Jump–diffusion Markov processes on orthogonal groups for object pose estimation

Anuj Srivastava\textsuperscript{a, *}, Ulf Grenander\textsuperscript{b}, Gary R. Jensen\textsuperscript{c},
Michael I. Miller\textsuperscript{d}

\textsuperscript{a}Department of Statistics, Florida State University Tallahassee, FL 32306, USA
\textsuperscript{b}Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
\textsuperscript{c}Department of Mathematics, Washington University, St. Louis, MO 63130, USA
\textsuperscript{d}Department of Electrical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

In the problem of recognizing targets from their observed images, the estimation of target orientations, as elements of the rotation group SO(3), plays an important role. For \( k \)-objects the unknown parameter is an element of \( \text{SO}(3)^k \). Since \( k \) may be unknown a priori, the parameter space is extended to \( \mathcal{X} = \bigcup_{k=0}^{\infty} \text{SO}(3)^k \). In this representation, both the target orientations and their numbers have to be estimated simultaneously. We present a Bayesian approach that builds a posterior probability measure on \( \mathcal{X} \). Then, utilizing a Markov jump–diffusion process \( X(t) \), we sample from this posterior to empirically generate the estimates. The two components of \( X(t) \), jumps and diffusions, are chosen in such a way that the resulting Markov process has the desired ergodic property: averages along its sample paths converge to the expectations under the posterior. Proper choice of the diffusion parameters and the jump intensities is demonstrated and the ergodic result associated with \( X(t) \) is proven. An example, involving the estimation of an airplane orientation, is used to illustrate this jump–diffusion algorithm. © 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 60J60; 60J75; secondary 93E23; 93E10

Keywords: Pattern theory; Jump–diffusion processes; Metropolis; sde on manifolds; Automated target recognition; MCMC sampling; Pose estimation

1. Introduction

Automated recognition of objects, from their observed images, has been an area of active research in recent years. One motivation for this research comes from the fact that the “human vision” is an impressive system with extraordinary processing power. During the last decade, progress has been made towards designing automated object...
recognition systems with similar performance. An emerging approach to address automated recognition has been the Bayesian technique based on statistical models of system components such as objects, sensors, and clutter. Since the probability models associated with these physical systems are complicated, general closed-form analytical solutions are ruled out, and hence, the computational approaches have become important. We explore a family of computational solutions applied to finding the unknowns associated with object recognition problems. A general automated target recognition (ATR) problem is as follows: a remote sensor (such as a camera, radar, or ladar) observes a scene containing a number of targets, either moving or stationary. This sensor produces observations, in the form of images or signals, which are then analyzed by computer algorithms to detect, track and recognize the targets of interest in that scene. A more detailed introduction can be found in Dugden and Lacoss (1993) or Srivastava et al. (1999).

One fundamental issue in ATR is the so-called variability in target pose. Consider a regular hand-held camera taking pictures of an object, say a car. Depending on the relative orientation between the camera and the car, and the distance between them, the car appears vastly different in different pictures. The possible variability in relative orientation, also called the pose, causes a tremendous variability in the images (pixel-values) of the targets as seen by a camera, or a sensor in general. This fact underlines a major difficulty in the design of automated algorithms for ATR. How to mathematically model the variation in the images (or the pixel-values) due to the variation in target pose? The task is further complicated by relative motion between the sensors and the targets, imperfections in sensor operations, and the presence of structured clutter in the scene which often obscures the targets. Following Grenander’s formulation of general pattern theory (Grenander, 1993), we have separated the two components. On one hand, we have developed detailed three-dimensional models for handling physical variation in target occurrences in a scene, and on the other hand, using physical considerations, we have mathematically modeled the transformations which incorporate the sensor operation. For modeling the target occurrences, we utilize ideas from the deformable template theory (Grenander et al., 1990; Grenander, 1993): we start from a three-dimensional CAD (for example a triangulated surface) model for each possible object we expect to find in the scene. These CAD models are called templates. Since the objects occur in a scene at arbitrary positions and orientations, these occurrences are then modeled by the actions of rotation and translation groups, on the appropriate templates. The physics of sensor operation and any prior knowledge on the object occurrence are combined to form a posterior density on these groups. In this scheme, given an observed image, the object recognition task reduces to: (i) estimating the target pose and location, and (ii) hypothesis selection of the most likely template. The object orientation is modeled by the elements of the special orthogonal group, \( \text{SO}(3) = \{ O \in \mathbb{R}^{3 \times 3} : O^T O = I, \det(O) = 1 \} \) and the translation by elements of \( \mathbb{R}^3 \). Together, their joint action is the special Euclidean group \( \text{SE}(3) \), the semi-direct product of \( \text{SO}(3) \) and \( \mathbb{R}^3 \). In this paper we will restrict to estimating the target orientations in order to isolate and focus on issues related to inferences on \( \text{SO}(3) \), a curved
Lie group. In case of multiple objects present in the scene, or articulated objects with many rotating components, the total pose variation is an element of $\text{SO}(3)^k$ for some positive integer $k$ (denote $\mathcal{X}_k = \text{SO}(3)^k$). Since the number $k$ is unknown a priori, the parameter space is generalized to include all values of $k$, i.e. define $\mathcal{X} = \bigcup_{k=0}^{\infty} \text{SO}(3)^k$.

Each $\mathcal{X}_k$ is then a subspace of $\mathcal{X}$.

As a first step, we will establish a posterior distribution on $\mathcal{X}$ with contributions from a prior distribution and the sensor data-likelihood function. For each subspace $\mathcal{X}_k$ ($\mathcal{X}_k$ is a Lie group), the Haar measure (denoted by $\gamma_k$), is chosen as the base measure with respect to which probability densities are defined. One representation of $\gamma_1$ using the exponential coordinates is given in Appendix A. Let $\pi_{0,k}$ be a prior density on $\mathcal{X}_k$ and $L_k$ be the likelihood function, then given $k$, the posterior probability measure on $\mathcal{X}_k$ is

$$
\mu_k(A) = \frac{\int_{A} \pi_{0,k}(s)L_k(s)\gamma_k(ds)}{\int_{\mathcal{X}_k} \pi_{0,k}(s)L_k(s)\gamma_k(ds)},
$$

for any measurable set $A \subset \mathcal{X}_k$. The prior density incorporates any contextual information available, which is often the case in ATR situations. For example, any knowledge about the past target motion puts a strong prior on its next location and pose. The likelihood function is a function of the target orientations given the observed images. It comes from the probability density of observing a given image conditioned on a value of target pose. To generate a posterior probability measure on the whole $\mathcal{X}$, call it $\mu$, we take a convex combinations of $\mu_k$’s weighted by the probability of $k$.

Next, similar to traditional estimation-theoretic contexts to generate inferences, we seek statistics under the posterior, as well as solutions of high posterior probability. Unlike the processing strategies for optimal detection and recognition in classical linear estimation problems, there is a nonlinear relationship between the target parameters and the image synthesis models. The posterior $\mu$ is often too complicated to analytically derive the classical estimators such as the conditional means and variances. Hence, taking a numerical approach, we will generate these estimators empirically. That is, we will simulate from the posterior $\mu$ in such a way that averaging the samples approximates, in an asymptotic sense, the conditional means and variances. The main tool presented here for simulating from $\mu$ is a Markov process having two distinct components: (i) diffusions generated by solving stochastic differential equations on the subgroups $\mathcal{X}_k$, and (ii) discrete jumps from one subspace to another to cover all of $\mathcal{X}$. This construction extends to Lie groups the jump–diffusion processes originally described in Grenander and Miller (1994) for vector spaces. A similar sampling process using reversible Markov jumps is presented in Green (1995) and the references therein. The diffusions contribute in sampling over the parameter values for a fixed parameter size while the jumps contribute in sampling over the parameter size. Note that here the jumps contribute in inference on parameter size but the framework can be generalized to select the target classes via jumps to achieve a full target recognition system (see for example Miller et al., 1995).

