DISTRIBUTED PARTICLE FILTERS FOR STATE TRACKING ON THE STIEFEL MANIFOLD USING TANGENT SPACE STATISTICS

Claudio J. Bordin Jr.[⋆] *Caio G. de Figueredo*[†] *Marcelo G. S. Bruno*[#]

[⋆]Universidade Federal do ABC, Santo André, 09210-580, Brazil
[†] Escola Naval, Rio de Janeiro, 20021-010, Brazil
[#]Instituto Tecnológico de Aeronáutica, São José dos Campos, 12228-900, Brazil
claudio.bordin@ufabc.edu.br,caiofigueredo@gmail.com,bruno@ita.br

ABSTRACT

This paper introduces a novel distributed diffusion algorithm for tracking the state of a dynamic system that evolves on the Stiefel manifold. To compress information exchanged between nodes, the algorithm builds a Gaussian parametric approximation to the particles that are previously projected onto the tangent space to the Stiefel manifold and mapped to real vectors. Observations from neighboring nodes are then assimilated for a general nonlinear observation model. Performance results are compared to those of competing linear diffusion Extended Kalman Filters and other particle filters.

Index Terms— Diffusion methods, Distributed particle filters, Stiefel manifold

1. INTRODUCTION

Several problems in engineering require the estimation of variables with built-in, nonlinear orthogonality constraints, which lead those variables to be naturally modeled as points on the Stiefel manifold $\mathcal{V}^{k,m}$, i.e., the set of $k \times m$, $k > m$, real matrices with orthonormal columns. Applications can be found in optimization [1], signal processing [2], robotics [3], image processing [4], attitude estimation and control [5], and machine learning [6]. Although signal processing methods developed for unconstrained variables can be used in such cases, they are suboptimal in those scenarios and the introduction of more principled approaches that properly consider the geometry of the state space may lead to significant performance gains. Moreover, in many applications, separate remote receivers often cooperate with each other via message exchanges over a partially connected wireless communication network to execute the state estimation task in a distributed fashion [7–10]. Despite the recent increase in interest in distributed filtering [11–13], most of the proposed algorithms found in previous literature apply only to the case in which both observations and hidden variables lie on Euclidean spaces.

This work was supported by grant #2018/26191-0, São Paulo Research Foundation (FAPESP).

In a previous work [14], we introduced fully distributed particle filter (PF) algorithms for cooperative estimation of hidden variables that are constrained to the Stiefel manifold. The algorithms in [14] employed both the Random Exchange (RndEx) [15] and the Adapt-then-Combine (ATC) [11], [16] diffusion techniques and used methods from directional statistics to represent probability density functions (p.d.f.'s) that are transmitted between nodes over the network by von Mises-Fisher (vMF) models.

In this paper, we follow a different approach stemming from Riemannian geometry and build instead Gaussian parametric approximations on tangent spaces to the Stiefel manifold to represent the p.d.f.'s that are transmitted over the network; for simplicity, only RndEx PFs are considered in this work. The novel algorithm in this paper extends to general Stiefel manifolds the RndEx diffusion PF developed in [17] for states that are constrained to the unit hypersphere, which is a particular instance of the Stiefel manifold. The difficulty in this case is to find a suitable parametrization taking into consideration the geometry of the tangent space for general Stiefel manifolds, which is more complex than in the particular case of the unit hypersphere. The proposed RndEx PF exchanges information between nodes in compressed form taking advantage of the aforementioned parametrization. The state estimates at each network node, on the other hand, are determined via the averaging algorithm introduced in [18], which approximates the Karcher mean of the particle set using orthographic retraction and lifting maps that are also based on the geometry of the tangent space to the manifold. The Karcher mean, in turn, generalizes to the manifold the notion of minimum mean square error estimates on Euclidean spaces.

Experimental results show that, by designing filters which explicitly take into account that the hidden state matrices evolve on the Stiefel manifold, we can achieve significant tracking performance gains compared to previous algorithms that ignored such constraints and tracked the states instead as unstructured real matrices, specifically, the ATC [11] and RndEx [19] extended Kalman filters, the Euclidean RndEx

PF of [15] and the Likelihood Consensus PF of [20]. Furthermore, the results also show the advantage of cooperative tracking algorithms based on multiple receivers compared to previous solutions that assumed states on the Stiefel manifold, but used a single, noncooperative PF to perform state tracking, e.g., [2].

