BULLETIN OF THE BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 39(1), 1-10

An information geometry algorithm for distribution control

Fengwei Zhong, Huafei Sun and Zhenning Zhang

Abstract. In this paper, we consider the problem of distribution control from the viewpoint of information geometry. Different from most existing models used in stochastic control, it is assumed that the control input directly affects the distribution of the system output in probability sense. Here, we set up a new manifold (S), meanwhile the B-spline manifold (B) and the system output manifold (M) can be referred to as its submanifolds. We give an information geometrical algorithm which can be called as geodesic-projection algorithm using the properties of manifold. In the geodesic step, we can obtain the geodesic equation from the initial point $V_0 = (\omega_{10}, \omega_{20}, \cdots, \omega_{(n-1)0})$ to the specified point $V_g = (\omega_{1g}, \omega_{2g}, \cdots, \omega_{(n-1)g})$ in B. This gives us an optimal trajectory for the points changing along in B. In the projection step, we project the sample points selected from the geodesic onto M. The coordinates of the projections in M give the trajectory of the control input u.

Keywords: geodesic, projection, distribution control, manifold.

Mathematical subject classification: 52B20, 62B10.

1 Introduction

Recently, in the control of stochastic system, scholars mainly concern with how to design control input in order to control the shape of the output probability density function of an unknown stochastic system, and make the output probability density function as close as possible to the target function. Several algorithms ([6], [7], [8], [9]) have been obtained and worked well for some stochastic systems. Among these algorithms, the B-spline function approximation algorithm ([10]) has been working particularly well to approximate the output probability density function of an unknown stochastic system. The B-spline function is formed by linear combination of all the pre-specified basis functions. When all

Received 26 August 2006.

the basis functions are fixed, the weights of the B-spline function can be regarded as directly related to the control input. Unfortunately all these algorithms have an extra constraint that the dynamics between the weights and the control inputs must be known. In the present paper, we give a new method of distribution control from information geometrical point of view to overcome this constraint.

Information geometry was proposed by Amari, until now it has been well applied into various fields, and it provides us new ways to solve some problems effectively. In this paper, we set the output probability density functions of an unknown stochastic system and the B-spline functions as two different manifolds, where control input (u) and weight vector (V) of B-spline functions play the role of coordinates in this two manifolds, respectively.

The purpose of distribution control is to adjust the control input u, such that the final system output probability density function $p(y; u_f)$ is as close as possible to the target function $g(y)$, and to study the trajectory of the parameter vector u turning from the initial distribution $p(y, u_0)$ to the finial distribution $p(y, u_f)$. We can get to our object under the help of B-spline function. We consider that $g(y)$ can be approximated by B-spline function $B(y; V)$, and at some point with the coordinate $V_g = (\omega_{1g}, \omega_{2g}, \cdots, \omega_{(n-1)g})$, they are infinitely close, that is,

$$g(y) \;=\; B(y; V_g) + e, \qquad y \in [a, b] \tag{1.1}$$

where e denotes the approximation error, and it is small enough under the choice of the pre-specified basis functions so that we can neglect it in the rest of this paper.

The key steps for designing the control input u are the following:

1. To formulate the relation between the weight vector (V) of the B-spline function and control input (u).

2. To choose the trajectory so that the points in B-spline manifold can change from the initial point V_0 to the specified point V_g smoothly and efficiently.

In the present paper, we use the projection between manifolds to give the relation between u and V, and use the geodesic to give an optimal trajectory from the initial point to the specified point in B-spline manifold. Then projecting the sample points selected from the geodesic on B onto M, we can get the trajectory of the control input u.

2 Preliminaries

Definition 2.1. *Suppose that $p(x;\theta)$ is a probability density function on a set X, where θ is the parameter of the probability density function. Let*

$$S = \{p(x;\theta) \mid \theta \in \Theta\} \tag{2.1}$$

be a set formed by probability density functions $p(x;\theta)$, then we call S a statistical manifold, where θ plays the role of coordinate system.

