

Mixture Reduction on Matrix Lie Groups

Josip Ćesić, *Member, IEEE*, Ivan Marković, *Member, IEEE*, and Ivan Petrović, *Member, IEEE**

Abstract—Many physical systems evolve on matrix Lie groups and mixture filtering designed for such manifolds represent an inevitable tool for challenging estimation problems. However, mixture filtering faces the issue of a constantly growing number of components, hence require appropriate mixture reduction techniques. In this letter we propose a mixture reduction approach for distributions on matrix Lie groups, called the concentrated Gaussian distributions (CGDs). This entails appropriate reparametrization of CGD parameters to compute the KL divergence, pick and merge the mixture components. Furthermore, we also introduce a multitarget tracking filter on Lie groups as a mixture filtering study example for the proposed reduction method. In particular, we implemented the probability hypothesis density filter on matrix Lie groups. We validate the filter performance using the optimal subpattern assignment metric on a synthetic dataset consisting of 100 randomly generated multitarget scenarios.

Index Terms—Mixture reduction, estimation on matrix Lie groups, multitarget tracking, probability hypothesis density filter.

I. INTRODUCTION

MANY statistical and engineering problems require modeling of complex multi-modal data, wherein mixture distributions became an inevitable tool [1], [2], primarily in traditional application domains like radar and sonar tracking [3], and later in different modern fields such as computer vision [4], speech recognition [5] or multimedia processing [6]. Approaches relying on mixture distributions often face the problem of large or an ever increasing number of mixture components, hence the growth of components must be controlled by approximating the original mixture with the mixture of a reduced size [7]–[9]. For example, in the case of multitarget tracking applications, by employing conventional Gaussian mixture based filters [10], [11], during the recursion process the number of components inevitably increases. This appears firstly due to appearance of newly birthed or spawned components, and secondly, due to inclusion of multiple measurements, which results in geometrical increase in the number of components.

Another important aspect of estimation is the state space geometry, hence many works have been dedicated to appropriate uncertainty modeling and estimation techniques for a wide range of applications [12]–[15], motivated by theoretical and implementation difficulties caused by treating a constrained problem naively with Euclidean tools. For example, Lie groups are natural ambient (state) spaces for description of the dynamics of rigid body mechanical

systems. In [16] it has been observed that the distribution of the pose of a differential drive mobile robot is not a Gaussian distribution in Cartesian coordinates, but rather a distribution on the special Euclidean group SE(2). Similarly, [17] discussed the uncertainty association with 3D pose employing the SE(3) group. Furthermore, attitude estimation arises naturally on the SO(3) group [15]. In [18] a feedback particle filter on matrix Lie groups was proposed, while in [19], [20] authors proposed an extended Kalman filter on matrix Lie groups (LG-EKF), building the theory upon the concentrated Gaussian distribution (CGD) on matrix Lie groups [21].

In this letter we address finite mixtures of distributions on matrix Lie groups. We propose a novel approach to CGD mixture reduction, which required finding solutions for computing Kullback-Leibler divergence of CGD components and CGD component merging. Furthermore, since previous methods require choosing the appropriate tangent space, we also provide an extensive analysis on the choice thereof. As a study example, we use the proposed reduction method in a multitarget tracking scenario. We introduce the probability hypothesis density filter (PHD) on matrix Lie groups with approximation based on a finite mixture of CGDs.

II. MATHEMATICAL PRELIMINARIES

We now introduce theoretical preliminaries concerning Lie groups; however, for a more rigorous introduction the reader is directed to [22]. A Lie group G is a group which has the structure of a smooth manifold; moreover, a tangent space $T_X(G)$ is associated to $X \in G$ such that the tangent space placed at the group identity, called Lie algebra \mathfrak{g} , is transferred by applying corresponding action to X . In this paper we are interested in matrix Lie groups which are usually the ones considered in engineering and physical sciences.