The remainder of the paper is organized as follows: the signal model is described in Sec. 2. In Sec. 3, we describe a method for representation of points on tangent spaces to the Stiefel manifold via real variables. The RndEx algorithm is presented in Sec. 4. The results of numerical simulations are shown in Sec. 5. Finally, our conclusions are left to Sec. 6.

2. SIGNAL MODEL

Let $\boldsymbol{S}_n \in \mathcal{V}^{k,m}$ denote an unknown state matrix[1] at time instant $n \geq 0$ that evolves in time according to a vMF random walk [2] model on $\mathcal{V}^{k,m}$, i.e.,

$$\boldsymbol{S}_n|\boldsymbol{S}_{n-1} \sim \mathrm{vMF}(\boldsymbol{S}_n|\kappa\boldsymbol{S}_{n-1}),\ n > 0, \tag{1}$$

with $p(\boldsymbol{S}_0) \propto 1$, where $\kappa \in \mathbb{R}^+$ denotes a known hyperparameter and $\mathrm{vMF}(\cdot)$ denotes a matrix vMF p.d.f., defined as [21]

$$\mathrm{vMF}(\boldsymbol{S}|\boldsymbol{F}) \triangleq \left[{}_0F_1\left(\frac{k}{2}, \frac{1}{4}\boldsymbol{F}^T\boldsymbol{F}\right)\right]^{-1} \mathrm{etr}\left(\boldsymbol{F}^T\boldsymbol{S}\right). \tag{2}$$

In (2), $\mathrm{etr}(\cdot)$ stands for the exponential of the trace of a square matrix and ${}_0F_1(\cdot)$ is the hypergeometric function with matrix argument [21].

Multiple nodes dispersed over a network and corresponding to the physical locations of each agent record at each instant n the observations

$$\boldsymbol{Y}_{n,r} = \boldsymbol{G}_r(\boldsymbol{H}_{n,r}\boldsymbol{S}_n) + \boldsymbol{V}_{n,r}, \quad n > 0, \tag{3}$$

where the index $r \in \{1, ..., R\}$ denotes the r-th node in the network, $\boldsymbol{G}_r(\cdot) : \mathbb{R}^{k\times m} \mapsto \mathbb{R}^{k\times m}$ is a generally nonlinear function, $\boldsymbol{H}_{n,r}$ is an $k \times k$ known regressor matrix, and $\{\boldsymbol{V}_{n,r}\}$ is a sequence of zero mean, independent Gaussian random matrices such that $\boldsymbol{V}_{n,r} \sim \boldsymbol{N}_{k,m}(\boldsymbol{V}_{n,r}|\boldsymbol{0}_{k,m}, \boldsymbol{\Omega}_r, \boldsymbol{\Gamma}_r)$, where $\boldsymbol{0}_{k,m}$ denotes a $k \times m$ matrix with null entries, $\{\boldsymbol{\Omega}_r, \boldsymbol{\Gamma}_r\}$ are (positive-definite) covariance matrices, and

$$\boldsymbol{N}_{k,m}(\boldsymbol{V}|\bar{\boldsymbol{V}}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}) \triangleq \frac{\mathrm{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}^{-1}(\boldsymbol{V}-\bar{\boldsymbol{V}})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{V}-\bar{\boldsymbol{V}})\right)}{(2\pi)^{km/2}|\boldsymbol{\Omega}|^{k/2}|\boldsymbol{\Gamma}|^{m/2}} \tag{4}$$

is the matrix Gaussian p.d.f. [2], such that $\boldsymbol{\Gamma} \in \mathbb{R}^{k\times k}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{m\times m}$ are positive-definite matrices and $|\cdot|$ denotes the determinant of a matrix.

[1] We use normal type letters to denote scalars, bold lowercase letters to denote vectors, and bold uppercase letters to denote matrices. For simplicity, the same notation is used for random variables and real variables with the distinction implied in context. We also denote by $p(\cdot)$ the p.d.f. of a random variable or vector.

