Extrinsic Mean of Brownian Distributions on Compact Lie Groups

Salem Said and Jonathan H. Manton, *Senior Member, IEEE*

***Abstract*—This paper studies Brownian distributions on compact Lie groups. These are defined as the marginal distributions of Brownian processes and are intended as a natural extension of the well-known normal distributions to compact Lie groups. It is shown that this definition preserves key properties of normal distributions. In particular, Brownian distributions transform in a nice way under group operations and satisfy an extension of the central limit theorem. Brownian distributions on a compact Lie group G belong to one of two parametric families $N_L(g, C)$ and $N_R(g, C)$—$g \in G$ and C a positive-definite symmetric matrix. In particular, the parameter g appears as a location parameter. An approach based on the extrinsic mean for estimation of the parameters g and C is studied in detail. It is shown that g is the unique extrinsic mean for a Brownian distribution $N_L(g, C)$ or $N_R(g, C)$. Resulting estimates are proved to be consistent and asymptotically normal. While they may also be used to simultaneously estimate g and C, it is seen this requires that G be embedded into a higher dimensional matrix Lie group. Going beyond Brownian distributions, it is shown the extrinsic mean can be used to recover the location parameter for a wider class of distributions arising more generally from Lévy processes. The compact Lie group structure places limitations on the analogy between normal distributions and Brownian distributions. This is illustrated by the study of multivariate Brownian distributions. These are introduced as Brownian distributions on some product group—e.g., $G \times G$. This paper describes their covariance structure and considers its transformation under group operations.**

***Index Terms*—Brownian motion, central limit theorem, compact Lie group, extrinsic and intrinsic mean, noncommutative harmonic analysis, Lévy process.**

I. Introduction

THIS paper considers the marginal distributions of Brownian processes in compact Lie groups, which it refers to as Brownian distributions. On a compact Lie group G, Brownian distributions belong to one of two parametric families $N_L(g, C)$ and $N_R(g, C)$. In general, $g \in G$ and C is a positive-definite symmetric matrix. The existence of two different parametric families reflects the fact that G may be noncommutative. The introduction of Brownian distributions is motivated by the desire to extend normal distributions to compact Lie groups. While the standard definition of normal distributions carries no explicit reference to Brownian motion, essential properties of normal distributions mirror those of Brownian motion. In particular, normal distributions can be obtained as marginal distributions of Brownian motion. In the literature of stochastic processes in Lie groups, Brownian processes are already well known and it is accepted they constitute the correct generalization of Brownian motion to Lie groups.

The analogy between normal distributions and Brownian distributions is developed in Section III. This section derives transformation properties of Brownian distributions under group operations. It also proves Brownian distributions arise through an extension of the central limit theorem—see Proposition 3. Transformation properties can be used to establish the role played by parameters g and C; in particular, g is seen to be a location parameter. As for the classical theorem, the desired extension of the central limit theorem is associated with the property of infinite divisibility.

The central problem studied in this paper is estimation of the parameters g and C. The proposed approach is detailed in Section IV, based on the extrinsic mean. In Section III, the role of g as a location parameter is established. As a result, estimation of g can be seen as a special case of the following problem. Let Z be a random variable with values in G whose distribution is known. It is required to estimate a parameter $g \in G$ from an observed $Y = gZ$. In Section IV, $Y \sim N_L(g, C)$ and it is possible to take $Z = N_L(e, C)$, where e is the identity element of G. In this case, Proposition 4 in Section IV-A shows g to be the unique extrinsic mean of the distribution $N_L(g, C)$ of Y. Section VIII is concerned with the generalization of this result. In particular, Proposition 9 shows g is the unique extrinsic mean of the distribution of Y whenever the distribution of Z is the marginal distribution of an inverse invariant Lévy process in G. Propositions 4 and 9 are the main results in this paper, with respect to the use of the extrinsic mean. For the definition of the extrinsic mean, as used below, see (22) in Section IV-A. In particular, this depends on a chosen unitary representation $U : G \to SU(v)$.

Section IV gives asymptotic properties of estimates based on the extrinsic mean, which it shown to be consistent and asymptotically normal. Improving the rate of convergence is possible by reducing the dimension v of U. However, v is bounded below as it must be ensured U is injective. This aspect is given in Proposition 5 and the ensuing discussion. Computation of the proposed estimates is particularly simple, requiring only a polar decomposition. In principle, this can be used to simultaneously estimate the matrix C. It is seen this requires that G be embedded into a higher dimensional matrix Lie group. This is explained in Section IV-A, based on an elementary dimension

Manuscript received February 04, 2011; revised September 29, 2011; accepted January 09, 2012. Date of publication January 31, 2012; date of current version May 15, 2012.

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Vic. 3010, Australia (e-mail: ssaid@unimelb.edu.au; jmanton@unimelb.edu.au).

Communicated by F. Hlawatsch, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2012.2185680

argument, and illustrated through the concrete example $G = SU(2)$ in Section VI. For this example, estimation of g and C follows only after taking $U : SU(2) \to SU(5)$.

Section V introduces multivariate Brownian distributions. These are defined as Brownian distributions on product groups and their covariance structure is given accordingly. Considering the product group $G \times G$, independence follows whenever the resulting covariance matrix is equal to zero. A precise statement is given in Proposition 8. Whereas this property is often exploited in the case of normal distributions, by applying appropriate linear transformations to obtain independence, a caveat is that similar tricks are in general hampered by the compact Lie groups structure. A novel aspect in the current paper is the introduction of the concept of joint characteristic function, for two random variables with values in G. This provides the technical basis for Section V, with Proposition 7 extending the classical Kac's theorem.

This paper complements a previous one [1], with both generalizing the results of [2]. In [1], a solution is provided for a problem of nonparametric estimation for compound Poisson processes in a compact Lie group. Brownian processes and compound Poisson processes are the two main types of Lévy processes in a Lie group [3]. An application of the example in Section VI to a signal processing situation is reviewed in Section VII. This provides a solution to the inverse problem of polarized light propagating in a dispersive optical fiber. The connection to inverse problems may be clarified by making the following observation. In Section II-A, g is first introduced as the initial condition of a Brownian process W; the distribution $N_L(g, C)$ is that of W_1—that is, of the same process W observed after one time unit.

The extrinsic mean, along with the complementary notion of intrinsic mean, has received sizable attention both from the statistics and the signal processing communities. In most cases, extrinsic and intrinsic means are considered on Riemannian manifolds. Here, the term intrinsic mean refers to the Karcher mean as considered in [4] and [5]. This is different from the established terminology of [6] and [7] (see Section IV-C). Specializing to compact Lie groups, a comprehensive framework for the computation of the sample intrinsic mean is given in [8]. This is applied to the problem of joint approximate diagonalization in [9]. In [6] and [7], the asymptotics of sample extrinsic and intrinsic means, on Riemannian manifolds, are given. Note that the intrinsic approach was first considered by Fréchet [10]

In most applications, the use of extrinsic means is more popular. This is due to nonuniqueness of the intrinsic mean and to the fact that the extrinsic mean is generally considerably easier to compute. In [11], Monte Carlo simulation is used to obtain extrinsic means in several classical compact Lie groups and quotient manifolds, which are of particular importance to signal processing—see [12] and [13]. In this paper, Section IV applies the extrinsic mean to parametric estimation of Brownian distributions. In addition to the parameter g, which coincides with the extrinsic mean, this can also allow the parameter C to be recovered. Section VIII generalizes the validity of the extrinsic mean beyond Brownian distributions. Instead of classical compact Lie groups, taken in their standard matrix form, the focus is on general abstract compact Lie groups. Section VI illustrates that this is pertinent even in an elementary applied context. Finally, Section IX investigates the possibility of computing estimates which improve upon those based on the extrinsic mean, while retaining their symmetry properties. Intrinsic means are briefly considered in Section IV-C. For a Brownian distribution $N_L(g, C)$, general conditions are discussed under which g is an intrinsic mean.

For the sake of coherence, all discussion in this paper is limited to the case where the underlying group G is compact. While this assumption is essential to Section IV, it can be overlooked, after some modifications, for the results of Sections III and V. The discussion in Section V may be considerably extended in this direction.

This paper concludes with an appendix, giving necessary lemmas for the proofs of Propositions 3 and 4.

II. Mathematical Background

In this section, mathematical background required for the following is captured. In Section II-A, Brownian processes in G are defined as solutions of (left or right) invariant Brownian stochastic differential equations. In Section II-B, the concept of characteristic function of a random variable with values in G is briefly recalled. This is further extended in Section V, to define the joint characteristic function of two random variables with values in G. Brownian processes are outlined in [14]. For a more self contained version of Section II-B, see [1]; a detailed and highly readable reference is [15] and a variety of engineering applications are discussed in [16]. A recent and comprehensive treatment of stochastic processes in Lie groups and their connection to information theory can be found in [17].

For computations relating to compact Lie group structure, as carried out in the following, see [18]. For general reference on stochastic differential equations on manifolds, see [19]. In reference to the Lévy property of Brownian processes, a self contained account can be found in [3].

A. Brownian Processes

Let G be a compact connected Lie group of dimension d with identity e and Lie algebra $\mathfrak{g}$. Since G is compact, it is possible to choose an Ad-invariant scalar product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$. Then, let $X_1, \ldots X_d$ be an orthonormal basis of $\mathfrak{g}$. In addition, suppose given a d-dimensional Brownian motion $B = (B_t)_{t \geq 0}$ with coordinates $(B^1, \ldots, B^d)$. Let $C = (C_{ij},\ 1 \leq i,\ j \leq d)$ be the covariance matrix of B. A left Brownian process in G is a process W verifying the Stratonovich equation

$$dW_t = \sum_{i=1}^{d} X_i^L(W_t) \circ dB_t^i, \qquad W_0 = g. \tag{1}$$

Here, $g \in G$ and L denotes the left invariant vector field. Precisely, (1) states that

$$df(W_t) = \sum_{i=1}^{d} [X_i^L f](W_t) \circ dB_t^i, \qquad f \in C^\infty(G). \tag{2}$$

The Itô form of this identity, obtained in the standard way, is the following:

$$df(W_t) = \sum_{i=1}^{d} [X_i^L f](W_t) dB_t^i + \frac{1}{2} \sum_{i,j=1}^{d} C_{ij} [X_i^L X_j^L f](W_t).$$

Taking expectations on both sides, the Brownian term disappears

$$\frac{d}{dt} \mathbb{E}[f(W_t)] = \mathbb{E}\left([\mathcal{D}f](W_t)\right) \tag{3}$$

where the operator $\mathcal{D}$ is left invariant, defined for $f \in C^\infty$ by

$$\mathcal{D}f = \frac{1}{2} \sum_{i,j=1}^{d} C_{ij} X_i^L X_j^L f. \tag{4}$$

A right Brownian process W' in G is defined as in (1). Here, R denotes the right invariant vector field

$$dW'_t = \sum_{i=1}^{d} X_i^R(W'_t) \circ dB_t^i, \qquad W'_0 = g. \tag{5}$$

The same development shows that (3) holds for the operator $\mathcal{D}$ which is right invariant and defined for $f \in C^\infty$ by

$$\mathcal{D}f = \frac{1}{2} \sum_{i,j=1}^{d} C_{ij} X_i^R X_j^R f. \tag{6}$$

If W is a left Brownian process then W^{-1} is a right Brownian process. Indeed, this verifies the equation:

$$dW'_t = -\sum_{i=1}^{d} [X_i^R f](W'_t) \circ dB_t^i, \qquad W'_0 = g^{-1}.$$

In this sense, left and right Brownian processes are equivalent. Since both are defined by Brownian stochastic differential equations, they can be assumed to have continuous paths. An explicit construction follows using multiplicative integration (see [19]).

