A new intrinsic mean-covariance estimator for Lie group observations: application to $SE(2)$

S. Labsir * **A. Renaux** **

* *IPSA/TéSA, 81, avenue de de Grande-Bretagne, 31300, Toulouse, France (e-mail:samy.labsir@ipsa.fr)*
** *University of Paris-Saclay, 3, rue Joliot Curie, 91190 Gif-sur-Yvette, France (e-mail: alexandre.renaux@universite-paris-saclay.fr)*

Abstract: In this communication, we propose to derive a novel estimator of both mean and covariance matrix of observations following a Gaussian distribution on Lie groups. The originality of the approach is to estimate the covariance by using its Lie group structure. To achieve this, we use an intrinsic descent gradient algorithm minimizing a criterion based on the log-likelihood. We derive novel expressions of the gradient of this criterion and we establish that, under suitable assumptions, it converges to a unique solution. Consistency of the proposed estimator is validated numerically by comparison with state-of-the-art approaches.

Keywords: Lie group, covariance matrices, Gaussian distribution, $SE(2)$.

1. INTRODUCTION

In a plethora of applications in control and signal processing, covariance matrix characterization is a significative challenge [Gupta and Nagar (2017), Fan et al. (2011)]. Classically, this matrix corresponds to the imperfection/uncertainty of the sensor measurements. It is common that it is unknown, due to the sensor imperfection but also due to the another (not amenable) phenomenon perturbing the measurements system. Estimating the covariance matrix is therefore crucial for accurately determining unknown parameters, particularly in tasks such as Kalman filtering [Zhang et al. (2020), Sasiadek (2001)]. Several covariance estimators have been intensively proposed in the literature; the well-known unbiased empirical covariance, the shrinkage estimator as well as M-robust in the case where the observations contain outlier [Ollila et al. (2021)]. Note that these estimators are always computed in the context of Euclidean representation of the data. Unfortunately, such assumption fails in the case where the sensor provides structured data (lying on manifold) which is classical in several applications. For instance, in the context of attitude satellite control, star sensor measures noisy attitude information corresponding to the satellite orientation belong to $SO(3)$ [da Silva et al. (2025), Xu et al. (2020)]. In autonomous robotics, relative transformation are measured by RGB-D camera between two instants are classically characterized by $SE(2)$ transformation [Bourmaud (2016)]. We can also cite the IMU (Inertial Measurement Unit), including gyrometer measurement orientation, typically hybridized with navigation control system [Brossard et al. (2022)]. Note that all these observations models belong to Lie group (LG), i.e. smooth manifold equipped with a group operation, and in this work, we specifically focus on this structure [Faraut (2008)]. Thus, the uncertainty of such observations can not be modeled by conventional Euclidean distribution, but by intrinsic distribution on LG such as the Gaussian distribution on LGs [Barfoot and Furgale (2014)]. In the same way as in the Euclidean case, this distribution is parametrized by the mean (LG-mean) and the covariance matrix (LG-covariance) characterizing the uncertainty. Dealing with such distribution involves several difficulties: first, the intrinsic estimation of the covariance can not theoretically be designed by the standard empirical covariance estimator and obtaining closed-form expression is not feasible. Second, the resulting covariance estimator obviously depends on the maximum likelihood estimator (MLE) of the mean. Whereas in the Euclidean case, well-known analytical expression is often available, in our case, the MLE of the LG-mean have to be determined by numerical algorithms or linear approximation [Labsir et al. (2021), Barrau and Bonnabel (2017)] due to the non-linear structure of the distribution. In this work, we propose a new maximum likelihood estimator for both the covariance matrix and the Lie group (LG) mean. It combines the space of covariance matrices and the Lie group of the mean, thus accounting for the fact that the LG-mean is also unknown. In most state-of-the-art approaches, covariance matrices are embedded within a Riemannian structure [Boumal (2023)], which makes the design of the joint estimator more challenging [Breloy et al. (2019)]. To establish a unified formalism adapted to the LG formalism, we propose here to endow the space of covariance matrices with a LG structure, based on the log-Euclidean metric [Arsigny et al. (2006)]. Then, the proposed estimator is defined on a LG product G of both LG-mean and LG-covariance. We first show that the latter must be determined by minimizing a nonlinear cost function on G requiring the use of numerical scheme. To achieve that, we design an intrinsic gradient descent-based algorithm on G [Taylor and Kriegman (1994)] in which we obtain new-closed form expressions of the intrinsic gradient with respect to the LG-covariance and the LG-mean. To the best of our knowledge, this have never been established in the state-of-the-art before. Furthermore, another contribution is the proof that our proposed estimator converges to a unique solution by demonstrating that the cost function is convex, under some realistic conditions. It is achieved by showing that the intrinsic second derivative remains positive. Validity and consistency of the covariance estimator are confirmed numerically by comparison with conventional empirical covariance estimator but also with the LG Cramér-Rao bound (LG-CRB) which is naturally implemented by extension of LG-CRB already established for Euclidean observations [Labsir et al. (2025)]. Specifically, we consider a scenario with observations lying on $SE(2)$ and we observe that our method allows us to obtain better precision for covariance matrix estimation in scenario with few measurements. It demonstrates a certain robustness of our approach. This communication is organized as follows: in the section II, we remind some background about Lie group theory in the section III, we present the proposed LG maximum likelihood estimator and prove convergence of the algorithm. In the section IV, simulation results of the proposed estimator are provided for the LG $SE(2)$.

