NRC Publications Archive Archives des publications du CNRC

Probability Distributions from Riemannian Geometry, Generalized Hybrid Monte Carlo Sampling and Path Integrals

Paquet, Eric; Viktor, Herna L.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1117/12.872862>

Proceedings. IS and T/SPIE International Symposium on Electronic Imaging, 78645, 2011

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=7f7a1018-88b9-4f92-be99-d7d8c0df9084>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=7f7a1018-88b9-4f92-be99-d7d8c0df9084>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Probability Distributions from Riemannian Geometry, Generalized Hybrid Monte Carlo Sampling and Path Integrals

Abstract. When considering probabilistic pattern recognition methods, especially methods based on Bayesian analysis, the probabilistic distribution is of the utmost importance. However, despite the fact that the geometry associated with the probability distribution constitutes essential background information, it is often not ascertained. This paper discusses how the standard Euclidian geometry should be generalized to the Riemannian geometry when a curvature is observed in the distribution. To this end, the probability distribution is defined for curved geometry. In order to calculate the probability distribution, a Lagrangian and a Hamiltonian constructed from curvature invariants are associated with the Riemannian geometry and a generalized hybrid Monte Carlo sampling is introduced. Finally, we consider the calculation of the probability distribution and the expectation in Riemannian space with path integrals, which allows a direct extension of the concept of probability to curved space.

Keywords: Bayesian, Distribution, Euclidian, Geometry, Information Retrieval, Lagrangian, Monte Carlo, Path Integral, Riemannian.

1 Introduction

One of the cornerstones of Bayesian analysis is the ability to incorporate, in a given probability distributions, the current knowledge as well as the underlying hypotheses [1]. It follows that the better the hypotheses, the better the statistical inference. Despite the fact that a wide range of background information has been considered, there is one that has mostly been neglected in the literature: the underlying geometry of the distribution. Let us consider, for example, a feature vector describing the shape of a three-dimensional object. This feature vector has a probability distribution, i.e. there is a probability associated with each possible value of the feature vector. In addition, one may associate a manifold with the feature vector which means that each value of the feature vector corresponds to a point in the manifold. Such a manifold may be characterized by its geometry. Most of the time, it is assumed, at least implicitly, that the geometry does not constraint the probability distribution. A common example is the Gaussian distribution which assumes that any feature vector may be connected with a straight line to any other feature vector and that the distance (Mahalanobis) does not depend on the underlying geometry (although it depends on the covariance of the feature vector which is a statistical property). It follows that a theoretical analysis of the underlying geometry with respect to the probability

distribution and the implications when calculating distances may provide us with novel insights.

The main objective of this communication is to perform such a general theoretical analysis of non-Euclidian distributions and more precisely, the generalization thereof to Riemannian or curved geometry. We are interested in determining how the underlying geometry affects the probability distribution thereof. The analysis requires some knowledge of differential geometry and tensor analysis. For more details the reader is invited to consult reference [2]. Our objective is to define the probability distribution in Riemannian space, to determine its characteristics, to provide a framework for its calculation and to be able to calculate quantities of interest such as the expectations (i.e. the averages). The main motivation is that the probability distribution should correspond as closely as possible to background information for statistical inference to be efficient; this is relevant especially for Bayesian-based pattern recognition methods. Consequently, the material is organized as follows. Section 2 provides a heuristic justification for the use of Riemannian geometry. In Section 3, a model is constructed in terms of a Lagrangian. This model is used in Sections 4 and 5, in order to show that the probability distribution may be sampled from the geometry with a hybrid Monte Carlo method and that expectations can be calculated from the Lagrangian within a path integral formalism. The main conclusions, including the implications for pattern recognition, and future work are presented in Section 6.

2 Riemannian Geometry and Probability Distributions

We are all familiar with Euclidian geometry; as a matter of fact, this is the geometry that it is implicitly assumed in most probability distributions. It is often supposed, given two feature vectors, that they may be compared to one another with some kind of metric, or distance, that may be defined *independently* of the underlying geometry associated with the feature vector distribution. It is true, as mentioned earlier, that the Mahalanobis distance, which is the distance associated with the Gaussian distribution, considers the statistical aspect of the observations. Nevertheless the underlying geometry is still assumed to be Euclidian.