For completeness, the Lévy property of Brownian processes is here indicated. This states that the increments of a left Brownian process W, defined as $W_s^{-1} W_t$ for $0 \leq s \leq t$, are independent and stationary. In other words, $W_s^{-1} W_t$ is independent of past values of W, up to time s, and its distribution depends only on $t - s$. In particular, W is a Markov process.

B. Characteristic Functions

The distribution of a random variable Y with values in G is completely determined by its characteristic function ϕ_Y. Moreover, ϕ_Y is compatible with the transformation of the distribution of Y under group operations. Characteristic functions are defined in terms of the irreducible representations of G. A general reference on representations of compact Lie groups is [20].

Let $\mathrm{Irr}(G)$ be the set of equivalence classes of irreducible representations of G. In particular, let $\delta_0 \in \mathrm{Irr}(G)$ contain the unit representation and $\mathrm{Irr}_+(G) = \mathrm{Irr}(G) - \{\delta_0\}$. The set $\mathrm{Irr}(G)$ is countable. Also, each equivalence class $\delta \in \mathrm{Irr}(G)$ contains a representation U^δ which is unitary and of finite dimension d_δ. That is, U^δ is an application $U^\delta : G \to SU(d_\delta)$ with the homomorphism property

$$U^\delta(gh) = U^\delta(g) U^\delta(h), \qquad U^\delta(g^{-1}) = U^\delta(g)^\dagger \tag{7}$$

for $g,\ h \in G$ and $\dagger$ denotes the Hermitian transpose. It is possible to choose U^δ to be smooth. Recall that $d_{\delta_0} = 1$ and, by definition, $U^{\delta_0}(g) = 1$ for $g \in G$. Given a fixed choice of smooth U^δ for $\delta \in \mathrm{Irr}_+(G)$, the characteristic function of Y is the sequence ϕ_Y of $d_\delta \times d_\delta$ complex matrices

$$\phi_Y(\delta) = \mathbb{E}[U^\delta(Y)], \qquad \delta \in \mathrm{Irr}_+(G) \tag{8}$$

where the expectation is applied to matrix elements. That the distribution of Y is indeed completely determined by ϕ_Y follows using the Peter–Weyl theorem. Indeed, this states that any continuous function f on G is, up to an additive constant, a uniform limit of linear combinations of the matrix elements U_{ab}^δ for $\delta \in \mathrm{Irr}_+(G)$ and $1 \leq a, b \leq d_\delta$. For a random variable X with values in G, $Y \stackrel{d}{=} X$ *iff* $\phi_Y = \phi_X$. If X, Y are independent and $Z = YX$, it follows from (7) that

$$\phi_Z(\delta) = \phi_Y(\delta) \phi_X(\delta). \tag{9}$$

In particular, for $h \in G$

$$\phi_{hY}(\delta) = U^\delta(h) \phi_Y(\delta), \qquad \phi_{Yh}(\delta) = \phi_Y(\delta) U^\delta(h). \tag{10}$$

Another consequence of (7) is that

$$\phi_{Y^{-1}}(\delta) = \phi_Y(\delta)^\dagger. \tag{11}$$

It is said that Y is uniformly distributed in G if $Y \stackrel{d}{=} gY$ for $g \in G$. This is equivalent to $\phi_Y(\delta) = U^\delta(g) \phi_Y(\delta)$ for $g \in G$ and $\delta \in \mathrm{Irr}_+(G)$. It is shown that Y is uniformly distributed *iff* $\phi_Y(\delta) = 0$ for $\delta \in \mathrm{Irr}_+(G)$. Recall that, then, the distribution of Y coincides with the Haar measure μ of G.

For later reference, (see (33) and discussion in Section IV), recall the following classical property of U^δ, $\delta \in \mathrm{Irr}_+(G)$. Namely, U^δ is an eigenfunction of $\Delta = \sum_{i=1}^{d} X_i^L X_i^L$ with eigenvalue $-\lambda_\delta$ where $\lambda_\delta > 0$. Precisely, matrix elements of U^δ span the corresponding eigenspace of Δ. A direct calculation shows that this is equivalent to

$$\sum_{i=1}^{d} X_i^\delta X_i^\delta = -\lambda_\delta I_\delta \tag{12}$$

where I_δ is the $d_\delta \times d_\delta$ identity matrix and $X_i^\delta = X_i U^\delta$. Note that X_i^δ is skew-Hermitian. This implies that the left-hand side of (12) is negative definite. It is a nonzero multiple of identity follows from Ad-invariance of $\langle \cdot, \cdot \rangle$ and irreducibility of U^δ, G being connected. The operator Δ is the Laplace operator associated to $\langle \cdot, \cdot \rangle$.

III. Brownian Distributions

Brownian distributions are the marginal distributions of Brownian processes. For processes W' and W' solving (1) and (5), $N_L(g, C)$ and $N_R(g, C)$ denote the distributions of W'_1 and

W_1'. These give rise to two different parametric families with the same parameters $g \in G$ and C a positive-definite symmetric $d \times d$ matrix. This section develops key properties of these two families. First, it obtains their transformation properties under group operations. It goes on to prove an extension of the central limit theorem, Proposition 3. This is discussed along with the property of infinite divisibility of Brownian distributions. Proposition 1 ensures that the notation $N_L(g, C)$, $N_R(g, C)$ is well defined. That is, either one of these distributions is completely determined by (g, C) and, moreover, different couples (g, C) determine different distributions.

In the following, $\exp$ will denote either the matrix exponential or the exponential map $\exp : \mathfrak{g} \to G$, according to context.

Proposition 1: For $\delta \in \mathrm{Irr}_+(G)$, there exists a Hermitian strictly positive-definite matrix $R_C(\delta)$, completely determined by C, such that

(i) If $Y \sim N_L(g, C)$, then $\phi_Y(\delta) = U^\delta(g) R_C(\delta)$.
(ii) If $Y \sim N_R(g, C)$, then $\phi_Y(\delta) = R_C(\delta) U^\delta(g)$.

Proof: The proofs of (i) and (ii) follow the same steps. In the case of (i), a stronger result is obtained from (3). For $\delta \in \mathrm{Irr}_+(G)$ consider the matrix

$$A_C(\delta) = \frac{1}{2} \sum_{i,j=1}^{d} C_{ij} X_i^\delta X_j^\delta.$$

This is a Hermitian matrix, since each X_i^δ is skew-Hermitian and C is symmetric. Note that for $1 \leq i, j \leq d$ and $h \in G$

$$[X_i^L X_j^L U^\delta](h) = U^\delta(h) X_i^\delta X_j^\delta.$$

For $t \geq 0$, let $\phi_t \equiv \phi_{W_t}$. Recalling the definition of $\phi_t(\delta)$ from (8), it follows by replacing in (3) that

$$\frac{d}{dt}\phi_t(\delta) = \phi_t(\delta) A_C(\delta) \phi_0(\delta) = U^\delta(g).$$

This has an explicit solution, $\exp$ being the matrix exponential

$$\phi_t(\delta) = U^\delta(g) \exp(t A_C(\delta)). \tag{13}$$

The proposition follows by putting $t = 1$ and recalling that $A_C(\delta)$ is Hermitian. ■

The parameter g, with C being fixed, acts as a location parameter for the distributions $N_L(g, C)$, $N_R(g, C)$. For $h \in G$ and $Y_1 \sim N_L(g, C)$, $Y_2 \sim N_R(g, C)$, it follows by applying (10) and Proposition 1 that

$$hY_1 \sim N_L(hg, C), \qquad Y_2 h \sim N_R(gh, C). \tag{14}$$

Also, both distributions are symmetric around g. This is in the sense that

$$Y_1 \stackrel{d}{=} g Y_1^{-1} g, \qquad Y_2 \stackrel{d}{=} g Y_2^{-1} g. \tag{15}$$

Note that the application $h \mapsto g h^{-1} g$ for $h \in G$ is an involution with fixed point g. From (11) and Proposition 1, it follows that

$$Y_1^{-1} \sim N_R(g^{-1}, C), \qquad Y_2^{-1} \sim N_L(g^{-1}, C). \tag{16}$$

Combining (14) and (16), it is possible to obtain (15). Note, from (16), that the two parametric families of $N_L(g, C)$ and $N_R(g, C)$ transform into each other by inversion.

The parameter C, with g being fixed, acts as a scale or concentration parameter for the distributions $N_L(g, C)$, $N_R(g, C)$. For $Y \sim N_L(g, C), Y = g$ almost surely *iff* $C = 0$. On the other hand, if C is strictly positive definite and $Y_t \sim N_L(g, tC)$, then Y_t is uniformly distributed in G, in the limit $t \uparrow \infty$. To see this, write $C = ODO^T$ where D is diagonal, O orthogonal and T stands for transpose. Let $Z_i = \sum_{j=1}^{d} O_{ji} X_j$, for $1 \leq i \leq d$. Again, Z_i is skew-Hermitian. In the notation of (13), $\phi_{Y_t} = \phi_t$. Moreover, $A_C(\delta)$ is strictly negative definite. This follows from (12) and the fact that O is orthogonal. Indeed

$$A_C(\delta) = \frac{1}{2} \sum_{i,j=1}^{d} C_{ij} X_i^\delta X_j^\delta = \frac{1}{2} \sum_{i=1}^{d} D_{ii} Z_i^\delta Z_i^\delta$$

where $D_{ii} > 0$. Thus, $\phi_{Y_t}(\delta) \to 0$ in the limit $t \uparrow \infty$. Note that, unlike g, C depends on the orthonormal basis $X_1, \ldots, X_d$.

Identities (14) and (16) show that C is unchanged under (left or right) translation and inversion. Lie group homomorphisms, on the other hand, transform C in a usual way. The following proposition gives the case $Y \sim N_L(g, C)$. For $Y \sim N_R(g, C)$, it is enough to note $\varphi(Y)^{-1} = \varphi(Y^{-1})$ and apply (16).

Proposition 2: Let $\varphi : G \to G$ be a Lie group homomorphism. Let $d\varphi$ be its derivative at e and Φ the matrix of $d\varphi$ in the basis $X_1, \ldots, X_d$. If $Y \sim \mathcal{N}_L(g, C)$ then $\varphi(Y) \sim \mathcal{N}_L(\varphi(g), \Phi C \Phi^T)$.

Proof: Let W satisfy (1) and W' the process $W' = \varphi(W)$. For $f \in C^\infty(G)$, by definition of $d\varphi$

$$[X_j^L (f \circ \varphi)](W_t) = [d\varphi(X_j)^L f](\varphi(W_t)) = \sum_{i=1}^{d} \Phi_{ij} [X_i^L f](W_t').$$

It now follows by applying (2) to the function $f \circ \varphi$ that

$$df(W_t') = \sum_{i=1}^{d} [X_i^L f](W_t') \circ dB_t'^i, \qquad B_t'^i = \sum_{j=1}^{d} \Phi_{ij} B_t^j.$$

The covariance matrix of $B' = (B'^1, \ldots, B'^d)$ is equal to $\Phi C \Phi^T$. The proposition follows by identifying Y and $\varphi(Y)$. ■

An important class of Lie group homomorphisms is that of conjugations (i.e., inner automorphisms). For $h \in G$, these are the applications $\mathrm{Ad}_h(g) = hgh^{-1}$. The derivative at e of Ad_h is represented in the basis $X_1, \ldots, X_d$ by an orthogonal matrix O_h. Applying Proposition 2, for $Y \sim N_L(g, C)$, it follows $\mathrm{Ad}_h(Y) \sim N_L(g', C')$, where

$$g' = \mathrm{Ad}_h(g), \qquad C' = O_h C O_h^{-1}. \tag{17}$$

If $g = e$ and $C = \sigma^2 I_d$, with I_d the $d \times d$ identity matrix, then $Y \stackrel{d}{=} \mathrm{Ad}_h(Y)$ for $h \in G$. If G is a simple group, then the converse holds. That is, if $Y \stackrel{d}{=} \mathrm{Ad}_h(Y)$ for $h \in G$, then $g = e$ and $C = \sigma^2 I_d$. When $Y \stackrel{d}{=} \mathrm{Ad}_h(Y)$ for $h \in G$, it is said that Y is conjugate invariant.