Definition 2.2. *Let us consider a family of probability distributions of a random variable x (which maybe a vector) whose probability density functions are specified by an n-dimensional parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$. When the probability density function can be written as*

$$p(x;\theta) = \exp\left\{\sum_{i=1}^{n} \theta_i r_i(x) + K(x) - \psi(\theta)\right\}, \tag{2.2}$$

where $r_i(x)$, $i = 1, 2, \ldots, n$, are functions of x, the family $S = \{p(x;\theta) \mid \theta \in \Theta\}$ is called an exponential family. $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ is natural coordinate system, and $\psi(\theta)$ is the potential function with respect to θ.

Definition 2.3. *The Fisher metric, α-connection and α-curvature tensor of a statistical manifold are defined by*

$$g_{ij} = \int \frac{\partial \log p(x;\theta)}{\partial \theta^i} \frac{\partial \log p(x;\theta)}{\partial \theta^j} p(x;\theta)dx, \tag{2.3}$$

$$\begin{aligned} \Gamma^{(\alpha)}_{ijk} &= E\left[(\partial_i \partial_j l(x;\theta))(\partial_k l(x;\theta))\right] \\ &\quad + \frac{1-\alpha}{2} E\left[(\partial_i l(x;\theta))(\partial_j l(x;\theta))(\partial_k l(x;\theta))\right], \end{aligned} \tag{2.4}$$

$$R^{(\alpha)}_{ijkl} = \left(\partial_j \Gamma^{(\alpha)s}_{ik} - \partial_i \Gamma^{(\alpha)s}_{jk}\right) g_{sl} + \left(\Gamma^{(\alpha)}_{jtl}\Gamma^{(\alpha)t}_{ik} - \Gamma^{(\alpha)}_{itl}\Gamma^{(\alpha)t}_{jk}\right), \tag{2.5}$$

where

$$l(x;\theta) = \log p(x;\theta), \quad \partial_i = \frac{\partial}{\partial \theta_i},$$

and E denotes the expectation with respect to $p(x;\theta)$.

When $\alpha = 0$, it is the Riemannian case.

Definition 2.4. *The geodesic equation $\theta(t)$ of a manifold is given by*

$$\ddot{\theta}^i + \Gamma^{(\alpha)i}_{jk}\dot{\theta}^j\dot{\theta}^k = 0, \tag{2.6}$$

where $\Gamma^{(\alpha)i}_{jk} = \Gamma^{(\alpha)}_{jkl}g^{li}$, (g^{li}) is the inverse of (g_{li}).

Next let us give a useful measure which is different from the general distance of two points in statistical manifold, Kullback divergence.

Definition 2.5. *Let $P = f(x)$ and $Q = g(x)$ be two points in a manifold where $x \in \mathcal{X}$, and the Kullback divergence of the two points is defined by*

$$D(P, Q) = \int_{\mathcal{X}} f(x) \log \frac{f(x)}{g(x)} dx, \tag{2.7}$$

and $D(P, Q) = 0$ if and only if $P = Q$.

Finally let us introduce an important theorem.

Theorem 2.1 (Projection Theorem, [2]). *Let M be a smooth submanifold in S. For a given point $P \in S$, let Q be the point that belongs to M and is closest to P in the sense of the Kullback divergence, that is,*

$$Q = \arg\min_{Q' \in M} D(P, Q'), \tag{2.8}$$

the point Q is given by the dual geodesic projection of P to M. Furthermore, the projection is unique when M is e-flat. (See Figure 1)

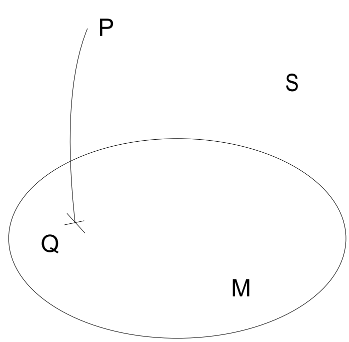

Figure 1: Projection Q from point P to submanifold M.