The Lie algebra $\mathfrak{g} \subset \mathbb{R}^{n \times n}$ associated to a p -dimensional matrix Lie group $G \subset \mathbb{R}^{n \times n}$ is a p -dimensional vector space. The matrix exponential \exp_G and matrix logarithm \log_G establish a local diffeomorphism between the two

$$\exp_G : \mathfrak{g} \rightarrow G \quad \text{and} \quad \log_G : G \rightarrow \mathfrak{g}. \quad (1)$$

Furthermore, a natural relation exists between \mathfrak{g} and the Euclidean space \mathbb{R}^p given through a linear isomorphism

$$[\cdot]_G^\vee : \mathfrak{g} \rightarrow \mathbb{R}^p \quad \text{and} \quad [\cdot]_G^\wedge : \mathbb{R}^p \rightarrow \mathfrak{g}. \quad (2)$$

For $x \in \mathbb{R}^p$ and $X \in G$ we use the following notation [23]

$$\exp_G^\wedge(x) = \exp_G([\cdot]_G^\wedge(x)) \quad \text{and} \quad \log_G^\vee(X) = [\log_G(X)]_G^\vee. \quad (3)$$

Lie groups are generally non-commutative, i.e., $XY \neq YX$. However, the non-commutativity can be captured by the so-called adjoint representation of G on \mathfrak{g} [24]

$$X \exp_G^\wedge(y) = \exp_G^\wedge(\text{Ad}_G(X)y)X, \quad (4)$$

*J. Ćesić, I. Marković and I. Petrović are with the University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia. E-mail: {name.surname@fer.hr}

This work has been supported from the Unity Through Knowledge Fund (no. 24/15) under the project Cooperative Cloud based Simultaneous Localization and Mapping in Dynamic Environments (cloudSLAM) and the European Union's Horizon 2020 research and innovation programme under grant agreement No 688117 (SafeLog).

which can be seen as a way of representing the elements of the group as a linear transformation of the group's algebra. The adjoint representation of \mathfrak{g} , ad_G , is in fact the differential of Ad_G at the identity. Another important result for working with Lie group elements is the Baker-Campbell-Hausdorff (BCH) formula, which enables representing the product of Lie group members as a sum in the Lie algebra. We will use the following BCH formulae [24], [25]

$$\log_G^\vee(\exp_G^\wedge(x) \exp_G^\wedge(y)) = y + \varphi_G(y)x + O(\|y\|^2), \quad (5)$$

$$\log_G^\vee(\exp_G^\wedge(x+y) \exp_G^\wedge(-x)) = \Phi_G(x)y + O(\|y\|^2), \quad (6)$$

where $\varphi_G(y) = \sum_{n=0}^{\infty} \frac{B_n \text{ad}_G(y)^n}{n!}$, B_n are Bernoulli numbers, and $\Phi_G(x) = \varphi_G(x)^{-1}$. For many common groups used in engineering and physical sciences closed form expressions for $\varphi_G(\cdot)$ and $\Phi_G(\cdot)$ can be found [17], [24]; otherwise, a truncated series expansion is used.

A. Concentrated Gaussian distribution

Herein we introduce the concept of the concentrated Gaussian distribution which is used to define random variables on matrix Lie group. A random variable $X \in G$ has a CGD with the mean μ and covariance Σ , i.e., $X \sim \mathcal{G}(X; \mu, \Sigma)$, if

$$X = \exp_G^\wedge(\xi)\mu, \quad (7)$$

where $\mu \in G$, and $\xi \sim \mathcal{N}(\xi; \mathbf{0}_{p \times 1}, \Sigma)$ is a zero-mean 'classical' Gaussian random variable with the covariance $\Sigma \subset \mathbb{R}^{p \times p}$ [17], [20]. Note that in this way, we are directly defining the CGD covariance in the pertaining Lie algebra \mathfrak{g} , while the mean is defined on the group G .

