Jump-Diffusion Processes on Matrix Lie Groups for Bayesian Inference

Anuj Srivastava
Florida State University
Tallahassee, FL 32308
anuj@stat.fsu.edu

Michael I. Miller
Johns Hopkins University
Baltimore, MD 21218
mim@cis.jhu.edu

Ulf Grenander
Brown University
Providence, RI 02912
ulf@dam.brown.edu

Abstract

A variety of engineering problems can be studied as inferences on constrained sets, Lie groups in particular. Additionally, the number of parameters to be estimated, namely the model-order, may also be unknown a-priori. We present a Bayesian approach by building a posterior probability distribution on a countable unions of Lie groups and utilizing the jump-diffusion processes to generate optimal estimators empirically, under this posterior. This approach is presented in the context of two well-known problems: pose estimation in object recognition and subspace estimation in signal processing.

1. Introduction

We present a Bayesian formulation and an inference technique to solve problems involving: (i) model-order variability in addition to the unknown parameters, and (ii) the parametric variations taking place on curved Lie groups. The inclusion of latter extends the approach introduced in Grenander-Miller [8] which focuses only on Euclidean parameters. An important aspect of this work is that a variety of engineering problems, when mathematically formulated, result in solving optimization problems on Lie groups. These manifolds may not be vector spaces even though they have some geometrical structure which can be used to solve problems intrinsic to them.

This technique is applicable to many problems in signal processing and image understanding. As an example, subspace estimation plays an important role in current techniques for blind channel identification in multiuser wireless communication systems ([19, 4, 13, 1, 20], code-division multiple access [21, 11] and direction of arrival tracking [2, 4]. In our approach, subspaces are represented as elements of Grassman manifold which itself is studied through the group action of the unitary group $\mathbf{U}(n)$. Since the subspace dimension m is unknown a-priori, the estimation is performed over all possible subspaces, with all possible values of m. Similarly, in image understanding, for example in automated object discovery and recognition, estimation of objects' pose (orientation with respect to the camera) is essential. In our approach, the object pose is represented by a unique rotation matrix, an element of the special orthogonal group $\mathbf{SO}(3)$. Again, since the number of objects m is unknown a-priori, the estimation space is augmented to include m.

In general, we will represent the unknown parameters in these problems by the elements of Lie groups (differentiable manifolds having a group structure), denoted by S_m. In addition, we will allow the number of parameters m to be unknown too, solving the parameter estimation and model-order estimation problems simultaneously. The complete estimation space becomes $S = \bigcup_m S_m$. Taking a Bayesian approach, we will define a posterior probability measure μ on S and will seek classical estimators such as conditional means and variances, with respect to that posterior. In most practical problems, the posterior is too complicated to support analytical solutions. Instead, we will seek numerical solutions using Markov Chain Monte Carlo (MCMC) type random sampling techniques. More precisely, we will construct a Markov process, $X(t)$, having two components, jumps and diffusions, in such a way that they visit the elements of S according to their posterior probability μ. Thus, integrations with respect to the posterior can be approximated by taking averages along the sample paths of this Markov process. In other words, for a continuous function f,

$$\lim_{t\to\infty} \frac{1}{t}\int_{t_0}^{t} f(X(\tau))d\tau = \int_S f d\mu \,. \tag{1}$$

Appropriate choice of f gives rise to classical Bayesian estimators.

The Markov process $X(t)$ performs discrete jumps at random, exponentially-separated times to search over the model order, while the (sample path) continuous diffusion process searches over a parameter space with fixed dimensions. On vector spaces, such as $\mathbb{R}^n$, it is well-known ([6, 7]) that the diffusions can be realized as solutions of

Langevin's stochastic differential equation (SDE):

$$dX(t) = -\nabla H(X(t))dt + \sqrt{2}dW(t) \ . \qquad (2)$$

This SDE generates a Markov process which has $\frac{\exp(-H(x))}{Z}$ as its stationary distribution. However, this equation is not valid for Lie groups with curved geometry, such as orthogonal and unitary groups, since addition is not a valid operation there. Using explicit evaluation of the tangent vector fields and integral curves, we will modify this equation to realize diffusions as stochastic gradient flows on these groups.

2. Bayesian Formulation on Lie Groups

We will explain our approach to Bayesian inference on Lie groups, with unknown model-order, in the context of two well-known problems: (i) automated object recognition, and (ii) subspace estimation and tracking. At first, we derive the Bayesian framework for addressing these two problems.

