates in a predict-update cycle analogous to the standard EKF:

- Prediction: The predictive distribution is obtained by propagating the previous estimate through the system dynamics. The mean is updated via the noiseless dynamics $\mu_{\text{pred}} = \exp_M(tA)\mu_{k-1}$, and the covariance is inflated to account for process noise, approximated here by the function η which maps ambient isotropic variance to intrinsic manifold variance.
- Update: The update step assimilates the new measurement z_m. This is performed in the tangent space of the predicted mean μ_{pred}. The measurement z_m is mapped to the tangent vector $y = \log_{\mu_{\text{pred}}}(z_m)$. The filter then computes the maximum *a posteriori* (MAP) estimate of the linearized model in the tangent space. This results in a Gaussian posterior on the tangent space $\text{N}(Ky, (1-K)P_{\text{pred}})$, which is then retracted back to the manifold to form the new state estimate $\mu^K = \exp_{\mu_{\text{pred}}}(Ky)$.

3.3 Error criterion

The update step explicitly minimizes the mean squared error (MSE) of the linearized error in the tangent space. Specifically, the Kalman gain K is derived to minimize the trace of the posterior covariance matrix in the tangent space. This corresponds to minimizing the quadratic cost function:

$$J(v) = v^T P_{\text{pred}}^{-1} v + (y-v)^T R^{-1}(y-v), \tag{7}$$

where $v \in T_{\mu_{\text{pred}}}\text{St}_{n,k}$ represents the correction to the predicted mean, and $R = \xi^2\text{id}$ is the measurement noise covariance. By minimizing the cost defined by Equation 7, the filter finds the optimal trade-off between the prior information (prediction) and the new information (measurement) consistent with the local Riemannian metric.

Remark 3.1. The update step in Algorithm 2 is theoretically justified by interpreting the filter as a recursive maximum *a*

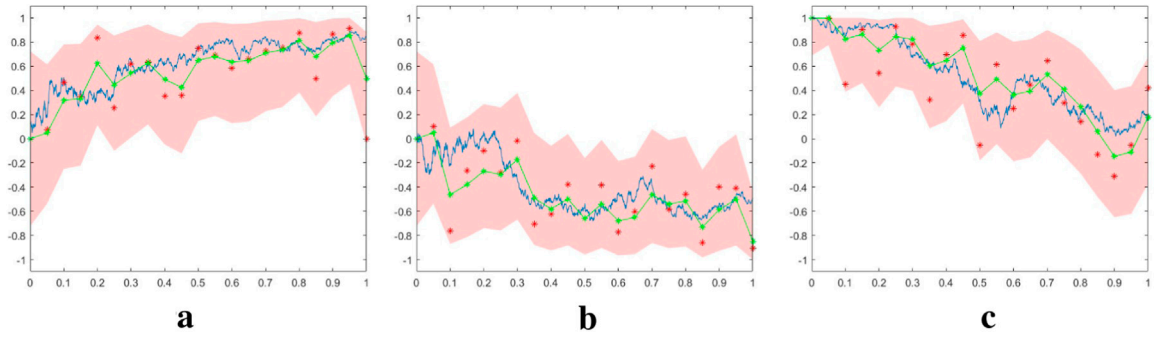

FIGURE 9
Simulated and filtered data over a time span of 1 time unit. In blue the (projected) underlying stochastic process X_t, in red, measurements Z_m, and in green, the mean filtered estimate. The pink filled regions represent a 95%-confidence intervals of the filtered estimate. To run this simulation the parameters A given in Equation 8, $\nu^2 = 1$, and $\xi^2 = 0.1$ were used. The initial distribution is given by $\Sigma_0 = 0.1I_3$ and μ_0 is given by Equation 9. **(a)** The x-coordinate. **(b)** The y-coordinate. **(c)** The z-coordinate.