Given that, the previous definition (7) then induces a pdf of X over G as follows [17], [20]

$$\begin{aligned} 1 &= \int_{\mathbb{R}^p} \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp_G^\wedge \left(-\frac{1}{2} \|\xi\|_\Sigma^2 \right) d\xi \\ &= \int_G \beta \exp_G^\wedge \left(-\frac{1}{2} \|\log_G^\vee(X\mu^{-1})\|_\Sigma^2 \right) dX \end{aligned} \quad (8)$$

where $\|x\|_\Sigma^2 = x^T \Sigma^{-1} x$. Therein the change of coordinates $\xi = \log_G^\vee(X\mu^{-1})$, with the pertaining differentials $dX = |\Phi(\xi)| d\xi$, resulted with the CGD normalizing constant

$$\beta = 1/\sqrt{(2\pi)^p |\Phi(\log_G^\vee(X\mu^{-1}))\Sigma\Phi(\log_G^\vee(X\mu^{-1}))^T|}. \quad (9)$$

Note that this change of variables is valid if all eigenvalues of Σ are small, i.e., almost all the mass of the distribution is concentrated in a small neighborhood around the mean value [20]. The pdf over X is now fully determined by (8) and (9).

III. CGD MIXTURE REDUCTION

With the theoretical preliminaries setup, we continue with mixture reduction on matrix Lie groups. A finite mixture of our present interest is given as the weighted sum of CGDs

$$\sum_{i=1}^N w_i \mathcal{G}(X; \mu_i, \Sigma_i), \quad (10)$$

where w_i are component weights and N is the total number of mixture components. An illustration of (10) is given in Fig. 1. Component reduction procedures typically require

three building blocks: (i) component distance measure, (ii) component picking algorithm, and (iii) component merging. While various solutions exist for 'classical' Gaussian mixtures [7]–[9], questions remain on how to approach the component number reduction for CGD mixtures on matrix Lie groups. Therefore, first we focus on the the fundamental question of how to measure the distance between two CGD components.

A. Component distance measure

Our aim is to use a standard information-theoretic measure between two CGD components and we propose to use the Kullback-Leibler (KL) divergence [26]. Let $\mathcal{G}_i = \mathcal{G}(X; \mu_i, \Sigma_i)$ and $\mathcal{G}_j = \mathcal{G}(X; \mu_j, \Sigma_j)$ be two mixture components with $p_i(X)$ and $p_j(X)$ as their respective pdfs. Since there is nothing intrinsic in the definition of KL divergence that requires the underlying space to be Euclidean, by definition

$$\mathcal{D}_{\text{KL}}(\mathcal{G}_i, \mathcal{G}_j) = \int_G p_i(X) \log \left(\frac{p_i(X)}{p_j(X)} \right) dX. \quad (11)$$

In order to evaluate the integral (11), we need to employ the change of coordinates as in (8), but this time from the direction of the group G , i.e., from $X \in G$ to $\xi \in \mathbb{R}^p$. Note that in (8) the change evolved around the distribution mean μ ; however, since in (11) generally $\mu_i \neq \mu_j$, we cannot apply the same approach. Hence, before evaluating (11), we first discuss how to change the coordinates on the level of a single distribution.

Let $\mathcal{G}(X; \mu, \Sigma)$ be a CGD, and if we change the coordinates using $X = \exp_G^\wedge(\xi)\mu_t$, $\mu_t \in G$, where $\mu_t \neq \mu$, we get

$$\begin{aligned} 1 &= \int_G \beta \exp_G^\wedge \left(-\frac{1}{2} \|\log_G^\vee(X\mu^{-1})\|_\Sigma^2 \right) dX \\ &\stackrel{\text{CoC}}{\approx} \int_{\mathbb{R}^p} \eta \exp_G^\wedge \left(-\frac{1}{2} \|\log_G^\vee(\exp_G^\wedge(\xi)\mu_t\mu^{-1})\|_\Sigma^2 \right) d\xi \\ &\stackrel{(6)}{\approx} \int_{\mathbb{R}^p} \eta \exp_G^\wedge \left(-\frac{1}{2} \|\Phi_G(r_t)(\xi - r_t)\|_\Sigma^2 \right) d\xi \\ &= \int_{\mathbb{R}^p} \eta \exp_G^\wedge \left(-\frac{1}{2} \|\xi - r_t\|_{\varphi_G(r_t)\Sigma\varphi_G^T(r_t)}^2 \right) d\xi, \end{aligned} \quad (12)$$

where $r_t = \log_G^\wedge(\mu\mu_t^{-1})$, η approximately evaluates to

$$\begin{aligned} \eta &= \beta |\Phi(\xi)| = \frac{|\Phi(\xi)|}{\sqrt{(2\pi)^p |\Sigma|} \cdot |\Phi(\log_G^\vee(\exp_G^\wedge(\xi)\mu_t\mu^{-1}))|} \\ &\approx \frac{1}{\sqrt{(2\pi)^p |\varphi_G(r_t)\Sigma\varphi_G^T(r_t)|}}, \end{aligned} \quad (13)$$

and we obtain $\xi \sim \mathcal{N}(\xi; r_t, \varphi_G(r_t)\Sigma\varphi_G^T(r_t))$.