As described in Geman and Geman (1984), Geman and Hwang (1987) and Grenander and Miller (1994), a diffusion process which samples from a given
probability density $\pi$ can be constructed using the Langevin’s stochastic differential equation (SDE)

$$dX(t) = -\nabla \log(\pi(X(t))) \, dt + \sqrt{2} \, dW(t),$$

where $W(t)$ is a standard Wiener process of appropriate dimensions. It has been shown that, under certain regularity conditions on $\log(\pi)$, $\pi$ is the unique stationary probability-density of the Markov process $X(t)$. Langevin’s SDE is valid only for vector spaces; for curved Lie groups (such as $\mathbb{X}_k$) this equation has to be modified according to the underlying geometry. From the differential geometry of SO(3), we specify the tangent vector space at each point on SO(3). Then, for each of the $k$-components in $\mathbb{X}_k$, the process is allowed to move only along the tangential directions, resulting in a flow which stays on $\mathbb{X}_k = SO(3)^k$. Adding a stochastic component, similar to Eq. (1), results in a stochastic gradient flow, or a diffusion, on that subspace. To extend to $\mathbb{X}$, a family of discrete moves, called the jump moves, which take the process from one subspace $\mathbb{X}_k$ to another $\mathbb{X}_k'$ are added. The two components are combined together to form a jump–diffusion process $X(t)$ in such a way that it satisfies the ergodic result: the empirical averages of functions evaluated on the sample paths converge to their expectations under the posterior $\mu$. This is a mechanism for generating statistics from the data (conditional means, covariances, and others) when the posterior is nonlinear in the parameter space. To make it rigorous we will study equilibrium measures and ergodicity properties. It is shown that by selecting the infinitesimal drifts and jump intensities properly, $\mu$ becomes the unique stationary probability measure of the jump–diffusion Markov process $X(t)$. From this, it is shown that empirical averages of the samples exhibit asymptotic convergence to their conditional expectations, i.e.

$$\frac{1}{T} \int_0^T f(X(t)) \, dt \overset{T \to \infty}{\longrightarrow} \int_{\mathbb{X}} f(s) \mu(ds),$$

for all bounded, measurable functions $f$ defined on $\mathbb{X}$.

There are essentially two parts to proving the ergodic results: (i) the posterior $\mu$ is a stationary measure of the jump–diffusion Markov process $X(t)$, and (ii) it is a unique stationary probability measure. Let the transition probability density of a Markov process, of transitioning from $x$ to $y$ in time $t$, be denoted by $P_t(x,y)$. Define an operator, indexed by $t$, according to

$$T_t f(y) = \int f(x) P_t(x,y) \, dy$$

for $f$ in some appropriate Banach space, call it $L$. The family of operators, $\{T_t : t \geq 0\}$ forms a semigroup that completely characterizes the behavior of a Markov process. The (infinitesimal) generator of a semigroup $\{T_t\}$ on $L$ is a linear operator defined by

$$A f = \lim_{t \to 0} \frac{1}{t} \{T_t f - f\}.$$ 

The domain $\mathcal{D}(A)$ is the set of all functions in $L$ for which this limit exists. We will characterize the generator of the semigroup associated with jump–diffusion processes, and essentially rely on the following result from Ethier and Kurtz (1986).
Theorem 1 (Ethier and Kurtz, 1986). Suppose an operator $A$ generates a strongly continuous, contraction semigroup on a closed subspace $L$, $L$ is separating, and the martingale problem for $A$ is well posed. Then, a probability measure $\mu$ is a stationary measure of the Markov process if and only if

$$\int A f(s) \mu(ds) = 0, \quad f \in D$$

for $D \subset \mathcal{D}(A)$, a core for $A$.

This theorem establishes the conditions for $\mu$ to be a stationary measure of $X(t)$. For establishing uniqueness, we will apply Harris recurrence. As observed in Doss et al. (1996), Harris recurrence is difficult to establish directly. We will utilize the existence of a stationary measure (namely the posterior) and the $\mu$-irreducibility to establish the Harris recurrence. For a $\mu$-irreducible process with Harris recurrence, Meyn and Tweedie (1993a,b) has proven the ergodic result (using ideas from the boundedness in probability).

There exists a vast literature on utilizing Markov chain Monte-Carlo-type algorithms for sampling from given probability distributions, see for example Green (1994), Meyn and Tweedie (1993a,b), Gelfand (1996), Gelfand and Smith (1990), Roberts and Tweedie (1995) and many more. Green (1994) has applied these techniques for pixel-based image analysis and object discovery. Comprehensive ergodic results associated with the discrete Markov chains and continuous Markov processes are proven in Doss et al. (1996), Meyn and Tweedie (1993a,b) and Athreya and Ney (1972). In addition, the role of stochastic processes on groups has been explored in the context of filtering applications by Duncan (1997, 1979, 1990), Brockett (1972) and others. Chang and Bingham (1996) have presented an analysis of the Bayesian spherical regression involving Fisher distribution on the orthogonal groups. In this paper, we utilize results from these references, and Grenander and Miller (1994), to develop empirical approaches for object pose estimation.

Section 2 describes the problem of object recognition viewed as a problem in parametric Bayesian inference. The next two sections describe a construction of jump–diffusion process $X(t)$ to sample from the posterior: Section 3 describes sampling on $\mathcal{X}_k$ for a fixed $k$ through a diffusion process only, and Section 4 adds a jump component to the diffusion process to sample over the complete $\mathcal{X}$. Section 5 presents a specific jump–diffusion algorithm to construct $X(t)$ and Section 6 illustrates this algorithm through an experiment on airplane pose estimation.

2. Bayesian approach to automated target recognition

As described in Srivastava et al. (1997, 1999) and Grenander et al. (1998, 2000), we take a model-based statistical approach to ATR. We will use ideas from the deformable
template theory (Grenander, 1993) to mathematically model the pose variation associated with the targets. Let \( z \) be a variable indexing the targets; for example \( z \) can be M60-tank or T72-tank or jeep, etc. The physical attributes of a target (such as shape, material, size, and polygonal surface reflection), indexed by \( z \), constitute a template, denoted by \( I^z \). As an example, for a rigid object, a CAD model consisting of surface patches, their vertices and normals, constitute its template. Since the targets occur in scenes at arbitrary positions and orientations, with respect to the sensors, we have to mathematically model this variability. The target-rotations are represented by elements of special orthogonal group, \( \text{SO}(3) \), while the translations are elements of \( \mathbb{R}^3 \). Together they form the special Euclidean group \( \text{SE}(3) \), the semi-direct product of \( \text{SO}(3) \) and \( \mathbb{R}^3 \). In this paper we will isolate and focus on the orthogonal group for target pose estimation. For an element \( s \in \text{SO}(3) \), \( sI^z \) denotes a target at a relative orientation of \( s \) with respect to a fixed frame of reference. The collection of all possible occurrences of this target forms an orbit \( J^z \), i.e. \( J^z = \{ sI^z : s \in \text{SO}(3) \} \). In this scheme of representing targets, given an observed image, the task is to estimate which target \( (z) \) and which transformation \( (s) \) best explain the observation. In the case of multiple targets, with the number of targets unknown a priori, the transformation space becomes \( \mathcal{X} = \bigcup_{k=0}^{\infty} \mathcal{X}_k \), for \( \mathcal{X}_k = \text{SO}(3)^k \).

A sensor (or multiple sensors) observing a scene gives rise to observations which can be in the form of images, sampled signal waveforms, or arrays of numbers. The characteristics of these observations are dependent upon the sensor mechanism which captures the image, and often its mode of operation. This mechanism can be thought of as a mapping \( T \) from the target space \( J^z \) to an observation space \( J^\mathcal{D} \), \( T : J^z \to J^\mathcal{D} \). In other words, \( TsI^z \) is the ideal (noiseless) image of the target \( z \) at the orientation \( s \) from the sensor. \( T \) generally constitutes nonlinear transformations such as projections, obscurations, and accumulations. For statistical object recognition and improved performance, the statistical behavior of \( T \) has to be included in the algorithms. \( T \) can be incorporated using sensor simulators such as PRISM, IRMA, XPATCH, etc., or by collecting real images of the targets of interest at sampled orientations and pre-storing for use in algorithms. For video images, we have utilized SGI workstations to simulate the projection imaging \( T \). The left panel in Fig. 1 depicts \( T \) as an orthographic projection scheme, while the right panel illustrates the perspective projection. Four sample images of an airplane template are shown in Fig. 2.