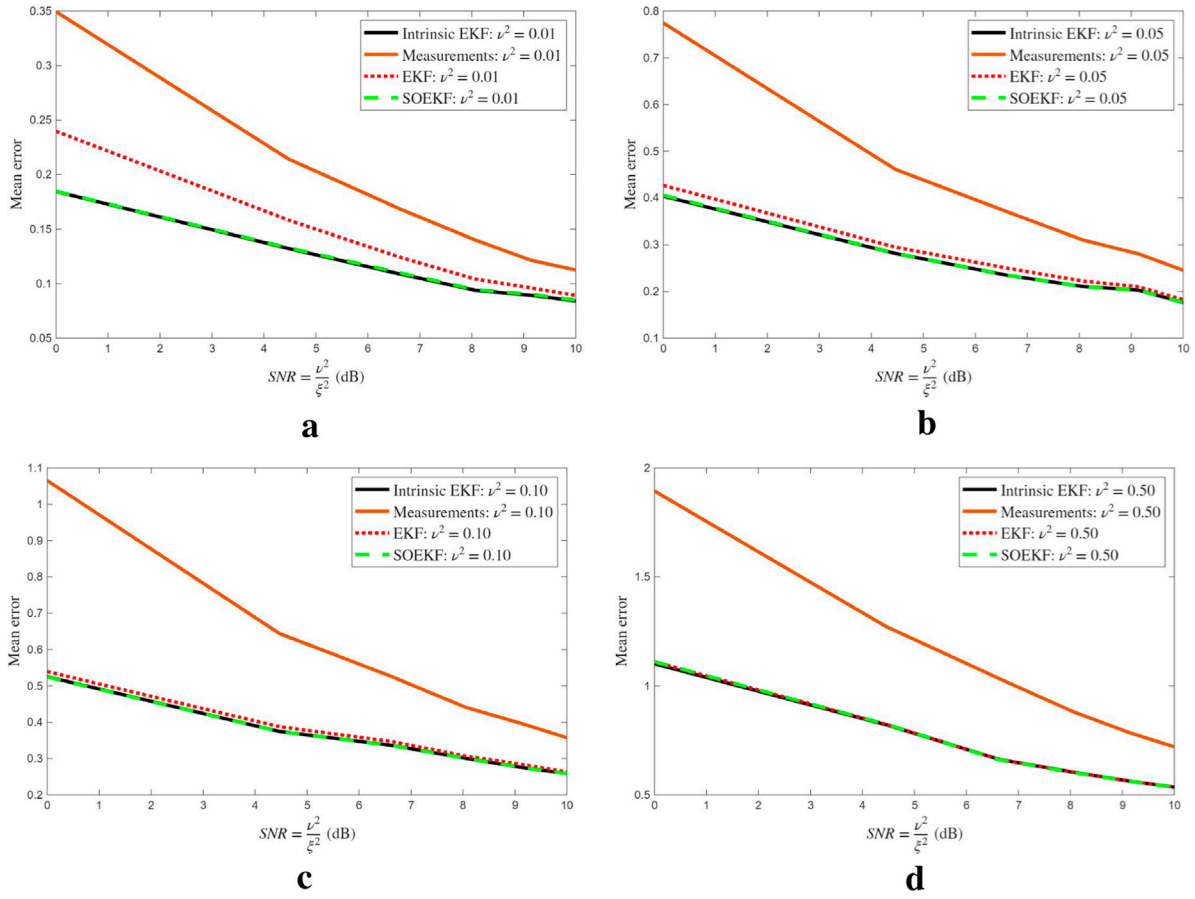

FIGURE 10
For different SNR defined by ν^2/ξ^2 given different values of ν^2 the following is plotted: The average error of the filtered mean is the average distance from μ_m^K to X_{t_m} in blue and the average error of the measurements is the average distance from Z_m to X_{t_m} in red. The errors are produced by averaging over 100 realisations of filtering over 20 measurements for each set of parameters ν, ξ. **(a)** SNR plot for $\nu^2 = 0.01$. **(b)** SNR plot for $\nu^2 = 0.05$. **(c)** SNR plot for $\nu^2 = 0.1$. **(d)** SNR plot for $\nu^2 = 0.5$.

posteriori (MAP) estimator acting on the tangent space $T_{\mu_{\text{pred}}}\text{St}_{n,k}$. By defining a quadratic cost function $J(v)$ that penalizes weighted deviations from both the predicted prior and the measurement innovation, the optimal state correction $v^\star$ is found at the point where the gradient of the log-posterior vanishes. Applying the Matrix Inversion Lemma to the resulting linear system transforms the solution into the standard Kalman gain form $v^\star = Ky$, where $K = \frac{\sigma_0^2 + t\nu^2}{\sigma_0^2 + t\nu^2 + \xi^2}$, confirming that the retraction of

TABLE 1 Tabulated data corresponding to Figure 10.

ν^2	SNR (dB)	0	4.47	6.63	8.06	9.14	10
0.01	Measurements	0.3495	0.2144	0.1683	0.1401	0.1215	0.1125
	Intrinsic EKF	0.1844	0.1323	0.1088	0.0940	0.0893	0.0841
	Extrinsic EKF	0.2398	0.1583	0.1242	0.1043	0.0961	0.0892
	Extrinsic SOEKF	0.1846	0.1328	0.1098	0.0945	0.0902	0.0848
0.05	Measurements	0.7746	0.4607	0.3698	0.3104	0.2806	0.2454
	Intrinsic EKF	0.4036	0.2812	0.2343	0.2107	0.2028	0.1764
	Extrinsic EKF	0.4268	0.2946	0.2496	0.2224	0.2102	0.1822
	Extrinsic SOEKF	0.4053	0.2815	0.2345	0.2108	0.2031	0.1767
0.1	Measurements	1.0653	0.6439	0.5258	0.4409	0.3963	0.3577
	Intrinsic EKF	0.5254	0.3747	0.3365	0.3008	0.2756	0.2590
	Extrinsic EKF	0.5393	0.3878	0.3471	0.3067	0.2839	0.2634
	Extrinsic SOEKF	0.5249	0.3747	0.3364	0.3010	0.2757	0.2590
0.5	Measurements	1.8940	1.2702	1.0338	0.8813	0.7856	0.7208
	Intrinsic EKF	1.1013	0.8211	0.6631	0.6039	0.5632	0.5360
	Extrinsic EKF	1.1101	0.8228	0.6636	0.6045	0.5637	0.5356
	Extrinsic SOEKF	1.1101	0.8228	0.6621	0.6031	0.5629	0.5351

the weighted tangent vector back to the manifold provides the optimal state estimate within the local Gaussian approximation.

Given prior $x_0 \overset{d}{=} \mathrm{N}(\mu_0, \sigma_0^2 \mathrm{id}_{\mathbb{R}^{n\times k}})$ in $\mathbb{R}^{n\times k}$ with $\mu_0 \in \mathrm{St}_{n,k}$ and a measurement $z_1 \in \mathrm{St}_{n,k}$.
$F_t = \exp_{\mathrm{M}}(tA)$
$\mu_{\text{pred.}} = F_t \cdot \mu_0$
$P_{\text{pred.}} = \eta(\sigma_0^2 + t\nu^2)$
$y = \log_{\mu_{\text{pred.}}}(z_1)$
$S = \sigma_0^2 + t\nu^2 + \xi^2$
$K = (\sigma_0^2 + t\nu^2)S^{-1}$
$\mu^K = \exp_{\mu_{\text{pred.}}}(Ky)$
$P^K = (1 - K)P_{\text{pred.}}$
The estimated Projected distribution is now $\mathrm{PrN}(\mu^K, P^K \mathrm{id}_{T_{\mu^K}\mathrm{St}_{n,k}})$ with corresponding Normal distribution $\mathrm{N}(\mu^K, \eta^{-1}(P^K)\mathrm{id}_{\mathbb{R}^{n\times k}})$

Algorithm 2. One step of the extended Kalman filter on a manifold after time t.