Turning to the property of infinite divisibility of Brownian processes, consider the distribution of the product Y_1Y_2 of independent random variables $Y_1 \sim \mathcal{N}_L(g_1, C_1)$ and $Y_2 \sim \mathcal{N}_L(g_2, C_2)$. In the case where C_1, C_2 commute with each other, this is given in the usual way. Precisely, it follows from (9) and Proposition 1 that

$$Y_1Y_2 \sim \mathcal{N}_L(g_1g_2, C_1 + C_2). \tag{18}$$

No similar conclusion can be made when C_1, C_2 do not commute. Infinite divisibility of the distribution $N_L(g, C)$ follows easily from (18). For $N \geq 1$, if $Z \sim N_L(e, C)$ and $Z_1, \ldots, Z_N$ are independent identically distributed (i.i.d.) with common distribution $N_L(e, C/N)$, then $Z \overset{d}{=} Z_1 \ldots Z_N$. Infinite divisibility can be associated with the fact that Brownian distributions arise as limit distributions of products of independent random variables with values in G. This is proved in Proposition 3, which can be seen as an extension of the central limit theorem. The proposition is stated in terms of triangular arrays of random variables. Precisely, a triangular array is a family Z_{Nn} of random variables with values in G defined for each $N \geq 1$ and $n = 1, \ldots, N$. For $N \geq 1$, these are such that $Z_{N1}, \ldots, Z_{NN}$ are i.i.d. The limit distribution of interest is that of the product $Y_N = Z_{N1\ldots} \ldots Z_{NN}$ as $N \uparrow \infty$.

Recall a property of the exponential map $\exp : \mathfrak{g} \to G$. Since the group G is connected, any $g \in G$ can be written $g = \exp(X)$ for some $X \in \mathfrak{g}$. Moreover, there exists $\mathcal{O} \subset G$, open with $\mu(\mathcal{O}) = 1$, such that X is unique for $g \in \mathcal{O}$ [21]. Let Z be a random variable with values in G, such that Z has a probability density with respect to μ. It follows that there exist a random variable ξ with values in $\mathfrak{g}$ such that $Z \overset{d}{=} \exp(\xi)$. Note that it is difficult to find an analytic relation between the distribution of Z and that of ξ [22]. Rather, the introduction of ξ should be understood in terms of the second-order theory in [23]. This is essential to the following statement.

Proposition 3: Given a triangular array of random variables as previously, assume Z_{Nn} has a probability density with respect to μ. Let ξ_{Nn} be a random variable with values in $\mathfrak{g}$ such that $Z_{Nn} \overset{d}{=} \exp(\xi_{Nn})$ and write $\xi_{Nn} = \sum_{i=1}^{d} \xi_{Nn}^i X_i$. Assume there exists a positive-definite symmetric matrix C such that

$$\mathbb{E}[\xi_{Nn}^i] = 0, \qquad \sum_{n=1}^{N} \mathbb{E}[\xi_{Nn}^i \xi_{Nn}^j] = C_{ij}$$

for $i, j = 1, \ldots, N$. Then, the limit distribution of Y_N is given by

$$Y_N \xrightarrow{d} Z, \qquad Z \in N_L(e, C). \tag{19}$$

Proof: Let ϕ_N be the characteristic function of Y_N and ϕ_Z the characteristic function of Z. Recall ϕ_Z is given as in (13). Let ϕ_{Nn} be the characteristic function of Z_{Nn}. By (9)

$$\phi_N(\delta) = \phi_{N1}(\delta) \ldots \phi_{NN}(\delta), \qquad \delta \in \mathrm{Irr}_+(G). \tag{20}$$

The aim will be to show that $\phi_N(\delta) \to \phi_Z(\delta)$. Recall the definition of ϕ_{Nn}, as in (8). This is estimated using the Taylor expansion of the matrix exponential

$$U^\delta(\exp \xi_{Nn}) =$$
$$I_\delta + \sum\nolimits_{i=1}^{d} \xi_{Nn}^i X_i^\delta + \frac{1}{2} \sum\nolimits_{i,j=1}^{d} \xi_{Nn}^i \xi_{Nn}^j X_i^\delta X_j^\delta + h(\xi_{Nn})$$

where $h(\xi_{Nn})$ is the error in the expansion to second-order terms. Applying conditions (i) and (ii), it follows from Lemma 1 in the appendix

$$\phi_{Nn}(\delta) = I_\delta + \frac{1}{2N} \sum_{ij}^{d} C_{ij} X_i^\delta X_j^\delta + o(1/N). \tag{21}$$

Replacing in (20) gives

$$\phi_N(\delta) = \left[I_\delta + \frac{1}{2N} \sum_{ij}^{d} C_{ij} X_i^\delta X_j^\delta + o(1/N) \right]^N.$$

This is a standard formula for the matrix exponential, so that (13) can be recovered. ■

Proposition 3 is interesting since its proof is based on the same steps as that of the classical central limit theorem and only requires a Taylor series expansion for the matrix exponential. A general theory of triangular arrays of random variables with values in Lie groups may be found in [24].

The product formula (18) and central limit theorem in Proposition 3 distinguish Brownian distributions, introduced in the current paper, from other extensions of normal distributions to compact Lie groups, proposed in the literature. One such extension is through so called wrapping of normal distributions, as in [25] and [26]. This definition yields transformation properties similar to the above and the same property as in Proposition 2. However, it does not satisfy (18) or Proposition 3.

IV. Estimation of g and C

This section is devoted to estimation of g using the extrinsic mean. Precisely, this is understood with respect to an injective unitary representation U of G. In Section IV-A, it is shown g is the unique extrinsic mean for Brownian distributions $N_L(g, C)$ or $N_R(g, C)$. This results in consistent estimates for g, which are simply computed by performing a polar decomposition. Under specific conditions on U, this is seen to lead to simultaneous consistent estimation of C. In Section IV-B, asymptotic normality of the extrinsic mean estimates of g is obtained by classical means. Finally, Section IV-C is concerned with intrinsic means, which it discusses using transformation properties obtained in Section III.

A. Consistency of the Extrinsic Mean

The extrinsic mean g_E of $Y \sim N_L(g, C)$ is defined, given an injective Lie group homomorphism $U : G \to SU(v)$, for some $v \geq 1$, by the following formula:

$$g_E = \mathrm{argmin}_{h \in G} \mathbb{E}[d_E^2(h, Y)] \tag{22}$$

where $d_E(\cdot,\cdot)$ is the extrinsic distance function corresponding to the Euclidean matrix norm $\|\cdot\|$

$$d_E(g,h) = \|U(g) - U(h)\|.$$

An injective Lie group homomorphism U exists whatever the underlying group G[20]. An intrinsic property of g is central in considering the extrinsic mean. Irrespective of the chosen U, g is the unique global minimum in (22). See Proposition 4.

Matrix functions in the following, i.e., square root$(\frac{1}{2})$ and logarithm (Log), denote unique Hermitian determinations. Let $M = \mathbb{E}[U(Y)]$. Applying (3), an expression similar to (13) can be obtained. Precisely, $M = VR$, with $V = U(g)$ and

$$R = \exp(A), \qquad A = \frac{1}{2}\sum_{i,j=1}^{d} C_{ij}\tilde{X}_i\tilde{X}_j \tag{23}$$

where $\tilde{X}_i = X_i U$ for $1 \le i \le d$. Again, the matrix A is Hermitian and R is Hermitian strictly positive definite. Note that M is nonsingular. It follows that

$$R = [M^\dagger M]^{\frac{1}{2}}, \qquad V = MR^{-1}. \tag{24}$$

In other words, V and R are the factors of the left polar decomposition of M. Properties used below, concerning polar decomposition, are based on [27].

Proposition 4: For (22), the unique global minimum is $g_E = g$.

Proof: For $h \in G$, a direct computation gives

$$\mathbb{E}[d_E^2(h,Y)] = \|M - U(h)\|^2 + (v - \|M\|^2) \tag{25}$$

where the second term does not depend on h. Using $M = VR$ and Lemma 2 in the appendix, $\|M-V\| < \|M-V'\|$ for unitary $V' \neq V$. Recall $V = U(g)$ and U is injective. Comparing (22) to (25), it follows that

$$\mathbb{E}[d_E^2(g,Y)] < \mathbb{E}[d_E^2(h,Y)] \tag{26}$$

for $h \neq g$. So $g_E = g$ is the unique global minimum. ■

Note that inequality (26) is a special case of the one used in [28], in order to establish a general lower bound for estimators in compact Lie groups.

Computation of V as in (24) also yields the Hermitian factor R. Equivalently, it is possible to consider $A = \mathrm{Log}(R)$. With U being fixed, this is determined by C. Using the fact that C is symmetric

$$A = \frac{1}{2}\sum_{i,j=1}^{d} C_{ij}(\tilde{X}_i\tilde{X}_j + \tilde{X}_j\tilde{X}_i). \tag{27}$$

For recovery of the covariance matrix C, the inverse situation is of interest. Precisely, A being given, (27) should be solved for the C_{ij}. Note A is Hermitian positive definite. Also, for $1 \le i,j \le d$, the matrix is $\tilde{X}_i\tilde{X}_j + \tilde{X}_j\tilde{X}_i$ is Hermitian. Since (27) is known to hold, it does have a solution for the C_{ij}. A sufficient condition for this solution to be unique is that the matrices $\tilde{X}_i\tilde{X}_j + \tilde{X}_j\tilde{X}_i$ be linearly independent. Since these are Hermitian, this implies $v^2 \ge d(d+1)/2$. Here, v^2 is the dimension of the space of $v \times v$ Hermitian matrices, $d(d+1)/2$ is the number of pairs (i,j) with $1 \le i \le j \le d$. If $G = SU(w)$ where $w > 1$, then $d = w(w-1)/2$ and it follows $v > w$. In other words, recovery of the covariance matrix C requires that G be embedded into a higher dimensional matrix Lie group. A detailed example is given in Section VI.

Consider now the consistency of estimates given by (24). Starting from i.i.d. observations of $Y, Z_1, \dots Z_N$, an empirical estimation step yields M. This is then followed by an identification step yielding g, and eventually also C. It is discussed below how observations $Z_1, \dots, Z_N$ may be obtained when observing an individual path of a process W, solving (1).

By the strong law of large numbers—for further discussion, see, e.g., [29]:

$$M_N = \frac{1}{N}\sum_{n=1}^{N} U(Z_n) \xrightarrow{a.s.} M \tag{28}$$

where $a.s.$ stands for almost sure convergence. Since the set of nonsingular $v \times v$ matrices is open in the standard topology, M_N is asymptotically nonsingular, again almost surely. Consider the estimates R_N and V_N by putting them to zero on the event where M_N is singular and writing on the complementary event

$$R_N = [M_N^\dagger M_N^\dagger]^{\frac{1}{2}}, \qquad V_N = M_N^- R_N^{-1}. \tag{29}$$

Recall that the factors of the polar decomposition are continuous at any nonsingular matrix. Accordingly, it is straightforward that $R_N \xrightarrow{a.s.} R$ and $V_N \xrightarrow{a.s.} V$. The following Proposition 5 bounds the sample size N used for a certain precision in approximating V. Also from a computational point of view, recall that it is advantageous to compute polar decompositions using singular value decomposition rather than directly applying (24).