3. REPRESENTATION ON THE TANGENT SPACE

For a smooth manifold, one can define retraction and lifting maps [18]. Retraction maps mimic the exponential map [1], without exhibiting, however, the same distance preserving properties. Lifting maps, in turn, mimic the logarithmic map [1], inverting their corresponding retraction maps. Denoting by $T_{\boldsymbol{M}}\mathcal{V}^{k,m}$ the tangent space to the manifold $\mathcal{V}^{k,m}$ at point $\boldsymbol{M}$, the retraction and lifting maps can be viewed respectively as a simplified projection from $T_{\boldsymbol{M}}\mathcal{V}^{k,m}$ onto $\mathcal{V}^{k,m}$, and its inverse, but their computation is far less expensive and well-posed than that for the exponential and logarithmic maps.

Let $\boldsymbol{S}$ and $\boldsymbol{M}$ be two sufficiently near points of $\mathcal{V}^{k,m}$ such that $\boldsymbol{X} = \mathcal{M}_{\boldsymbol{M}}^{-1}(\boldsymbol{S})$, the lifting map of $\boldsymbol{S}$ at point $\boldsymbol{M}$, is well-defined. Then, $\boldsymbol{X} \in T_{\boldsymbol{M}}\mathcal{V}^{k,m}$ must satisfy $\boldsymbol{M}^T\boldsymbol{X} + \boldsymbol{X}^T\boldsymbol{M} = \boldsymbol{0}_{m,m}$ [18]. As a consequence, $\boldsymbol{X}$ can be parametrized as $\boldsymbol{X} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{M}^{\perp}\boldsymbol{B}$ [1], where $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ is a skew-symmetric matrix, i.e., $\boldsymbol{A}^T = -\boldsymbol{A}$, $\boldsymbol{B} \in \mathbb{R}^{(k-m)\times m}$ is an unstructured matrix, and $\boldsymbol{M}^{\perp} \in \mathbb{R}^{k\times(k-m)}$ is a matrix with orthonormal columns that spans the orthogonal complement of the range of $\boldsymbol{M}$, such that $\boldsymbol{M}^T\boldsymbol{M}^{\perp} = \boldsymbol{0}_{m\times k-m}$, and $(\boldsymbol{M}^{\perp})^T\boldsymbol{M}^{\perp} = \boldsymbol{I}_{(k-m)\times(k-m)}$. In fact, it suffices to see that

$$\begin{aligned}\boldsymbol{X}^T\boldsymbol{M} + \boldsymbol{M}^T\boldsymbol{X} = \boldsymbol{A}^T\underbrace{\boldsymbol{M}^T\boldsymbol{M}}_{\boldsymbol{I}_m} + \boldsymbol{B}^T\underbrace{(\boldsymbol{M}^{\perp})^T\boldsymbol{M}}_{\boldsymbol{0}_{(k-m)\times m}} \\ + \underbrace{\boldsymbol{M}^T\boldsymbol{M}}_{\boldsymbol{I}_m}\boldsymbol{A} + \underbrace{\boldsymbol{M}^T\boldsymbol{M}^{\perp}}_{\boldsymbol{0}_{m\times(k-m)}}\boldsymbol{B} = -\boldsymbol{A} + \boldsymbol{A} = \boldsymbol{0}_{m\times m}\end{aligned} \tag{5}$$

as $\boldsymbol{A}$ is skew-symmetric and $\boldsymbol{M} \in \mathcal{V}^{k\times m}$.

Note that $\boldsymbol{A}$ and $\boldsymbol{B}$ have, respectively, $m(m-1)/2$ and $(k-m)m$ free entries; such entries can be collected in an arbitrary order in a real, unconstrained vector $\boldsymbol{\mathfrak{x}}$ of dimension $km - \frac{m}{2} - \frac{m^2}{2}$. Exploiting the facts that $\boldsymbol{A} = \boldsymbol{M}^T\boldsymbol{X}$ and $\boldsymbol{B} = (\boldsymbol{M}^{\perp})^T\boldsymbol{X}$, we can establish a map $(\boldsymbol{S}, \boldsymbol{M}) \mapsto \boldsymbol{\mathfrak{x}}$, which is useful for performing statistical approximations as required in Sec. 4. Finally, note that the parametrization method described in this section shares some resemblance, but is substantially different from that introduced in [22].