In addition to the projection mechanism, the sensing process includes a random noise \( \varepsilon \), whose characteristics depend upon the sensor and its mode of operation. For additive noise models, the image obtained from the sensor is modeled by the equation

\[
I^D = \sum_{i=1}^{k} Ts_i I^{x_i} + \varepsilon \in J^\mathcal{D}, \quad s_i \in \text{SO}(3)
\]

for \( k \) targets. In practice, the contributions from different objects do not appear additively in the image, but rather through nonlinear maps (involving projections and obscurations). To keep the discussion simple here, we will assume additive contributions. Assuming additive Gaussian noise, the likelihood of observing an image \( I^D \) given
Fig. 1. Illustration of the imaging process as orthographic and perspective projections.

Fig. 2. Left panel: Three-dimensional rendering of an airplane. Other panels: noisy images of this airplane at four different orientations.

the orientation $s$ is:

$$L_k(s) = \frac{1}{(2\pi \sigma^d)^{d/2}} \exp \left\{ -\frac{1}{2 \sigma^2} \left\| T^D - \sum_{i=1}^{k} T_{s_i} I_{s_i} \right\|^2 \right\}, \quad d = \text{dim}(\mathcal{X}),$$

(6)

where $s = [s_1, s_2, \ldots, s_k] \in \mathcal{X}_k$ for some $k$.

Working towards a classical Bayesian approach, we first derive a posterior probability distribution $\mu$ on $\mathcal{X}$ in the following way. According to Bayes’ rule, the posterior probability is the product of the prior probability and the data-likelihood function. Let $\gamma_k$ be the Haar measure on the subspace $\mathcal{X}_k$ and let $\pi_{0,k}$ be the prior density on $\mathcal{X}_k$, for each $k$. Then, the posterior density, for a fixed $k$, is given by

$$\pi_k(s|T^D) = \frac{\pi_{0,k}(s)L_k(s)}{P(T^D)} = \frac{e^{-H_k(s)}}{\mathcal{Z}_k}, \quad s \in \mathcal{X}_k,$$

(7)

where $H_k: \mathcal{X}_k \rightarrow \mathbb{R}_+$ is called the posterior energy. $\mathcal{Z}_k$ is the normalizer on $\mathcal{X}_k$. This posterior is extended to the union set $\mathcal{X}$ by taking convex combinations of $\mu_k$’s. Let $\gamma$ be the base measure on $\mathcal{X}$, according to the definition, $\gamma(A) = \sum_k \gamma_k(A \cap \mathcal{X}_k)$. For all $\gamma$
measurable sets \( A \subset \mathcal{X} \), define the posterior probability-measure \( \mu \) according to
\[
\mu(A) = \sum_{k=0}^{\infty} \mu_k(A \cap \mathcal{X}_k) = \sum_{k=0}^{\infty} \int_{A \cap \mathcal{X}_k} \frac{1}{Z} e^{-H_k(s)} \gamma_k(ds)
\]
with the normalizer \( Z = \sum_{k=0}^{\infty} \int_{\mathcal{X}_k} e^{-H_k(s)} \gamma_k(ds) \).

Having derived a posterior probability, we seek classical Bayesian estimators such as the conditional-means and variances under this posterior. Due to the complicated posterior, an analytical derivation of these quantities is not feasible and we take a numerical approach. We generate samples from \( \pi \) in such a way that their sample averages converge to the conditional expectations, in an asymptotic fashion. Our method to generate samples utilizes a jump–diffusion Markov process \( X(t) \), taking values in \( \mathcal{X} \), and having suitable ergodic properties. In the next two sections, we construct and analyze this Markov process, starting with only the diffusions and later introducing the jumps.

3. Diffusions on subspace (SO(3)\(^k\))

In this section, we focus on constructing a diffusion Markov process taking values on \( \mathcal{X}_k \), for some fixed \( k \), in such a way that \( \mu_k \) is its unique stationary probability measure. \( \mathcal{X}_k = \text{SO}(3)\(^k\) \) is the space of \( k \) rotations, \( k \) assumed fixed in this section. Without loss of generality we assume \( k = 1 \). The posterior density on \( \mathcal{X}_1 = \text{SO}(3) \) is given by
\[
\pi_1(s) = \frac{e^{-H_1(s)}}{Z_1}, \quad s \in \text{SO}(3),
\]
where \( \gamma_1 \) is the Haar measure and \( Z_1 \) is the normalizer on \( \text{SO}(3) \). It is well known that for flat vector spaces, the diffusions are generated as solutions of Langevin’s SDE, Eq. (1) (see for example Geman and Geman, 1984; Geman and Hwang, 1987). This equation can also be viewed as stochastic gradient equation with the infinitesimal mean given by the gradient of log-posterior and infinitesimal variance being constant.

Using that idea, we generate diffusions as stochastic gradient processes, on \( \text{SO}(3) \), for the energy function \( H_1 \). Let \( Y_1, Y_2 \), and \( Y_3 \) be the three orthonormal, smooth vector fields on \( \text{SO}(3) \) (refer to Appendix A for details). Then, for any point \( s \in \text{SO}(3) \), \( Y_i H_1 \), \( i = 1,2,3 \), is the directional derivative of the posterior energy \( H_1 \) in the direction of \( Y_i \) (at point \( s \in \text{SO}(3) \)). The vector, \( \sum_{i=1}^{3} (Y_i H_1) Y_i, \) is the gradient vector, or the tangential direction of maximum change in the value of \( H_1 \), at \( s \). A diffusion Markov process on \( \text{SO}(3) \), which samples from the posterior \( \mu_1 \) on \( \text{SO}(3) \), is a solution of the SDE
\[
dX(t) = - \sum_{i=1}^{3} (Y_i H_1) Y_{i,X(t)} dt + \sum_{i=1}^{3} Y_{i,X(t)} \circ dW_i(t).
\]

\( W_i \)'s are real-valued, independent, standard Wiener processes and \( \circ \) denotes the Stratonovich’s integral.

For characterizing the stationary measure of this \( X(t) \), we use Theorem 1. Let \( A \) be the generator (as defined in Eq. (3)), with the domain \( \mathcal{D}(A) \), generating \( X(t) \).
Theorem 2. Assume that posterior energy \( H_1 \) in Eq. (9) is a smooth function on SO(3). Then, for the diffusion process \( X(t) \in \text{SO}(3) \), the solution of Eq. (10), 
\[
\mu_1(ds) = (e^{-H_1(s)}/Z_1) \gamma_1(ds)
\]
on SO(3) is a stationary measure.

Proof. To use Theorem 1, let \( X(t) \) have the transition measure, \( P_t(s_1, ds_2) = \Pr \{ X(t) \in ds_2 | X(0) = s_1 \} \), and with it associate a semigroup \( \{ T_t : t \geq 0 \} \) of bounded linear operators on the Banach space \( L \). Here, \( L \equiv \hat{C}(\text{SO}(3)) \) is the space of continuous functions on SO(3); \( L \) is closed and separating. The action of \( T_t \) on \( L \) is defined by Eq. (2). The infinitesimal generator of the semigroup on \( L \) is a linear operator \( A^d \) defined by Eq. (3), with the domain \( \mathcal{D}(A^d) \subset L \). Take as the core for \( A^d \), \( D(A^d) = C^\infty(\text{SO}(3)) \subset \mathcal{D}(A^d) \) (the set of all the smooth functions), and define \( \mathcal{P}(\text{SO}(3)) \) as the set of all the probability measures on SO(3).