2. BACKGROUND ON LIE GROUPS

2.1 Definition

A matrix Lie Group G is a mathematical structure that combines the properties of a group with those of a differentiable manifold. Its key feature is that the group operations (multiplication and inversion) are differentiable. This allows the definition of a tangent space at each point $\mathbf{M} \in G$. The tangent space $T_{\mathbf{M}}G$ is a vector space whose dimension m matches that of the manifold.

At the identity element $\mathbf{I}$, the tangent space $T_{\mathbf{I}}G$ is better known as *Lie algebra*, denoted by $\mathfrak{g}$. This space is in bijection with $\mathbb{R}^m$, as illustrated in Figure 2.1. The Lie group G and its Lie algebra $\mathfrak{g}$ are locally connected via the exponential map $\mathrm{Exp}_G(\cdot)$ and its inverse, the logarithm map.

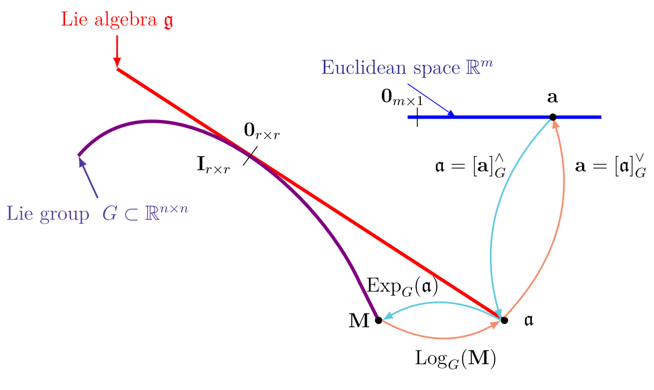

Fig. 1. Relation between G, $\mathfrak{g}$ and $\mathbb{R}^m$. We denote the exponential and logarithm applications such as $\forall\, \mathbf{a} \in \mathbb{R}^m$, $\mathrm{Exp}^\wedge_G(\mathbf{a}) = \mathrm{Exp}([\mathbf{a}]^\wedge_G)$ with the bijection $[.]^\wedge : \mathbb{R}^m \mapsto \mathfrak{g}$ and $\forall\, \mathbf{M} \in G, [\mathrm{Log}_G(\mathbf{M})]^\vee_G = \mathrm{Log}^\vee_G(\mathbf{M})$ with the bijection $[.]^\vee : \mathfrak{g} \mapsto \mathbb{R}^m$

2.2 Non-commutativity

In general, the composition of two elements of an LG is not commutative. Indeed, it is not trivial to find $\mathbf{z}$ such that $\mathrm{Exp}^\wedge_G(\mathbf{z}) = \mathrm{Exp}^\wedge_G(\mathbf{a})\ \mathrm{Exp}^\wedge_G(\mathbf{b})$, with $\mathbf{z}$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$. The Baker-Campbell-Hausdorff (BCH) formula [Miller (1973)] provides an analytical solution to this problem.

$$\mathrm{Log}^\vee_G\left(\mathrm{Exp}^\wedge_G(\mathbf{a})\ \mathrm{Exp}^\wedge_G(\mathbf{b})\right) = \mathbf{b} + \mathbf{\Psi}_G(\mathbf{b})\mathbf{a} + O(\|\mathbf{a}\|^2), \tag{1}$$

where $\mathbf{\Psi}_G(\mathbf{b}) = \mathbf{\Phi}_G(\mathbf{b})^{-1}$ is the inverse of the left Jacobian of the LG G given by

$$\mathbf{\Phi}_G(\mathbf{b}) = \sum_{n=0}^{+\infty} \frac{ad_G(\mathbf{b})^n}{(n+1)!}, \quad \forall \mathbf{b} \in \mathbb{R}^m, \tag{2}$$

and $ad_G(\mathbf{a})\mathbf{b} = \mathbf{ba} - \mathbf{ab}$ is the adjoint operator on the Lie algebra. Note that the closed-form expression of $\mathbf{\Phi}_G(.)$. and $\mathbf{\Psi}_G(.)$ are generally known for most of LG of interest (for instance $SO(n)$ and $SE(n)$). We will work under this assumption in our following developments.

2.3 Estimation on Lie groups

Gaussian distribution on Lie groups To perform inference on LGs, it is essential to define probability density functions (pdfs) directly on these manifolds. Among the various models proposed in the literature, we focus on the *Gaussian distribution on LGs* (LG-GD), which generalizes the classical multivariate Gaussian distribution to the Lie group setting. This distribution exhibits several useful properties, such as enabling entropy minimization under specific constraints [Chirikjian (2011)].

Samples from LG-GD can be generated straightforwardly. Let $\boldsymbol{\epsilon}$ be a zero-mean Gaussian vector in $\mathbb{R}^m$ with covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{m\times m}$, and let $\mathbf{M}$ be a reference element on a LG G of intrinsic dimension m. The LG-GD of $\mathbf{Z} = \mathbf{M}\mathrm{Exp}^\wedge_G(\boldsymbol{\epsilon}) \in G$ is given by:

$$p(\mathbf{Z}|\mathbf{M}, \mathbf{\Sigma}) = \frac{a(\mathbf{Z}, \mathbf{M})}{\sqrt{(2\pi)^m}} \exp\left(-\frac{1}{2}\, \|\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}})\|^2_{\mathbf{\Sigma}}\right), \tag{3}$$

where $\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}} \triangleq \mathrm{Log}^\vee_G\left(\mathbf{M}^{-1}\mathbf{Z}\right)$ and $\|\cdot\|^2_{\mathbf{\Sigma}}$ denotes the Mahalanobis distance. The multiplicative constant is defined as

$$a(\mathbf{M}, \mathbf{Z}) = \left|\mathbf{\Phi}_G(-\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}})\, \mathbf{\Sigma}\, \mathbf{\Phi}_G(-\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}})^\top\right|^{-\frac{1}{2}} \tag{4}$$

where $|\cdot|$ denotes the determinant.