The following inequalities (30) and (31) are required [27]. For $a = 1,2$, let M_a be a nonsingular $v \times v$ matrix with polar decomposition $M_a = V_a R_a$, where V_a is unitary and M_a is Hermitian. In (31), $k \ge 0$ such that the singular values $\varsigma_a^1 \le \cdots \le \varsigma_a^v$ of M_a are all $\ge k$:

$$\sum_{i=1}^{v}(\varsigma_1^i - \varsigma_2^i)^2 \le 2\|M_1 - M_2\|^2 \tag{30}$$

$$\|V_1 - V_2\| \le k^{-1}\|M_1 - M_2\|. \tag{31}$$

Proposition 5: There exist $L > 0$ and $0 < k \le 1$ such that for $N \ge 1$

$$\mathbb{E}\|V_N - V\|^2 \le \frac{L}{N}\left(\frac{v}{k}\right)^2.$$

Proof: Let k be the smallest singular value of M. Since M is nonsingular, $0 < k$. Using Jensen's inequality, it also holds that $k \le 1$. Indeed, $k = \|Mz\|$ for some complex vector z.

For $N \ge 1$, let k_N be the smallest singular value of M_N. Consider the event $T_N = \{k_N \ge k/2\}$. After using (30) and (31), respectively:

$$\mathbb{E}[\|V_N - V\|^2; T_N^c] \le 4v\mathbb{P}\left(\|M_N - M\|^2 > k^2/16\right)$$

$$\mathbb{E}[\|V_N - V\|^2; T_N] \le 4k^{-2}\mathbb{E}\|M_N - M\|^2$$

where $T_N^c = \{k_N < k/2\}$. Note that $\mathbb{E}\|M_N - M\|^2 \leq 4v/N$. The proposition is obtained by applying Chebychev's inequality and summing. ■

Proposition 5 only uses the weak law of large numbers, so that independence of the observations $Z_1, \ldots, Z_N$ is not a necessary condition. The given rate of convergence is controlled by the classical central limit theorem for $M_N \to M$ and also conditioned by k, the smallest singular value of M. The singular value k measures the dispersion of $U(Y)$ away from V

$$(1-k) \leq \|R - I_v\| = \|M - V\| \leq \sqrt{v}(1-k)$$

where I_v is the $v \times v$ identity matrix. Values of k closer to 1 imply $U(Y)$ is distributed closer to V and the rate of convergence is improved.

Dispersion of $U(Y)$ away from V is also given by the covariance matrix C. In the case $C = \sigma^2 I_d$, there is a straightforward relationship between k and σ^2. Clearly, U defines a v-dimensional unitary representation of G and can be decomposed into an orthogonal sum, $\delta_1, \ldots, \delta_r \in \mathrm{Irr}_+(G)$

$$U = U^{\delta_1} \oplus \cdots \oplus U^{\delta_r}. \tag{32}$$

Let $\lambda = \max\{\lambda_{\delta_1}, \ldots, \lambda_{\delta_r}\}$. Then

$$k = e^{-\sigma^2\lambda/2}. \tag{33}$$

The representations $\delta_1, \ldots, \delta_r$ need not be different. However, since U is injective

$$\mathrm{Ker}(\delta_1) \cap \cdots \cap \mathrm{Ker}(\delta_r) = \{e\}$$

which imposes a sufficient choice of irreducible representations. On the other hand, Proposition 5 indicates that U should be chosen to minimize λ. If G is a simple group, any representation $\delta \in \mathrm{Irr}_+(G)$ is injective. An obvious choice is $\Phi = U^{\delta_1}$ where $\lambda_{\delta_1} = \min\{\lambda_\delta, \delta \in \mathrm{Irr}_+(G)\}$.

For $Y \sim N_R(g, C)$, the unique global minimum $g_E = g$ continues to hold in (22). This follows from (16) and since $\|\cdot\|$ is invariant under Hermitian transposition. Moreover, $V = U(g)$ and R defined in (23) are the factors of the right polar decomposition of M.

The estimates R_N and V_N of (29) can be computed when observing an individual path of a process W solving (1). The key is to use the Lévy property of W. Suppose given a path $t \mapsto \hat{W}(t)$ over some interval $t \in [0, T]$. If $N \leq T$, then it is possible to construct the following observations:

$$Z_n = \hat{W}(0)\hat{W}^{-1}(n-1)\hat{W}(n), \qquad 1 \leq n \leq N.$$

The Lévy property of W ensures that $Z_1, \ldots, Z_N$ are independent and with the distribution of W_1. Moreover, in this case $g = \hat{W}(0)$ is immediately available. In this context, the importance of the Markov property of W becomes clear. Empirical estimation of M is aimed at obtaining the transition operator of W, whose spectral properties are used for the identification step.

B. Asymptotic Normality

Proposition 5 in Section IV-A established the consistency of estimates V_N given by (29). This results from consistency of M_N, due to the strong law of large numbers, and continuity of the factors of the polar decomposition. Here, asymptotic normality of V_N will be obtained in a similar way. Indeed, M_N is asymptotically normal, due to the central limit theorem, and the factors of the polar decomposition are differentiable, in addition to being continuous. The reasoning used below is related to that of [7], also employed by [11].

For nonsingular $v \times v$ matrix M let $\mathcal{V}(M)$ be its left polar factor. For instance, (24) and (29) give $V = \mathcal{V}(M)$ and $V_N = \mathcal{V}(M_N)$. The application $\mathcal{V}$ is defined on a set of matrices which is open in the standard topology. Moreover, it has its values in $SU(v)$. $\mathcal{V}$ is continuously differentiable that may be found, along with an evaluation of the derivative, in [30]. In particular, the derivative at M is a linear map $d\mathcal{V}_M$ from the set of $v \times v$ complex matrices to the tangent space of $SU(v)$ at V.

Consider the asymptotic distribution of $M_N - M$. Since U is unitary, it is bounded. In particular, the $U(Z_n)$ in (28) have finite second-order moments. Clearly, $\mathbb{E}[U(Z_n)] = M$. It follows from the central limit theorem, in its simplest form [29], that the distribution of $\sqrt{N}(M_N - M)$ converges to a centered normal distribution. The corresponding asymptotic covariances are the same as those of the elements of $U(Z_n)$. Interestingly, these can be evaluated using only the characteristic function ϕ_Y. Indeed, note that all products of elements of $U(Z_n)$ are contained in the elements of the Kronecker product $U(Z_n) \otimes U(Z_n)$. However, $(U \otimes U) : G \to SU(v^2)$, where $(U \otimes U)(g) = U(g) \otimes U(g)$ is a finite-dimensional unitary representation of G. As such, it can be written as an orthogonal sum of a finite number of irreducible unitary representations U^δ for $\delta \in \mathrm{Irr}_+(G)$ [20]. In short, second-order moments of $U(Z_n)$ are linear combinations of first order moments, given by the characteristic function.

Asymptotic normality of $V_N - V$ is obtained from the following general reasoning, applied to each matrix element. Let f be a continuously differentiable function, defined on nonsingular $v \times v$ matrices and with values in $\mathbb{R}$. Recall M is nonsingular. Using the fact that $M_N \to M$ almost surely, it is possible to assume any convex combination of M_N and M is nonsingular. From the mean value theorem, $f(M_N) - f(M) = df_{M'}(M_n - M)$, where df is the derivative of f and M' is a convex combination of M_N and M. In particular

$$\sqrt{N}[f(M_N) - f(M)] = df_{M'}[\sqrt{N}(M_N - M)].$$

Since $M_N \to M$ almost surely, the same holds for M'. By continuity of df, it follows that $df_{M'} \to df_M$ almost surely. Note df_M is a constant, i.e., nonrandom, linear map. From convergence in distribution of $\sqrt{N}(M_N - M)$ it follows that the asymptotic distribution of $\sqrt{N}[f(M_N) - f(M)]$ is centered normal. Using this result, it is immediate to conclude the following. If w has the asymptotic distribution of $\sqrt{N}(M_N - M)$, then

$$\sqrt{N}(V_N - V) \to dd\mathcal{V}_M(w). \tag{34}$$

As expected, the asymptotic distribution has its support in the tangent space to $SU(v)$ at V.

The derivative of $\mathcal{V}$ has no straightforward analytic expression. On the other hand, under simplifying assumptions, $\mathcal{V}$ can be replaced by a more tractable application. Precisely, assume $C = \sigma^2 I_d$. For simplicity, assume also that G is simple. In particular, it is possible to take $U = U^\delta$ for some $\delta \in \mathrm{Irr}_+(G)$. In the notation of (24), it follows that $R = e^{-\sigma^2\lambda_\delta/2} I_\delta$. The relation between M and V becomes

$$V = \left[\mathrm{tr}(M^\dagger M)/d_\delta\right]^{\frac{-1}{2}} M \tag{35}$$

where tr denotes trace. Let $\mathcal{P}$ be the application where $V = \mathcal{P}(M)$ according to this formula. This is defined for all nonsingular M but does not in general take its values in $SU(v)$. It is possible to evaluate its derivative and see that it is continuously differentiable. Precisely, for complex $v \times v$ matrix h

$$d\mathcal{P}_M(h) = [\mathrm{tr}(M^\dagger M)]^{\frac{-1}{2}} \left[h - d_\delta^{-1/2}\mathrm{tr}(hV^\dagger)V\right].$$

Recalling the expression for R. This can further be written as

$$d\mathcal{P}_M(h) = e^{\sigma^2\lambda_\delta/2} \left[h - d_\delta^{-1/2}\mathrm{tr}(hV^\dagger)V\right]. \tag{36}$$

If estimates V_N are given as in (35) instead of (24), i.e., $V_N = \mathcal{P}(M_N)$, then the asymptotic distribution of $V_N - V$ follows as in (34). Precisely

$$\sqrt{N}(V_N - V) \to de^{\sigma^2\lambda_\delta/2} \left[w - d_\delta^{-1/2}\mathrm{tr}(wV^\dagger)V\right]. \tag{37}$$

It should be noted that there is an abuse of notation in writing $V_N = \mathcal{V}(M_N)$ and $V_N = \mathcal{P}(M_N)$, at the same time. However, this only concerns (37).

While (37) is more explicit than (34), the resulting asymptotic distribution does not have its support in the tangent space of $SU(v)$ at V. The asymptotic distribution of (34) can further be identified with a distributing with support in the tangent space of G at g, giving it an intrinsic character. Since U is injective, its derivative can be used to identify the tangent space of G at g with a subspace of the tangent space of $SU(v)$ at V. A centered normal distribution with support in the tangent space of G at g can then be obtained by projecting the asymptotic distribution of (34).

C. Transformation Properties and Intrinsic Means

To clarify the role played by transformation properties of $Y \sim N_L(g, C)$, a general version of (22) is now considered. Proposition 6 uses these transformation properties to give a first insight into using intrinsic means. Let $d(\cdot,\cdot)$ be a distance function with

$$d(g,h) = d(kg, kh) = d(gk, hk) \tag{38}$$
$$d(g,h) = d(g^{-1}, h^{-1}) \tag{39}$$

for $g, h, k \in G$. In the following, it is assumed that C is strictly positive definite. In this case, Y has a smooth probability density p with respect to μ. Consider the problem of minimizing the cost function

$$F(h) = \mathbb{E}[d^2(h, Y)] = \int d^2(h,k)p(k)d\mu(k). \tag{40}$$

Conditions (38) and (39) both hold for the distance function $d_E(\cdot,\cdot)$, leading to (22). They also hold for the distance function $d_R(\cdot,\cdot)$ arising from the Riemannian metric induced by $\langle\cdot,\cdot\rangle$. With this latter choice for $d(\cdot,\cdot)$, the problem of minimizing F leads to the notion of intrinsic mean of Y. Here, an intrinsic mean is any local minimum of F—in [6], [7], a local minimum is called a Karcher mean and the term intrinsic mean is reserved for a unique global minimum.