2.1. Automated Object Recognition

Given arbitrary images of a scene, our task is to discover and recognize objects of interest in that scene. An important issue is to have tractable mathematical models for handling the variety associated with arbitrary scenes. The images can contain an unknown number of objects imaged at random positions and orientations with respect to the camera. The utilization of models, for representations of object shapes, supporting recognition invariant to orientations and locations, is crucial. An emerging paradigm for object shape representation is the **deformable template** theory. In this approach, the starting point is to select a standard template for each of the objects and then define a family of transformations to account for the variability associated with target occurrences. For example, a rigid object, occurring at a particular orientation, is modeled by applying a rotation to each point on its template. This rotation is represented by a 3×3 rotation matrix; the set of all such matrices is denoted $\mathbf{SO}(3)$, the special orthogonal group. Let m be the number of objects of interest present in the scene. If m is already known, then the posterior density is proportional to

$$\exp(-H_m(s)) \ , \quad s = [s_1 \ s_2 \ \ldots \ s_m] \in \mathbf{SO}(3)^m \ ,$$

where $H_m(\cdot)$ contains terms representing: prior on s and the likelihood of observing an image given the rotations s. For details on the choice of H_m, please refer to [12, 9]. In a more general situation when m is unknown, it has to be estimated along with the rotations. The inference is then extended to the space, $S = \bigcup_{m=0}^{\infty} \mathbf{SO}(3)^m$, and the posterior measure becomes

$$\mu(A) = \frac{\sum_{m=0}^{\infty} \int_{A \bigcap \mathbf{SO}(3)^m} p_m \exp(-H_m(s)) ds_m}{\sum_{m=0}^{\infty} \int_{\mathbf{SO}(3)^m} p_m \exp(-H_m(s)) ds_m} \ ,$$

for $A \subset S$ measurable and where p_m is a prior on m.

Given this posterior measure our goal is to derive classical estimators (least-square, conditional mean, conditional variance etc). Using a jump-diffusion Markov process, we will sample from the posterior μ over S: the diffusions will result in sampling process over $\mathbf{SO}(3)^m$ for fixed m and jumps change the value of m. Averages computed along the sample paths of $X(t)$ will converge asymptotically to the optimal pose estimators, stated as expectations under μ.

2.2. Subspace Estimation and Tracking

Subspace estimation and tracking is of interest in many signal processing applications. Given the received signal, our task is to estimate and track the principal signal subspace associated with the observations. We will use the following representation: (uniquely) represent an m-dimensional complex subspace of an n-dimensional observation space by a (rank m, $n \times n$ complex) projection matrix ($m \leq n$). Denote by $\mathbb{P}_m$ the set of all such matrices. $\mathbb{P}_m$ is a compact, differentiable manifold (called complex Grassman-manifold) of $m(n-m)$ complex-dimensions, with interesting geometric properties, as described in Kobayashi & Nomizu [10]. In order to estimate and track subspaces on $\mathbb{P}_m$, we utilize the fact that the space of unitary matrices forms a *transitive group action* on $\mathbb{P}_m$. The unitary group is defined by $\mathbf{U}(n) = \{U \in \mathbb{C}^{n \times n} : U^\dagger U = I\}$. $\mathbf{U}(n)$ is a $n^2/2$ complex-dimensional, compact, Lie group with matrix multiplication being the group operation (refer to Kobayashi & Nomizu [10], pg. 130). Let Q_m be a $n \times n$ diagonal matrix with the first m entries being 1 and the rest 0. Any element of $\mathbb{P}_m$ can be written as $U^\dagger Q_m U$ for some appropriate $U \in \mathbf{U}(n)$ (this U may not be unique). The significance of this mapping is that instead of estimating on $\mathbb{P}_m$, we shift the estimation to $\mathbf{U}(n)$ and exploit its Lie group structure. This is another illustration of the **deformable template** theory: study the variability on a manifold through the transitive action of some Lie group on it.