Cramér-Rao bound Let us consider an LG estimator $\widehat{\mathbf{M}}$ of $\mathbf{M} \in G$, constructed from a set of independent observations $\mathbf{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_N\} \in G^N$. Analogously to the Euclidean case, the LG-MSE is given by $\mathbb{E}\left(\|\boldsymbol{l}^{\widehat{\mathbf{M}}}_{\mathbf{M}}\|^2\right)$ and under unbiasedness condition, i.e. $\mathbb{E}\left(\boldsymbol{l}^{\widehat{\mathbf{M}}}_{\mathbf{M}}\right) = \mathbf{0}$, it is bounded, in the sense of the Loewner ordering, by the following Cramér-Rao bound on LG, called LG-CRB [Labsir et al. (2023)].

Definition 2.1. (**LG-CRB**). The generic expression of the LG-CRB is given by

$$\mathbf{P}_{\text{LG-CRB}} = \mathbb{E}_{p(\mathbf{Z}|\mathbf{X})}\left(\mathbf{s}(\mathbf{Z}|\mathbf{X})\mathbf{s}(\mathbf{Z}|\mathbf{X})^\top\right)^{-1} \tag{5}$$

where $\mathbf{s}(\mathbf{Z}|\mathbf{X}) = \left.\frac{\partial \log p(\mathbf{Z}|\mathbf{X}\mathrm{Exp}^\wedge_G(\boldsymbol{\delta}))}{\partial \boldsymbol{\delta}}\right|_{\boldsymbol{\delta}=\mathbf{0}}$.

In the case of the LG-GD, each observation verifies

$$\mathbf{Z}_i = \mathbf{M}\,\mathrm{Exp}_G^{\wedge}(\boldsymbol{\epsilon}_i)\ ,\ \ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}) \tag{6}$$

and we have the following closed-form expression.

Definition 2.2. (**LG-CRB for LG-GD**). *The LG-CRB on* $\mathbf{M}$ *for the observation model* (6) *is given by:*

$$\mathbf{P}_{\text{LG-CRB}} = \mathcal{I}^{-1} \tag{7}$$

$$\mathcal{I} = \sum_{i=1}^{N} \boldsymbol{\Psi}_G(\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}})^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Psi}_G(\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}}) \tag{8}$$

2.4 Lie group product with $\mathbf{M}$ *and* $\boldsymbol{\Sigma}$

Let us consider a matrix LG $(G \subset \mathbb{R}^{n\times n}, *)$ with dimension m equipped with the matrix multiplication $*$. In what follows, we will omit this symbol as it is trivial. Let us also consider the LG of the symmetric positive definite (SPD) matrices of size s, equipped with the law $\odot$ defined as [Arsigny et al. (2006)]:

$$\boldsymbol{\Sigma}_1 \odot \boldsymbol{\Sigma}_2 = \mathrm{Expm}\left(\mathrm{Logm}(\boldsymbol{\Sigma}_1) + (\mathrm{Logm}(\boldsymbol{\Sigma}_2)\right) \tag{9}$$

$$= \mathrm{Expm}\left(sym(\boldsymbol{\epsilon}_1^{\boldsymbol{\Sigma}}) + sym(\boldsymbol{\epsilon}_2^{\boldsymbol{\Sigma}})\right), \tag{10}$$

where *sym* refers to the symmetric operator transforming $\boldsymbol{\epsilon} \in \mathbb{R}^p$ with $p = \dfrac{s\,(s+1)}{2}$ to a symmetric matrix. Expm and Logm are respectively the matrix exponential and logarithm. The product space $G' = G \times \mathcal{P}^+(s)$ is also a LG equipped with law $\oplus$ such as $\forall\ \mathbf{X}_i = \begin{bmatrix}\mathbf{M}_i & \mathbf{0}_{n\times s}\\ \mathbf{0}_{s\times n} & \boldsymbol{\Sigma}_i\end{bmatrix} \in G \times \mathcal{P}^+(s)$, with $i \in \{1,2\}$

$$\mathbf{X}_1 \oplus \mathbf{X}_2 = \begin{bmatrix}\mathbf{M}_1\mathbf{M}_2 & \mathbf{0}_{n\times s}\\ \mathbf{0}_{s\times n} & \boldsymbol{\Sigma}_1 \odot \boldsymbol{\Sigma}_2.\end{bmatrix}. \tag{11}$$

The group logarithm is then defined as

$$\mathrm{Log}_{G'}^{\vee}(\mathbf{X}_1 \oplus \mathbf{X}_2) = \begin{bmatrix}\mathrm{Log}_G^{\vee}(\mathbf{M}_1\,\mathbf{M}_2)\\ sym^{-1}(\mathrm{Logm}(\boldsymbol{\Sigma}_1)) + sym^{-1}(\mathrm{Logm}(\boldsymbol{\Sigma}_1))\end{bmatrix} \tag{12}$$

where sym^{-1} is the reciprocal of *sym*.