The function F is smooth. If $g = e$, then $F(h) = F(h^{-1})$ for $h \in G$. If, moreover, $C = \sigma^2 I_d$, then $F(khk^{-1}) = F(h)$ for $h, k \in G$. Smoothness of F follows since F can be written as a convolution of a continuous function d^2 with the smooth density p. It follows from (38) that

$$F(h) = \int d^2(hk^{-1})p(k)d\mu(h) \tag{41}$$

where $d(h) = d_E(e, h)$. The symmetry properties of F also follow from (38) and (39). If $g = e$, then $C = \sigma^2 I_d$. By the discussion after (17), Y is conjugate invariant. Accordingly, for $h, k \in G$

$$\mathbb{E}[d_E^2(khk^{-1}, Y)] = \mathbb{E}[d_E^2(h, k^{-1}Yk)] = \mathbb{E}[d^2(h, Y)].$$

Here, the first step uses (38) and the second step uses conjugate invariance of Y. Using the definition of F, it follows $F(khk^{-1}) = F(h)$.

The unique global minimum (22) is a specific property of the extrinsic distance function $d_E(\cdot,\cdot)$. For a general distance function $d(\cdot,\cdot)$, multiple local minima may arise in (40). To decide whether the parameter g of Y is such a local minimum, it is enough to consider the case $Y \sim N_L(e, C)$. The general case is then obtained by (14) and (38). Assuming $Y \sim N_L(e, C)$, the symmetry properties of F can be used to simplify its derivatives of first and second order at e. These then characterize the existence of a local minimum.

For Proposition 6, $H(h) = (H_{ij}(h), 1 \le i, j \le d)$, is the Hessian matrix of F at $h \in G$. In particular, this is calculated with respect to the basis $X_1, \ldots, X_d$ of $\mathfrak{g}$

$$H_{ij}(h) = \frac{1}{2}[(X_i^L X_j^L + X_j^L X_i^L)F](h).$$

Proposition 6: The following hold.

(i) If $g = e$, then for $X \in \mathfrak{g}$, $XF = 0$.

(ii) If $C = \sigma^2 I_d$, then for $k \in G$, $H(e) = O_k H(e) O_k^{-1}$.

Proof: For (i), let $X \in \mathfrak{g}$ and $X^I = (1/2)(X^L + X^R)$. For $h \in G$

$$\begin{aligned} X^I F(h^{-1}) &= (1/2)\tfrac{d}{dt}\left[F(h^{-1}e^{tX}) + F(e^{tX}h^{-1})\right]_{t=0} \\ &= (1/2)\tfrac{d}{dt}\left[F(e^{-tX}h) + F(he^{-tX})\right]_{t=0} \\ &= -[X^I F](h) \end{aligned}$$

the second step uses $F(h) = F(h^{-1})$. In particular, (i) follows since $[X^I F](e) = 0$.

For (ii), note that for $h, k \in G$ and $1 \le i, j \le d$

$$X_i^L X_j^L F(h) = \tfrac{\partial^2}{\partial t \partial s}\left[F(he^{tX_i}e^{sX_j})\right]_{(0,0)}$$

$$= \frac{\partial^2}{\partial t \partial s}\left[F(khk^{-1}e^{tX'_i}e^{sX'_j})\right]_{(0,0)}$$
$$= [X'^L_i X'^L_j F](khk^{-1})$$

where the notation X'_i stands for $\mathrm{Ad}_k X_i$. Here, the second step uses $F(h) = F(khk^{-1})$ for $h, k \in G$. Now

$$H_{ij}(h) = \frac{1}{2}[(X'^L_i X'^L_j + X'^L_j X'^L_i)F](khk^{-1}).$$

Replacing the fact that $X'_i = \sum_{j=1}^{d} O_k^{ji} X_i$, it follows in matrix form that

$$O_k H(h) O_k^T = H(khk^{-1})$$

putting $h = e$, (ii) follows immediately. ■

If G is simple, it follows from (ii) of Proposition 6 that $H(e)$ is a multiple of identity. Moreover[1], if $d^2 \in C^2(\mathcal{O})$ where $\mathcal{O} \subset G$ is open with $\mu(\mathcal{O}) = 1$

$$H(e) = (1/d)\,[\Delta F](e) I_d = (1/d)\,\mathbb{E}\left([\Delta d^2](W_1)\right) I_d. \quad (42)$$

If $d^2 \in C^2(G)$, then this is the right-hand side of (3) for a process W such that $W_1 \overset{d}{=} Y$. Intuitively, it may be said $H(e)$ is positive definite if W is, at time 1, is moving away from e. Positive definiteness of $H(e)$ is a sufficient condition for e to be a local minimum of F [31], i.e., an intrinsic mean of Y.

If G is not simple, the situation of (42) continues to hold if the function d^2 can be decomposed into a sum of similar functions (squared distance to e) for each simple or one dimensional factor of G. This aspect will not be detailed here.

V. Multivariate Brownian Distributions

Multivariate Brownian distributions are Brownian distributions on some product group $G \times \cdots \times G$. The current section aims to describe the covariance structure of multivariate Brownian distributions. For this, it will be enough to consider the product $G \times G$. This is again a compact connected group, so that a Brownian process $W = (W^1, W^2)$ may be introduced as in (1). Given random variables Y_1, Y_2 with values in G, these are said to be jointly Brownian if $(Y_1, Y_2) \overset{d}{=} (W^1_1, W^2_1)$ for some Brownian process W. The main result here is Proposition 8. This states that Y_1, Y_2 are independent whenever $C^{12} = 0$, where C^{12} is the covariance matrix of Y_1, Y_2 to be defined below. In order to prove this proposition, the concept of joint characteristic function is introduced. Independence is then characterized in Proposition 7, which is an extension of the classical Kac's theorem. A further question considered is transformation of the covariance structure of jointly Brownian random variables under Lie groups homomorphisms. It is seen the compact Lie group structure does not, in general, allow statistical analysis by means of Lie group homomorphisms.

To write down (1) on $G \times G$, recall the product Lie group structure. Note that $G \times G = G_1 G_2$ where $G_1 = G \times \{e\}$ and $G_2 = \{e\} \times G$. G_1, G_2 are Lie subgroups commuting with each other. The Lie algebra of $G \times G$ is a direct sum $\mathfrak{g}_1 \oplus \mathfrak{g}_2$ where $\mathfrak{g}_1 = \mathfrak{g} \oplus \{0\}$ and $\mathfrak{g}_2 = \{0\} \oplus \mathfrak{g}$. Here, $\mathfrak{g}_1, \mathfrak{g}_2$ are the Lie algebras of G_1, G_2 and also commute with each other. Consider the chosen basis $X_1, \ldots, X_d$ of $\mathfrak{g}$. This gives a basis of $\mathfrak{g}_1$, $X^1_1 = (X_1, 0), \ldots, X^1_d = (X_d, 0)$ and a basis of $\mathfrak{g}_2$, $X^2_1 = (0, X_1), \ldots, X^2_d = (0, X_d)$. It follows that $X^1_1, \ldots, X^1_d, X^2_1, \ldots, X^2_d$ is a basis of the Lie algebra of $G_1 \times G_2$. Of course, $\dim(G \times G) = 2d$.

[1]In particular, this is the case for both $d_E(\cdot,\cdot)$ and $d_R(\cdot,\cdot)$.

A Brownian process W in $G \times G$ verifies the following Stratonovich equation, where $W_0 = (g_1, g_2)$ and $B = (B^1_1, \ldots, B^d_1, B^1_2, \ldots, B^d_2)$ is a Brownian motion with values in $\mathbb{R}^{2d}$:

$$dW_t = \sum_{i=1}^{d} X^{1\,L}_i(W_t) \circ dB^i_{1,t} + \sum_{i=1}^{d} X^{2\,L}_i(W_t) \circ dB^1_{2,t}. \quad (43)$$

Let C be the covariance matrix of B. This has a block structure where C^{11}, C^{22} are positive-definite symmetric $d \times d$ matrices and C^{12} is a $d \times d$ off diagonal block. If Y_1, Y_2 are such that $(Y_1, Y_2) \overset{d}{=} (W^1_1, W^2_1)$, then C^{12} is referred to as the covariance matrix of Y_1, Y_2.

The joint statistics of Y_1, Y_2 are given by the characteristic function of the random variable $Y = (Y_1, Y_2)$ which has its values in $G \times G$. In order to construct this characteristic function, recall that the set $\mathrm{Irr}(G \times G)$ can be identified with $\mathrm{Irr(G)} \times \mathrm{Irr}(G)$. For $\delta_1, \delta_2 \in \mathrm{Irr}(G)$, there exists $\delta \in \mathrm{Irr}(G \times G)$ of dimension $d_\delta = d_{\delta_1} \times d_{\delta_2}$ and where it is possible to choose $U^\delta : G \times G \to SU(d_\delta)$, given by

$$U^\delta(g_1, g_2) = U^{\delta_1}(g_2) \otimes U^{\delta_2}(g_2), \qquad g_1,\ g_2 \in G \quad (44)$$

where, as already defined in Section IV-B, $\otimes$ is the Kronecker product. Any $\delta \in \mathrm{Irr}(G \times G)$ is of the above form.

According to (8) and (44), the characteristic function of Y appears as a joint characteristic function of Y_1, Y_2

$$\phi_Y(\delta_1, \delta_2) = \mathbb{E}\left[U^{\delta_1}(Y_1) \otimes U^{\delta_2}(Y_2)\right]. \quad (45)$$

Here and in the following, $(\delta_1, \delta_2) \neq (\delta_0, \delta_0)$.

Proposition 8 uses the following Proposition 7. This is an extension of the classical Kac's theorem (see, for example, [32]) and essentially also holds when G is any compact topological group.

Proposition 7: Let Y_1, Y_2 be random variables with values in G and define $Y = (Y_1, Y_2)$. Y_1, Y_2 are independent *iff*

$$\phi_Y(\delta_1, \delta_2) = \phi_{Y_1}(\delta_1) \otimes \phi_{Y_2}(\delta_2). \quad (46)$$

Proof: Y_1, Y_2 are independent *iff*, for all continuous functions f_1, f_2, $\mathbb{E}[f_1(Y_1) f_2(Y_2)] = \mathbb{E}[f_1(Y_1)]\mathbb{E}[f_2(Y_2)]$. The *only if* part follows since the functions $U^{\delta_i}_{ab}$ are continuous, $i = 1, 2$ and $1 \leq a, b \leq d_{\delta_i}$. The *if* part follows by the Peter–Weyl theorem, applied to f_1, f_2, and dominated convergence. ■

Proposition 8: Let Y_1, Y_2 be jointly Brownian and C^{12} their covariance matrix. If $C^{12} = 0$, then Y_1, Y_2 are independent.

Proof: Let W be a process verifying (43) such that $(Y_1, Y_2) \overset{d}{=} (W^1_1, W^2_1)$. By Proposition 1, the left-hand side of (46) is equal to

$$\phi_Y(\delta_1, \delta_2) = [U^{\delta_1}(g_1) \otimes U^{\delta_2}(g_2)] R_C(\delta_1, \delta_2) \quad (47)$$

where $Y = (Y_1, Y_2)$. Here, $R_C(\delta_1, \delta_2) = \exp[A_C(\delta_1, \delta_2)]$, where

$$A_C(\delta_1,\delta_2) = \tfrac{1}{2}[A_{C^{11}}(\delta_1)\otimes I_{\delta_2} + I_{\delta_2}\otimes A_{C^{22}}(\delta_2)] + \tfrac{1}{2}\sum_{i,j=1}^{d} C^{12}_{ij}[(X_i^{\delta_1}\otimes X_j^{\delta_2}) + (X_j^{\delta_2}\otimes X_i^{\delta_1})].$$

Putting $\delta_2 = \delta_0$ and then $\delta_1 = \delta_0$ in (47), it follows that

$$\phi_{Y_i}(\delta_i) = U^{\delta_i}(g_i)\exp[A_{C^{ii}}(\delta_i)]$$

where $i = 1, 2$. This allows for the right-hand side of (46) to be evaluated. Applying Proposition 7 and taking Hermitian matrix logarithms, Y_1, Y_2 are independent *iff*

$$\sum_{i,j=1}^{d} C^{12}_{ij}[(X_i^{\delta_1}\otimes X_j^{\delta_2}) + (X_j^{\delta_2}\otimes X_i^{\delta_1})]. \tag{48}$$

The proposition follows immediately. ■

Let Y_1, Y_2 be jointly Brownian random variables. It appears natural to attempt to apply Proposition 2 in order to obtain Y_1', Y_2' jointly Brownian and with a prescribed covariance matrix. For example, this could be aimed at obtaining an independent couple Y_1', Y_2' from the original Y_1, Y_2. The compact Lie group structure of G imposes strong restrictions on this approach.