4. RNDEX DIFFUSION PARTICLE FILTER ON THE STIEFEL MANIFOLD

At time instant $n-1$, assume that node l has a weighted particle set $\{w_{n-1,l}^{(q)}, \boldsymbol{S}_{n-1,l}^{(q)}\}$ that represents the posterior p.d.f. $p(\boldsymbol{S}_{n-1}|\tilde{\boldsymbol{Y}}_{1:n-1,l})$, where $\tilde{\boldsymbol{Y}}_{1:n-1,l}$ denotes all measurements that have been assimilated by node l up to instant $n-1$. The RndEx diffusion method, as explained in a different context in [15], is divided into two steps, known respectively as the Random Exchange and the Data Assimilation steps, which are further explained in the sequel.

RndEx Step In the RndEx Step, each node l transmits its posterior p.d.f. $p(\boldsymbol{S}_{n-1}|\tilde{\boldsymbol{Y}}_{1:n-1,l})$ to another random node r in the network. That is done by a sequence of p.d.f. exchanges with randomly chosen neighboring nodes using a protocol that is described in detail in [15] and not discussed here for conciseness. The nature of the proposed random exchange protocol, however, is such that, at the end of the protocol, the node r to which $p(\boldsymbol{S}_{n-1}|\tilde{\boldsymbol{Y}}_{1:n-1,l})$ has been moved is not necessarily in the immediate neighborhood of l.

To reduce the internode communication cost, node l, after randomly choosing another neighboring node to perform an exchange, sends it a parametric representation of its posterior. In this paper, we propose to use a Gaussian representation on the tangent plane based on the mapping discussed in Sec. 3. Initially, given the weighted particle set $\{w_{n-1,l}^{(q)}, \boldsymbol{S}_{n-1,l}^{(q)}\}$ we compute its centroid $\hat{\boldsymbol{S}}_{n-1,r}$ using the method introduced in [18]:

$$\hat{\boldsymbol{S}}_{n-1,l}^{<i+1>} = \mathcal{M}_{\hat{\boldsymbol{S}}_{n-1,l}^{<i>}}\left[\sum_{q=1}^{Q} w_{n-1,l}^{(q)}\, \mathcal{M}_{\hat{\boldsymbol{S}}_{n-1,l}^{<i>}}^{-1}\left(\boldsymbol{S}_{n-1,l}^{(q)}\right)\right],\ i \geq 0 \tag{6}$$

where $\hat{\boldsymbol{S}}_{n-1,l}^{<i>}$ denotes the i-th estimate of the weighted average, with $\hat{\boldsymbol{S}}_{n,l}^{<0>}$ chosen as a random element of the particle set, and $\mathcal{M}$ and $\mathcal{M}^{-1}$ are the orthographic retraction and lifting maps [18], respectively, see also Appendix A. The algorithm in (6) was run until $\|\boldsymbol{S}_{n-1,l}^{<i+1>} - \boldsymbol{S}_{n-1,l}^{<i>}\|_F < 10^{-6}$, where $\|\cdot\|_F$ stands for the Frobenius norm; when the algorithm meets this condition, the estimate $\hat{\boldsymbol{S}}_{n-1,l}$ is taken as $\boldsymbol{S}_{n-1,l}^{<i+1>}$.

Given $\hat{\boldsymbol{S}}_{n-1,l}$, we evaluate $\boldsymbol{X}_{n-1,l}^{(q)} = \mathcal{M}_{\hat{\boldsymbol{S}}_{n-1,l}}^{-1}(\boldsymbol{S}_{n-1,l}^{(q)})$, compute $\boldsymbol{A}_{n-1,l}^{(q)} = (\hat{\boldsymbol{S}}_{n-1,l})^T \boldsymbol{X}_{n-1,l}^{(q)}$ and $\boldsymbol{B}_{n-1,l}^{(q)} = (\hat{\boldsymbol{S}}_{n-1,l}^{\perp})^T \boldsymbol{X}_{n-1,l}^{(q)}$, and form the vector $\boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)}$ by collecting the free-varying entries of $\boldsymbol{A}_{n-1,l}^{(q)}$ and $\boldsymbol{B}_{n-1,l}^{(q)}$. The weighted particle set $\{w_{n-1,l}^{(q)}, \boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)}\}$ can then be approximated by a Gaussian p.d.f. adjusted via moment matching, i.e.,