Definition 2.3. (**Cramér-Rao bound with SPD matrices**). If the observations follow the model (6), the LG-CRB on $G' = G \times \mathcal{P}^+(s)$ which bounds

$$\mathbb{E}\left(\|\mathrm{Log}_{G'}^{\vee}\left(\mathbf{X}^{-1} \oplus \widehat{\mathbf{X}}\right)\ \|^2\right) \tag{13}$$

is

$$\mathbf{P}'_{\text{LG-CRB}} = \mathcal{I}_{G'}^{-1} \tag{14}$$

$$\mathcal{I}_{G'} = \begin{bmatrix}\mathcal{I} & \mathbf{0}_{m\times p}\\ \mathbf{0}_{p\times m} & \mathcal{I}_\Sigma\end{bmatrix}, \tag{15}$$

where $\mathcal{I}$ is given by (8) and

$$\mathcal{I}_\Sigma = \frac{N}{2}\,\mathrm{diag}\left[\underbrace{1,\ldots,1}_{s},\underbrace{2,\ldots,2}_{\frac{s\,(s-1)}{2}}\right]. \tag{16}$$

This proposed LG structure including SPD matrices will be the heart of the following section.

3. NEW MAXIMUM LIKELIHOOD ESTIMATOR

In this section, we first develop the proposed maximum likelihood estimator for both LG-mean and LG-covariance of observations modeled by a Lie group Gaussian distribution. We then prove that, under realistic assumptions, the resulting numerical estimator converges to a unique solution. Note that due to space limitations, the proofs of proposed theorems will be given in an extended version of this work.

3.1 Problem formulation

Let us consider a set of independent LG observations $\{\mathbf{Z}_i\}_{i=1}^N$ ($\mathbf{Z}_i \in G$ with dimension m) drawn from the following Gaussian model with $\mathbf{M} \in G \subset \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathcal{P}^+(s)$

$$p(\mathbf{Z}_i|\mathbf{M},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\,\pi)^m |\boldsymbol{\Phi}_G(-\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}})\boldsymbol{\Sigma}\,\boldsymbol{\Phi}_G(-\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}})^\top|}} \exp\left(-\frac{1}{2}\|\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}}\|_{\boldsymbol{\Sigma}}^2\right) \tag{17}$$

where we define $\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}} \triangleq \mathrm{Log}_G^{\vee}\left(\mathbf{M}^{-1}\mathbf{Z}_i\right)$ [1].

We aim to determine the maximum likelihood estimator of $\mathbf{M}$ and $\boldsymbol{\Sigma}$. To obtain an unified formalism taking into account both geometry of each parameter, we propose to embed them on the LG $G' = G \times \mathcal{P}^+(s)$. Thus, G' is with dimension $p' = m + \frac{m(m+1)}{2}$. It results

$$\mathbf{X} = \begin{bmatrix}\mathbf{M} & \mathbf{0}_{n\times s}\\ \mathbf{0}_{s\times n} & \boldsymbol{\Sigma}\end{bmatrix} \in G',$$

$$\widehat{\mathbf{X}} = \underset{\mathbf{X}\in G'}{\mathrm{argmax}} \prod_{i=1}^{N} p(\mathbf{Z}_i|\mathbf{M},\boldsymbol{\Sigma}). \tag{18}$$

Unfortunately, the maximization problem (18) is not resolvable analytically. To overcome that, we can rewrite it as the following minimization problem by using the logarithm on (18)

$$\widehat{\mathbf{X}} = \underset{\mathbf{X}\in G'}{\mathrm{argmin}}\ J(\mathbf{X}) \tag{19}$$

with

$$J(\mathbf{X}) = J(\mathbf{M},\boldsymbol{\Sigma}) = F(\mathbf{M},\boldsymbol{\Sigma}) + H(\mathbf{M},\boldsymbol{\Sigma}) \tag{20}$$

and

$$F(\mathbf{M},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \log|\boldsymbol{\Phi}_G(-\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}})\boldsymbol{\Sigma}\,\boldsymbol{\Phi}_G(-\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}})^\top| \tag{21}$$

$$H(\mathbf{M},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \|\boldsymbol{l}_{\mathbf{Z}_i}^{\mathbf{M}}\|_{\boldsymbol{\Sigma}}^2. \tag{22}$$

Several solutions are available to find the minimum: we propose to focus on the descent gradient for its simplicity of implementation and its low computational cost. Analogously to the standard gradient descent algorithm, the principle is to build a set of iterates $\{\mathbf{X}^{(t)} \in G'\}$.

$$\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)} \oplus \mathrm{Exp}_{G'}^{\wedge}\left(\alpha\,\boldsymbol{\delta}^{(t)}\right) \tag{23}$$

where $\alpha \in \mathbb{R}^+$ is the step size and $\boldsymbol{\delta}^{(t)} \in \mathbb{R}^{p'}$ is the descent direction. It is given by the opposite of the LG gradient of $J(\mathbf{X}^{(t)})$:

[1] In the following, we will also define $\boldsymbol{l}_{\mathbf{A}}^{\mathbf{B}} = \mathrm{Log}_G^{\vee}\left(\mathbf{A}^{-1}\mathbf{B}\right)\ \forall \mathbf{A},\mathbf{B} \in G$.

$$\boldsymbol{\delta}^{(t)} = - \left. \frac{\partial J(\mathbf{X}^{(t)} \oplus \mathrm{Exp}^{\wedge}_{G'}(\boldsymbol{\epsilon}))}{\partial \boldsymbol{\epsilon}} \right|_{\boldsymbol{\epsilon}=\mathbf{0}} \tag{24}$$

that we propose to compute analytically in the following subsection.