4 Simulations on the 2-sphere

Throughout we consider system parameters

$$A = \begin{pmatrix} 0 & 0.263 & 0.036 \\ -0.263 & 0 & -0.653 \\ -0.036 & 0.653 & 0 \end{pmatrix}, \tag{8}$$

$\nu^2 = 1$, and $\xi^2 = 0.1$ with initial covariance $\Sigma_0 = 0.1I_3$ and initial mean

$$\mu_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{9}$$

Now consider stochastic process X_t solving Equation 1. The system process X_t is simulated as follows: Consider constant interval discrete times $0 = t_0, \ldots t_N$ where $\Delta t = t_j - t_{j-1}$ for $1 \le j \le N$. We choose $N = 2000$. The initial point is drawn from $X_0 = \mathrm{N}(\mu_0, \Sigma_0)$. Now each point X_{t_j} is simulated by drawing V_j from $\mathrm{N}(0, \Delta t\nu^2 I_3)$ and computing

$$X_{t_j} = \exp_{\mathrm{M}}(\Delta t A) X_{t_{j-1}} + V_j.$$

The projected system process $\mathrm{pr}(X_t) = \frac{X_t}{\|X_t\|}$ can be seen in Figure 8 in blue.

At even time intervals $\tau_1, \ldots, \tau_L = 1$ with $L = 20$ we simulate measurements of X_{τ_j} as follows: For each τ_j we pseudo-randomly draw W_j from $\mathrm{N}(0, \xi^2 I_3)$ and then compute $Z_{\tau_j} = \mathrm{pr}(\mathrm{pr}(X_{\tau_j}) + W_j)$. These measurements can be seen in Figure 8 in red. The predictions filtered that arise from the application of Algorithm 2 to the measurements Z_{τ_j} are visualized in Figure 8 in black.

To illuminate this simulation into something more easily readable, we have also provided Figure 9a which shows the $x-$coordinate of the same data as in Figure 8 over time but including a 95%-confidence interval around the filtered mean. Similarly, the y-coordinate is shown in Figure 9b and the z-coordinate in Figure 9c.

As can be seen in Figures 8, 9a–c, the filtered mean is on average closer to the underlying process compared to the distance from the measurements to the underlying process. This tendency will be