Fig. 1. Euclidian star shaped (*left*) and Riemannian star shaped (*right*) [3].

Perhaps the best way to introduce the problem is through an example. Consider Figure 1 which represents, schematically, two distributions. For example, it may be the distribution of feature vectors associated with the shapes of 3D objects. The distribution on the left represents the scenario we are the most familiar with. It is

convex and any two points belonging to the distribution may be connected with a straight line which is located entirely within the distribution. The situation is quite different for the distribution shown on the right: points A and B may be connected by a straight line, but the latter does not lie entirely within the distribution. In this case, it is not coherent to define a distance on points that do not exist, that is to say, that do not belong to the distribution, within which distances are calculated. Rather, the most natural solution is to find the closest distance in between the two points, for which the path joining them is situated entirely within the distribution as illustrated in Figure 1 (*right*). It follows that the associated distribution is not Euclidian, because there are points that cannot be connected with a straight line. This is due to the *curvature* of the underlying geometry. During a similarity search, it follows that using one of the standard distance measures such as the Euclidian, Manhattan or, in the case of the Gaussian distribution, the Mahalanobis distance, is thus not appropriate. Even if, for practical reasons, one does not wish to take the geometry of the distribution into account, it still remains important to determine if the Euclidian approximation may be justified and what its domain of validity is. Consequently, a shift of paradigm from flat space to curved space is required. To this end, the modelling of the geometry of feature vector spaces with such a curved geometry is addressed in the next section.

3 Modelling of the Geometry: Curvatures, Surface Terms and Lagrangian

A Riemannian space is characterized by the local curvature, i.e. the Riemann tensor. (Note that the Ricci tensor and scalar are obtained by contraction of the Riemann tensor). This implies that the probability distribution should be defined in terms of the Riemannian tensor, since the latter characterizes the underlying Riemannian geometry [2]. However, the Riemann tensor is not invariant under general coordinate transformations. It is clear that the probability distribution should be invariant under general coordinate transformations. Otherwise, it would be arbitrary, in the sense that it would be defined only up to a general coordinate transformation. This implies that one must construct quantities, from the Riemannian tensor, that are invariant under general coordinate transformations. This may be achieved by constructing scalar quantities from the Riemannian tensor [2] by contracting the indices of the various powers of the curvature with the metric, e.g. as in equation (4).

The various invariants constructed from the curvature are used to construct a Lagrangian. The main motivation for this choice is that it allows us to make use of the full machinery of hybrid Monte Carlo sampling [1]. Also, it provides us with a framework for its generalization in Riemannian space. This is due to the fact that the Hamiltonian formalism, on which a Lagrangian is based, may be generalized to the Riemannian geometry [2, 4]. Generally speaking, the Lagrangian is the difference between the kinetic and potential energy. The Lagrangian is constructed from the curvature which uniquely characterizes the Riemannian geometry. Since the curvatures are local, a Lagrangian density must be defined as

$$L = \int d^d x \mathcal{L}(g^{\mu\nu}(x)) \quad (1)$$

The Lagrangian density is built from scalar functions constructed with the curvatures, the reason being that scalar functions are invariant under coordinate transformation. This is unlike a tensor, for instance the curvatures, which is not invariant. The Lagrangian density is built as

$$\mathcal{L}(g^{\mu\nu}(x)) = \sqrt{|\det \mathbf{g}(x)|} \left(c_0 R(x) + c_1 \sum_{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}(x) R^{\mu\nu\rho\sigma}(x) + \dots \right) \quad (2)$$

where $\sqrt{|\det \mathbf{g}(x)|}$ is the Jacobian of the transformation (if it would not be present, equation (2) would not be invariant under coordinate transformation because the bare infinitesimal volume element is not), where $R(x)$ is the Ricci scalar and where $\{c_i\}$ are model dependent constants multiplying the scalar invariants. Equation (2) is invariant under general coordinate transformations.