3.2 Closed-form expression of the intrinsic gradient

Note that the gradient can be divided into two parts

$$\left. \frac{\partial J(\mathbf{X} \oplus \mathrm{Exp}^{\wedge}_{G'}(\boldsymbol{\epsilon}))}{\partial \boldsymbol{\epsilon}} \right|_{\boldsymbol{\epsilon}=\mathbf{0}} = \begin{bmatrix} \left. \frac{\partial J(\mathbf{M}\mathrm{Exp}^{\wedge}_{G}(\boldsymbol{\epsilon}_M), \boldsymbol{\Sigma})}{\partial \boldsymbol{\epsilon}_M} \right|_{\boldsymbol{\epsilon}_M=\mathbf{0}} \\ \left. \frac{\partial J(\mathbf{M}, \boldsymbol{\Sigma} \odot \mathrm{Exp}^{\wedge}_{\mathcal{P}+(s)}(\boldsymbol{\epsilon}_\Sigma))}{\partial \boldsymbol{\epsilon}_\Sigma} \right|_{\boldsymbol{\epsilon}_\Sigma=\mathbf{0}} \end{bmatrix} \tag{25}$$

The following theorem provides the associated expressions

Theorem 1. **(Computation of the intrinsic gradient)** *The gradient defined by* (25) *is computed analytically by*

$$\left. \frac{\partial J(\mathbf{M}\,\mathrm{Exp}^{\wedge}_{G}(\boldsymbol{\epsilon}_M), \boldsymbol{\Sigma})}{\partial \boldsymbol{\epsilon}_M} \right|_{\boldsymbol{\epsilon}_M=\mathbf{0}} = \sum_{i=1}^{N} \boldsymbol{\Psi}_G(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i})^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i} - \boldsymbol{\Psi}_G(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i}) \frac{\partial}{\partial \mathbf{w}} \log|\boldsymbol{\Phi}_G(-\mathbf{w})| \,|_{\mathbf{w}=\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i}} \tag{26}$$

$$\left[\frac{\partial J(\mathbf{M}, \boldsymbol{\Sigma} \odot \mathrm{Exp}^{\wedge}_{\mathcal{P}+(s)}(\boldsymbol{\epsilon}_\Sigma))}{\partial \boldsymbol{\epsilon}_\Sigma} \right]_l \Bigg|_{\boldsymbol{\epsilon}_\Sigma=\mathbf{0}} = - \left(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i}\right)^\top \mathbf{G}_l \boldsymbol{\Sigma}^{-1} \boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i} + \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{G}_l\right) \tag{27}$$

with $\{\mathbf{G}_l\}_{l=1}^p$ *a basis of the Lie algebra of* $\mathcal{P}^+(s)$.

3.3 Convergence

In the following, we provide condition of convergence of the proposed gradient descent algorithm (26).

Theorem 2. (**Condition of convergence**). *The set of values* $\{\mathbf{X}^{(t)}\}$ *following the recursion* (23) *converges to a minimum* $\mathbf{X}^\star$ *of the criterion* J, *i.e.*

$$\lim_{t \to +\infty} J(\mathbf{X}^{(t)}) = J(\mathbf{X}^\star) \tag{28}$$

This minimum is unique under the following assumption

$$|\mathbf{D} - \mathbf{B}^\top \mathbf{A}^{-1}\mathbf{B}| > 0 \;\; \forall \mathbf{X} \in G' \tag{29}$$

with

$$[\mathbf{D}]_{k,l} = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}\mathbf{G}_k\mathbf{G}_l\right) \forall (k,l) \in \{1,\ldots,p\}^2 \tag{30}$$

$$\tilde{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i} \boldsymbol{l}^{\mathbf{M}\,\top}_{\mathbf{Z}_i} \tag{31}$$

$$\left[\mathbf{B}^\top \mathbf{A}^{-1}\mathbf{B}\right]_{i,j} = \sum_{r=1,s=1,q=1}^{N} \sum_{k=1}^{p}\sum_{l=1}^{p} b^r_{i,k}\, a^s_{k,l} b^q_{l,j} \quad \forall (i,j) \in \{1,\ldots,p\}^2, \tag{32}$$

and

$$b^r_{i,k} = \left[\boldsymbol{\Psi}_G(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i})\boldsymbol{\Sigma}^{-1}\mathbf{G}_k \boldsymbol{l}_G\right]_i \tag{33}$$

$$a^s_{k,l} = \left[\boldsymbol{\Psi}_G(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i})^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Psi}_G(\boldsymbol{l}^{\mathbf{M}}_{\mathbf{Z}_i})\right]_{k,l} \tag{34}$$

Corollary 3.1. (**Convexity condition: particular case**) The assumption in (29) is always satisfied under the following conditions:

(1) $\boldsymbol{\Sigma} \approx \tilde{\boldsymbol{\Sigma}}$,
(2) the eigenvalues of $\boldsymbol{\Sigma}$ are not excessively large.