To utilize Theorem 1 for proving stationarity, we must establish that the diffusion operator \( A^d \) satisfies the conditions in Theorem 1. Our approach is to consider SO(3) as a collection of neighborhoods, each having an associated local coordinate system, called local charts. The generator \( A^d \), of the diffusion process \( X(t) \), takes the form (for a derivation refer to Kunita, 1984, p. 260, Eq. (13), or to the Hörmander form in Kliemann, 1987)
\[
A^d = -\sum_{i=1}^{3} (Y_i H_1) Y_i + \sum_{i=1}^{3} Y_i^2.
\]

On a coordinate neighborhood, using the local charts for Euclidean representation, \( A^d \) is given by
\[
A^d = -\sum_{i=1}^{3} a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{3} b_{ij} \frac{\partial^2}{\partial x_i \partial x_j},
\]
where \( x_i \)'s are the local coordinates of \( s \in \text{SO}(3) \) and \( a_i \) and \( b_{ij} \) are the infinitesimal means and the covariances of the components of \( X(t) \), respectively, in these local coordinates. Note, \( a_i \) are smooth functions of the partial derivatives of \( H_1 \). Since the directional derivatives of \( H_1 \) are Lipschitz continuous (\( H_1 \) is assumed \( C^\infty \)) and \( a_i \)'s and \( b_{ij} \)'s are representation through smooth charts, they are also Lipschitz. Then, Eq. (10) has a unique solution, in particular, a unique stopped solution in each coordinate neighborhood. Since the coefficients are bounded, by Ethier and Kurtz (1986, Theorem 3.6, p. 296 together with Corollary 3.4, p. 295), the stopped martingale problem for \( A^d \) in each coordinate neighborhood has a unique solution with continuous sample paths. Consequently, using Ethier and Kurtz (1986, Theorem 6.2, p. 217) there exists a unique solution to the martingale problem for \( A^d \) on SO(3) with continuous sample paths and this Markov process generates a contraction semigroup.

Next we show that the semigroup \( \{ T_t \} \) is also strongly continuous, or in other words, \( \lim_{t \to 0} T_t = \text{identity} \). It is given that the infinitesimal drifts (\( \sum (Y_i H_1) \)) are smooth functions on SO(3). For small \( t \), the transition probability \( P_t \) can be expressed in local coordinates. \( P_t(x, dy) \) is essentially normal with mean \( x \) and variance a function of \( t \), integrated over \( dy \). Since, for \( t \to 0 \), \( P_t(x, dy) \) converges to \( \delta_x(y) dy \), the result is immediate.
Summarizing these results, the operator \( A^d \) generates a strongly continuous, contraction semigroup on \( L \) and the martingale problem for \( A^d \) is well posed. Theorem 1 applies for \( A^d \) and we need only demonstrate that \( \mu_1 \) is a stationary measure for \( X(t) \). For stationarity of \( \mu_1 \), this requires (from Eq. (4))

\[
\sum_{i=1}^{3} \left[ \int_{SO(3)} \left( Y_i^2 f(s) - Y_i H_1(s) Y_i f(s) \right) \frac{e^{-H_1(s)}}{\mathcal{F}_1} \gamma_1(ds) \right] = 0, \quad f \in C^\infty(SO(3)).
\]

(13)

Now, use the following results (see Appendix B for proof):

1. If \( Z \) is a smooth tangent vector field on \( SO(3) \), then for any \( f, g \in C^\infty(SO(3)) \)

\[
\int_{SO(3)} Zf(s)g(s) \gamma_1(ds) = - \int_{SO(3)} f(s)Zg(s) \gamma_1(ds),
\]

where \( Zf, Zg \) are treated as elements of \( C^\infty(SO(3)) \).

2. If \( Z \) is a tangent vector field on \( SO(3) \) and if there is a function \( f \in C^\infty(SO(3)) \), then

\[
Ze^{-f} = -(Zf)e^{-f}.
\]

Let \( Z = Y_i \) (as defined in Appendix A) then, for \( i = 1, 2, 3 \),

\[
\int_{SO(3)} Y_i^2 f(s) \frac{e^{-H_1(s)}}{\mathcal{F}_1} \gamma_1(ds) \overset{(a)}{=} - \int_{SO(3)} Y_i f(s) Y_i \frac{e^{-H_1(s)}}{\mathcal{F}_1} \gamma_1(ds)
\]

\[
\overset{(b)}{=} \int_{SO(3)} Y_i f(s) Y_i H_1(s) \frac{e^{-H_1(s)}}{\mathcal{F}_1} \gamma_1(ds)
\]

with (a) following Result 1 and (b) from Result 2 proving Eq. (13). Hence, \( \mu_1 \) is a stationary measure of \( X(t) \) generated as a solution of Eq. (10).

The next result relates to the uniqueness of \( \mu_1 \).

**Theorem 3.** The posterior \( \mu_1(ds) = (e^{-H(s)}/\mathcal{F}_1) \gamma_1(ds) \) is the unique stationary probability measure of \( X(t) \).

**Proof.** To establish uniqueness of a stationary measure and to derive ergodic statements, one has to establish Harris recurrence (Athreya and Ney, 1972; Revuz, 1975) (Proposition 2.10, p. 76). As observed in Doss et al. (1996), Harris recurrence is difficult to establish directly. Instead, for a given stationary measure, irreducibility implies the uniqueness of the stationary measure as shown in Doss et al. (1996) and Revuz (1975, Exercise 2.20(1), p. 81). Using results from Meyn and Tweedie (1993a,b), \( \mu_1 \)-irreducibility and Harris recurrence lead to the ergodic result. Therefore, we only need to show that \( X(t) \) is \( \mu_1 \)-irreducible. \( X(t) \in SO(3) \) with transition probability \( P_t(s, \cdot) \) is \( \mu_1 \)-irreducible if for all measurable \( F \subset SO(3) \), \( \mu_1(F) > 0 \) implies \( P_t(s, F) > 0 \), for some \( t > 0 \) and for all \( s \in SO(3) \).
Lemma 1. The diffusion process $X(t) \in SO(3)$ is $\mu_1$-irreducible.

Proof. Consider the diffusion operator $A^d$ (Eq. (12)) generating diffusions on $SO(3)$. Since the functions $a_i$ and $b_{ij}$ are smooth, and $SO(3)$ is compact, using results from Karatzas and Shreve (1987), the fundamental solution of an elliptical partial differential equation (the Kolmogorov forward equation) exists. This solution, $P_t(s_1, s_2)$, the transition probability density of the diffusion process $X(t)$, is positive for all $t > 0$ and is continuous in all its arguments. Hence, for $\mu_1(F) > 0$, $\int_F P_t(s_1, s_2) \gamma_1(ds_2) > 0$ for some $t > 0$. □

Therefore, $\mu_1$ is the unique stationary probability measure of the diffusion process $X(t)$ and the ergodicity holds.

4. Jump–diffusion process on $\mathcal{X} = \bigcup_{k=0}^\infty \mathcal{X}_k$

Next we extend the sampling space to $\mathcal{X}$, to allow for simultaneous estimation of the number of targets, along with their orientations. In Eq. (8), we have already defined a posterior probability measure $\mu$ on $\mathcal{X}$ and we want to generate estimates under $\mu$ empirically. To enable this Markov process to move from one subspace to another, we add a jump component to the diffusion process constructed earlier in Section 3. Before we analyze the generator of a composite jump–diffusion process, we introduce an additional notation.

Let $B(\mathcal{X})$ denote the set of bounded, measurable functions on $\mathcal{X}$. Let $\hat{C}(\mathcal{X})$ be the set of continuous functions $f = \sum_{k=1}^\infty f_k 1_{\mathcal{X}_k}$ where $f_k \in \hat{C}(\mathcal{X}_k)$ and $\|f_k\|_\infty \to 0$ as $k \to \infty$. $\| \cdot \|$ denotes the supremum norm. Let $e_k$ be the identity element in $\mathcal{X}_k$ for each $k$ (it is the concatenation of $k$, $3 \times 3$ identity matrices). We define a metric $d$ on $\mathcal{X}$, which makes it complete and separable (analogous to Problem 28, p. 266 of Ethier and Kurtz, 1986) in the following way. For $s_1 \in \mathcal{X}_{k_1}, s_2 \in \mathcal{X}_{k_2}$, then

$$d(s_1, s_2) = \begin{cases} d_k(s_1, s_2) & \text{for } s_1, s_2 \in \mathcal{X}_k \ (k_1 = k_2 = k), \\ d_{k_1}(s_1, e_{k_1}) + d_{k_2}(s_2, e_{k_2}) + 1 & \text{for } s_1 \in \mathcal{X}_{k_1}, s_2 \in \mathcal{X}_{k_2} \ (k_1 \neq k_2), \end{cases}$$

where $d_k$ is the Frobenious metric in $\mathcal{X}_k$. Also, for $s \in \mathcal{X}$, let $k(s)$ denote the model order, i.e. $s \in \mathcal{X}_{k(s)}$ and $l(s) = d_k(s, e_k)$ be the distance of $s$ from the identity element in its subspace, also referred to as the size of $s \in \mathcal{X}_k$. For example, a $3 \times 3$ rotation matrix has model-order 1 and a $3 \times 3$ identity matrix has size 0.