Remark 1. Covariance of a CGD represents the uncertainty relevant only to the tangent space of its own mean. In [24] authors studied how the covariance changes if looked at from the perspective of a value which is different than the distribution mean. They dubbed this procedure 'distribution unfolding'. For example, if we unfold $\mathcal{G}(X; \mu, \Sigma)$ around an arbitrary $\mu_t \in G$, using (5) and following [24] we get

$$\begin{aligned} \xi_t &= \log_G^\vee(\exp_G^\wedge(\xi)\mu\mu_t^{-1}) \\ &\approx \log_G^\vee(\mu\mu_t^{-1}) + \varphi_G(\log_G^\vee(\mu\mu_t^{-1}))\xi. \end{aligned} \quad (14)$$

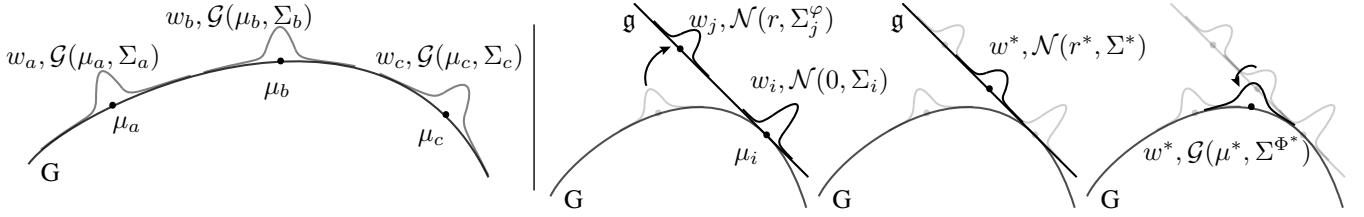


Fig. 1: Illustration of a finite mixture of CGDs (left) and the component merging procedure (right).

By computing the expectation and covariance of (14), we obtain a reparametrized distribution, $\xi_t \sim \mathcal{N}(\xi_t; r_t, \Sigma^\varphi)$, where

$$r_t = \log_G^\vee(\mu \mu_t^{-1}) \quad (15)$$

$$\Sigma^\varphi = \varphi_G(r_t) \Sigma \varphi_G^T(r_t). \quad (16)$$

This pdf is equal to the one obtained through the change of coordinates in (12). However, obtaining this result by using the procedure of coordinates change through a pdf is important from the perspective of KL divergence evaluation. An illustration of unfolding a component j around μ_i , using (15) and (16), is given in Fig. 1.

The KL divergence between two CGDs $\mathcal{G}_i = \mathcal{G}(\mu_i, \Sigma_i)$ and $\mathcal{G}_j = \mathcal{G}(\mu_j, \Sigma_j)$ can now be evaluated as

$$\mathcal{D}_{\text{KL}}(\mathcal{G}_i, \mathcal{G}_j) \approx \int_{\mathbb{R}^p} p_i(\xi) \log \left(\frac{p_i(\xi)}{p_j(\xi)} \right) d\xi = D_{\text{KL}}(\mathcal{N}_i, \mathcal{N}_j),$$

$$p_i(\xi) \sim \mathcal{N}_i = \mathcal{N}(\xi; r_i, \Sigma_i^\varphi), \quad r_i = \log_G^\wedge(\mu_i \mu_t^{-1}), \quad (17)$$

$$p_j(\xi) \sim \mathcal{N}_j = \mathcal{N}(\xi; r_j, \Sigma_j^\varphi), \quad r_j = \log_G^\wedge(\mu_j \mu_t^{-1}),$$