$$\bar{\boldsymbol{\mathfrak{x}}}_{n-1,l} = \sum_{q=1}^{Q} w_{n-1,l}^{(q)} \boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)}, \tag{7}$$

$$\boldsymbol{\Sigma}_{n-1,l} = \left\{\sum_{q=1}^{Q} w_{n-1,l}^{(q)} \boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)} [\boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)}]^T\right\} - \bar{\boldsymbol{\mathfrak{x}}}_{n-1,l} [\bar{\boldsymbol{\mathfrak{x}}}_{n-1,l}]^T. \tag{8}$$

Therefore, at the end of the Random Exchange Step, node r receives $\{\bar{\boldsymbol{\mathfrak{x}}}_{n-1,l}; \boldsymbol{\Sigma}_{n-1,l}; \hat{\boldsymbol{S}}_{n-1,l}\}$ and rebuilds the particle set as

$$\begin{aligned} \boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)} &\sim \mathcal{N}\left(\bar{\boldsymbol{\mathfrak{x}}}_{n-1,l}; \boldsymbol{\Sigma}_{n-1,l}\right), \\ \boldsymbol{S}_{n-1,l}^{(q)} &= \mathcal{M}_{\hat{\boldsymbol{S}}_{n-1,l}}\left(\hat{\boldsymbol{S}}_{n-1,l}\boldsymbol{A}(\boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)}) + \hat{\boldsymbol{S}}_{n-1,l}^{\perp}\boldsymbol{B}(\boldsymbol{\mathfrak{x}}_{n-1,l}^{(q)})\right) \\ w_{n-1,l}^{(q)} &= \frac{1}{Q}, \end{aligned} \tag{9}$$

where $q \in \{1, ..., Q\}$, $\mathcal{M}$ denotes the retraction map (Sec. 3; see also Appendix A), $\boldsymbol{A}(\cdot)$ and $\boldsymbol{B}(\cdot)$ form, respectively, a skew-symmetric and an arbitrary matrix given the entries in the argument, and $\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\Xi})$ denotes a (vector) multivariate Gaussian p.d.f. with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Xi}$.

Data Assimilation Step In the sequel, at time instant n, node r resamples

$$\boldsymbol{S}_{n,r}^{(q)} \sim \mathrm{vMF}(\boldsymbol{S}_n | \kappa \boldsymbol{S}_{n-1,l}^{(q)}), \tag{10}$$

and updates the particle weights assimilating the measurements that are available in the closed neighborhood of r, $\tilde{N}(r) \triangleq N(r) \cup \{r\}$, where $N(r)$ denotes the open neighborhood of r.

Specifically, after receiving $\{\boldsymbol{Y}_{n,u}, \boldsymbol{\Omega}_u, \boldsymbol{\Gamma}_u\}$, $u \in \tilde{N}(r)$, node r computes the updated weights

$$\begin{aligned} w_{n,r}^{(q)} &\propto w_{n-1,l}^{(q)} \left[\prod_{u \in \tilde{N}(r)} p(\boldsymbol{Y}_{n,u} | \boldsymbol{S}_{n,r}^{(q)})\right] \\ &= w_{n-1,l}^{(q)} \left[\prod_{u \in \tilde{N}(r)} \boldsymbol{N}_{k,m}(\boldsymbol{Y}_{n,u} | \boldsymbol{G}_u(\boldsymbol{H}_{n,u}\boldsymbol{S}_{n,r}^{(q)}), \boldsymbol{\Omega}_u, \boldsymbol{\Gamma}_u)\right], \end{aligned} \tag{11}$$

since, from the model in (3), the observations received at each node are conditionally independent of each other given the state $\boldsymbol{S}_n$. In (11), the symbol $\propto$ means that the term on the right-hand side of the first line must be multiplied by a proportionality constant such that $\sum_{q=1}^{Q} w_{n,r}^{(q)} = 1$.