Now, to the diffusion process add a jump operator in such a way that the resulting jump–diffusion process $X(t)$: (i) jumps across subspaces $\mathcal{X}_k$ at random exponentially separated times (similar to a Poisson counting process) according to a predefined transition probability, and (ii) follows the SDE (of dimensions appropriate to the current subspace) generating diffusions, in between the jumps. $X(t)$ is then a Markov process whose generator is a superposition of the two generators, one for jump and one
for diffusion. Let $Q(s, \cdot)$ be a transition probability measure and let $q(s) \in B(\mathcal{X})$ be a nonnegative function. Then, for any $s_1, s_2 \in \mathcal{X}$,

$$A^i f(s_1) = q(s_1) \int (f(s_2) - f(s_1))Q(s_1, ds_2)$$

(14)

defines a bounded, linear operator $A^i$ on $B(\mathcal{X})$, and $A^i$ is the generator for a Markov jump process (Ethier and Kurtz, 1986).

For appropriate values of $Q$ and $q$, $\mu$ becomes the unique stationary probability measure associated with the process $X(t)$. As described later, appropriately implies that $Q$ and $q$ are such that $A^i$ satisfies the detailed balance equation and the jump moves are restricted to neighbouring subspaces, in a particular fashion. First, we will prove the stationarity and then establish its uniqueness. In order to apply Theorem 1, we have to verify its condition that $A = A^d + A^i$, the generator of $X(t)$, generates a strongly continuous contraction semigroup on a closed subspace $L$, $L$ is separating and the martingale problem for $A$ is well posed. We do so by using the following theorem built from the first half of problem 28, p. 266 of Ethier and Kurtz (1986).

**Theorem 4.** Suppose the closure of $A^d (\subset \mathcal{C}(\mathcal{X}_k) \times \mathcal{C}(\mathcal{X}_k))$, generates a strongly continuous, contraction semigroup on $L_k \equiv \mathcal{C}(\mathcal{X}_k)$, that $L_k$ is separating, and that for each probability measure $\nu$ on the initial condition $X(0)$, there exists a solution of the martingale problem. Suppose the jump component has a generator of the form as in Eq. (14), $s \rightarrow Q(s, \cdot)$ from $\mathcal{X}$ to $\mathcal{P}(\mathcal{X})$ is continuous, and $q(s)$ is nonnegative with $q(s) \in \mathcal{C}(\mathcal{X})$ such that $A^i : \mathcal{C}(\mathcal{X}) \rightarrow \mathcal{C}(\mathcal{X})$.

1. Then, the closure of $A^d \subset \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$

$$A^d = \left\{ \left( \sum_{k=0}^{m} 1_{\mathcal{X}_k} f_k, \sum_{k=0}^{m} 1_{\mathcal{X}_k} A_k^d f_k \right), f_k \in C^\infty(\mathcal{X}_k), \forall m \geq 0 \right\},$$

(15)

generates a strongly continuous, contraction semigroup on $L \equiv D(A^d)$ which is single valued, and there exists a solution of the martingale problem for $A^d$.

2. Also, the closure of $A = A^d + A^i$ generates a strongly continuous, contraction semigroup on $L \equiv D(A^d)$, and there exists a unique solution of the martingale problem for $A^d + A^i$, for any initial probability measure $\nu \in \mathcal{P}(\mathcal{X})$.

**Proof.** The proof comes from a collection of technical results taken from Ethier and Kurtz (1986). For part (1) see first half of problem 28, p. 266 of Ethier and Kurtz (1986), with the single-valuedness following from its dissipativity and Lemma 5.2, p. 21 of Ethier and Kurtz (1986). For part (2), from the same Lemma 5.2 of p. 21, Ethier and Kurtz (1986), $A^i$ is single valued. Also, for $f \in \mathcal{C}(\mathcal{X})$ such that $\sup_{x \in \mathcal{X}} f(x) = f(x_0)$ for some $x_0 \in \mathcal{X}$, and $f(y) - f(x_0) \leq 0$ for all $y \in \mathcal{X}$. Therefore, $A^i f(x_0) \leq 0$ and $A^i$ satisfies the positive maximum principle on $B(\mathcal{X})$, and furthermore, Lemma 2.1, p. 165 (Ethier and Kurtz, 1986) implies that $A^i$ is dissipative. Since $\|A^i f\| \leq \|f\|$, and $D(A^d) \subset D(A^i)$, the conditions of Theorem 7.1 (Ethier and Kurtz, p. 37) are satisfied implying that the closure of $A = A^d + A^i$ generates a strongly continuous,
contraction semigroup on $L$. That there exists a solution to the martingale problem for $A = A^l + A^d$ and $v$ follows from Theorem 10.2, p. 256 of Ethier and Kurtz (1986), with the uniqueness proven from Theorem 4.1 (p. 182) using the conditions verified in Theorem 2.6 (p. 13). For a more detailed discussion refer to Srivastava (1996). 

For a jump–diffusion process $X(t)$ to satisfy the requirements of this theorem, we shall need two additional properties: (i) define a jump process to be local if for all $s_1 \in \mathcal{X}$ there exists a $\lambda$ such that $|l(s_1) - l(s_2)| \leq \lambda$ and $|k(s_1) - k(s_2)| \leq \lambda$ for all $s_2$ is the support of $Q(s_1, \cdot)$, and (ii) define the state space to be connected under the jump moves if for all $k_1$, $k_2$, there exists a finite sequence of jumps, which starts in $\mathcal{X}_{k_1}$ and ends up in $\mathcal{X}_{k_2}$. Recall that $k(s)$ is the model-order associated with $s \in \mathcal{X}$ and $l(s)$ is its size. These properties ensure that the process does not move very far from its current state in one jump and still can travel from one subspace to another in a finite number of jumps. With these definitions, we can make our construction of $X(t)$ more precise.

**Theorem 5.** If the jump–diffusion process $X(t)$ has the properties that the

1. the jump component is local, has a generator given by Eq. (14) with the parameters: $q(\cdot) \in \hat{C}(\mathcal{X})$ nonnegative, and $Q(\cdot, \cdot)$ the transition probability, satisfying

$$q(s_1)\mu(ds_1) = \int_{\mathcal{X}} q(s_2)Q(s_2, ds_1)\mu(ds_2),$$

(16)

2. diffusion $X(t)$ within any of the subspaces $\mathcal{X}_k$ has $k$-components, each following Eq. (10), then, $\mu$ is a stationary measure of the jump–diffusion process $X(t)$.

**Proof.** With locality and from the results of Theorem 4, the generator $A = A^l + A^d$ satisfies the conditions for Theorem 1, and hence, only the associated stationary measure condition (Eq. (4)), $\int_{\mathcal{X}} (A^d + A^l)f(s)\mu(ds) = 0$, for $f \in D$, must be verified. For the diffusion component, we need to show that $\int_{\mathcal{X}} A^d f(s)\mu(ds) = 0$. This is the same as proving

$$\sum_{k=0}^{m} \int_{\mathcal{X}_k} A_k^d f_k e^{-H(s)} \gamma_k(ds) = 0,$$

since $f$ is of the form $f = \sum_{k=1}^{m} f_k 1_{\mathcal{X}_k}$, for $f_k \in C^\infty(\mathcal{X}_k)$ and $m \geq 0$, and, by definition, $A^d f = \sum_{k=0}^{m} 1_{\mathcal{X}_k} A_k^d f_k$. By item 2 of this theorem the diffusion within each subspace follows Eq. (10) implying, as shown in Theorem 2, that each integral (for a fixed $k$) is zero. For the jump operator, substituting for $A^l$ in Eq. (4) gives

$$\int_{\mathcal{X}} q(s_1) f(s_1)\mu(ds_1) = \int_{\mathcal{X}} q(s_1)\mu(ds_1) \int_{\mathcal{X}} Q(s_1, ds_2) f(s_2),$$

$$\Rightarrow \int_{\mathcal{X}} f(s_1) (q(s_1)\mu(ds_1)) = \int_{\mathcal{X}} f(s_1) \left( \int_{\mathcal{X}} q(s_2)Q(s_2, ds_1)\mu(ds_2) \right),$$

where interchanging variables of integration gives the last equality. This is satisfied since the two measures on both sides are equal as provided in the detailed balance condition Eq. (16).
Hence, the generator of the Markov jump–diffusion process \( X(t), A = A^d + A^l \) satisfies the condition for \( \mu \) to be a stationary measure of \( X(t) \). \( \square \)

Next, we prove that \( \mu \) is the unique stationary probability measure of \( X(t) \).