As a matter of fact, equation (2) is certainly not the only Lagrangian that may be built. Rather, the Ricci tensor and higher power of the curvatures may also be used if a model of higher complexity is required; the higher the power of the curvature, the higher the order of derivation in the Lagrangian. One must keep in mind, though, that for some particular dimensionality, curvature invariants may be related to each other.

For instance, in four dimensions one has the identity

$$\sum_{\mu\nu\rho\sigma=1}^4 {}^{(4)}R_{\mu\nu\rho\sigma}(x) {}^{(4)}R^{\mu\nu\rho\sigma}(x) - \sum_{\mu\nu=1}^4 {}^{(4)}R_{\mu\nu}(x) {}^{(4)}R^{\mu\nu}(x) + {}^{(4)}R^2(x) + k = 0 \quad (3)$$

where k is a constant.

So far, all the terms in the Lagrangian involve the metric; for instance the second term is of the form

$$R_{\mu\nu\rho\sigma}(x) R^{\mu\nu\rho\sigma}(x) = \sum_{\mu\nu\rho\sigma} \sum_{\eta\mu\kappa\lambda} g_{\eta\mu}(x) g_{\nu\omega}(x) g_{\kappa\rho}(x) g_{\lambda\sigma}(x) R^{\eta\mu\kappa\lambda}(x) R^{\mu\nu\rho\sigma}(x) \quad (4)$$

One may consider if it may not be possible to define invariant terms that do not involve the metric. That is, terms defined only in terms of differential forms; the so-called topological terms. Actually, they are of the form:

$$\int \text{tr}(\mathbf{R}(x) \wedge \mathbf{R}(x) \wedge \dots \wedge \mathbf{R}(x)) \equiv \int \text{tr}(\mathbf{R}^m(x)) \quad (5)$$

where the trace is on the Latin indices (remember that $\mathbf{R}(x)$ is a rotation matrix in flat space as well as a differential 2-form) and where

$$\text{tr}(\mathbf{R}^m(x)) = d\omega_{2m-1}(x) \quad (6)$$

where

$$\omega_{2m-1}(x) = m \int_0^1 d\lambda \lambda^{m-1} \text{tr}(\Omega(x; \lambda)^{m-1}) \quad (7)$$

$$\Omega(x; \lambda) = \omega(x) \wedge (d\omega(x) + \lambda \omega(x) \wedge \omega(x)) \quad (8)$$

and where $\omega(x)$, the connection, is defined as

$$\omega_{\mu ab}(x) = \frac{1}{2}(-\Omega_{\mu ab}(x) + \Omega_{ab\mu}(x) - \Omega_{b\mu a}(x)) \quad | \quad \Omega_{\mu\nu}^a(x) = 2(\partial_\mu e_\nu^a(x) - \partial_\nu e_\mu^a(x)) \quad (9)$$

and where the mobile frame is related to the metric by

$$g_{\mu\nu}(x) = \sum_{ab} e_\mu^a(x) e_\nu^b(x) \delta_{ab} \quad (10)$$

Here the Greek indices refer to curved coordinates while the Latin indices refer to the mobile or local frame. The point of interest here is that the topological term is an exact differential form, i.e. the exterior derivative of another differential form. Suppose the distribution is star-shaped. That is, given a point within the distribution, any other point may be connected to this point with a path that lies entirely within the distribution; then Stokes' theorem can be invoked

$$\int_{\mathcal{M}} \text{tr}(\mathbf{R}^m(x)) = \int_{\mathcal{M}} d\omega_{2m-1}(x) = \int_{\partial\mathcal{M}} \omega_{2m-1}(x) \quad (11)$$

where $\mathcal{M}$ and $\partial\mathcal{M}$ are the Riemannian space associated with the distribution and its boundary respectively. One may distinguish in between two cases of interest. If the distribution is not bounded, one may put the boundary at infinity and equation (5) disappears. If the distribution is bounded, topological terms such as (5) must be added to the Lagrangian, in order to take a possible boundary effect into account.