and $\Sigma^\varphi = \varphi_G(r) \Sigma \varphi_G^T(r)$. By employing the change of coordinates, we can evaluate the KL divergence of two CGDs similarly as in the case of ‘classical’ Gaussian distributions, but with reparametrized means and covariances. The KL divergence is then equal to

$$\begin{aligned} \mathcal{D}_{\text{KL}}(\mathcal{N}_i, \mathcal{N}_j) = & \frac{1}{2} \left(\text{tr}(\Sigma_j^{\varphi^{-1}} \Sigma_i^\varphi) - K + \log_{\mathbb{R}} \frac{|\Sigma_j^\varphi|}{|\Sigma_i^\varphi|} \right) \\ & + (r_j - r_i)^T (\Sigma_j^\varphi)^{-1} (r_j - r_i), \end{aligned} \quad (18)$$

where $\text{tr}(\cdot)$ and $|\cdot|$ designate matrix trace and determinant, respectively, while K is the mean vector dimension. Finally, for mixture components it is necessary to use the scaled symmetrized KL divergence [27], which also takes component weights into account

$$\begin{aligned} \mathcal{D}_{s\text{KL}}(w_i \mathcal{N}_i, w_j \mathcal{N}_j) = & \frac{1}{2} \left((w_i - w_j) \log_{\mathbb{R}} \frac{w_i}{w_j} + \right. \\ & \left. w_i \mathcal{D}_{\text{KL}}(\mathcal{N}_i, \mathcal{N}_j) + w_j \mathcal{D}_{\text{KL}}(\mathcal{N}_j, \mathcal{N}_i) \right). \end{aligned} \quad (19)$$

B. Component picking algorithm

Now that we know how to compute a distance measure between two CGD mixture components, we need to choose an appropriate component picking algorithm which will tell us how to screen the whole mixture and which components to pick for merging. However, with CGD mixtures there is also another momentum. If we have N components in the mixture with different weights, how should we approach the problem of measuring distance, i.e., choosing μ_t for the

change of coordinates? Should all the distances be calculated with respect to the mean of the component with the highest weight or the lowest weight? Or should we ‘reparametrize’ each component on a pairwise basis? In this letter we study the following five scenarios: (i/ii) all components are reparametrized about the mean of the component with the highest/lowest weight, (iii) the reparametrization about the identity element, and (iv/v) components are reparametrized on a pairwise basis by choosing the mean of the component pair with the higher/lower weight. For analyzing the five scenarios we use two common component picking strategies; (i) Exhaustive pairwise [28], and (ii) West’s [29] algorithms. The Exhaustive pairwise algorithm determines distances between all components and merges the closest pair, while West’s algorithm sorts the components according to their respective weights, then finds and merges the component most similar to the first one.

C. Merging the components

A component merging algorithm for Gaussian components in \mathbb{R}^p was proposed in [28]:

$$r^* = \frac{1}{w^*} \sum_i w_i r_i, \quad \Sigma^* = \frac{1}{w^*} \sum_i \left(w_i (\Sigma_i + r_i r_i^T) \right) - r^* (r^*)^T$$

where $w^* = \sum_i w_i$, $w_i \mathcal{N}(r_i, \Sigma_i)$ represents the i -th component, and $w^* \mathcal{N}(r^*, \Sigma^*)$ is the resulting component. Although merging works for an arbitrary number of components, in our case we will always merge two.

However, the previous expressions are defined for Gaussians in \mathbb{R}^p and the question arises how to apply the same approach for CGD mixtures? We propose to use the same principle as for computing the KL divergence described in Section III-A, i.e., the components to be merged need to be first reparametrized about the tangent space of the same mean, since covariances are only relevant with respect to their own mean. Once we compute the resulting component, $w^* \mathcal{N}(r^*, \Sigma^*)$, we need to map it back to the group G . Given a lemma from [20] and following convention (7), the procedure evaluates to

$$\mu^* = \exp_G^\wedge(r^*) \mu_t, \quad \Sigma^{\Phi^*} = \text{Ad}_G^{r^*} \Phi_G(r^*) \Sigma^* (\text{Ad}_G^{r^*} \Phi_G(r^*))^T, \quad (20)$$

where $\text{Ad}_G^{r^*} = \text{Ad}_G(\exp_G^\wedge(r^*))$. We can notice that covariance reparametrization was necessary to make it relevant from the perspective of the tangent space of the newly computed μ^* . An illustration of merging and reparametrization (20) of component j with respect to μ_i is given in Fig. 1.