3.4 Discussion

Note that the conditions assumed in Corollary (3.1) are realistic when sensor measurements are sufficiently accurate. Indeed, $\boldsymbol{\Sigma}$ is close to $\tilde{\boldsymbol{\Sigma}}$ if the residual between $\mathbf{M}$ and each $\mathbf{Z}_i$ is very low. From a numerical standpoint, when this assumption is well justified, our criterion can be considered convex i.e. convergence of our algorithm can be assumed. However, if the observation variance is excessively large, the criterion to optimize may contain several local minima. Consequently, there is no guarantee that the obtained solution corresponds to the global minimum. Finally, from a numerical perspective, it is important to pay attention to the variance of the simulated data to obtain an efficient estimator

4. SIMULATION RESULTS

In this section, we propose to numerically analyze the proposed maximum likelihood estimator by considering LG observations simulated on the LG $SE(2)$. This LG is the semi-product of the LG of rotation matrix $SO(2)$ and $\mathbb{R}^2$ i.e.

$$SE(2) = \left\{ \mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix} | \mathbf{R} \in SO(2), \mathbf{p} \in \mathbb{R}^2 \right\}. \tag{35}$$

Using $SE(2)$ is crucial from an application point of view because it models unknown affine transformation in robots, vision or control [Remsing (2011)].

4.1 Simulation protocol

We assume a sensor (for instance a star sensor) measuring several noisy attitude and position of a satellite. It ensues that the unknown parameter $\mathbf{M} \in SE(2)$ and the measurement model is $\forall i \in \{1,\ldots,N\}$

$$\mathbf{Z}_i = \mathbf{M}\mathrm{Exp}^{\wedge}_{SE(2)}(\boldsymbol{\epsilon}_i) \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \boldsymbol{\Sigma} \in \mathcal{P}^+(3) \tag{36}$$

First, we simulate synthetic observations according to the model (36) with $\mathbf{M} = \begin{bmatrix} \mathbf{I}_{2\times 2} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ with $\mathbf{U} = \mathrm{Exp}^{\wedge}_{SO(3)}([0.1, 0.1, 0.1])$ and $\mathbf{D} = \mathrm{diag}\left(\left[0.4.^2, 0.1^2, 0.1^2\right]\right)$. Each value of $\mathbf{D}$ represents the variance of each components of $\mathbf{Z}_i$. Note that $\boldsymbol{\Sigma}$ is chosen to satisfy the assumptions given by the corollary (3.1).

Then, we implement our proposed estimator of $\mathbf{X} = \begin{bmatrix} \mathbf{M} & \mathbf{0}_{4\times 3} \\ \mathbf{0}_{3\times 4} & \boldsymbol{\Sigma} \end{bmatrix}$ called $\widehat{\mathbf{X}}_{LG} = \begin{bmatrix} \widehat{\mathbf{M}}_{LG} & \mathbf{0}_{4\times 3} \\ \mathbf{0}_{3\times 4} & \widehat{\boldsymbol{\Sigma}}_{LG} \end{bmatrix}$ on the LG $G' = SE(2) \times \mathcal{P}^+(3)$, It is compared to three state-of-the-art estimators of $\mathbf{X}$, all using the same gradient descent algorithm on $SE(2)$ to estimate $\mathbf{M}$ but updating $\boldsymbol{\Sigma}$ in a different manner at each iteration t:

- the first naive approach consists in computing the empirical covariance by considering an Euclidean approximation of $SE(2)$,

$$\widehat{\boldsymbol{\Sigma}}^{(t)}_{Eucl} = \frac{1}{N}\sum_{i=1}^{N} \left(\mathrm{Log}^{\vee}_{SE(2)}(\mathbf{Z}_i) - \mathrm{Log}^{\vee}_{SE(2)}\left(\widehat{\mathbf{M}}^{(t)}_{Eucl}\right)\right) * \left(\mathrm{Log}^{\vee}_{SE(2)}(\mathbf{Z}_i) - \mathrm{Log}^{\vee}_{SE(2)}\left(\widehat{\mathbf{M}}^{(t)}_{Eucl}\right)\right)^\top \tag{37}$$

where $\widehat{\mathbf{M}}^{(t)}_{Eucl}$ is the current estimation of $\mathbf{M}$ depending on $\widehat{\mathbf{\Sigma}}^{(t-1)}_{Eucl}$.

- the second considers another empirical covariance integrating the LG structure of $SE(2)$

$$\widehat{\mathbf{\Sigma}}^{(t)}_{emp} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{l}^{\mathbf{Z}_i}_{\mathbf{M}^{(t)}_{emp}} \boldsymbol{l}^{\mathbf{Z}_i}_{\mathbf{M}^{(t)}_{emp}}{}^{\top} \tag{38}$$

where $\widehat{\mathbf{M}}^{(t)}_{emp}$ is the current estimation of $\mathbf{M}$ depending on $\widehat{\mathbf{\Sigma}}^{(t-1)}_{emp}$.

- the third more probably more competitive with our approach, consists in using the Riemannian structure of the covariance matrix. It ensues that $\mathbf{\Sigma}$ is optimized with a Riemannian gradient descent algorithm [Zhou et al. (2022)].

$$\widehat{\mathbf{\Sigma}}^{(t+1)}_{Riem} = \widehat{\mathbf{\Sigma}}^{(t)}_{Riem}{}^{1/2} * \\ \mathrm{Expm}\left(-\alpha\, \widehat{\mathbf{\Sigma}}^{(t)}_{Riem}{}^{-1/2}\, \boldsymbol{\nabla} J(\widehat{\mathbf{\Sigma}}^{(t)}_{Riem})\, \widehat{\mathbf{\Sigma}}^{(t)}_{Riem}{}^{-1/2}\right) \widehat{\mathbf{\Sigma}}^{(t)}_{Riem}{}^{1/2} \tag{39}$$

where where α is a step size. $\widehat{\mathbf{\Sigma}}^{(t)}_{Riem}$ depends on $\widehat{\mathbf{M}}^{(t)}_{Riem}$ the current estimation of $\mathbf{M}$. $\boldsymbol{\nabla} J(\mathbf{\Sigma}^{(t)}_{Riem})$ is the Riemannian gradient of J computed at $\mathbf{\Sigma}^{(t)}_{Riem}$.