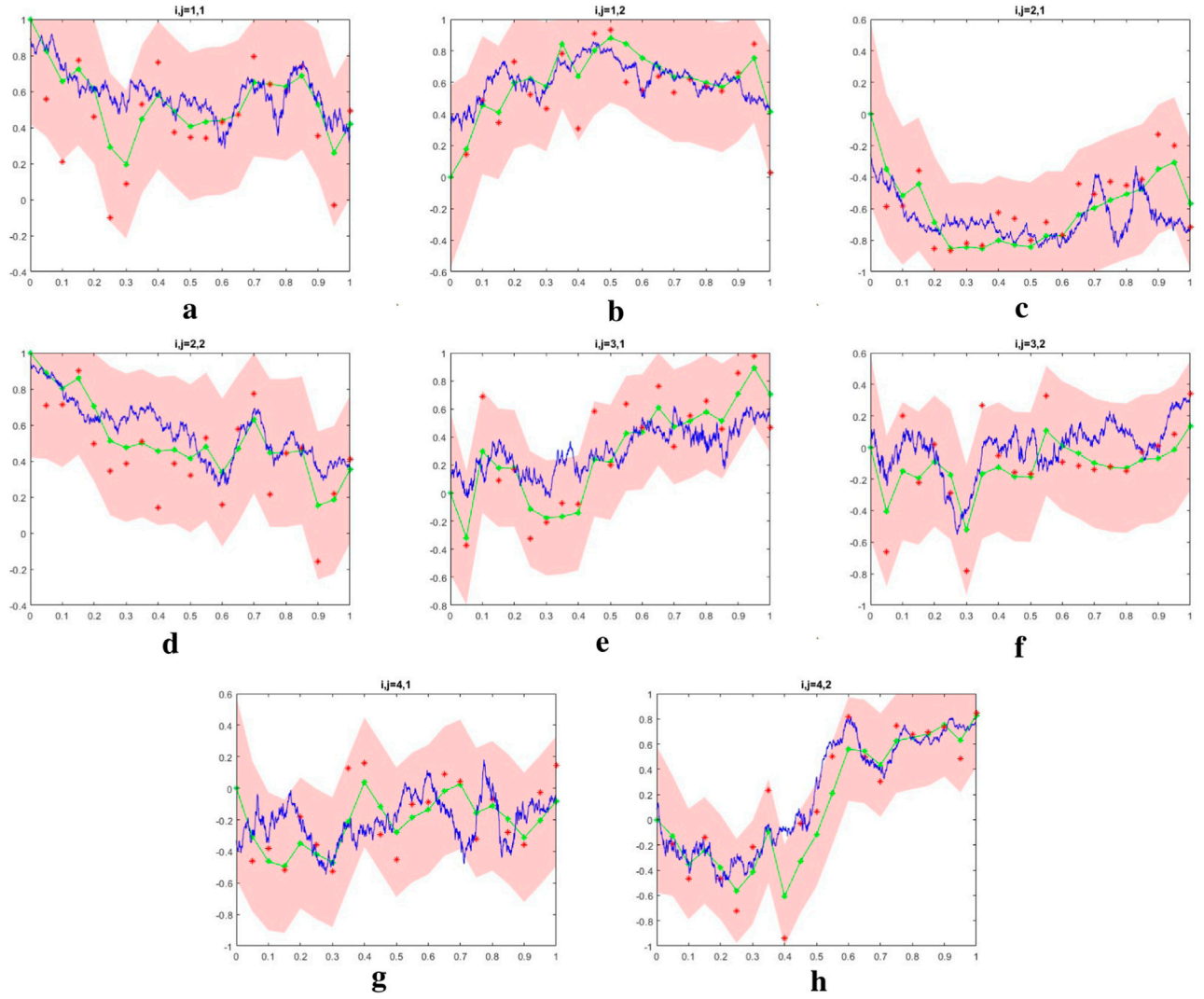

FIGURE 11
Simulated and filtered data over a time span of 1 unit of time. Throughout, the (projected) underlying stochastic process X_t is depicted in blue, the measurements are depicted in red, the filtered averages are depicted in green together with pink filled regions representing an approximated 95%-confidence interval for each filtered mean estimate. The parameter A is given by Equation 10, $\nu^2 = 1$, and $\xi^2 = 0.1$. The initial mean μ_0 is given by Equation 11 and the initial covariance is $\Sigma_0^2 = 0.1\mathrm{id}_{\mathbb{R}^{4\times 2}}$. **(a)** Coordinate X_{11}. **(b)** Coordinate X_{12}. **(c)** Coordinate X_{21}. **(d)** Coordinate X_{22}. **(e)** Coordinate X_{31}. **(f)** Coordinate X_{32}. **(g)** Coordinate X_{41}. **(h)** Coordinate X_{42}.

made clearer in the next set of simulations. We redo the simulation above for $\nu^2 = 0.1, 0.2, 0.5, 1.0$. Then, for each such ν^2 we simulate the filter with 100-fold repetitions for different choices of ξ^2. This is done for

$$\xi^2 = \nu^2, \frac{\nu^2}{2.8}, \frac{\nu^2}{4.6}, \frac{\nu^2}{6.4}, \frac{\nu^2}{8.2}, \frac{\nu^2}{10}.$$

In Figures 10a–d, we plot, for different signal to noise ratios in units decibel, i. e $10\log_{10}(\frac{\nu^2}{\xi^2})$ (dB), the average distance from true point X_t to the measurements, and to the filtered means, respectively. As can be seen for all Figures in Figure 10 there is a significant improvement on the average error of the filtered mean compared to the error of the raw measurements. The proposed algorithm (here referred to as the intrinsic EKF) is compared to a naive extrinsic EKF and a second order EKF (in short SOEKF) implemented analogously to Sorenson and Stubberud (1968). All the values of the mean errors in Figure 10 is also tabulated in Table 1.

5 Simulations on the 4 by 2 Stiefel manifold

Here we consider the filtering problem in Equation 6 on $\mathrm{St}_{4,2}$. We choose system parameters

$$A = \begin{pmatrix} 0 & 0.173 & 0.267 & -0.288 \\ -0.173 & 0 & -0.279 & 0.122 \\ -0.267 & 0.279 & 0 & 0.316 \\ 0.288 & -0.122 & -0.316 & 0 \end{pmatrix} \tag{10}$$