IV. STUDY EXAMPLE - PHD FILTER ON LIE GROUPS

MTT is a complex problem consisting of many challenges and PHD filter presents itself as one of the solutions to MTT.

The reason why PHD filter is interesting for the present letter is because one of its implementations is based on Gaussian mixtures (GM-PHD) [10]. Besides Gaussians, other distributions can be used and in our previous work [30] we proposed a mixture approximation of the PHD filter based on the von Mises distribution on the unit circle. In this letter, as a study example, we implement a PHD filter tailored for Lie groups (LG-PHD), based on the mixture of CGDs and the reduction schemes presented in the previous section. The LG-PHD can be potentially applied in MTT scenarios where the target state is modelled as a pose in $SE(2)$ or $SE(3)$.

The PHD filter propagates the *intensity function* D_{k-1} , and operates by evaluating two steps—prediction and update. By assuming D_{k-1} and birth intensity being Gaussian mixtures [10], the GM-PHD prediction results with another Gaussian mixture (Prop. 1 in [10]). Similarly, if D_{k-1} and birth intensity are given with CGD mixtures, the LG-PHD prediction results with another CGD mixture, relying on the LG-EKF prediction applied to each mixture component [23].

The product of two Gaussians evaluates to a scaled Gaussian, hence the update step of GM-PHD can be calculated analytically (Prop. 2 in [10]). In contrary, the product of two CGDs, occurring in LG-PHD update, cannot be evaluated directly. Hence, we apply approximations following the same train of thought as in LG-EKF prediction [23] where given posterior $p(X_{k-1}|Z_{1:k-1})$ and motion model $p(X_k|X_{k-1})$, it approximates the joint distribution $p(X_k, X_{k-1}|Z_{1:k-1})$, and then marginalizes obtaining $p(X_k|Z_{1:k-1})$. Similarly, given $p(X_k|Z_{1:k-1})$ and likelihood $p(Z_k|X_k)$, we approximate the joint distribution $p(X_k, Z_k|Z_{1:k-1})$, and then marginalize obtaining $p(X_k|Z_k)$. Final LG-PHD formulae are nearly identical to GM-PHD, except for Jacobian matrices.

A. Experiments

In order to validate the performance of the proposed LG-PHD filter, and compare different reduction approaches that are applied after update steps, we devised appropriate Monte Carlo simulation scenarios. We applied two component picking strategies, namely the West's algorithm and the pairwise component picking algorithm. For each we applied the reparametrization approaches as discussed in Section III-B, including the mapping to tangent space of (i) pairwise larger component \mathcal{T}_L , (ii) pairwise smaller component \mathcal{T}_S , (iii) identity element \mathcal{T}_{Id} , (iv) largest component \mathcal{T}_{Max} , (v) smallest component \mathcal{T}_{Min} (West's algorithm always merges the smallest component hence (ii) and (v) are the same). We generated 100 examples of an MTT scenario and compared the performance of the approaches. The initial number of targets in the scene was a random integer $N_{0|0} \in [5, 7]$, while the probability of survival was $p_S = 0.975$ and birth rate was $\lambda_b = 0.25$. All measurements were corrupted with white noise variance $\sigma_{xy}^2 = 0.5^2 \text{ m}^2$ in distance and $\sigma_\phi = 0.1 \text{ rad/s}$ in orientation, while clutter was governed by the Poisson distribution with intensity $\lambda_Z = 5$. The state $X = (X^{\text{pos}}, X^{\text{vel}}) \in SE(2) \times \mathbb{R}^3$ contains position and velocity components. Here we apply the constant velocity motion model [31] given as

$$f(X_{k-1}) = X_{k-1} \exp_G^{\wedge} \begin{bmatrix} T X_{k-1}^{\text{vel}} \\ \mathbf{0} \end{bmatrix}. \quad (21)$$