Despite the extensive mathematical apparatus used, the conclusions and implications are quite straightforward. The distribution is described by a scalar function, the Lagrangian, which is invariant under general coordinate transformation. This means that it does not depend on the particular coordinate system that is used to represent the distribution. This is a "sine qua non" condition if one does not want the representation to be arbitrary, in the sense that it would be valid only in a particular reference frame. In other words, the distribution should not depend on the way it is represented, but on its sole intrinsic geometrical structure irrespectively of the reference frame chosen to describe it. The Lagrangian is constructed from the curvatures, which uniquely characterized the Riemannian space. If the distribution is compact, boundary terms of the form (5) have to be included. Otherwise, if the distribution is not bounded; they all disappear by Stokes' theorem.

4 Generalized Hybrid Monte Carlo Method

In this section, the probability distribution is defined and sampled in terms of a generalized hybrid Monte Carlo method. The choice of the hybrid Monte Carlo framework is justified by the fact that it is based on the Hamiltonian formalism. The Hamiltonian is provided with the Legendre transform (see below) of the Lagrangian [4] defined in the previous section, which is an invariant defined in terms of the Riemann tensor that uniquely characterizes the underlying geometry [2]. The main idea of the standard hybrid Monte Carlo method is to associate a Hamiltonian, which is related to the total energy of the distribution. This Hamiltonian is obtained by transforming the Lagrangian with a Legendre transform

$$H(x^i, \pi_i) = \sum_i \pi_i \frac{dx^i}{d\tau}(x^i, \pi_i) - L\left(x^i, \frac{dx^i}{d\tau}(x^i, \pi_i)\right) \quad (12)$$

where x^i are the generalized coordinates (here as defined in Euclidian space), and π_i the conjugated momentum associated with the coordinate x^i . Here, τ is an evolution parameter, called the pseudo-time, that is introduced in order to eventually replace the spatial sampling by a sequential (one may say temporal) sampling. This parameter is necessary to introduce the Hamiltonian formalism and will be necessary later in order to sample the distribution in terms of a Markov process. The evolution of the system, in terms of the pseudo-time, is related to Hamilton's equations:

$$\begin{aligned} \pi_i &= \frac{dx_i}{d\tau} \\ \frac{dx_i}{d\tau} &= \frac{\partial H(x^i, \pi_i)}{\partial \pi_i} \\ \frac{d\pi_i}{d\tau} &= -\frac{\partial H(x^i, \pi_i)}{\partial x^i} \end{aligned} \quad (13)$$

Invoking statistical physics [5], the probability of occurrence of a given observation (x^i, π_i) is given by

$$p(x^i, \pi_i) = \frac{1}{Z} \exp(-H(x^i, \pi_i)) \quad (14)$$

where

$$Z = \sum_i \exp(-H(x^i, \pi_i)) \quad (15)$$

the partition function associated with the Hamiltonian, is responsible for the normalization. Then, the calculation of the transition probability is isomorphic to a Markov process. From a previous state $({}^\tau x^i, {}^\tau \pi_i)$ one evolves the system to a new state $({}^{\tau+\Delta\tau} x^i, {}^{\tau+\Delta\tau} \pi_i)$ with Hamilton's equations (13) and the transition probability in between the two states, which has a Metropolis [6] form, is given by

$$\min\left(1, \exp\left\{-H({}^\tau x^i, {}^\tau \pi_i) - H({}^{\tau+\Delta\tau} x^i, {}^{\tau+\Delta\tau} \pi_i)\right\}\right) \quad (16)$$