$\nu^2 = 1$ and $\xi^2 = 0.1$, with initial mean

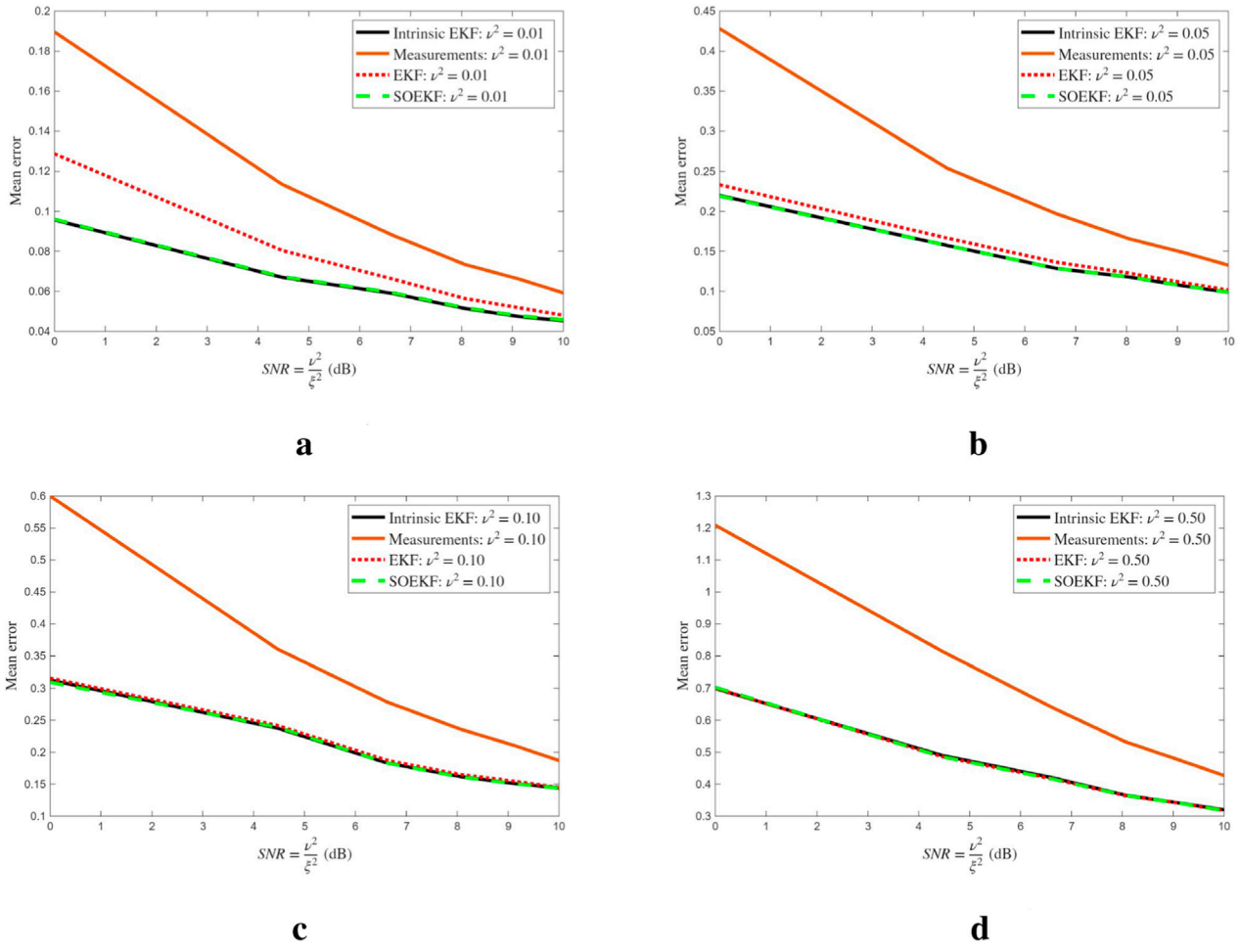

FIGURE 12
For different SNR defined by ν^2/ξ^2 given different values of ν^2 the following is plotted: The average error of the filtered mean is the average distance from μ_m^K to X_{t_m} in blue. The average error of the measurements is the average distance from Z_m to X_{t_m} in red. The errors are produced by averaging over 100 realizations of filtering over 20 measurements for each set of parameters ν, ξ. **(a)** SNR plot for $\nu^2 = 0.01$. **(b)** SNR plot for $\nu^2 = 0.05$. **(c)** SNR plot for $\nu^2 = 0.1$. **(d)** SNR plot for $\nu^2 = 0.5$.

$$\mu_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{11}$$

and initial covariance $\Sigma_0 = 0.1\mathrm{id}_{\mathbb{R}^{4\times 2}}$.

The system process $X_t \in \mathbb{R}^{4,2}$ is done equivalently to how it is done for $\mathbb{S}^2$ in the previous section. Each coordinate of the projected system process $\mathrm{pr}(X_t) = X_t (X_t^T X_t)^{-1/2}$ can be seen in Figure 11 in blue.

Similarly to before, given even time intervals $\tau_1, \ldots, \tau_L = 1$ we simulate measurements of X_{τ_j} as follows: For each τ_j we pseudo-randomly draw W_j from $\mathrm{N}(0.\xi^2 \mathrm{id}_{\mathbb{R}^{4\times 2}})$ then we compute $Z_{\tau_j} = \mathrm{pr}(\mathrm{pr}(X_{\tau_j}) + W_j)$, here $L = 20$. These measurements can be seen in red in Figure 11. By directly applying Algorithm 2 with the measurements Z_{τ_j} the filtered predictions is visualized in green in Figure 11. Note that here we roughly give 95-% confidence bounds as intervals of the form $[\mu^K - 1.96\sqrt{P^K}, \mu^K - 1.96\sqrt{P^K}]$.

As can be seen in Figure 11, on average, the filtered mean is closer to the underlying process compared to the measurements. Again, we shall verify this in the following simulations. We once more define SNR as $10\log_{10}(\frac{\nu^2}{\xi^2})$ (dB). In Figure 12 one can see the average error of the filtered mean and the average error of the measurements. This is done with the same choices of ν and ξ as was done for $\mathbb{S}^2$ in the previous section. As can be seen in Figures 12a–d, the error of the filtered mean is significantly smaller compared to the error of the measurements. Again, the proposed algorithm (the intrinsic EKF) is compared to a naive extrinsic EKF and a second order EKF (SOEKF). All the values of the mean errors in Figure 12 is also tabulated in Table 2.

6 Application to a scenario of spacecraft attitude determination

Here we shall give a physically grounded simulation of the orientation of a spacecraft in orbit around the Earth with inspiration following Markley (1998). For example, given the direction of a star, $x_1 \in \mathbb{R}^3$ and the (linearly independent) direction of Earth's magnetic field $x_2 \in \mathbb{R}^3$. It is possible to determine the attitude (meaning orientation) of a spacecraft relative to a fixed coordinate system on

TABLE 2 Tabulated data corresponding to Figure 12.