4.2 Results

The performance of each estimator is assessed by computing the LG-MSE of $\mathbf{M}$ and $\mathbf{\Sigma}$ given by:

$$\text{LG-MSE}(\mathbf{M}, \widehat{\mathbf{M}}) = \mathbb{E}\left(\|\mathrm{Log}^{\vee}_{SE(2)}\left(\mathbf{M}^{-1}\widehat{\mathbf{M}}\right)\|^2\right) \tag{40}$$

$$\text{LG-MSE}(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}) = \mathbb{E}\left(\|\mathrm{Logm}(\mathbf{\Sigma}) - \mathrm{Logm}(\widehat{\mathbf{\Sigma}})\|^2\right) \tag{41}$$

and are compared to the LG-CRB of $\mathbf{M}$ and $\mathbf{\Sigma}$ given by (14).

In the Fig. 2, we first plot the LG-MSE of $\mathbf{M}$ for different set of observation N. We observe that our approach is the best in the terms of accuracy. Whereas the estimators $\widehat{\mathbf{M}}_{Eucl}$ and $\widehat{\mathbf{M}}_{emp}$ diverge brutally for a number of observations close to 5, our estimator of $\widehat{\mathbf{M}}_{LG}$ is robust and provides high precision for few observations. Additionally, it remains slightly better that the Riemannian estimator $\widehat{\mathbf{M}}_{Riem}$ and provides a precision always better as the number of observations increases. We also draw the LG-CRB of $\mathbf{M}$ given by (14). Note that estimators does not achieve the LG-CRB for high number of observations, but tends to be closer. This is due to the correlation induced by joint estimation of $\mathbf{\Sigma}$ degrading the performance compared to the case where $\mathbf{\Sigma}$ would be known. In the Fig. 3, we plot the MSE on $\mathbf{\Sigma}$ for the four approaches as a function of the number of observation along with the LG-CRB. We observe an interesting behavior regarding our approach in the case where few measurements are available. Indeed, the precision tends to be better whereas the other approaches tends to clearly diverges. Regarding the Riemannian estimation, it remains comparable but less precise for high number of observations, in the same way as for the estimation $\mathbf{M}$. This behavior is explained by the fact that our approach not only preserves the geometric structure, but also better captures the correlation between the mean and the covariance contrary to the the Riemannian case. Note the LG-CRB of $\mathbf{\Sigma}$ is relatively distant from the estimators which is not surprising because the proposed estimators are sub-optimal in the sense that it does not minimize the mean square error. Additionally, the LG-CRB is assumed unbiased, whereas in the the case where only $\mathbf{M}$ should be estimated, this assumption is respected, it is probably not the case for the joint estimation of $\mathbf{M}$ and $\mathbf{\Sigma}$. Such observation is consistent with the Riemmnian estimator of the covariance matrices developed in [Smith (2005)] where the intrinsic CRB is never attained and an explicit analytical expression for the bias is provided. We propose also to verify numerically the condition of convergence established in 3.1 by drawing the optimization criterion computed for each considered estimator. In the Fig. 4, we remark that the criterion converges only with our proposed approach ensuring that we achieve global minimum as proved in section which is not the case for the other estimators. Such behavior is crucial because it guarantees that the obtained solution is stable and reliable.

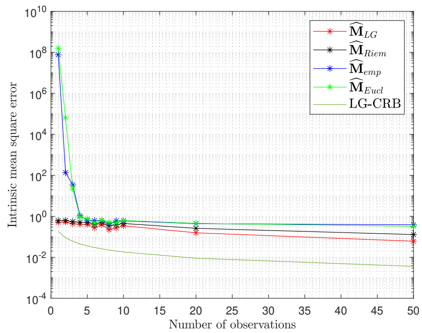

Fig. 2. LG-MSE of $\widehat{\mathbf{M}}$ for each estimator $\widehat{\mathbf{M}}_{\text{Riem}}$, $\widehat{\boldsymbol{M}}_{\text{Eucl}}$, $\widehat{\mathbf{M}}_{\text{LG}}$ and $\widehat{\mathbf{M}}_{\text{emp}}$

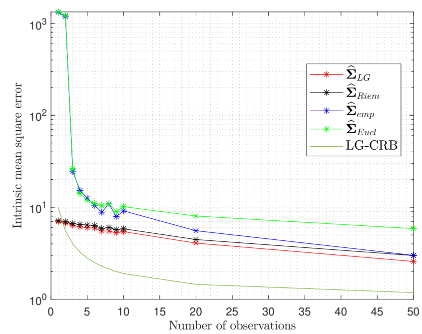

Fig. 3. LG-MSE of $\widehat{\mathbf{\Sigma}}$ for each estimator $\widehat{\mathbf{\Sigma}}_{\text{Riem}}$, $\widehat{\mathbf{\Sigma}}_{\text{Eucl}}$, $\widehat{\mathbf{\Sigma}}_{\text{LG}}$ and $\widehat{\mathbf{\Sigma}}_{\text{emp}}$.

Remark 1. Note that the proposed method has low computational complexity and is compatible with real-time implementation. Additionally, it can be directly applied to real-world datasets; experimental validation on such data is left for future work.