A sufficient, although not necessary condition for this equation to be valid is that the detailed balance equation holds; the latter being

$$\begin{aligned} \frac{1}{Z} \exp(-H({}^\tau \Omega)) \delta V \min\left(1, \exp\left\{H({}^\tau \Omega) - H[{}^{\tau+\Delta\tau} \Omega]\right\}\right) &= \dots \\ \frac{1}{Z} \exp(-H({}^{\tau+\Delta\tau} \Omega)) \delta V \min\left(1, \exp\left\{H({}^{\tau+\Delta\tau} \Omega) - H[{}^\tau \Omega]\right\}\right) \end{aligned} \quad (17)$$

where ${}^\tau\Omega$ refers to the previous state in phase space. This is the state built from the generalized coordinates and conjugated momentum, ${}^{\tau+\Delta\tau}\Omega$ refers to the evolved state, as obtained from Hamilton's equations, δV is the elementary volume in phase space. Such a volume is an invariant by Liouville's theorem. Indeed, it may be shown that the Lie derivative in phase space of the differential form associated with the elementary phase volume vanishes, which means that such a volume is an invariant. Such an equation is valid in Euclidian as well as in Riemannian space, where

$$\begin{aligned} {}^i\Lambda^2 &= d\pi_i \wedge dq^i \\ \mathcal{L}_X \Lambda^{2n} &= \mathcal{L}_X ({}^1\Lambda^2) \wedge \dots \wedge ({}^n\Lambda^2) = 0 \end{aligned} \quad (18)$$

Equation (18) has far-reaching practical implications; for instance in the numerical implementation of Hamilton's equations. It implies that the numerical integration algorithm must preserve, by construction, the elementary volume in phase space. If this is not the case, detailed balance does not hold and the Markov process is not valid. For instance, the leapfrog algorithm is an integration algorithm for which detailed balance holds [1]. In conclusion, there is a Markov process which alternate in between a stochastic updating from the partition function and a dynamical updating from Hamilton's equations.

The extension of the hybrid Monte Carlo method to a Riemannian geometry is far from trivial. Because all quantities are local, the Hamiltonian should be replaced by the Hamiltonian density which is obtained by transforming the Lagrangian with a Laplace transform

$$\mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x)) = \sum_{\mu\nu} \sqrt{\det \mathbf{g}(x)} \Pi_{\mu\nu}(x) \frac{dg^{\mu\nu}}{d\tau}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x)) - \mathcal{L}(g^{\mu\nu}(x)) \quad (19)$$

It should be noticed here that the coordinate x is a continuous index. The generalized coordinates are the metric which, as stated earlier, is the fundamental quantity associated with the Riemannian geometry (note that all other quantities e.g. curvatures, connections can be expressed in terms of the metric in a Riemannian space), as well as the corresponding conjugate momentum. This implies that there are an infinite number of generalized coordinates associated with a Riemannian geometry. As for the Lagrangian, the Hamiltonian may be obtained from the Hamiltonian density

$$H = \int dx^d \mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x)) \quad (20)$$

Hamilton's equations have to be generalized to

$$\begin{aligned} \Pi_{\mu\nu}(x) &= \frac{dg_{\mu\nu}(x)}{d\tau} \\ \frac{dg^{\mu\nu}(x)}{d\tau} &= \frac{\delta \mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x))}{\delta \Pi_{\mu\nu}(x)} \\ \frac{d\Pi_{\mu\nu}(x)}{d\tau} &= - \frac{\delta \mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x))}{\delta g^{\mu\nu}(x)} \end{aligned} \quad (21)$$

where $\frac{\delta}{\delta g^{\mu\nu}(x)}$ is the Fréchet or functional derivative [4] which is defined as

$$\frac{\delta}{\delta g^{\mu\nu}(x)} \triangleq \frac{\partial}{\partial g^{\mu\nu}(x)} - \frac{\partial}{\partial \tau \left(\frac{\partial g^{\mu\nu}(x)}{\partial \tau} \right)} - \sum_{\rho} \frac{\partial}{\partial x_{\rho} \left(\frac{\partial g^{\mu\nu}(x)}{\partial x^{\rho}} \right)} + \dots \quad (22)$$