ν^2	SNR (dB)	0	4.47	6.63	8.06	9.14	10
0.01	Measurements	0.1896	0.1135	0.0883	0.0734	0.0660	0.0592
	Intrinsic EKF	0.0957	0.0670	0.0590	0.0514	0.0473	0.0453
	Extrinsic EKF	0.1287	0.0804	0.0663	0.0564	0.0517	0.0480
	Extrinsic SOEKF	0.0959	0.0673	0.0593	0.0517	0.0476	0.0457
0.05	Measurements	0.4282	0.2538	0.1967	0.1654	0.1482	0.1325
	Intrinsic EKF	0.2197	0.1573	0.1285	0.1177	0.1066	0.0988
	Extrinsic EKF	0.2331	0.1665	0.1363	0.1226	0.1104	0.1015
	Extrinsic SOEKF	0.2191	0.1573	0.1285	0.1179	0.1067	0.0988
0.1	Measurements	0.6000	0.3609	0.2783	0.2361	0.2101	0.1871
	Intrinsic EKF	0.3129	0.2376	0.1832	0.1619	0.1509	0.1441
	Extrinsic EKF	0.3157	0.2419	0.1872	0.1651	0.1545	0.1448
	Extrinsic SOEKF	0.3097	0.2389	0.1832	0.1615	0.1508	0.1440
0.5	Measurements	1.2088	0.8141	0.6395	0.5325	0.4747	0.4271
	Intrinsic EKF	0.6990	0.4895	0.4203	0.3657	0.3409	0.3200
	Extrinsic EKF	0.7000	0.4855	0.4170	0.3644	0.3407	0.3174
	Extrinsic SOEKF	0.7025	0.4856	0.4167	0.3650	0.3405	0.3183

TABLE 3 The mean error for the Intrinsic EKF given by Algorithm 2, the extrinsic second order extended Kalman filter (SOEKF) and measurements for the attitude determination of a spacecraft.

Method	Intrinsic EKF	SOEKF	Measurements
Mean error	0.0040	0.0040	0.040

Earth. For technical reasons, it is only the orthogonal 2 directions spanned by x_1, x_2 that is necessary in order to determine the attitude of the spacecraft. Therefore, the spacecraft performs measurements that are of the type

$$\mathrm{pr}\begin{pmatrix} x_1 & x_2 \end{pmatrix} \in \mathrm{St}_{3,2}.$$

Moreover, as the spacecraft is in orbit of the Earth, the coordinate system is continuously rotated relative to a reference frame on Earth. Therefore, since x_1 and x_2 are fixed points in the fixed reference frame, the observed point $\mathrm{pr}(x_1, X_2) \in \mathrm{St}_{3,2}$ is being continuously rotated by that same rotational motion. This rotational motion is exactly the 1-parameter Lie group generated by the system parameter A in Equation 1. Measurements are performed at the frequency of once every 10 s. If the orbital time for the spacecraft is at about 90 min, and if the unit of time for t is in seconds, we then consider

$$A = \begin{pmatrix} 0 & 0.0012 & 0 \\ -0.0012 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and initial mean

$$\mu_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We consider uncertainty quantities $\nu^2 = 10^{-8}$ and $\xi^2 = 10^{-3}$, and initial uncertainty $\Sigma_0 = 10^{-5}\mathrm{id}_{\mathbb{R}^{(3\times 2)}}$. Running Algorithm 2 for this scenario over 100 repetitions gives a mean error similar to the extrinsic second order Kalman filter, see Table 3.

7 Discussion

As can be seen in Figures 10 and 12 depending on the situation, one can expect at optimal conditions about a 60% filtering gain. Note, however, that if ν is too large, then the projected distribution will be close to the uniform distribution. In this case the extended Kalman filtering algorithm is not feasible as the predictive power of the dynamics becomes unreliable. This is also true for the error of the measurements, if the error of the measurements are large. Indeed, then there is little information gained from the measurements, naturally.

Moreover, the error of the proposed filtering algorithm seem to be equal in precision to the second order as can be seen in Figures 10 and 12. The proposed algorithm has the upside that no second order derivative needs to be computed, which can be costly to compute as the dimensions go up. Also, for large noise there seem to be little gain of using the proposed algorithm instead of the more simple classical EKF. This is expected, as with large noise the observations does not capture the curvature of the underlying manifold but rather seem like observations from the ambient space (consider, for example $\mathbb{S}^3$ where large noise throws observations around the ball).

Some inaccuracies might originate from the logarithm map used in step 5 of Algorithm 2. The logarithm map is only defined for points close to the base point, therefore one can not guarantee that the algorithm is reliable if the measurements appear outside what is

called the injectivity radius (the radius which guarantees that the logarithm map is well defined). It is still an open question what the injectivity radius is for the Stiefel manifold. However, very recently there has been some advances in the following direction: there is a lower bound $\sqrt{\frac{4}{5}}\pi$ for the injectivity radius of the Stiefel manifold (under the canonical metric), see Stoye and Zimmermann (2024). Therefore, as long as points are never farther away from each other than $\sqrt{\frac{4}{5}}\pi$, the injectivity of the logarithm map is not an issue.

One obvious limitation with Algorithm 2 is that it requires quite restrictive assumptions on the noise, i.e. we require that the noise is isotropic. Indeed, already on $\mathbb{S}^2$, if we allow for general covariance matrices Σ then the statement of Conjecture 2.4 is rarely true. Thus, if the filtering problem is set up so that future points may not allow for one-to-one correspondence between the covariance of a normal random variable and it's projected counterpart, our proposed filter is useless. In the case of the sphere, as long as the covariance matrix Σ of the ambient normal random variable has the mean as eigenvector, then Algorithm 2 may still be tractable.