Following the data assimilation in (11), node r has at instant n a Monte Carlo representation for the posterior p.d.f. $p(\boldsymbol{S}_n|\tilde{\boldsymbol{Y}}_{n,r}, \tilde{\boldsymbol{Y}}_{1:n-1,l}) \triangleq p(\boldsymbol{S}_n|\tilde{\boldsymbol{Y}}_{1:n,r})$, where $\tilde{\boldsymbol{Y}}_{n,r}$ is the collection of all assimilated observations $\{\tilde{\boldsymbol{Y}}_{n,u}\}$, $u \in \tilde{N}(r)$.

State Estimate To compute the state estimate $\hat{\boldsymbol{S}}_{n,r}$ given $\{w_{n,r}^{(q)}, \boldsymbol{S}_{n,r}^{(q)}\}_{q=1}^{Q}$, we employ again the iterative algorithm described in (6).

5. SIMULATION RESULTS

We ran Monte Carlo simulations consisting of 300 independent runs to evaluate the performance of the proposed algorithm. We deployed a network with five nodes: nodes 1 to 4 are on the vertices of a square and node 5 is at its center and is connected to all other nodes; this is the same topology used in [14] and [17]. Nodes 1 to 5 are subject to the following signal-to-noise ratios (SNRs): 3, 6, 10, 13 and 20 dB, respectively. We assumed that $k = 3$, $m = 2$, $\kappa = 50$, and $Q = 200$. The observation function $\boldsymbol{G}_r$ in (3) was defined as $[\boldsymbol{G}_r(\boldsymbol{H}_{n,r}\boldsymbol{S}_n)]_{i,j} = g\left([\boldsymbol{H}_{n,r}\boldsymbol{S}_n]_{i,j}\right)$, where $[\boldsymbol{A}]_{i,j}$ denotes the entry of matrix $\boldsymbol{A}$ with the given indexes and $g(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ is a pointwise nonlinear function. We set $g(x) = 0.5x + 0.5x^2$, which corresponds to a strong nonlin-

earity at each receiving node. This formulation can be used to tackle the problem of *equalization* of the downlink of a TDD (time division duplexed) MIMO communication system using SVD precoding [14, 23], assuming that the base station has perfect knowledge of the downlink channel state information matrix, or, alternatively, to solve supervised subspace tracking problems.

For comparison, we ran in the same setup the following competing methods: i) the density-assisted PF algorithm of [24] on the Stiefel manifold operating as a data fusion center, i.e., with access to all the measurements of the network (Centralized DAPF), and adapted versions of: ii) the likelihood consensus-based PF (LC Euclid PF), using the same configurations of [20], iii) the RndEx PF [15] (RndEx Euclid PF), iv) the RndEx EKF method [19] (RndEx EKF), and v) the diffusion EKF of [25] (ATC EKF). The adapted methods ii)-v) were based on a vectorized version of the model in (1) and (3): the state evolves according to an empirically adjusted Gaussian random walk on the Euclidean space $\mathbb{R}^{km\times 1}$ (rather than on the Stiefel manifold), and their estimates were projected onto the Stiefel manifold by determining their polar factors. The centralized DAPF i) uses in turn a parametric vMF representation of the weighted particle set at each time instant, which is built using the maximum likelihood estimation algorithm described in [14].

As shown in Fig. 1, the proposed algorithm of Sec. 4 (RndEx GTS PF) performs closely to the centralized DAPF. Its transient performance, however, is worse than those of both the centralized DAPF and the RndEx method of [14]. The RndEx GTS PF also outperforms both the likelihood consensus-based PF and the Euclidean RndEx PF in the steady state, despite showing a slower transient response. Furthermore, the proposed algorithm performs better than the diffusion EKFs while incurring in similar communication cost (Table 1), at the expense, however, of increased computational complexity.