3 Model representation

Suppose that the unknown stochastic system has m inputs, so the control input u is an m-dimensional vector. And $y \in [a, b]$ denotes the output of the unknown stochastic system, so the output probability density functions of the unknown stochastic system is referred to as $p(y; u)$, where u plays a role of an m-dimensional vector-valued parameter. (See Figure 2)

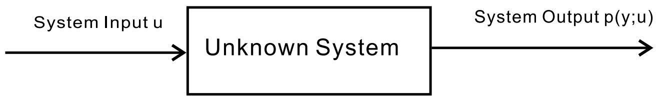

Figure 2: The considered unknown system.

We adjust u to control the shape of the output probability density functions $p(y; u)$ as close as possible to the target function $g(y)$. Suppose that the output probability density functions $p(y; u)$ is continuous and finite in $[a, b]$, then use the B-spline function formed as $\sum_{i=1}^{n} \omega_i^{'}(u) B_i(y)$ to approximate $p(y, u)$, where $B_i(y)$ are the pre-specified basis functions and $\omega_i^{'}(u)$ are the corresponding weight components.

Since $\int_a^b p(y; u) = 1$, the B-spline function has the constraint

$$\int_a^b \sum_{i=1}^{n} \omega_i^{'}(u) B_i(y) dy = 1 \,,$$

this means that only $n - 1$ weights of B-spline function are independent. To guarantee such a constraint, we define

$$\begin{aligned}
b_i &= \int_a^b B_i(y) dy \qquad (i = 1, 2, \ldots, n), \\
L(y) &= b_n^{-1} B_n(y), \\
C(y) &= \Bigg(B_1(y) - \frac{B_n(y)}{\int_a^b B_n(y)dy} \int_a^b B_1(y)dy, \\
&\qquad \vdots \\
&\qquad\qquad B_{n-1}(y) - \frac{B_n(y)}{\int_a^b B_n(y)dy} \int_a^b B_{n-1}(y)dy \Bigg),
\end{aligned}$$

so B-spline function can also be written as

$$B(y; V) = C(y) \cdot V + L(y), \tag{3.1}$$

and satisfies $\int_a^b B(y; V)dy = 1$, where $V = (\omega_1, \omega_2, \ldots, \omega_{n-1})$ is the weight vector corresponding to the re-constructed basis $C(y)$ which is an $(n-1)$-dimensional vector.

We use the B-spline function (3.1) to approximate the output probability density functions $p(y; u)$ of the unknown stochastic system, and e denotes the approximation error, that is,

$$p(y; u) = B(y; V) + e. \tag{3.2}$$

The purpose of adjusting the control input u is to make the final output probability density function $p(y; u_f)$ as close as possible to a specified distribution $g(y)$, where

$$g(y) = C(y) \cdot V_g + L(y), \tag{3.3}$$

and can be regarded as a point in B-spline manifold with coordinate $V_g = (\omega_{1g}, \omega_{2g}, \ldots, \omega_{(n-1)g})$.

Once such an idea is formulated, the next question arising from many practical situations is to see how the parameter of a distribution can be selected so that the actual distribution is made as close as possible to the target distribution, and study the trajectory of the parameter turning, this is a control problem, and has many applications in particulate processing.

4 Information geometrical algorithm for distribution control

In this section, we may redescribe the distribution control problem in the context of information geometry. It is assumed that the initial system output distribution is characterized by u_0 and the desired distribution is characterized by u_f which should be as close as possible to V_g in the sense of Kullback divergence, then focus will be made on the evaluation of how the weight vector V will behave in tuning the initial distribution $p(y, u_0)$ to the final distribution $p(y, u_f)$. Thus an effective trajectory of V should be chosen. Then the coordinate of the projections from V onto M will give the trajectory of control input u.