5. CONCLUSIONS AND PERSPECTIVES

In this communication, we propose a novel approach to determine the maximum likelihood of the LG-mean and the LG-covariance of a Gaussian distribution on Lie groups. It is achieved by designing a gradient descent algorithm on Lie groups. Theoretical gradient are computed then convergence and uniqueness of the solution are proved. Validity and relevance of the approach have been analyzed numerically by considering experiments on the LG $SE(2)$ Perspectives of these works are numerous: first it would be interesting to validate the approach for others LGs of interest. Second, framework of the proposed method should also be envisaged in a dynamic context, for instance with a Kalman-based filtering on LGs integrating the proposed algorithm at each instant.

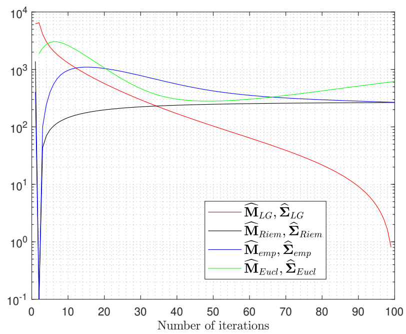

Fig. 4. Evolution of the criterion (20) for each estimator of $\mathbf{M}$ and $\mathbf{\Sigma}$ with $N = 50$.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author used CoPilot/Microsoft365 in order to verify the orthograph of the sentences. After using this tool the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-Euclidean framework for statistics on diffeomorphisms. volume 9, 924–31.

Barfoot, T.D. and Furgale, P.T. (2014). Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Transactions on Robotics*, 30, 679–693.

Barrau, A. and Bonnabel, S. (2017). The invariant extended Kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4), 1797–1812.

Boumal, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press.

Bourmaud, G. (2016). Online variational Bayesian motion averaging. In B. Leibe, J. Matas, N. Sebe, and M. Welling (eds.), *Computer Vision – ECCV 2016*, 126–142. Springer International Publishing, Cham.

Breloy, A., Ginolhac, G., Renaux, A., and Bouchard, F. (2019). Intrinsic Cramér-Rao Bounds for Scatter and Shape Matrices Estimation in CES Distributions. *IEEE Sig. Proc. Letters*, 26(2), 262–266.

Brossard, M., Barrau, A., Chauchat, P., and Bonnabel, S. (2022). Associating uncertainty to extended poses for on Lie group IMU preintegration with rotating earth. *IEEE Transactions on Robotics*, 38(2), 998–1015.

Chirikjian, G.S. (2011). *Stochastic models, information theory, and Lie groups, volume 2: Analytic methods and modern applications*, volume 2. Springer Science & Business Media.

da Silva, L.C., Dias, S.S., and da Silva Bruno, M.G. (2025). Error-State Kalman Filter for Autonomous Celestial-Based Navigation. In *2025 28th International Conference on Information Fusion (FUSION)*, 1–8.

Fan, J., Liao, Y., and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39, 3320–3356.

Faraut, J. (2008). *Analysis on Lie Groups: An Introduction*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Gupta, A. and Nagar, D. (2017). On testing structure of covariance matrix and mean vector of a complex multivariate Gaussian model. *Demonstratio Mathematica [electronic only]*, 18.

Labsir, S., Giremus, A., Yver, B., and Benoudiba–Campanini, T. (2021). Joint shape and centroid position tracking of a cluster of space debris by filtering on Lie groups. *Signal Processing*, 183, 108027.

Labsir, S., El Bouch, S., Renaux, A., Vilà-Valls, J., and Chaumette, E. (2025). Cramér-Rao bound for Lie group parameter estimation with euclidean observations and unknown covariance matrix. *IEEE Transactions on Signal Processing*, 73, 130–141.

Labsir, S., Renaux, A., Vilà-Valls, J., and Chaumette, E. (2023). Barankin, McAulay–Seidman and Cramér–Rao bounds on matrix Lie groups. *Automatica*, 156, 111199.

Miller, W. (1973). *Symmetry groups and their applications*, volume 50. Academic Press.

Ollila, E., Palomar, D.P., and Pascal, F. (2021). Shrinking the eigenvalues of M-estimators of covariance matrix. *IEEE Transactions on Signal Processing*, 69, 256–269.

Remsing, C.C. (2011). Control and stability on the Euclidean group SE(2). In *Proceedings of ICAEM 2011*, 225–230. London, UK.

Sasiadek, J. (2001). Sensor fusion and adaptive Kalman filtering for guidance, navigation, and control. In *IFAC Professional Brief*. International Federation of Automatic Control.

Smith, S. (2005). Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Transactions on Signal Processing*, 53(5), 1610–1630.

Taylor, C.J. and Kriegman, D.J. (1994). Minimization on the Lie group SO(3) and related manifolds. Technical Report Technical Report No. 9405, Yale Center for Systems Science.

Xu, H., Pei, F., and Jiang, N. (2020). An attitude estimation method based on Lie group representation for deep space probe. *Journal of Deep Space Exploration*, 7(1).

Zhang, L., Sidoti, D., Bienkowski, A., Pattipati, K.R., Bar-Shalom, Y., and Kleinman, D.L. (2020). On the identification of noise covariances and adaptive Kalman filtering: A new look at a 50 year-old problem. *IEEE Access*, 8, 59362–59388.

Zhou, J., Said, S., and Berthoumieu, Y. (2022). Riemannian information gradient methods for the parameter estimation of ECD. *Signal Processing*, 192, 108376.