Again, we will take a Bayesian approach by deriving a posterior density on $\mathbf{U}(n)$ and will utilize empirical techniques to derive classical estimators such as conditional mean and variance. As earlier, m may not be known beforehand and we have to allow for all possible values of m according to a chosen prior p_m. In this situation, the posterior probability measure on $\mathbf{U}(n)$ becomes:

$$\mu(A) = \frac{\sum_{m=0}^{\infty} \int_{A \bigcap \mathbf{U}(n)} p_m \exp(-H_m(s)) ds}{\sum_{m=0}^{\infty} \int_{\mathbf{U}(n)} p_m \exp(-H_m(s)) ds} \ , \qquad (3)$$

for some $A \subset \mathbf{U}(n)$, measurable. γ is the translation-invariant reference measure (also called Haar measure, see Folland [5]) on $\mathbf{U}(n)$. In the context of uniform array signal processing, an expression for H_m is derived in [15]. Notice, the posterior is a convex-combination of densities each determined by a value of m. This case differs from the previous application in that the change in m does not change the dimensionality of parameter space but it only changes the posterior density.

3. Jump-Diffusion Processes

To keep the notation simple we will restrict to the first application, namely that of object pose estimation. Let S_m be a Lie group with dimensions specified by the index m. A simultaneous model-order selection and parameter estimation approach seeks stochastic inference over the space $S = \bigcup_{m=0}^{\infty} S_m$. Let μ be a posterior probability measure on S.

Our approach is empirical in that we construct a Markov process $X(t)$, called a jump-diffusion process, which visits the elements of S with frequencies proportional to their posterior. This process has two components: (i) a diffusion process realized as a solution to Langevin's type SDE defined on S_m for each m, and (ii) a jump component which moves the process from one value of m to another. The jumps are performed at random, exponentially separated times and, in between the jumps, the process diffuses according to a SDE of appropriate dimension. The choice of diffusion parameters (infinitesimal mean and variance) and the jump transition probabilities are such that the process $X(t)$ has the desired ergodic properties (Eqn. 1 is satisfied). This result requires establishing two conditions: first that the posterior μ is a stationary probability measure of the Markov process $X(t)$ on S, and second that it is unique. Given these two conditions the ergodic result follows [16]. For the jump-diffusion process constructed below, these conditions are verified in [16]. These processes have been used to derive optimal estimators and lower bounds on expected errors, in case of object recognition problems for various sensors [12, 16, 18, 9] and subspace estimation problem [15].

3.1. Diffusions: Stochastic Flows on Lie Groups

A stochastic gradient process, obtained as a solution of Langevin's SDE (Eqn. 2), has been well-utilized ([6, 7]) to sample from given probabilities on Euclidean spaces. Since our parameter spaces are not vector-spaces, we will modify Eqn. 2 to construct an appropriate stochastic gradient process using the intrinsic geometry of these groups.

Let $Y_i, i = 1, 2, \ldots$ be a set of orthonormal, smooth vector fields on S_m (please refer to [3, 10, 17] for geometry of orthogonal and unitary groups). For a positive function H defined on S_m, $Y_i H_m$ denotes the directional derivative of H_m in the direction of Y_i. The quantity $-\sum_i (Y_i H_m) Y_i$ is then called the gradient vector of H_m; it is the direction of maximal decrease in the value of H_m. Interpreting diffusions as stochastic gradient processes, we generate them as solutions of the following equation:

$$\begin{aligned} dX(t) &= -\sum_{i=1}^{\dim(S_m)} (Y_{i,X(t)} H_m) Y_{i,X(t)} dt \\ &\quad + \sqrt{2} \sum_{i=1}^{\dim(S_m)} Y_{i,X(t)} \circ dW_i(t) \, . \end{aligned} \quad (4)$$

where $W_i(t)$ are real-valued, independent standard Wiener processes and $\circ$ denotes the Stratonovich integral. This equation is the analog of Eqn. 2 on Lie group S_m. If the last term is left out, then $X(t)$ is simply a gradient process seeking a local-minimum of H_m on S_m.

3.2. Jumps: Model Order Estimation

Jump moves result in the process $X(t)$ changing its dimensions, i.e. the process goes from S_{m_1} to S_{m_2}. These jumps are performed at random, exponentially separated times. In order to keep the jumps simple, we allow to process to go from an element of S_m to only the elements of S_{m-1}, S_m, and S_{m+1}. The jumps are performed according to a transition probability $Q(s_1, ds_2)$ for $s_1 \in S_m$ and $s_2 \in (S_m \bigcup S_{m-1} \bigcup S_{m+1})$. The exact choice of Q depends on the posterior energy functions $\{H_m\}$ in such a way that certain properties are satisfied (to ensure that μ is the unique stationary probability measure of $X(t)$).