Next we will introduce three manifolds.

Definition 4.1. *Let*

$$B = \left\{ B(y; V) \;\middle|\; B(y; V) = C(y) \cdot V + L(y),\, V = (\omega_1, \omega_2, \ldots, \omega_{n-1}) \right\}$$

be B-spline manifold, and $V = (\omega_1, \omega_2, \ldots, \omega_{n-1})$ *plays the role of coordinate system.*

Definition 4.2. *Let*

$$M = \Big\{ p(y;u) \;\Big|\; p(y;u) \textit{ is the output probability density function of unknown stochastic system.} \Big\}$$

be system output manifold, and u plays the role of coordinate system. Here u is an m-dimensional vector-valued parameter, denoted by $u = (u_1, \ldots, u_m)$.

Definition 4.3. *Let S be the set of all smooth non-zero probability distributions of y, that is*

$$S = \big\{ r(y) \mid r(y) > 0 \big\}. \tag{4.1}$$

This is an infinite-dimensional manifold, and has dual flat structure (see [11]*).*

In this paper, we take the case where the system output probability density functions are subject to the exponential family distribution as an example, that is, M can be written as

$$M = \big\{ p(y;u) \big| p(y;u) = \exp\{u \cdot y - \psi(u)\} \big\}, \tag{4.2}$$

where u is the control input, and plays the role of coordinate system.

Any parametric model is just a finite-dimensional submanifold of S in the framework of [11], and we can get

Proposition 4.1. *B is a -1-flat submanifold of S, and M is a 1-flat submanifold of S.*

The manifolds we consider here have so good properties as described in Proposition 4.1 that given a point $P \in M \subset S$, the point $\hat{Q} \in B \subset S$ that minimizes $D(Q, P)$ is given by the e-projection of P onto B. The e-projection is given by the e-geodesic connecting $\hat{Q}$ and P which is orthogonal to B at $\hat{Q}$.

Dually, for a given point $Q \in B \subset S$, the point $\hat{P} \in M \subset S$ that minimizes $D(Q, P)$ is given by the m-projection of $Q \in B$ onto M. The m-projection is given by the m-geodesic connecting $\hat{P}$ and Q which is orthogonal to M at $\hat{P}$.

We know the initial point $P_0 \in M \subset S$ with the coordinate u_0 and the target point $Q_g \in B \subset S$ with the coordinate V_g. We want to get P_f in M which is an optimal estimation of $Q_g \in B$ in the sense of Kullback divergence, and a trajectory $\widehat{P_0 P_f}$ of the control input u turning from P_0 to P_f. We project P_0 onto B, obtain $Q_0 \in B$ with the coordinate $V_0 = (\omega_{10}, \ldots, \omega_{(n-1)0})$. In manifold, generally, geodesic is considered as a "straight" curve connecting the two points, so the geodesic connecting Q_0 and Q_g can provide an optimal trajectory in B-spline manifold. And this geodesic can be obtain by (2.6).

Then we select some sample points on $\widehat{Q_0 Q_g}$ in a fixed learning rate, and make m-projection onto M, also P_f is obtained by projecting Q_g onto M.

From the above consideration, we can formulate the geodesic-projection algorithm as follows. (See Figure 3)

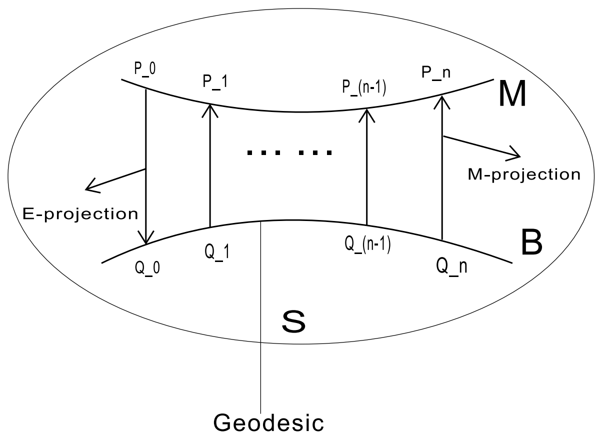

Figure 3: The geodesic-projection algorithm.