One more subtle issue with Algorithm 2 is that the update steps are only accurate as long as the actual process X_t does not significantly drift away from it's starting point in the radial direction. By choosing A to be anti-symmetric, $\exp_{\mathrm{M}}(tA) \in SO(n)$, then it is always true that $\exp_{\mathrm{M}}(tA)\mu_0 \in St_{n,k}$ if $\mu_0 \in \mathrm{St}_{n,k}$. Therefore on average, the process X_t does not drift away along a radial direction. However, that does not mean that it's realization does not drift away. In order to bootstrap the filtering process, one way periodically measure the process noise term $\eta(\nu^2)$, of $\mathrm{pr}(X_t)$, since $\eta(\nu^2)$ also depends on the symmetric part S of the polar decomposition $X_t = \mathrm{pr}(X_t)S$.

Finally let us mention some ideas for future improvements. Firstly, observe that the true distribution of $p_{X_{t_\ell}|Z_\ell}$ does not follow the projected normal distribution (though it is close in the L^1-norm) but it is more complicated. Using the true distribution of $p_{X_{t_\ell}|Z_\ell}$ would allow one to adapt the function η to the true distribution that should give performance, although such η might be difficult to attain. Secondly, studying other than Stiefel manifolds would be an interesting direction of future research. For example, one could consider other homogeneous spaces including Grassmann manifolds or more exotic ones such as higher dimensional tori. The latter appears as a model for collection of angles in a protein chain or robotic motion.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

J-LF: Writing – original draft, Writing – review and editing, Formal Analysis, Methodology, Supervision. AP: Formal Analysis, Methodology, Writing – original draft, Writing – review and editing, Conceptualization, Investigation, Software. LV: Formal Analysis, Methodology, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This project has been partially funded by Huawei technologies. The funder was not involved in the study design, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. During the preparation of this manuscript, the author used Google Gemini (version 3 Pro, January 2026) to assist with literature summarization and text coherence. The author reviewed and edited the output and takes full responsibility for the content.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Astolfi, A., and Lovera, M. (2002). "Global spacecraft attitude control using magnetic actuators," in Proceedings of the 2002 American Control Conference (IEEE), 1331–1335.

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., and van der Vorst, H. (2000). *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Philadelphia: Society for Industrial and Applied Mathematics.

Bansal, N., Chen, X., and Wang, Z. (2018). Can we gain more from orthogonality regularizations in training deep cnns? *Adv. Neural Inf. Process. Syst.* 31, 4266–4276.

Barrau, A., and Bonnabel, S. (2016). The invariant extended kalman filter as a stable observer. *IEEE Trans. Automatic Control* 62, 1797–1812. doi:10.1109/tac.2016.2594085

Barrau, A., and Bonnabel, S. (2018). Invariant kalman filtering. *Annu. Rev. Control, Robotics, Aut. Syst.* 1, 237–257. doi:10.1146/annurev-control-060117-105010

Bloch, A. M., and Marsden, J. E. (1990). Stabilization of rigid body dynamics by the energy-casimir method. *Syst. Control Lett.* 14, 341–346. doi:10.1016/0167-6911(90)90055-y

Boulanger, J., Said, S., Le Bihan, N., and Manton, J. H. (2016). Filtering from observations on stiefel manifolds. *Signal Process.* 122, 52–64. doi:10.1016/j.sigpro.2015.11.012

Brossard, M., Barrau, A., Chauchat, P., and Bonnabel, S. (2021). Associating uncertainty to extended poses for on lie group imu preintegration with rotating earth. *IEEE Trans. Robotics* 38, 998–1015. doi:10.1109/tro.2021.3100156

Chahbazian, C., Merlinge, N., Dahia, K., Winter-Bonnet, B., Blanc, A., and Musso, C. (2022). "Generalized laplace particle filter on lie groups applied to ambiguous Doppler navigation," in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 2387–2394.

Chepuri, S. P., Geus, G., and Van der Veen, A. J. (2014). Rigid body localization using sensor networks. *IEEE Trans. Signal Process.* 62, 4911–4924. doi:10.1109/tsp.2014.2336621

Craig, B. M., and Oppe, M. (2010). From a different angle: a novel approach to health valuation. *Soc. Sci. Med.* 70, 169–174. doi:10.1016/j.socscimed.2009.10.009

de Figueredo, C. G., Bordin, Jr C. J., and Bruno, M. G. (2022). Nonlinear distributed state estimation on the stiefel manifold using diffusion particle filters. *Digit. Signal Process.* 122, 103354. doi:10.1016/j.dsp.2021.103354

Demir, Y., and Bilgin, O. C. (2019). Application of circular statistics to life science data. *Med. Sci. Discov.* 6, 63–72. doi:10.17546/msd.507582

Edelman, A., Arias, T. A., and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM Journal Matrix Anal. Appl.* 20, 303–353. doi:10.1137/s0895479895290954

Figueras, J. L., Persson, A., and Viitasaari, L. (2025a). On the convergence of the extended kalman filter on stiefel manifolds when observing a constant particle with measurement errors. *arXiv Preprint arXiv:2511.02680.* doi:10.48550/arXiv.2511.02680

Figueras, J. L., Persson, A., and Viitasaari, L. (2025b). On parameter estimation for $N(\mu, \sigma^2 I_3)$ based on projected data into $\mathbb{S}^{2}$. *Mod. Stoch. Theory Appl.* 12, 407–431. doi:10.15559/25-VMSTA279

Fiori, S. (2001). A theory for learning by weight flow on stiefel-grassman manifold. *Neural Comput.* 13, 1625–1647. doi:10.1162/089976601750265036

Fiori, S. (2008). Geodesic-based and projection-based neural blind deconvolution algorithms. *Signal Process.* 88, 521–538. doi:10.1016/j.sigpro.2007.08.014

Hussien, M. T., Seddik, K. G., Gohary, R. H., Shaqfeh, M., Alnuweiri, H., and Yanikomeroglu, H. (2014). "Multi-resolution broadcasting over the grassmann and stiefel manifolds," in 2014 IEEE International Symposium on Information Theory (IEEE), 1907–1911.