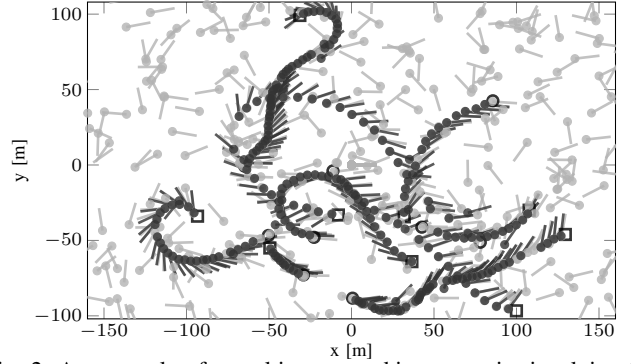


Fig. 2: An example of a multitarget tracking scenario, involving 10 objects, out of which 5 appeared at the beginning, and 5 more were born during the 100 steps long sequence (gray arrows—measurements including false alarms, black arrows—estimated states, black circles—true object birth place, black square—true object death place).

TABLE I: Average OSPA over 100 multitarget scenarios.

	Exhaustive pairwise					West			
	\mathcal{T}_L	\mathcal{T}_S	\mathcal{T}_{Id}	\mathcal{T}_{Max}	\mathcal{T}_{Min}	\mathcal{T}_L	\mathcal{T}_S	\mathcal{T}_{Id}	\mathcal{T}_{Max}
\mathcal{D}_t	2.445	2.515	2.764	2.912	3.082	1.910	1.924	2.060	2.125
\mathcal{D}_d	2.100	2.163	2.419	2.558	2.695	1.415	1.420	1.537	1.605
\mathcal{D}_c	0.594	0.613	0.627	0.653	0.745	0.737	0.746	0.797	0.792

We derive the pertaining Jacobian

$$F_{k-1} = -\frac{d}{ds} \left(\log_G^{\vee} \left(f(\mu_{k-1}) f(\exp_G^{\wedge}(s) \mu_{k-1})^{-1} \right) \right) \Big|_{s=0} \quad (22)$$

$$= \begin{bmatrix} \mathbf{I} & T \Phi_{SE(2)} \left(T \text{Ad}_{SE(2)}(\mu_{k-1}^{\text{pos}}) \mu_{k-1}^{\text{vel}} \right) \text{Ad}_{SE(2)}(\mu_{k-1}^{\text{pos}}) \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

where $\mu_{k-1} = (\mu_{k-1}^{\text{pos}}, \mu_{k-1}^{\text{vel}}) \in SE(2) \times \mathbb{R}^3$ is the mean value, and T is discretization time. The probability of measurement detection was $p_D = 0.975$ and the measurements were arising as $SE(2)$, hence $h(X_k) = X_k^{\text{pos}}$ and the measurement Jacobian was $H_k = [\mathbf{I} \ \mathbf{0}]$. For illustration purposes, an example of a multitarget scenario with tracking in total 10 targets on $SE(2)$ is given in Fig. 2 together with LG-PHD results.

As a performance metric we used the optimal subpattern assignment (OSPA) metric [32]. In Table I we present the results where for each of the 100 multitarget trajectories the cumulative OSPA \mathcal{D}_t , and its localization component \mathcal{D}_d and cardinality component \mathcal{D}_c were calculated. For both Exhaustive pairwise and West's picking strategies, relying on mapping to the tangent space of pairwise larger components \mathcal{T}_L generally outperformed the other approaches.

V. CONCLUSION

In this letter we have studied the problem of mixture reduction on matrix Lie groups. We have particularly dealt with the manipulation of CGD components to compute the KL divergence, pick and merge the mixture components. As a study example, we implemented the LG-PHD filter, a mixture approximation of the PHD filter tailored for MTT with states evolving on matrix Lie groups. Using the OSPA metric we analyzed the performance of the LG-PHD filter with respect to mixture component number reduction.