**Lemma 2.** If the space \( \mathcal{X} \) is connected under the jump moves, then

1. \( \mu \) is the unique, stationary probability measure of the Markov jump–diffusion process \( X(t) \),
2. For any initial condition \( \nu \), the time-varying measure \( \nu_t(ds_2) = \int P_t(s_1, ds_2)\nu(ds_1) \) converges in variational norm to \( \mu \) as \( T \to \infty \), and for any bounded measurable function \( f \)
\[
\frac{1}{T} \int_0^T f(X(t)) \, dt \to \int_{\mathcal{X}} f(s)\mu(ds), \quad \text{a.s. } \nu(ds).
\] (17)

**Proof.**

1. As earlier, \( \mu \)-irreducibility results in uniqueness of the stationary probability measure via Harris recurrence. So we only have to prove that \( X(t) \) is \( \mu \) irreducible. That is, for any initial condition the probability of \( X(t) \) reaching a set of positive \( \mu \) measure is nonzero for all \( t > 0 \). In the previous section, we have already shown that if the starting point and the target set are both in the same subspace then this probability is nonzero. For them to be in different subspaces, say \( \mathcal{X}_{k_1} \) and \( \mathcal{X}_{k_2} \), the process can reach from one to another in finite jumps (since \( \mathcal{X} \) is assumed connected). The time \( X(t) \) stays in any subspace is random and is a finite sum of random number of exponential random variables. Therefore, for any \( t > 0 \) the probability that the process \( X(t) \) reaches from \( \mathcal{X}_{k_1} \) to \( \mathcal{X}_{k_2} \) in time \( t \) is nonzero. Putting this together with the previous result, \( P_t(s_1, F) > 0 \) for some \( t > 0 \) and \( \mu(F) > 0 \).
2. Given Harris recurrence and \( \mu \)-irreducibility of \( X(t) \), the variational norm convergence follows directly from Meyn and Tweedie (1993a,b, Theorem 6.1) and the ergodic statement follows from the ergodic theorem in Azema et al. (1967), as stated in Amit and Miller (1992). A similar result is obtained in Theorem 8.1 in Meyn and Tweedie (1993a,b) using results from Azema et al. (1967, p. 169). \( \square \)

### 5. Jump–diffusion algorithm

In this section, we state a jump–diffusion algorithm and utilize results from the previous sections to verify the convergence of this algorithm.

In the scenario of sampling from \( \mu \) over \( \mathcal{X} \), there are two kinds of jump moves performed. Let the process \( X(t) \) be at point \( s \in \mathcal{X} \) and let \( k \) be the model order of \( s \), i.e. \( s \in \mathcal{X}_k \). Then, the first type of jump allowed corresponds to removing (death) the last component from \( s \), and the second type corresponds to adding (birth) a component to \( s \). To make this precise, we introduce this notation: for \( s \in \mathcal{X}_k \), \( s' \in \mathcal{X}_1 \) let \([s, s'] \in \mathcal{X}_{k+1}\)
be the vector concatenation of $s$ and $s'$, and let $\tilde{s}$ be the projection of $s$ onto $X_{k-1}$ by removing the $k$th component. Also, let $s^k \in X_1$ be the last component of $s$ such that $s = [\tilde{s}, s^k]$. Using this notation, the transition probabilities for these jump moves can be specified in the following way. Define the transition measure by, for $s_1 \in X_k$,

$$q(s_1, ds_2) = \begin{cases} 
q_b(s_1, s_2) \delta_{s_1}(\tilde{s}_2) \gamma_{k+1}(ds_2) & \text{for } s_2 \in X_{k+1}, \\
q_d(s_1, s_2) \delta_{s_1}(\tilde{s}_2) \gamma_{k-1}(ds_2) & \text{for } s_2 \in X_{k-1}, \\
0 & \text{otherwise}, 
\end{cases}$$

(18)

where $\delta_{s_1}(s_2) = 1$ if $s_1 = s_2$, 0 otherwise. This measure is then normalized to provide the transition probability according to,

$$Q(s_1, ds_2) = \frac{q(s_1, ds_2)}{\int_{X} q(s_1, ds_2)}. \quad (19)$$

The total jump intensity of a move is given by

$$q(s_1) = \int_{X_1} q_b(s_1, [s_1, s']) \gamma_1(ds') + q_d(s_1, \tilde{s}_1), \quad s_1 \in X_k. \quad (20)$$

Using delta functions, the jump measures have been made singular with respect to the base measure by keeping most of $s_1$ fixed, allowing only incremental changes. In other words, at a time only one jump move, either the birth or the death, is allowed. Furthermore, the death move is restricted to removal of only the last component and birth move is restricted to adding only one component, keeping all the previous components unchanged. It must be noted that this restriction can be relaxed to allow for more complicated jump moves, which can possibly improve the speed of convergence. The weights $q_b, q_d$ (also called the birth and death intensities, respectively), can be derived from the posterior measure $\mu$ in multiple ways, as described in Grenander and Miller (1994); here we choose a metropolis-based technique (Metropolis et al., 1953). For convenience, we split the posterior energy $H_k(s)$ into two components: $P_k(s)$ the prior energy and $E_k(s)$, the likelihood energy ($E_k(s) \propto -\log(L_k(s))$ and $P_k(s) \propto -\log(\pi_{0,k}(s))$). Often, there is a convenient form available for the prior in form a conditional prior such that $P_k(s) = P_{k-1}(\tilde{s}) + P_1(s'|\tilde{s})$. Now, define the intensities such that for $s_1 \in X_k$, $s' \in X_1$,

$$q_b(s_1, [s_1, s']) = \frac{1}{\mathcal{W}_1 \gamma_1} e^{-[E_{k+1}([s_1, s']) - E_k(s_1)]}, \quad e^{-P_1(s'|s_1)},$$

$$q_d(s_1, \tilde{s}_1) = \frac{1}{\mathcal{W}_1 \gamma_1} e^{-[E_{k-1}(\tilde{s}_1) - E_k(s_1)]}. \quad (21)$$

In this construction we have utilized a conditional probability on the $k$th component, given last $(k-1)$ components, denoted by $\exp\{ -P_1(s'|s_1) \}/\mathcal{W}_1$. In case no such prior exists then it can be replaced by the Haar measure $\gamma_1$ on $X_1 = \text{SO}(3)$. $\mathcal{W}_1$ is the normalizer for the conditional prior on $X_1$. 
The jump–diffusion process, \( \{X(t), t \geq 0\} \), jumping at random times \( t_1, t_2, \ldots \), is constructed according to the following algorithm:

**Algorithm 1.** Let \( i = 0, \ t_0 = 0, \ X(0) = X_0 \in \mathcal{X} \) be any initial condition.

1. Generate a sample \( u \) of an exponential random variable with mean 1, a constant.
2. Follow the stochastic differential equation generating diffusion in the subspace determined by \( X(t_i) \in \mathcal{X}_k \) for time interval \( t \in [t_i, t_{i+1}) \), \( t_{i+1} = t_i + u \):

\[
dX(t) = \sum_{i=1}^{3k} (Y_i H)Y_{i,X(t)} \, dt + \sum_{i=1}^{3k} Y_{i,X(t)} \circ dW_i(t). \tag{22}
\]

There are \( k \)-rotations in this diffusion, each flows on the corresponding component of \( \mathcal{X}_k \).
3. At \( t = t_{i+1} \), perform a jump move, from \( X(t_{i+1}) \) to an element of \( \text{supp}\{Q(X(t_{i+1}), \cdot)\} \), according to the transition probability measure \( Q(s_1, ds_2) = q(s_1, ds_2)/q(s_1) \) given by Eqs. (18), (20) and (21).
4. Set \( i \leftarrow i + 1 \), go to step 1.