Kálmán, R. E. (1960a). A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45. doi:10.1115/1.3662552

Kálmán, R. E. (1960b). Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana* 5, 102–119.

Kálmán, R. E., and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *J. Basic Eng.* 83, 95–108. doi:10.1115/1.3658902

Karaibrahimoglu, A., Ayhan, S., Karaagac, M., and Artac, M. (2021). Circular analyses of dates on patients with gastric carcinoma. *J. Appl. Statistics* 48, 2931–2943. doi:10.1080/02664763.2021.1977259

Lang, M., and Feiten, W. (2012). "MPG-fast forward reasoning on 6 dof pose uncertainty," in ROBOTIK 2012; 7th German Conference on Robotics (VDE), 1–6.

Lenglet, C., Rousson, M., Deriche, R., and Faugeras, O. (2006). Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor mri processing. *J. Math. Imaging Vis.* 25, 423–444. doi:10.1007/s10851-006-6897-z

Li, K., Frisch, D., Radtke, S., Noack, B., and Hanebeck, U. D. (2018). "Wavefront orientation estimation based on progressive Bingham filtering," in 2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF), 1–6.

Liu, X., Srivastava, A., and Gallivan, K. (2004). Optimal linear representations of images for object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 662–666. doi:10.1109/TPAMI.2004.1273986

Lui, Y. M. (2012). Tangent bundles on special manifolds for action recognition. *IEEE Trans. Circuits Syst. Video Technol.* 22, 930–942. doi:10.1109/tcsvt.2011.2181452

Mahony, R., Van Goor, P., and Hamel, T. (2022). Observer design for nonlinear systems with equivariance. *Annu. Rev. Control, Robotics, Aut. Syst.* 5, 221–252. doi:10.1146/annurev-control-061520-010324

Markley, F. L. (1998). Attitude determination using two vector measurements. *Tech. Rep. NASA*.

Murangira, A., Musso, C., and Dahia, K. (2016). A mixture regularized rao-blackwellized particle filter for terrain positioning. *IEEE Trans. Aerosp. Electron. Syst.* 52, 1967–1985. doi:10.1109/taes.2016.150089

Nuñez-Antonio, G., Gutiérrez-Peña, E., and Escarela, G. (2011). A Baye-sian regression model for circular data based on the projected normal distribution. *Stat. Model.* 11, 185–201. doi:10.1177/1471082x1001100301

Nuñez-Antonio, G., Ausín, M. C., and Wiper, M. P. (2015). Bayesian nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions. *J. Agric. Biol. Environ. Statistics* 20, 47–64. doi:10.1007/s13253-014-0193-y

Park, W., Liu, Y., Zhou, Y., Moses, W., and Chirikjian, G. S. (2008). Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map. *Robotica* 26, 419–434. doi:10.1017/S0263574708004475

Pitaval, R. A., and Tirkkonen, O. (2013). Joint Grassmann-Stiefel quantization for MIMO product codebooks. *IEEE Transactions Wirel. Commun.* 13, 210–222. doi:10.1109/twc.2013.111313.130208

Schwarz, S., and Rupp, M. (2014). Predictive quantization on the stiefel manifold. *IEEE Trans. Signal Process.* 62, 6459–6471. doi:10.1109/LSP.2014.2354258

Seddik, K. G., Gohary, R. H., Hussien, M. T., Shaqfeh, M., Alnuweiri, H., and Yanikomeroglu, H. (2017). Multi-resolution multicasting over the grassmann and stiefel manifolds. *IEEE Trans. Wirel. Commun.* 16, 5296–5310. doi:10.1109/twc.2017.2707549

Seifert, L., Orth, D., Boulanger, J., Dovgalecs, V., Hérault, R., and Davids, K. (2014). Climbing skill and complexity of climbing wall design: assessment of jerk as a novel indicator of performance fluency. *J. Appl. Biomechanics* 30, 619–625. doi:10.1123/jab.2014-0052

Snoussi, H., and Mohammad-Djafari, A. (2006). Particle filtering on riemannian manifolds. *AIP Conf. Proc.* 872, 219–226. doi:10.1063/1.2423278

Sorenson, H., and Stubberud, A. (1968). Recursive filtering for systems with small but non-negligible non-linearities. *Int. J. Control* 7, 271–280. doi:10.1080/00207176808905606

Stoye, J., and Zimmermann, R. (2024). On the injectivity radius of the Stiefel manifold: Numerical investigations and an explicit construction of a cut point at short distance. *arXiv Preprint arXiv:2403.03782.* doi:10.48550/arXiv.2403.03782

Sveier, A., and Egeland, O. (2018). Pose estimation using dual quaternions and moving horizon estimation. *IFAC-PapersOnLine* 51, 186–191. doi:10.1016/j.ifacol.2018.07.275

Turaga, P., Veeraraghavan, A., Srivastava, A., and Chellappa, R. (2011). Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2273–2286. doi:10.1109/TPAMI.2011.52

Wang, J., Chen, Y., Chakraborty, R., and Yu, S. X. (2020). "Orthogonal convolutional neural networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11505–11515.

Zimmermann, R. (2017). A matrix-algebraic algorithm for the riemannian logarithm on the Stiefel manifold under the canonical metric. *SIAM J. Matrix Anal. Appl.* 38, 322–342. doi:10.1137/16m1074485