Consider the basic case where G is a simple group. Let $\varphi : G\times G \to G\times G$ be a Lie group homomorphism and $\varphi_i = \pi_i\circ\varphi$, for $i = 1, 2$, where $\pi_1, \pi_2 : G\times G\to G$ are the canonical projections. It can be shown that, for $g_1, g_2 \in G$, either $\varphi_i(g_1, g_2) = e$ or $\varphi_i(g_1, g_2) = \mathrm{Ad}_h(g_j)$ for some $h\in G$, where $j = 1, 2$. Assume φ_i nontrivial and let $(Y_1', Y_2') = \varphi(Y_1, Y_2)$. It follows that Y_1', Y_2' are independent only if Y_1, Y_2 are independent. Thus, the class of Lie group homomorphisms φ is not sufficiently rich in order for the desired statistical analysis to be defined.

The problem of obtaining an independent couple Y_1', Y_2' from the original Y_1, Y_2 can be considered from an information theoretic point of view. Tentatively, a Lie group homomorphism $\varphi : G\times G\to G\times G$ may be sought which minimizes the mutual information of Y_1', Y_2'. In this connection, a broad review of information theoretic inequalities relevant to compact Lie group structure can be found in [33].

VI. Example: The Special Unitary Group, $su(2)$

The current section illustrates the situation described in Section IV-A, with regard to estimation of C, through the example $G = SU(2)$. Although $SU(2)$ is naturally a matrix group, it will be necessary to embed it into the higher dimensional $SU(5)$ in order to estimate both g and C.

For relevant properties of $SU(2)$, see [34]. In the following, $I_{\frac{1}{2}}$ and I_1 denote the 2×2 and 3×3 identity matrices. Recall that $SU(2)$ is the group of 2×2 complex matrices g, which are unitary, $gg^\dagger = I_{\frac{1}{2}}$, and have unit determinant $\det(g) = 1$. The corresponding Lie algebra $\mathfrak{su}(2)$ is the vector space of 2×2 complex matrices X, which are skew-Hermitian and with zero trace, considered with the additional bracket operation $[X, X'] = XX' - X'X$. The dimension of $SU(2)$ is $d = 3$. It is usual to choose the following basis of $\mathfrak{su}(2)$:

$$X_1 = \begin{bmatrix} & \frac{-i}{2} \\ \frac{-i}{2} & \end{bmatrix},\quad X_2 = \begin{bmatrix} & \frac{1}{2} \\ \frac{-1}{2} & \end{bmatrix},\quad X_3 = \begin{bmatrix} \frac{-i}{2} & \\ & \frac{i}{2} \end{bmatrix}.$$

From bilinearity, the operation $[\cdot,\cdot]$ is completely determined from

$$[X_i, X_j] = \epsilon_{ikj}X_k, \qquad 1\le i,j,k\le 3 \tag{49}$$

where ϵ_{ijk} is totally antisymmetric, $\epsilon_{123} = 1$. The matrices J_i of the linear maps $X\mapsto [X_i, X]$—in the basis X_1, X_2, X_3—are the following:

$$J_1 = \begin{bmatrix} & & \\ & & -1 \\ & 1 & \end{bmatrix},\quad J_2 = \begin{bmatrix} & & 1 \\ & & \\ -1 & & \end{bmatrix},\quad J_3 = \begin{bmatrix} & -1 & \\ 1 & & \\ & & \end{bmatrix}.$$

These satisfy the same relations (49). An immediate computation shows that the matrices $X_iX_j,\ 1\le i,j\le 3$, are not linearly independent, while the matrices $J_iJ_j,\ 1\le i,j\le 3$, are linearly independent.

The scalar product $\langle\cdot,\cdot\rangle$ on $\mathfrak{su}(2)$ is obtained by declaring the basis X_1, X_2, X_3 to be orthonormal. This is Ad-invariant that will be verified below. Note here that, by (49), $SU(2)$ is simple, so that there exists essentially one Ad-invariant scalar product on $\mathfrak{su}(2)$.

Let $B = (B^1, B^2, B^3)$ be a 3-D Brownian motion with covariance matrix C. Equation (1) may be copied:

$$dW_t = \sum_{i=1}^{3} X_i^L(W_t)\circ dB_t^i, \qquad W_0 = g. \tag{50}$$

According to Sections II-A and III, W is a left Brownian process in $SU(2)$ and the distribution of W_1 is $N_L(g, C)$. The approach of Section IV will be applied to estimation of g, C from \textbackslash\ observations $Z_1, \ldots Z_N$ of $Y \sim N_L(g, C)$. This consists in three main steps. The first is to choose an injective Lie group homomorphism $U : SU(2)\to SU(v)$. The second is to apply (24) and the third is to retrieve the covariance matrix C from the resulting Hermitian factor. The first step controls the precision of the estimates, as discussed after Proposition 5. It also determines the possibility of recovering C. As discussed in Section IV, this requires $v > 2$. The proposed choice is $v = 5$. Precisely

$$U = U^{\frac{1}{2}}\oplus U^1 \tag{51}$$

where $U^{\frac{1}{2}} : SU(2)\to SU(2)$ is the identity representation and $U^1 : SU(2)\to SU(3)$ is the adjoint representation. It is clear that $d_{1/2} = 2, d_1 = 3$ and $U : SU(2)\to SU(5)$. For $g\in SU(2)$, $U^1(g) = O_g$, the matrix of Ad_g in the basis X_1, X_2, X_3. For $X\in\mathfrak{su}(2)$, where $X = x_1X_1 + x_2X_2 + x_3X_3$

$$U^1(e^X) = \exp(x_1J_1 + x_2J_2 + x_3J_3).$$

Since this is an orthogonal matrix, $\langle\cdot,\cdot\rangle$ is indeed Ad-invariant. In the notation of (12), $X_i^{\frac{1}{2}} = X_i$ and $X_i^1 = J_i$ for $1 \leq i \leq 3$. Moreover, $\lambda_{1/2} = 3/4$ and $\lambda_1 = 2$.

Note that $U^{\frac{1}{2}}$ and U^1 are irreducible representations of $SU(2)$. In other words, (51) is a special case of (32). By Proposition 1

$$\phi_Y(1/2) = U^{\frac{1}{2}}(g) R_C(1/2), \qquad \phi_Y(1) = U^1(g) R_C(1)$$

where in particular

$$U^{\frac{1}{2}}(g) = g, \qquad R_C(1) = \exp\left[\frac{C - \mathrm{tr}(C) I_1}{2}\right]. \tag{52}$$

Letting $M = \mathbb{E}[U(Y)]$, (24) gives the polar factors

$$R = \exp\left[-\frac{\mathrm{tr}(C) I_{\frac{1}{2}}}{4} \oplus \frac{C - \mathrm{tr}(C) I_1}{2}\right] V = g \oplus O_g. \tag{53}$$

Thus, using $U^{\frac{1}{2}}$ by itself allows for g, but not for C, to be recovered. On the other hand, U^1 does not give g but, according to (52), it is necessary in obtaining C. The choice of U simply consists in combining both $U^{\frac{1}{2}}$ and U^1. Recovering C in addition to g, from observations $Z_1, \ldots, Z_n$, by forming R_N and V_N as in (29), comes at the price of a slower convergence. Indeed, increasing dimensionality from $SU(2)$ to $SU(5)$ will increase both v and k in Proposition 5. If $C = \sigma^2 I_3$, then $U^{\frac{1}{2}}$ by itself allows for both g and C to be recovered.

VII. Application: Polarized Light in a Dispersive Fiber

The problem of estimating g, for the example of Section VI, has an immediate application to a signal processing situation. Precisely, this concerns the inverse problem of polarized light propagating in a dispersive optical fiber. In this section, the application is discussed and the inverse problem formulated in terms of physical quantities. The use of a comprehensive formalism for polarization optics, based on the tools of Section II-A, with $G = SU(2)$, was advanced in [2] and [35]. This has been successful in interpreting recent experimental results.

In a variety of signal processing situations, wave propagation is the physical support of signals of interest. Lightwaves have a vector nature, so that measurements made on them create a state of polarization, whose determination is of great physical significance. On the other hand, from the point of view of communication, this vector nature seems to offer a desirable redundancy. Unfortunately, it may happen that different components of a vector wave propagate at different speeds or with different attenuation. Formulation of a signal model which takes into account polarization-dependent effects arising for lightwaves naturally involves classical Lie groups, in particular compact Lie groups.

Transverse lightwaves are described using the two components of an electric field, in the plane orthogonal to the direction of propagation. An optical fiber acts as a waveguide so that the direction of propagation is constant. The desired signal model is obtained from the Stokes formalism of polarization optics. The following development relies on [35] and [36]. A general background may be found in [37].

Let l denote distance measured along the fiber, in the direction of propagation. The state of polarization is given by the so called Jones vector, $E(l)$ which has two complex components $E_1(l), E_2(l)$. These are the analytic signals corresponding to the components of the electric field. Let L be the total length of the fiber. The resulting inverse problem is to estimate $E(0)$ based on observation of $E(L)$. When the fiber is predominantly dispersive, this problem is resolved by an immediate application of the example of Section VI.

Physically, the situation is the following. In a perfectly circular fiber, the two components of the electric field propagate at the same speed. In terms of the Jones vector, the phase difference between $E_1(l)$ and $E_2(l)$ and the physical orientation of the vector $E(l)$ are constant. Small departures from circularity occurring throughout the length of the fiber lead to two random effects. Namely, phase difference between $E_1(l)$ and $E_2(l)$ and change of the physical orientation of $E(l)$. The word "random" here has a straightforward meaning. Repetition of an experiment, with the same input to the fiber, leads to different observed outputs belonging to some well defined probability distribution. The output of the fiber is $E(L)$. Iteration of the two effects, in general, leads to a random linear relation between $E(0)$ and $E(L)$

$$E(L) = T_L E(0) \tag{54}$$

where $T_L \in SU(2)$ is a random unitary matrix, known as the Jones matrix. Equation (54) constitutes the signal model to be used in solving the inverse problem. The unknown $E(0)$ appears as a parameter of the probability distribution of the observed $E(L)$. Indeed, $E(0)$ is usually prepared in a given deterministic value.

The physical origin of T_L suggests that it may be described as the terminal value of a diffusion T_l with values in $G = SU(2)$. Further simplifying assumptions show this to be a Brownian process. Precisely, independent stationary increments are justified by the local nature of physical effects and homogeneity of the fiber. It is clear that $T_0 = I_{1/2}$—recall the notation of Section VI. It follows that $T_L \sim N(I_{\frac{1}{2}}, TC)$, where C is a covariance matrix. The results of [36] suggest that $C = \sigma^2 I_1$. From (13) of Proposition 1, the average of $E(L)$ is a real multiple of $E(0)$. This relation is further refined by removing the redundancy present in $E(0)$.