A brief algorithm to construct a jump-diffusion sampling process is as follows:

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

1. *Generate a sample u of an exponential random variable with mean λ, a constant.*
2. *Follow the stochastic differential equation generating diffusion in S_m, with m determined by $X(t_i)$, for the time interval $t \in [t_i, t_{i+1})$, $t_{i+1} = t_i + u$: according to Eqn. 4.*
3. *At $t = t_{i+1}$, perform a jump move, from $X(t_{i+1})$ to an element of supp$\{Q(X(t_{i+1}), \cdot)\}$, according to the transition probability measure $Q(x, dy)$.*
4. *Set $i \leftarrow i + 1$, go to step 1.*

Averages along the sample paths of $X(t)$ converge asymptotically to the expectations under the posterior.

3.3. Computational Issues

We have constructed a stochastic process $X(t)$ taking values on S and having desired ergodic properties; the computer implementation of the diffusion component involves discretizing Eqn. 4 and implementing the resulting stochastic difference equation. A fundamental issue at this stage is the choice of step size for discretization. It has been shown by Roberts et. al. [14] that incorrect step-size may lead to sampling from a distribution which is not μ. One solution, as proposed in [14], is to utilize Metropolis-adjusted Langevin's (MAL) algorithm. This algorithm utilizes the new state proposed by the stochastic difference equation but accepts it only with certain probability.

Results from implementation of a jump-diffusion algorithm, in the context of object recognition and subspaces estimation are presented in [17, 16, 15].

References

[1] K. Abed-Meraim, J.-F. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. Moulines. On subspace methods for blind identification of single-input multiple-output fir systems. *IEEE Trans. Signal Processing*, 45(1):42–55, January 97.

[2] K. V. S. Babu, Y. Yoganandam, and V. U. Reddy. Adaptive estimation of eigensubspace and tracking the directions of arrival. *Signal Processing*, 68(3):317–339, Aug. 1998.

[3] W. M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. Academic Press, Inc., 1986.

[4] J. P. Delmas and J. F. Cardoso. Performance analysis of an adaptive algorithm for tracking dominant subspaces. *IEEE transactions on signal processing*, 46(11):3045–3057, November 1998.

[5] G. B. Folland. *Real Analysis: Modern Techniques and their applications*. John-Wiley & Sons, 1984.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[7] S. Geman and C.-R. Hwang. Diffusions for global optimization. *SIAM J. Control and Optimization*, 24(24):1031–1043, 1987.

[8] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society*, 56(3), 1994.

[9] U. Grenander, M. I. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix lie groups for atr. *IEEE Transactions on PAMI*, 20(8):790–802, August, 1998.

[10] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry, vol 2*. Interscience Publishers, 1969.

[11] H. Liu and G. Xu. A subspace method for signature waveform estimation in synchronous cdma systems. *IEEE Trans. on Communications*, 44(10):1346–54, October 96.

[12] M. I. Miller, A. Srivastava, and U. Grenander. Conditional-expectation estimation via jump-diffusion processes in multiple target tracking/recognition. *IEEE Transactions on Signal Processing*, 43(11):2678–2690, November 1995.

[13] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel fir filters. *IEEE Trans. Signal Processing*, 43(2):516–525, February 95.

[14] G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to langevin's diffusions. *MCMC website*, June, 1995.

[15] A. Srivastava. A bayesian approach to geometric subspace estimation. *submitted to IEEE Transactions Signal Processing*, January. 1999.

[16] A. Srivastava. *Inferences on Transformation Groups Generating Patterns on Rigid Motions*. D. Sc. Thesis, Washington University, St. Louis, Missouri, July 1996.

[17] A. Srivastava, M. Miller, and U. Grenander. *Ergodic Algorithms on Special Euclidean Groups for ATR*. Systems and Control in the Twenty-First Century: Progress in Systems and Control, Volume 22. Birkhauser, 1997.

[18] A. Srivastava, M. I. Miller, and U. Grenander. Multiple target direction of arrival tracking. *IEEE Transactions on Signal Processing*, 43(5):1282–85, May 1995.

[19] L. Tong and S. Perreau. Multichannel blind estimation: From subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[20] A.-J. van der Veen, S. Talwar, and A. Paulraj. A subspace approach to blind space-time signal processing for wireless communication systems. *IEEE Trans. Signal Processing*, 45(1):173–190, January 97.

[21] X. Wang and H. V. Poor. Blind adaptive multiuser detection in multipath cdma channels based on susbspace tracking. *IEEE Transactions Signal Processing*, 46(11):3030–3044, Nov. 1998.