plus eventually additional terms if derivatives or higher order are present. It is important to stress that the generalized coordinates are not coordinates anymore, but functions of the later, and this is why the functional derivative has to be introduced. The main consequence, from a pattern recognition perspective, is that it is not the indexes that are the fundamental quantity anymore, but the metrics associated with them. This implies that probabilities should be formulated in terms of the metric. Once more, invoking statistical physics [5], one can define a probability density and a partition function

$$p(g^{\mu\nu}(x), \Pi_{\mu\nu}(x)) = \frac{1}{\mathcal{Z}} \exp\left(-\int dx^d \mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x))\right) \quad (23)$$

where

$$\mathcal{Z} = \int \mathcal{D}g(x) \mathcal{D}\Pi(x) \exp\left\{-\int dx^d \mathcal{H}(g^{\mu\nu}(x), \Pi_{\mu\nu}(x))\right\} \quad (24)$$

Here $p(g^{\mu\nu}(x), \Pi_{\mu\nu}(x))$ is the probability of having a certain metric, $\mathcal{Z}$ is the partition function and where the measure $\mathcal{D}q(x)$ which is either the metric or the momentum is defined as

$$\int \mathcal{D}q \equiv \lim_{N \rightarrow \infty} \int \dots \int \prod_{\mu} \prod_{\nu} \prod_{l=1}^N dq^{\mu\nu}[l] \quad (25)$$

and where $dq^{\mu\nu}[l]$ is the differential of either the metric or the momentum at a particular discrete location l [7]: this is a generalization of the summation expressed in equation (15). In practice, a finite number of samples are taken and the integration is performed with a Monte Carlo technique [6] (not to be confused with the hybrid Monte Carlo method). Equation (24) is a generalization of equation (15) from a discrete number of coordinates to a continuum of coordinates.

Equation (16), the transition probability, becomes in Riemannian space

$$\min\left\{1, \frac{1}{\mathcal{Z}} \exp\left\{\int dx^d \mathcal{H}(\tau g^{\mu\nu}(x), \tau \Pi_{\mu\nu}(x)) - \mathcal{H}(\tau + \Delta\tau g^{\mu\nu}(x), \tau + \Delta\tau \Pi_{\mu\nu}(x))\right\}\right\} \quad (26)$$

where $(\tau g^{\mu\nu}(x), \tau \Pi_{\mu\nu}(x))$ is the previous state and where $(\tau + \Delta\tau g^{\mu\nu}(x), \tau + \Delta\tau \Pi_{\mu\nu}(x))$ is the evolved state obtained from Hamilton's equations. It can be shown that detailed balance hold as in equation (17) which is a sufficient condition for the validity of the Markov approach.

5 Calculation of Expectations with Path Integrals

In this section, the probability distribution as introduced in Section 4 is defined directly in terms of the Lagrangian, which simplifies the formalism and its implementation. It is shown how the expectations (averages) may be obtained in such a case. A number of functions $f(x)$ defined on the coordinates are of prime importance, e.g. the expectation and the covariance, i.e.

$$\begin{aligned} f(x) = x &\Rightarrow \mathbb{E}(x) = \mu \\ f(x) = (x - \mu)^2 &\Rightarrow \mathbb{E}\left((x - \mu)^2\right) = \Sigma \end{aligned} \quad (27)$$

Furthermore, it would be more convenient to work directly with the Lagrangian and to spare the additional work of a Laplace transform. Because equation (19) is quadratic in the momentum, one can integrate over the momentum, at least for the regularized integral (due to the finite number of elements in equation (25)), with the help of the following matrix identity:

$$\int \dots \int dq_1 \dots dq_N \exp\left[-\frac{1}{2} \mathbf{q}^T \mathbf{A} \mathbf{q}\right] = (2\pi)^{N/2} \exp\left[-\frac{1}{2} \text{tr}(\ln(\mathbf{A}))\right] \quad (28)$$