This resulting process \( X(t) \in \mathcal{X} \) is characterized by the following corollary.

**Corollary 1.** This algorithm results in a Markov jump–diffusion process \( X(t) \) whose unique stationary probability measure is given by the posterior \( \mu \). Furthermore, it satisfies Eq. (17) in Lemma 2.

**Proof.** To prove this corollary, we have to show that \( X(t) \) satisfies the conditions stated in the Theorems 2, 5 and Lemma 2:

1. Since \( H \) is smooth, and by the choice of tangent vector fields in Eq. (22), the conditions in Theorem 2 are satisfied.

2. Next we show that \( q \in \hat{C} (\mathcal{X}) \). Assuming that both the likelihood energy and the prior energy are smooth functions, \( q \) is a continuous function. Additionally, it is a bounded function since, from Eqs. (20), (21) and (18),

\[
q(s_1) = \frac{1}{\mathcal{W}} \int_{\mathcal{X}} e^{-[E_{k+1}(s_1, s') \circ E_k(s_1)]} e^{-P(s' | s_1)} \gamma_1(ds') + e^{-[E_{k+1}(s_1) \circ E_k(s_1)]}
\]

\[
\leq \left( 1 + \frac{\gamma_1(\text{SO}(3))}{\mathcal{W_1}} \right) < \infty.
\]

3. Locality: Only unit changes are allowed, i.e. addition/removal of only one orientation at a time implies that the support set of \( Q(s_1, \cdot) \) is given by \( \mathcal{X}_{k+1} \cup \mathcal{X}_{k-1} \), for \( s \in \mathcal{X}_k \). By the choice of jump measures, for any \( s_1 \in \mathcal{X} \), \( |n(s_1) - n(s_2)| \leq 1 \) and \( |l(s_1) - l(s_2)| < 2 \sup_{s \in \mathcal{X}_1} l(s) \) for all \( s_2 \in \text{supp}(Q(s_1, \cdot)) \).
4. Detailed balance: Assume \( s \in \mathcal{X}_k \), and substituting for \( q(\cdot), Q(\cdot, \cdot) \) in the RHS of Eq. (16), we obtain

\[
\gamma_k(ds_1) \frac{1}{\mathcal{W}_1} \left( \int_{\mathcal{X}_1} e^{-[E_k(s_1) - E_{k+1}([s_1, s'])]} \frac{e^{-E_{k+1}([s_1, s']) - P([s_1, s'])}}{\mathcal{F}} \gamma_1(ds') + e^{-[E_k(s_1) - E_{k-1}(\tilde{s}_1)]} e^{-P(s_1|\tilde{s}_1)} e^{-E_{k-1}(\tilde{s}_1) - P(\tilde{s}_1)} \mathcal{F} \right).
\]

Since \( P([s_1, s']) = P(s_1) + P(s'_1|s_1) \) and \( P(s_1) = P(\tilde{s}_1) + P(s_1|\tilde{s}_1) \), this term is equal to \( q(s_1)\mu(ds_1) \), which is the left-hand side of Eq. (16).

5. Connectedness: Starting from a subspace \( \mathcal{X}_k \), any another subspace \( \mathcal{X}_{k'} \) can be reached in \(|k - k'|\) steps, hence the connectedness of \( \mathcal{X} \).

Based on these conditions, Theorems 2, 5 and Lemma 2, the posterior \( \mu \) is the unique stationary probability measure of the jump–diffusion process constructed in the algorithm and the ergodic result is verified.

**Computer Implementation:** This jump–diffusion process when implemented on a discrete computer raises additional issues. The jumps, scheduled for random exponentially separated times, are now performed at times that are the nearest integers in terms of the machine cycle. Similarly, the diffusions are approximated by the solutions of corresponding stochastic difference equations. Define \( \xi_i(t), -\infty < t < \infty, i = 0, 1, \ldots, 3k \) be the integral flows generated by each of the vector-fields \( Y_i \), as defined in Appendix A. Let \( w_1, w_2, \ldots, w_{3k} \) be independent standard Brownian motions. Define a composite flow,

\[
\Gamma(t_1, t_2)s = \xi_{3k}(w_{3k}(t_2) - w_{3k}(t_1)) \ldots \xi_1(w_1(t_2) - w_1(t_1)) \xi_0(t_2 - t_1)s.
\]

Choose \( \Delta > 0 \) and consider the discrete time Markov process

\[
X(n\Delta)s = \Gamma((n - 1)\Delta, n\Delta)X((n - 1)\Delta)s,
\]

where \( X(0)s = s \). In Amit (1991), it is shown that this approximation approaches the diffusion in Eq. (22) when \( \Delta \to 0 \), over finite time intervals. For \( \Delta \) comparable to the machine precision, this discrete time Markov chain will be too slow to converge to the posterior. On the other hand, it is shown by Roberts and Tweedie (1995) that larger step sizes may vary the sampling process to such an extent that the stationary probability associated with the discrete process is not the intended one. That is, the sequence \( \{X(n\Delta): n = 1, 2, \ldots\} \) may sample from a distribution other than the posterior. To handle this problem, Roberts and Tweedie (1995) have proposed a Metropolis-Adjusted-Langevin’s algorithm which adjusts the discrete process at every sampling step. As an illustration, we have utilized this correction procedure to sample from a given posterior on the group of unitary matrices in Srivastava (2000) for problems dealing with statistical array signal processing.

Another issue frequently discussed in the MCMC literature is the rate of convergence of these sampling algorithms. For a certain class of target probability distributions,
preliminary convergence results have been established (Roberts and Tweedie, 1995; Amit and Grenander, 1991; Roberts and Rosenthal, 1996). In this paper, we have not investigated the convergence rates issue, but the experiments presented here and in other papers (Srivastava, 1993; Miller et al., 1995; Lanterman et al. 1995; Srivastava, 2000) display reasonable convergence times.

6. Experimental results

In this section we illustrate a jump–diffusion algorithm using a simple example from target pose estimation on SO(3). Keeping $k = 1$ fixed, given a noisy image of a target, say an airplane, our task is to estimate its orientation in SO(3). Additionally, detailed algorithms on jump–diffusion sampling for estimating complete target motions in \( \bigcup_{k=0}^{\infty} \text{SE}(3)^k \) are listed in Srivastava et al. (1997), and sampling results are presented in Srivastava (1993), Miller et al. (1995) and Srivastava (1996). A similar problem in subspace estimation for array signal processing is analyzed using diffusion sampling techniques in Srivastava (2000), sampling results at different noise levels are presented there.

We illustrate a jump–diffusion algorithm from a posterior \( \mathcal{E} \) on a single orientation in SO(3). The jump–diffusion algorithm is used to search in space \( \mathcal{X} = \text{SO}(3) \) with the uniform (Haar) measure as prior and the likelihood function as given in Eq. (6). The jumps, in this case, result in transition from one point in SO(3) to another point in SO(3), since the model order is fixed here \( (k = 1) \), i.e. \( \text{supp}(Q(s, \cdot)) = \text{SO}(3) \) for all \( s \in \text{SO}(3) \), The diffusions follow the SDE (Eq. (10)) between jumps to generate stochastic gradient flows on SO(3). The algorithm seeks to generate a minimum mean squared error (MMSE) estimate of the target orientation given this image. To make that precise, recall that \( d_1(s_1, s_2) \) is the regular Euclidean distance between \( s_1 \) and \( s_2 \) considering them as \( 3 \times 3 \) matrices. Then, the MMSE estimate is given by

\[
\hat{s} = \arg \min_{s \in \text{SO}(3)} \int_{\text{SO}(3)} d_1(s, s_1)^2 \mu(ds_1|I^D).
\]