If one is interested in the underlying state of polarization, then $E(0)$ contains information on the intensity of the lightwave, which is irrelevant. The state of polarization is uniquely given by the Stokes vector $S(0)$, which belongs to the unit sphere in $\mathbb{R}^3$. This is defined as $S(0) = (s_1/s, s_2/s, s_3/s)$, where s, s_1, s_2, s_3 are real numbers such that

$$E(0) \otimes E^\dagger(0) = s I_{\frac{1}{2}} + i \sum_{i=1}^{3} s_i \sigma_i. \tag{55}$$

The matrix on the left-hand side is known as the coherency matrix. Equation (54) is equivalent to

$$S(L) = M_L S(0), \qquad M_L = O_{T_L}. \tag{56}$$

In other words, $S(0)$ transforms under the adjoint representation of $SU(2)$, denoted U^1 in Section VI. From (13)

$$\langle S(L)\rangle = e^{-\sigma^2 L} S(0) \tag{57}$$

where angular brackets denote averaging over repeated experiments. Since $S(0)$ belongs to the unit sphere, it becomes possible to solve the inverse problem, identifying $S(0)$ and also σ^2 which is a physical constant of the fiber, related to its birefringent strength.

As mentioned in Section II-A, explicit construction of Brownian processes follows using multiplicative integration. The resulting numerical scheme is common in the literature, for example, [12], [23]. In the present situation, this can be particularly helpful. While it is easy to observe $S(L)$, the Jones matrix T_l—or the corresponding M_l, as in (56)—are "inside the fiber" and not easily accessible.

VIII. Validity of the Extrinsic Mean

In Section IV-A, it was shown g is the unique extrinsic mean of Brownian distributions $N_L(g, C)$ or $N_R(g, C)$. Here, the validity of the extrinsic mean for a wider class of distributions is shown. The proposed generalization will retain properties (14) and (15) of Brownian distributions. Let Z be a random variable with values in G. Assume also $Z \stackrel{d}{=} Z^{-1}$; Z is then said to be inverse invariant. Consider the problem of estimating $g \in G$ from an observed $Y = gZ$. The distribution of Y, parametrized by g, belongs to a location model, with g as location parameter. Proposition 9 below gives a sufficient conditions for g to be the unique extrinsic mean of the distribution of Y. This condition captures the importance of the Lévy property of Brownian processes, indicated in Section II-A. Recall, for instance from [1], that Z is inverse invariant *if* $\phi_Z(\delta)$ is Hermitian for all $\delta \in \mathrm{Irr}_+(G)$. The characteristic function of Y is given by

$$\phi_Y(\delta) = U^\delta(g)\phi_Z(\delta), \qquad \delta \in \mathrm{Irr}_+(G). \tag{58}$$

In particular, $U^\delta(g)$ is a left polar factor of $\phi_Y(\delta)$. If $\phi_Z(\delta)$ is strictly positive definite, for at least some $\delta \in \mathrm{Irr}_+(G)$, then g can be recovered as in (24).

For Brownian distributions, Proposition 1 showed that $\phi_Z(\delta)$ is strictly positive definite for all $\delta \in \mathrm{Irr}_+(G)$ whenever $Z \sim N_L(e, C)$. Other distributions that lead to the same property were encountered in [1]. If $(Z_n)_{n\geq 1}$ are i.i.d. random variables with values in G and N, a Poisson random variable, independent from $(Z_n)_{n\geq 1}$, then Z defined as follows is said to have a compound Poisson distribution, (putting $Z_0 = e$):

$$Z = \prod_{n=0}^{N} Z_n. \tag{59}$$

If Z_1 is inverse invariant then $\phi_Z(\delta)$ is strictly positive definite for all $\delta \in \mathrm{Irr}_+(G)$. Precisely

$$\phi_Z(\delta) = \exp\left[\tau(\phi(\delta) - I_\delta)\right] \tag{60}$$

where τ is the parameter of N and ϕ denotes the characteristic function of Z_1. Analytically, what is in common between (13) and (60) is the matrix exponential leading to strict positive definiteness. The same is true for any marginal distribution of a Lévy process in G.

Proposition 9: Let W be a Lévy process in G with $W_0 = e$, $Z = W_1$, $g \in G$ and $Y = gZ$. Let $U : G \to SU(v)$ be an injective unitary representation. If Z is inverse invariant then g is the unique global minimum for (22).

Proof: By a well-known reasoning (see, for example, [3]), there exist $d_\delta \times d_\delta$ matrices A_δ such that

$$\phi_t(\delta) = \exp(tA_\delta), \qquad \delta \in \mathrm{Irr}_+(G) \tag{61}$$

where $\phi_t \equiv \phi_{W_t}$. Indeed, it is straightforward to show $\phi_t(\delta)$ is continuous in t and

$$\phi_{t+s}(\delta) = \phi_t(\delta)\phi_s(\delta), \qquad \phi_0(\delta) = I_\delta.$$

This only uses the Lévy property of W, with $W_0 = e$, and implies the general form (61). It follows that if Z is inverse invariant, then $\phi_Z(\delta)$ is strictly positive definite for $\delta \in \mathrm{Irr}_+(G)$. Note here that $\phi_Z = \phi_1$.

Let U be an injective unitary representation, $V = U(g)$ and $M = \mathbb{E}[U(Y)]$. Writing U as in (32) and applying (58) and (61), it follows that

$$M = VR, \qquad R = \phi_Z(\delta_1) \oplus \cdots \oplus \phi_Z(\delta_r). \tag{62}$$

In particular, M is nonsingular and $M = VR$ is its left polar decomposition. The proposition now follows by the same reasoning as Proposition 4. ■

It may be interesting to recall the concept of positive-definite function on G[38]. Assume Z has a smooth density p with respect to μ. It is said that p is a positive-definite function on G if

$$\sum_{i,j=1}^{r} p(g_i^{-1}g_j)c_i\bar{c}_j \geq 0 \tag{63}$$

for all $r \geq 1$, $g_1, \ldots, g_r \in G$ and $c_1, \ldots, c_r \in \mathbb{C}$, with the bar denoting complex conjugation. In fact, p is positive definite *iff* $\phi_Z(\delta)$ is positive definite for all $\delta \in \mathrm{Irr}_+(G)$. Unfortunately, this falls short of the desired strict positive definiteness.

IX. Equivariant Estimators

The extrinsic mean estimator of Section IV-A can be regarded as an equivariant estimator, for the risk function arising from (22). The current section investigates the possibility of computing an estimator which improves upon that of Section IV-A, while retaining its property of equivariance. It is argued this could be carried out numerically, using the development of Section V.

Consider again the case $Y \sim N_L(g, C)$, for a fixed sample size N. The notation V_N and $Z_1, \ldots, Z_N$ is the same as for (29). With a slight abuse of notation, write $V_N = V(Z_1, \ldots, Z_N)$. Based on (22), the risk function associated to V_N is

$$R_N(g) = \mathbb{E}_g\|V(Z_1, \ldots, Z_N) - U(g)\|^2 \tag{64}$$

where $\mathbb{E}_g$ indicates $Z_1, \ldots, Z_N$ are taken from $N_L(g, C)$. The function R_N is constant

$$\begin{aligned} R_N(hg) &= \mathbb{E}_{hg} \| V(Z_1, \ldots, Z_N) - U(hg) \|^2 \\ &= \mathbb{E}_g \| V(hZ_1, \ldots, hZ_N) - U(hg) \|^2 = R_N(g). \end{aligned}$$

The second equality uses (14). The third equality uses unitary invariance of the Euclidean matrix norm and the fact that

$$V(hZ_1, \ldots, hZ_N) = U(h) V(Z_1, \ldots, Z_N) \tag{65}$$

which follows from an elementary property of Polar decomposition. According to a usual terminology, (65) states that V is an equivariant estimator of g [39]. The property of Equivariance does not uniquely define V. If V' is an equivariant estimator, then

$$V'(Z_1, \ldots, Z_N)^{-1} V(Z_1, \ldots, Z_N) = Q(Z_N^{-1} Z_1, \ldots, Z_N^{-1} Z_{N-1})$$

where Q is a $SU(v)$-valued function of $N-1$ variables, with values in G. A usual approach for improving upon V_N is to search for an equivariant estimator with minimum risk, by optimizing over the function Q. This leads to minimum risk equivariant (MRE), estimation (again, see [39]). Following a usual transformation, let $V'(Z_1, \ldots, Z_N) = U(Z_N)$. This is again equivariant, i.e., verifying (65). The MRE estimator is $V^*(Z_1, \ldots, V_N) = V'(Z_1, \ldots, Z_N) Q(Z_N^{-1} Z_1, \ldots, Z_N^{-1} Z_{N-1}^{-})$, where Q is given by

$$Q(Z_N^{-1} Z_1, \ldots, Z_N^{-1} Z_N) = \mathbb{E}_e[U(Z_N^{-1}) | Z_N^{-1} Z_1, \ldots, Z_N^{-1} Z_{N-1}].$$

In this formula, evaluation of the right-hand side requires knowledge of the joint distribution of Z_N^{-1} and $Z_N^{-1} Z_1, \ldots, Z_N^{-1} Z_N$. This is provided using the development of Section V. For simplicity, assume $N = 2$ so that the goal is to obtain

$$Q(Z_2^{-1} Z_1) = \mathbb{E}_e[U(Z_2^{-1}) | Z_2^{-1} Z_1]. \tag{66}$$

In this expression, Z_1 and Z_2^{-1} are independent and, by (11) and Proposition 1, both have distribution $N_L(e, C)$. Moreover, from the definition in Section V, the couples Z_2^{-1}, Z_2^{-1} and e, Z_1 are jointly Brownian, with joint characteristic functions written down as in (47). Since these couples are also independent, (9) can be used to calculate the joint characteristic function of $Z_2^{-1}, Z_2^{-1} Z_1$.

The last remark completely describes the joint distribution of $Z_2^{-1}, Z_2^{-1} Z_1$. This could be exploited to evaluate numerically (66), e.g., using Bayes formula.

X. Conclusion

This paper has considered marginal distributions of Brownian processes in compact Lie groups, which it has called Brownian distributions. These appear as belonging to parametric families, which display certain similarities to the well-known normal families. For Brownian distributions, as well as a more general class of distributions, the paper has justified the use of the extrinsic mean for the estimation of location parameters. Indeed, for the class of distributions in question, the location parameter was shown to be the unique extrinsic mean. Moreover, at least for Brownian distributions, resulting estimates were seen to have adequate asymptotic properties and to be computationally uncostly. A further aspect of the paper was the study of multivariate Brownian distributions. Several improvements to the presented material are possible. With respect to the extrinsic mean, as used for parametric estimation of Brownian distributions, two issues arise. First, it was shown that extrinsic estimation can simultaneously recover the location parameter (i.e., parameter g) and the concentration parameter (i.e., parameter C). However, conditions required on the embedding used for the extrinsic mean, in order to allow recovery of both parameters, were only given in a general form. It seems particularly interesting to be able to obtain an analytic or at least numerical construction of an embedding verifying these conditions—possibly, this could arise from a generalization of the example in Section VI.

Second, it should be noted the extrinsic mean does not, in a specific way, lead to optimal estimates for the location parameter. For finite sample size, it was argued that optimal equivariant estimates could be computed numerically, after a simplification due to the development of multivariate Brownian distributions. An open problem remains as to asymptotically optimal estimates. It should be noted that this is quite separate from the problem for finite sample size. While Brownian distributions can be generalized to arbitrary sample size, in the form of multivariate Brownian distributions, they do not seem to admit closed-form simplification independent of sample size. The discussion of multivariate Brownian distributions presented in this paper does not depend on the hypothesis that the underlying group is compact. However, compact Lie group structure was seen to impose strong restrictions on statistical analysis of multivariate Brownian distributions. It would be interesting, and relatively straightforward, to see if this situation is changed when considering Abelian or noncompact, e.g., nilpotent, Lie groups.