Algorithm	Communication Cost (Real Numbers)	Complexity (per sample/node)
RndEx GTS	$2(km+(km)^2)+km\lvert\tilde{N}(r)\rvert$	$\mathcal{O}(Qk^3m^3)$
PF-LC-Euclid	$\mathcal{O}(km)\lvert\tilde{N}(r)\rvert$	$\mathcal{O}(Qk^3m^3)$
ATC EKF	$3km\lvert\tilde{N}(r)\rvert$	$\mathcal{O}(k^3m^3)$
RndEx EKF	$3(km+(km)^2)\lvert\tilde{N}(r)\rvert$	$\mathcal{O}(k^3m^3)$

Table 1. Communication cost (for a given node r, per processed sample) and Computational Complexity of the considered distributed algorithms.

6. CONCLUSIONS

In this paper, we introduced a new distributed RndEx diffusion algorithm to track a sequence of hidden random matrices that evolve on the Stiefel manifold. This algorithm included a novel Gaussian parametric approximation of the weighted particle set on the tangent space, in which we extended a previous method applied to the unit hypersphere [17]. Simulation results show that, in a scenario with strongly nonlinear receivers and state vectors defined on the Stiefel manifold, the proposed algorithm approaches the performance of a centralized DAPF algorithm and outperforms both distributed EKFs in the literature and state-o-the-art distributed Euclidean PFs such as the LC-PF and the RndEx-PF.

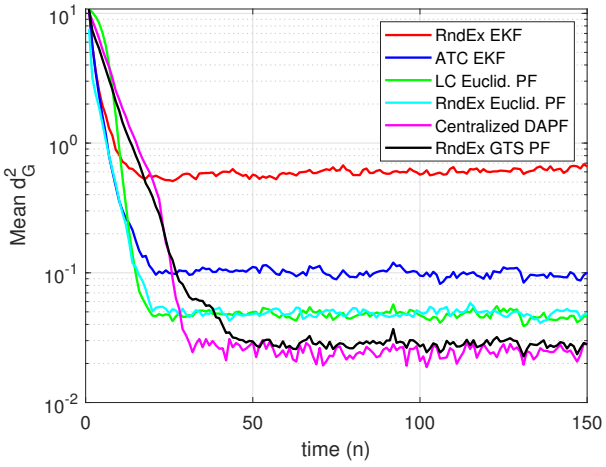

Fig. 1. Squared geodesic distance $d_G^2\left(\hat{\boldsymbol{S}}_{n,r},\boldsymbol{S}_n\right) \triangleq \left\|\mathrm{Log}_{\boldsymbol{S}_n}(\hat{\boldsymbol{S}}_{n,r})\right\|_F^2$ (averaged over all nodes r) provided by the proposed method (RndEx GTS) and by competing algorithms described in the text as functions of the number of processed samples n.

APPENDIX A: RETRACTION AND LIFTING MAPS

Given $\boldsymbol{M}\in\mathcal{V}^{k,m}$ and $\boldsymbol{X}\in T_{\boldsymbol{M}}\mathcal{V}^{k,m}$, the orthographic retraction map can be computed as follows [18]:
1) Compute $\boldsymbol{D}=\boldsymbol{M}^T\boldsymbol{X}+\boldsymbol{I}_m$.
2) Solve for $\boldsymbol{Z}$ the algebraic Riccati equation $\boldsymbol{Z}^2+\boldsymbol{Z}\boldsymbol{D}+\boldsymbol{D}^T\boldsymbol{Z}+\boldsymbol{X}^T\boldsymbol{X}=\boldsymbol{0}$.
3) Compute the retraction map as $\mathcal{M}_{\boldsymbol{M}}(\boldsymbol{X})=\boldsymbol{M}(\boldsymbol{I}_m+\boldsymbol{Z})$.

Given $\boldsymbol{M},\boldsymbol{S}\in\mathcal{V}^{k,m}$, the orthographic lifting map is computed in turn as [18]

$$\mathcal{M}_{\boldsymbol{M}}^{-1}(\boldsymbol{S})=\boldsymbol{S}-\frac{1}{2}\boldsymbol{M}\left(\boldsymbol{S}^T\boldsymbol{M}+\boldsymbol{M}^T\boldsymbol{S}\right).$$

7. REFERENCES

[1] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[2] F. Tompkins and P. J. Wolfe, "Bayesian Filtering on the Stiefel Manifold," in *IEEE Intl. Workshop Comput. Adv.*

Multi-Sensor Adaptive Process. (CAMPSAP), 2007, pp. 261–264.