It is shown in Grenander et al. (1998), that the MMSE solution is given by

\[
\hat{s} = U \begin{bmatrix} 1 & \cdots & 0 \\ 0 & \cdots & 0 \\ 0 & \cdots & \det(UV^T) \end{bmatrix} V^T,
\]

where \( A_v = U \Sigma V^T \) is the singular-value decomposition of the average matrix

\[
A_v = \int_{\text{SO}(3)} s \mu(ds|I^D).
\]

This average matrix is computed via jump–diffusion random sampling from the posterior \( \mu \) over \( \text{SO}(3) \). The results from computing the MMSE estimate \( \hat{s} \) are shown in Fig. 3. To understand the values plotted, let \( \{X_i; i = 1,2,\ldots\} \) be the sample generated
using the jump–diffusion sampling algorithm. Shown in Fig. 3 are two distance plots: the thin line plots the distance between $X_i$ and the true target orientation as function of $i$ ($d_1(X_i,s_i)$); the thicker line plots the distance $\hat{s}_i$ (computed using the first $i$ samples) as the function of $i$ ($d_1(\hat{s}_i,s_i)$). From the plot, it can be seen that jumps are made at $i=1,13,39,59$ reflecting the random times when other orientation $X_{i+1}$ is better matched to the data image than the present state $X_i$. Shown in the upper panels of Fig. 4 is the target rendered at the orientations $X_i \in \text{SO}(3)$ for $i=1,41,81,121,401$. The lower panels represent the difference images generated by removing the contribution from target at the current sample $X(j\Delta)$, from the data image while the middle panels display the evolution of MMSE estimates $\hat{s}_i$ for the same times. The noise level used here represents a high signal-to-noise ratio, the posterior energy is peaked around the true orientation, and hence, the sampling process converges quickly to that point and stays there.

7. Conclusion

We have presented a random sampling based on constructing a Markov jump–diffusion process for sampling from a given posterior $\mu$ defined on $\mathcal{X}$, which is a countable union of subgroups $\mathcal{X}_k = \text{SO}(3)^k$. Langevin’s stochastic differential equation,
modified to account for the curved geometry of SO(3), results in the diffusion component. A jump component, designed to move from one subspace to another, is added in such a way that the jump–diffusion process has the given posterior as its unique stationary probability measure. This paper extends the pattern theoretic techniques presented first in Grenander and Miller (1994) to the representations taking values on Matrix Lie groups. An algorithm for constructing jump–diffusion process is presented. Experimental results are presented for estimating the orientation of an airplane from its noisy image.

**Acknowledgements**

We thank Prof. Yali Amit, University of Chicago, for providing us with the guidelines for a general ergodic analysis of jump–diffusion processes on Lie groups through the monograph (Amit and Miller, 1992). This work was supported in part by ARO DAAG55-98-1-0102, NSF-9871196, ARO CIS DAA-H04-95-1-0494 and DAAD19-99-1-0267.

**Appendix A. Geometry of SO(3)**

SO(3) is a three-dimensional, compact, Lie group with matrix multiplication being the group operation. To study its geometry in detail refer Boothby (1986), Warner
(1994), or Spivak (1979). In order to study the variational calculus on Lie groups, or manifolds in general, one has to establish two important components: the tangent structure consisting of the derivatives, and the integration using the volume element or the differential forms. Starting with the tangent structure, the space of all vectors tangent to SO(3) at identity, denoted by $T_e(SO(3))$, is known to be the set of all $3 \times 3$ skew-symmetric matrices. Let $Y_1,e, Y_2,e, Y_3,e$ be the canonical orthonormal basis elements of $T_e(SO(3))$. They can be extended to all points on SO(3) using the left translation in the following way. Define the left translation on SO(3), parameterized by $g$, according to $g \cdot h$, the $3 \times 3$ matrix product, for all $h \in SO(3)$. This mapping results in the left translation on the space of tangent vectors in such a way that for each $Y_i,e \in T_e(SO(3))$ we get a vector field over all of SO(3). By a simple chain rule for derivatives, the tangent vector at $g \in SO(3)$ can be shown to be $Y_i,g = g \cdot Y_i,e$ establishing a 1-to-1 equivalence between the elements of $T_e(SO(3))$ and the vector fields on SO(3). Define the inner product on the tangent spaces by

$$\langle Y_1,Y_2 \rangle = \text{tr}(Y_1^T Y_2)$$

for any $Y_1,Y_2 \in T_g(SO(3)) \quad \forall g \in SO(3)$,

where tr() is the matrix trace. By example (8.6) p. 352 (Boothby, 1986) this inner product determines a bi-invariant Riemannian metric on SO(3). Under this metric the tangent vectors $Y_{1,g}, Y_{2,g}, Y_{3,g}$ form a frame field of orthonormal vectors, i.e. $\langle Y_i,g, Y_j,g \rangle = \delta_{ij}$. A mapping $\xi: \mathbb{R}_+ \to SO(3)$ is said to be the integral curve of a vector field if $d\xi/dt$ is equal to the vector field at each point on the curve. For example, for the vector field $Y_i, i=1,2,3$, the curve

$$\xi_i(t) = \exp(-Y_i t),$$

is an integral curve. If the starting point is $s \in SO(3)$, then $\xi_i(t)s = s \exp(-Y_i t)$ is the integral curve.

To tackle the second part, i.e. the definition of volume element or integration on SO(3), we utilize the Haar measure on SO(3). The Haar measure can be expressed in different forms depending upon the choice of local coordinates. Since we are interested in estimating target position and orientations, our choice is expressing orientations through exponential coordinates. The exponential coordinates correspond to the angular velocities associated with the rotating object, and have physical interpretation attached to them. It has been shown by Jensen (1995) that the Haar measure on SO(3) when expressed in the exponential coordinates $(q_1,q_2,q_3)$ takes the form

$$\frac{\sin^2(\rho)}{\rho^2} \, dq_1 \wedge dq_2 \wedge dq_3 \quad \text{where} \quad \rho = \sqrt{q_1^2 + q_2^2 + q_3^2}.$$

Appendix B. Proofs of results 1 and 2

(i): Let $\xi_t, t \in [-\varepsilon, \varepsilon]$ be the integral curve associated with the vector-field $Y$ on SO(3), i.e. $Zf(s) = d\xi(s)/dt |_{t=0}$. For small enough $t$, $\xi_t^{-1} = \xi_{-t}$ and $\xi_t: SO(3) \to$
SO(3) is a diffeomorphism, and hence,
\[
\int_{\mathbb{R}} Zf(s)g(s)\gamma_1(ds) = \int_{SO(3)} \frac{df(\xi_t(s))}{dt} g(s)\gamma_1(ds)
\]
\[
= \frac{d}{dt} \int_{\xi_{-t}(SO(3))} f(\xi_t(\xi_{-t}(s)))g(\xi_{-t}(s))\xi_{-t}^*(\gamma_1(ds))
\]
\[
= \int_{SO(3)} f(s)\frac{d}{dt} g(\xi_{-t}(s))\gamma_1(ds)
\]
\[
= -\int_{SO(3)} f(s)Zg(s)\gamma_1(ds)
\]
the second equality follows Theorem 2.2(iv) (Boothby, 1986) and the facts, \(\xi_{-t}(SO(3)) = SO(3), \xi_{-t}^*(\gamma_1(ds)) = \gamma_1(ds)\) (since \(\gamma_1\) is the Haar measure on SO(3)).

(ii) Let \(Z\) be a vector field on SO(3), in local coordinates \(Z = \sum_{j=1}^9 \mathcal{X}_j\partial / \partial y_j\). For \(s \in SO(3)\), let \((U, \phi)\) be a coordinate neighborhood, for \(f \in C^\infty_s(SO(3)) (f\) is a smooth function defined in a neighborhood of \(s\)),
\[
Z(e^{-f(s)}) = \sum_{j=1}^9 \mathcal{X}_j(s) \partial e^{-f(\phi^{-1}(y))}/\partial y_j \bigg|_{y=\phi(s)}
\]
\[
= -\sum_{j=1}^m \mathcal{X}_j(s) \partial f(\phi^{-1}(y))/\partial y_j \bigg|_{y=\phi(s)} e^{-f(s)} = -Zf(s)e^{-f(s)}. \tag*{\Box}
\]

References