1. For a given initial point $P_0 \in M \subset S$ with the coordinate u_0, projecting P_0 onto B, we get $Q_0 = \Pi_B P_0 \in B$ that minimizes $D(Q, P_0)$, where $\Pi_B P_0$ denotes the e-geodesic-projection from P_0 onto B;

2. Select some sample points along the geodesic which connects $Q_0 (= V_0)$ and $Q_g (= V_g)$ in a fixed learning rate, we get the point $Q_1 (= V_1), \ldots,$ $Q_n (= V_n)$, $Q_g (= V_g)$;

3. Projecting these points onto M, we get

$$P_1 = \Pi_M Q_1, \ldots, P_n = \Pi_M Q_n, P_f = \Pi_M Q_g \,,$$

those minimize $D(Q_1, P), \ldots, D(Q_n, P)$ and $D(Q_g, P)$ respectively, where $\Pi_M Q$ denotes the m-geodesic-projection from Q onto M.

The coordinates of $P_1, P_2, \ldots, P_f$ are $u_1, u_2, \ldots, u_f$. Connecting points P_0, $P_1, P_2, \ldots, P_f$ using a smooth curve, we get the trajectory of the control input u and the optimal approximation of $g(y)$: $p(y; u_f)$.

When M is an exponential family manifold, because it is e-flat, or at least an α-convex manifold, the projection from $Q \in B$ to M is unique; because B is m-flat, the projection from $P \in M$ to B is unique. This uniqueness will avoid the existence of the local minima and guarantee the stability of controller design.

Acknowledgment. The authors would like to thank the anonymous referees for their valuable comments and suggestions.

References

[1] S. Amari. *Differential geometrical methods in statistics*. Springer lecture Notes in Statistics, **28** (1985).

[2] S. Amari. *Methods of Information Geometry*. TMM 191, Oxford University Press, Oxford, 2000.

[3] S. Amari. *Information Geometry of the EM and em Algorithm for Neural Networks*. Neural Networks, **8** (1995), 1379–1408.

[4] S. Amari, Koji Kurata and Hiroshi Nagaoka. *Information Geometry of Boltaman machines*. IEEE Transactions on Neural Networks, **3** (1992), 260–271.

[5] S. Amari. *Information Geometry on Hierarchy of Probability Distributions*. IEEE Transactions on Information Theory, **47** (2001), 1701–1711.

[6] C.T. Donson and H. Wang. *Iterative approximation of statistical distribution and relation to information geometry*. J. Statistical Inference for Stochastic Process, **147** (2001), 307–318.

[7] H. Wang. *Model reference adaptive control of the output stochastic distributions for unknown linear stochastic systems*. International Journal of System Science, **30** (1999), 707–715.

[8] H. Wang. *Robust control of the output probability density functions for multivariable stochastic systems with guaranteed stability*. IEEE Transaction on Automatic control, **44** (1999), 2103–2107.

[9] H. Wang, H. Baki and P. Kabore. *Control of bounded dynamic stochastic distributions using squre root models: an applicability study in papermaking systems*. Transactions of the Institute of Measurement and Control, **23** (2001), 51–68.

[10] M. Brown and C.J. Harris. Neurofuzzy Adaptive Modelling and Control (Hemel-Hempstead: Prentice-Hall), 1994.

[11] G. Pistone and C. Sempi. *An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one*. Annals of Statistics, **23** (1995), 1543–1561.

Fengwei Zhong,
Huafei Sun
and
Zhenning Zhang
Department of Mathematics
Beijing Institute of Technology
Beijing, 100081
CHINA

E-mails: killowind@sina.com /
sunhuafei@263.net /
ningning0327@163.com