[3] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. S. Chirikjian, "Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map," *Robotica*, vol. 26, pp. 419, 2008.

[4] Y. M. Lui, "Tangent bundles on special manifolds for action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 930–942, 2011.

[5] J. Boulanger, S. Said, N. Le Bihan, and J. H. Manton, "Filtering from observations on Stiefel manifolds," *Signal Process.*, vol. 122, pp. 52–64, 2016.

[6] J. Tian, J. Zhao, and C. Zheng, "Clustering of cancer data based on Stiefel manifold for multiple views," *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–15, 2021.

[7] R. F. Bordley, "The combination of forecasts: a Bayesian approach," *J. Operational Research Soc.*, vol. 33, no. 2, pp. 171–174, 1982.

[8] S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch, "Bayes and big data: The consensus Monte Carlo algorithm," *Intl. J. Manag. Sci. Eng. Manag.*, vol. 11, no. 2, pp. 78–88, 2016.

[9] D. Luengo, L. Martino, V. Elvira, and M. Bugallo, "Efficient linear fusion of partial estimators," *Digital Signal Process.*, vol. 78, pp. 265–283, 2018.

[10] L. Martino and V. Elvira, "Compressed Monte Carlo with application in particle filtering," *Information Sciences*, vol. 553, pp. 331–352, 2021.

[11] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Automat. Contr.*, vol. 55, no. 9, pp. 2069–2084, 2010.

[12] S. A. Alghunaim, K. Yuan, and A. H. Sayed, "A proximal diffusion strategy for multiagent optimization with sparse affine constraints," *IEEE Trans. Automat. Contr.*, vol. 65, no. 11, pp. 4554–4567, 2020.

[13] D. Jin, J. Chen, C. Richard, J. Chen, and A. H. Sayed, "Affine combination of diffusion strategies over networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 2087–2104, 2020.

[14] Caio G de Figueredo, Claudio J Bordin Jr, and Marcelo GS Bruno, "Nonlinear distributed state estimation on the Stiefel manifold using diffusion particle filters," *Digital Signal Process.*, vol. 122, pp. 103354, 2022.

[15] M. G. S. Bruno and S. S. Dias, "Collaborative emitter tracking using Rao-Blackwellized random exchange diffusion particle filtering," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, pp. 19, 2014.

[16] K. Dedecius and P. M. Djurić, "Sequential estimation and diffusion of information over networks: A Bayesian approach with exponential family of distributions," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1795–1809, 2016.

[17] C. G. de Figueredo, C. J. Bordin, and M. G. S. Bruno, "Cooperative parameter estimation on the unit sphere using a network of diffusion particle filters," *IEEE Signal Process. Lett.*, vol. 27, pp. 715–719, 2020.

[18] T. Kaneko, S. Fiori, and T. Tanaka, "Empirical arithmetic averaging over the compact Stiefel manifold," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 883–894, 2012.

[19] M. G. S. Bruno and S. S. Dias, "A Bayesian Interpretation of Distributed Diffusion Filtering Algorithms [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 118–123, 2018.

[20] O. Hlinka, O. Slučiak, F. Hlawatsch, P. M. Djurić, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, 2012.

[21] C. G. Khatri and K. V. Mardia, "The Von Mises–Fisher matrix distribution in orientation statistics," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 39, no. 1, pp. 95–106, 1977.

[22] Z. Wang and V. Solo, "Numerical Solution of Stochastic Differential Equations in Stiefel Manifolds via Tangent Space Parametrization," in *IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2021, pp. 5125–5129.

[23] S. K. Mohammed, E. Viterbo, Y. Hong, and A. Chockalingam, "MIMO Precoding with X- and Y-Codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3542–3566, 2011.

[24] P. M. Djurić, M. F. Bugallo, and J. Míguez, "Density assisted particle filters for state and parameter estimation," in *IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, vol. 2, pp. ii–701.

[25] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion Recursive Least-Squares for Distributed Estimation over Adaptive Networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.