REFERENCES

- [1] D. Alspach and H. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.
- [2] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.
- [3] S. S. Blackman, *Multiple-target tracking with radar applications*, Gilmour, Ed., 1986.
- [4] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, Aug 2000.
- [5] J. Goldberger and H. Aronowitz, "A distance measure between GMMs based on the unscented transform and its application to speaker recognition," in *Proceedings of Interspeech*, 2005, pp. 1985–1989.
- [6] A. Nikseresht and M. Gelgon, "Gossip-based computation of a gaussian mixture model for distributed multimedia indexing," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 385–392, 2008.
- [7] D. Salmond, "Mixture reduction algorithms for target tracking," in *IEEE Colloquium on State Estimation in Aerospace and State Estimation*, 1989, pp. 1–4.
- [8] A. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, 2007.
- [9] M. Bukal, I. Marković, and I. Petrović, "Composite distance based approach to von Mises mixture reduction," *Information Fusion*, vol. 20, no. 1, pp. 136–145, 2014.
- [10] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.
- [11] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553–3567, 2007.
- [12] Y. M. Lui, "Advances in matrix manifolds for computer vision," *Image and Vision Computing*, vol. 30, no. 6-7, pp. 380–388, 2012.
- [13] G. Loianno, M. Watterson, and V. Kumar, "Visual inertial odometry for quadrotors on SE(3)," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1544–1551.
- [14] J. Česić, V. Joukov, I. Petrović, and D. Kulić, "Full body human motion estimation on Lie groups using 3D marker position measurements," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 826–833.
- [15] A. Barrau and S. Bonnabel, "Intrinsic filtering on Lie groups with applications to attitude estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 436–449, 2015.
- [16] A. W. Long, K. C. Wolfe, M. J. Mashner, and G. S. Chirikjian, "The banana distribution is Gaussian: A localization study with exponential coordinates," in *Proceedings of Robotics: Science and Systems (RSS)*, Sydney, Australia, 2012.
- [17] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Transactions on Robotics*, vol. 30, no. 3, pp. 679–693, Jun. 2014.
- [18] C. Zhang, A. Taghvaei, and P. G. Mehta, "Feedback particle filter on matrix Lie groups," in *Proceedings of the American Control Conference*, 2016, pp. 2723–2728.
- [19] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu, "Discrete extended Kalman filter on Lie groups," in *European Signal Processing Conference (EUSIPCO)*, 2013, pp. 1–5.
- [20] G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus, "Continuous-discrete extended Kalman filter on matrix Lie groups using concentrated Gaussian distributions," *Journal of Mathematical Imaging and Vision*, vol. 51, no. 1, pp. 209–228, 2015.
- [21] K. C. Wolfe, M. Mashner, and G. S. Chirikjian, "Bayesian fusion on Lie groups," *Journal of Algebraic Statistics*, vol. 2, no. 1, pp. 75–97, 2011.
- [22] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups*. Birkhäuser, 2012, vol. 2.
- [23] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu, "From intrinsic optimization to iterated extended Kalman filtering on Lie groups," *Journal of Mathematical Imaging and Vision*, vol. 55, no. 3, pp. 284–303, 2016.
- [24] G. Bourmaud, "Estimation de paramètres évoluant sur des groupes de Lie : Application à la cartographie et à la localisation d'une caméra monoculaire," Ph.D. dissertation, University of Bordeaux, 2015.
- [25] W. Miller, *Symmetry Groups and their Applications*, 1972.
- [26] S. Kullback, *Information Theory and Statistics*. New York: Dover Publications, 1997.
- [27] S. Amari, "Alpha-divergence is unique, belonging to both f-divergence and Bregman divergence classes," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4925–4931, 2009.
- [28] D. J. Salmond, "Mixture reduction algorithms for point and extended object tracking in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 667–686, 2009.
- [29] M. West, "Approximating posterior distributions by mixtures," *Journal of Royal Statistical Society, Series B*, vol. 55, no. 2, pp. 409–442, 1993.
- [30] I. Marković, J. Česić, and I. Petrović, "Von Mises Mixture PHD Filter," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 2229–2233, 2015.
- [31] J. Česić, I. Marković, and I. Petrović, "Moving object tracking employing rigid body motion on matrix Lie groups," in *International Conference on Information Fusion (FUSION)*, 2016, p. 7.
- [32] D. Schuhmacher, B. T. Vo, and B. N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3447–3457, 2008.