Then, from equation (23), the probability density reduces to

$$p(x) = \frac{\exp\left[-\int dx^d \mathcal{L}(g^{\mu\nu}(x))\right]}{\int \mathcal{D}g(x) \exp\left[-\int dx^d \mathcal{L}(g^{\mu\nu}(x))\right]} \quad (29)$$

and the expectation becomes

$$\mathbb{E}(f(x)) = \int \mathcal{D}g(x) f(x) p(x) \quad (30)$$

Equation (29) has a form that is reminiscent of a Gaussian distribution. However, this is as far as the comparison goes. Rather, the measure, cf. equation (25), is completely different by nature and the convergence of the integral is only guaranteed if the integral is regularized. Furthermore, equivalent configurations should not be counted twice in the integral. This implies either choosing unambiguous curved coordinates for the space, or alternatively introducing a constraint within the integral that ensures that equivalent configurations are not counted twice. The first approach seems to be easier to implement and shall be part of our future work.

6 Conclusions

In this communication, it has been shown that the underlying geometry of the observations is an important Bayesian hypothesis, which has far-reaching fundamental and practical consequences, especially for pattern recognition algorithms based on Bayesian analysis. An analysis has been performed in the framework of Riemannian geometry. This is due to the fact that the Riemannian geometry is much more general than the Euclidian geometry and also because locally, the Riemannian geometry is approximately Euclidian. This local isomorphism is of prime importance

if one wishes to establish a correspondence in between the two geometries. The postulate of metricity inherent to Riemannian space is a “sine qua non” condition in order to be able to define the very concepts of distance and similarity. It has been shown that, in such a case, the metric is local and the “Euclidian distance” should be replaced by the geodesic distance. Furthermore, the geometry may be modelled by a Lagrangian or a Hamiltonian constructed from curvature invariants. We also showed how hybrid Monte Carlo sampling may be extended to Riemannian geometry by introducing functional derivatives and path integrals. Finally, by integrating out the momentum, it is possible to express the expectation in terms of a path integral defined over the Lagrangian.

In conclusion, we mention some related work of high interest in which, at least implicitly, the curvature of the underlying space is considered. This includes [8] in which the Mahalanobis distance is replaced by the geodesic distance for principal component analysis, [9] which maintains the use of the Mahalanobis distance but consider the later as only valid locally on an open subset, [10] which defines the metric in terms of “ad hoc” constraints on the geometry transformations, [11] which calculates the Eigen vector in terms of the Laplacian in curved coordinates (the Laplacian is related to the differential forms: the solution of the Laplace equations are the same than the cohomology of the associated Riemannian space) and finally, [12] which performs local dimensionality reduction based on factor analysers. Future work will focus on the applications of the theory as well as further analysis of the regularization of the path integrals.

References

1. Bishop, C. M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
2. Frankel, T.: The Geometry of Physics: An Introduction. Cambridge University Press, Cambridge (2003)
3. Removed for revision
4. Goldstein, H.: Classical Mechanics. Addison-Wesley, Reading (1980)
5. Greiner, W., Neise, L. Stocker, H and Rischke, D.: Thermodynamics and Statistical Mechanics. Springer, Berlin (2001)
6. Robert, C. P. and Casella, G.: Monte Carlo Statistical Methods. Springer, New York (1999)
7. Kleinert, H. Path Integrals in Quantum Mechanics, Statistics Polymer Physics and Financial Market. World Scientific Publishing, Singapore (2010)
8. Tenenbaum, J. B., da Silva, V. and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290** (2000) 2319-2323
9. Roweis, S. T. and Saul, L. K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **290** (2000) 2323-2326
10. Kilian, M., Mitra, N. J. and Pottmann, H.: Geometric Modeling in Shape Space, *ACM Transactions on Graphics* **26** (3) (2007) DOI=<http://doi.acm.org/10.1145/1276377.1276457>
11. Sorkine, O.: Differential Representations for Mesh Processing. *Computer Graphics Forum* **25** (4) (2006) 789-807
12. Ghahramani, Z. and Beal, M. J.: Variational Inference for Bayesian Mixtures of Factor Analysers. *Advances in Neural Information Processing Systems* **12** (2000) 449-455