Appendix

Lemmas 1 and 2 give technical aspects of the proofs of Proposition 3 and 4, respectively. These lemmas state results from matrix analysis which were used in the proofs, but were skipped in order to maintain a direct flow of ideas.

Lemma 1 provides an estimate for the rest in the Taylor development of the matrix exponential of an antisymmetric matrix. Its use in the proof of Proposition 3 is explained below. As for lemma 2, it states the approximation property of Polar decomposition, with respect to the Euclidean matrix norm. It is a special case of a more general result given in [27]. This lemma was used in the proof of Proposition 4.

Lemma 1: Let X be a $v \times v$ complex matrix, $X^\dagger + X = 0$. For $n \geq 0$ let $X_n = \sum_{k=0}^{n} X^k / k!$. Then

$$\| \exp(X) - X_n \| \leq 2\sqrt{v} \left[\|X\|^n \wedge \|X\|^{n+1} \right] \tag{67}$$

where $\wedge$ indicates the minimum.

Proof: For $t \in [0, 1]$ and $n \geq 0$, let $X(t) = \exp(tX)$ and $X_n(t) = \sum_{k=0}^{n} (tX)^k / k!$. Let $h_n(t) = X(t) - X_n(t)$. Recall the identity

$$h_n(t) = \int_0^t h_{n-1}(t) X dt \tag{68}$$

which can be shown by induction. This holds for $n = 0$ if $h_{-1}(t) = X(t)X$. Using norm inequalities

$$\|h_n(1)\| \le \|h_{n-1}(1)\| \|X\|. \tag{69}$$

Since X is skew-Hermitian, $X(t)$ is unitary. It follows that

$$\|h_{-1}(1)\| \le \sqrt{v}\|X\|, \qquad \|h_0(1)\| \le 2\sqrt{v}.$$

The lemma follows by induction in (69), using both of these estimates. Note that $X(1) = \exp(X)$ and $X_n(1) = X_n$. ∎

Lemma 1 is used in the proof of Proposition 3 in order to obtain (21). Note that, according to this lemma, in the notation of the proposition

$$\|h(\xi_{Nn})\| \le 2\sqrt{v}\|\xi^\delta_{Nn}\|^2 \left[1 \wedge \|\xi^\delta_{Nn}\|\right]$$

where $\xi^\delta_{Nn} = \sum_{i=1}^d \xi^i_{Nn} X^\delta_i$. By condition (ii), the average of the squared norm is a constant multiple of $1/N$. Then, by Jensen's inequality, the factor in brackets is $O(1/\sqrt{N})$. Finally

$$\mathbb{E}\|h(\xi_{Nn})\| = O\left(1/N^{\frac{3}{2}}\right)$$

which is enough for (21), by a second application of Jensen's inequality.

Lemma 2: Let M be a nonsingular $v \times v$ complex matrix with polar decomposition $M = VR$, where V is unitary and R Hermitian strictly positive definite. All unitary $V' \neq V$ verifies the inequality $\|M - V\| < \|M - V'\|$.

Proof: Since V, V' are both unitary, $V^\dagger V = V'V' = I_v$. Letting $E = V' - V$, it follows that

$$V^\dagger E + E^\dagger V + E^\dagger E = 0. \tag{70}$$

Indeed, the left-hand side is equal to $V^\dagger V - V'V'$.

A straightforward computation gives

$$\|M - V\|^2 = \|M\|^2 - \|V\|^2 - 2\mathrm{tr}[(M - V)^\dagger V].$$

Subtracting from a similar identity for V' and using $\|V\| = \|V'\|$ yields

$$\|M - V'\|^2 - \|M - V\|^2 = -2\mathrm{tr}(ME). \tag{71}$$

Using (70) and $M = VR$, it follows that

$$2\mathrm{tr}(ME) = -\mathrm{tr}(ERE^\dagger).$$

Replacing in (71), the strict inequality is seen to hold since R is strictly positive definite. ∎

In [27], it is shown that if the Euclidean matrix norm is replaced by a general unitary invariant matrix norm, the lemma continues to hold, but with strict inequality replaced by plain inequality ($\le$). This shows the importance of Euclidean matrix norm to Proposition 4.

REFERENCES

[1] S. Said, C. Lageman, N. Le Bihan, and J. H. Manton, "Decompounding on compact lie groups," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2766–2777, Jun. 2010.

[2] S. Said, "Estimation and filtering of processes in matrix lie groups," Ph.D. dissertation, Institut Polytechnique de Grenoble, Grenoble, France, 2009.

[3] M. Liao, *Lévy Processes on Lie Groups*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Commun. Pure Appl. Math.*, vol. 30, pp. 409–541, 1977.

[5] W. Kendall, "Probability, convexity and harmonic maps with small image, I: Uniqueness and fine existence," *Proc. Lond. Math. Soc.*, vol. S3-61, no. 2, pp. 371–406, 1990.

[6] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds. I," *Ann. Statist.*, vol. 31, no. 1, pp. 1–29, 2003.

[7] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds. II," *Ann. Statist.*, vol. 33, no. 3, pp. 1225–1259, 2005.

[8] J. H. Manton, "A globally convergent numerical algorithm for computing the centre of mass on compact lie groups," in *Proc. Int. Conf. Control, Autom., Robot. Vis.*, 2004, pp. 2211–2216.

[9] J. H. Manton, "A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case," *Digital Signal Process.*, vol. 16, pp. 468–478, 2006.

[10] M. Fréchet, "Les elements aleatoires de nature quelconque dans un espace distancie," *Annales de l'Institut Henri Poincare*, vol. 10, no. 4, pp. 215–310, 1948.

[11] A. Srivastava and E. Klassen, "Monte Carlo extrinsic estimators of manifold-valued parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 299–208, 2000.

[12] A. Srivastava, "A Bayesian approach to geometric subspace estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1390–1400, May 2000.

[13] A. Srivastava, M. I. Miller, and U. Grenander, "Jump-diffusion processes on matrix lie groups for Bayesian inference," in *Proc. IEEE Signal Process. Workshop High-Order Statist.*, 1999, pp. 126–129.

[14] D. Applebaum, "Lévy processes in stochastic differential geometry," in *Lévy Processes. Theory and Applications*, O. Barndorff-Nielsen, T. Mikosch, and S. Resnick, Eds. Cambridge, MA: Birkhäuser, 2001.

[15] U. Grenander, *Probabilities on Algebraic Structures*. New York: Wiley, 1963.

[16] G. S. Chirikjian and A. B. Kyatkin, *Engineering Applications of Noncommutative Harmonic Analysis*. Boca Raton, FL: CRC Press, 2000.

[17] G. S. Chirikjian, *Stochastic Models, Information Theory and Lie Groups*. Cambridge, MA: Birkhäuser, 2011, vol. 1,2.

[18] J. Faraut, *Analysis on Lie Groups, an Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[19] J. M. Elworthy, *Stochastic Differential Equations on Manifolds*. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[20] T. Bröcker and T. D. Tom, *Representations of Compact Lie Groups*. New York: Springer-Verlag, 1985.

[21] I. Chavel, *Riemannian Geometry: A modern Introduction*. Cambridge, MA: Cambridge Univ. Press, 2006.

[22] S. Xavier and J. H. Manton, "On the generalization of AR processes to riemannian manifolds," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2006, pp. 468–478.

[23] Y. Wang and G. S. Chirikjian, "Nonparametric second order theory of error propagation on motion groups," *Int. J. Robot. Res.*, vol. 27, no. 11/12, pp. 1258–1273, 2008.

[24] H. Heyer, *Probability Measures on Locally Compact Groups*. New York: Springer-Verlag, 1977.

[25] T. P. Fletcher, S. Joshi, C. Lu, and S. M. Pizer, "Gaussian distributions on lie groups and their application to statistical shape analysis," in *Proc. Inf. Process. Med. Imag.*, 2003, pp. 450–462.

[26] T. P. Fletcher, S. Joshi, C. Lu, and S. M. Pizer, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 995–1005, Aug. 2004.

[27] N. J. Higham, *Functions of Matrices, Theory and Computation*. Philadelphia, PA: SIAM, 2008.

[28] U. Grenander, M. I. Miller, and A. Srivastava, “Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 1258–1273, Aug. 1998.

[29] O. Kallenberg, *Foundations of Modern Probability*. New York: Springer-Verlag, 2002.

[30] T. Papadopoulo and M. Lourakis, Estimating the Jacobian of the singular value decomposition: Theory and applications. Grenoble, France, 2000, INRIA, Tech. Rep. 3961.

[31] C. Udriste, *Convex Functions and Optimisation Methods on Riemannian Manifolds*. Norwell, MA: Kluwer, 1994.

[32] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] G. S. Chirikjian, “Information-theoretic inequalities on unimodular lie groups,” *J. Geometr. Mech.*, vol. 2, no. 2, pp. 119–158, 2010.

[34] S. L. Altmann, *Rotations, Quaternions and Double Groups*. New York: Dover, 2005.

[35] S. Said and N. Le Bihan, “Higher order statistics of stokes parameters in a random birefringent medium,” *Waves Random Complex Media*, vol. 18, no. 2, pp. 275–292, 2008.

[36] A. Vanucci and A. Bononi, “Statistical characterization of the jones matrix of long fibers affected by polarization mode dispersion (PMD),” *J. Lightw. Technol.*, vol. 20, no. 5, pp. 811–821, May 2002.

[37] C. Brosseau, *Fundamentals of Polarized Light: A Statistical Approach*. New York: Wiley, 1998.

[38] R. E. Edwards, *Integration and Harmonic Analysis on Compact Groups*. Cambridge, U.K.: Cambridge Univ. Press, 1972.

[39] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York: Springer-Verlag, 1998.

Salem Said completed his Ph.D. at Institut Polytechnique de Grenoble in 2009, under the supervision of Drs. Nicolas Le Bihan and Stephen J. Sangwine. His Ph.D. theses was titled “Estimation and filtering of processes in matrix Lie groups”. Since 2010 he has been working at the Department of Electrical and Electronic Engineering of the University of Melbourne, under the supervision of Pr. Jonathan H. Manton. His research focuses on stochastic processes with values in manifolds, related estimation and filtering problems and their application to wave physics. His broad area of expertise is in stochastic calculus and its application to signal processing and statistics.

Jonathan H. Manton (M’02–SM’03) received the B.Sc. degree in mathematics and the B.Eng. degree in electrical engineering in 1995 and the Ph.D. degree in 1998, all from the University of Melbourne, Melbourne, Vic., Australia. He holds a Distinguished Chair at the University of Melbourne with the title Future Generation Professor. He is also an Adjunct Professor in the Mathematical Sciences Institute at the Australian National University. From 1998 to 2004, he was with the Department of Electrical and Electronic Engineering at the University of Melbourne. During that time, he held a Postdoctoral Research Fellowship then subsequently a Queen Elizabeth II Fellowship, both from the Australian Research Council. In 2005 he became a Full Professor in the Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE), Australian National University. From July 2006 to May 2008, he was on secondment to the Australian Research Council as Executive Director, Mathematics, Information, and Communication Sciences. His traditional research interests range from pure mathematics (e.g., commutative algebra, algebraic geometry, differential geometry) to engineering (e.g., signal processing, wireless communications). Recently though, led by a desire to participate in the convergence of the life sciences and the mathematical sciences, he has started to apply his expertise to research in neuroscience. He also has extensive experience in software development. Prof. Manton has served recently as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Committee Member for the IEEE Signal Processing for Communications (SPCOM) Technical Committee, and a Committee Member on the Mathematics Panel for the ACT Board of Senior Secondary Studies in Australia. He is currently on the Editorial Board of Systems and Control Letters. He has recently received a 2009 Future Summit Australian Leadership Award in recognition of “outstanding achievements and commitment to playing a leading role in shaping the future of Australia.