Dynamics of the Fisher information metric

Xavier Calmet[1,*] and Jacques Calmet[2,†]

[1]*University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA*
[2]*Institute for Algorithms and Cognitive Systems (IAKS) University of Karlsruhe (TH), D-76131 Karlsruhe, Germany*

We present a method to generate probability distributions that correspond to metrics obeying partial differential equations generated by extremizing a functional $J[g^{\mu\nu}(\theta^i)]$, where $g^{\mu\nu}(\theta^i)$ is the Fisher metric. We postulate that this functional of the dynamical variable $g^{\mu\nu}(\theta^i)$ is stationary with respect to small variations of these variables. Our approach enables a dynamical approach to the Fisher information metric. It allows one to impose symmetries on a statistical system in a systematic way.

I. INTRODUCTION

The Fisher information metric can be calculated once a probability distribution has been chosen. In this work we wish to present a method to generate probability distributions that correspond to metrics obeying Euler-Lagrange partial differential equations generated by extremizing a functional $J[g^{\mu\nu}(\theta^i)]$, where $g^{\mu\nu}(\theta^i)$ is the Fisher metric. We will postulate that this functional of the dynamical variable $g^{\mu\nu}(\theta^i)$ is stationary with respect to small variations of these variables. Our approach enables a dynamical approach to the Fisher information metric. It allows one to impose symmetries on a statistical system in a systematic way. We will show how to obtain partial differential equations for the probability distributions. Imposing that the functional remains invariant under certain transformations of the coordinates θ^i will constrain the class of probability distributions. These symmetries can also be broken in a consistent way using methods borrowed from theoretical particle physics and general relativity. As an example we will require that the functional is covariant under general transformations of the coordinates θ. This will lead to nonlinear partial differential equations corresponding to Einstein's equations [1] of general relativity (see, e.g., [2] for a nice introduction to general relativity). The requirement that the functional remains invariant under infinitesimal general coordinate transformations $\theta^i \rightarrow \theta^{i'}$ leads to Bianchi identities. We shall nevertheless emphasize that variations of the Fisher information functional to obtain equations of dynamics have already been considered in another framework [3].

A set of distributions $p_\theta = p(\theta)$ parametrized by θ^i with $i \in \{1, \ldots, n\}$ is a manifold. The Riemanian metric on this manifold is the Fisher information metric defined by

$$g_{\mu\nu} = \int_X d^4x \, p_\theta(x) \left(\frac{1}{p_\theta(x)} \frac{\partial p_\theta(x)}{\partial \theta^\mu} \right) \left(\frac{1}{p_\theta(x)} \frac{\partial p_\theta(x)}{\partial \theta^\nu} \right). \quad (1)$$

In the following we shall first briefly introduce the concept of the Fisher information metric insisting on the connection to the concepts of Shannon entropy and of the Kullback-Leibler distance. We shall then introduce the dynamics in the Fisher information metric. We shall argue that requiring the covariance of the functional $J[g^{\mu\nu}(\theta^i)]$ under general transformations of θ^i is natural since the Fisher information metric is itself invariant under reparametrization of the manifold [4]. We note that the Fisher information metric is also invariant under transformations of the random variable $x \in X$. We will concentrate on general coordinate invariance, but in principle, depending on the application, any type of symmetry could be imposed on the functional.

II. REVIEW OF FISHER INFORMATION METRIC

There are many excellent reviews and books on the Fisher information metric; a nice introduction can be found in [3]. Fisher information is a useful concept to judge the quality of statistical estimates. The Fisher information metric has been applied in different fields. This concept appears in such different fields as, e.g., instanton calculus [5], ontology [6], models for short distance modifications of space-time [7], or econometrics [8]. Symbolic computations of Fisher information matrices have also been considered [9]. We shall give a brief review of the derivation of the information metric (see [10,11]).

A distance $d(P_1, P_2)$ between two points P_1 and P_2 has to satisfy the following three axioms:
(1) Positive definiteness: $\forall \, P_1, P_2, \, d(P_1, P_2) \geq 0$;
(2) Symmetry: $d(P_1, P_2) = d(P_2, P_1)$;
(3) Triangle inequality:
$\quad \forall \, P_1, P_2, P_3, \, d(P_1, P_2) \leq d(P_1, P_3) + d(P_2, P_3)$.

It is often useful to introduce the concept of distance between elements of a more abstract set. For example, one could ask what is the distance between two distributions, between, e.g., the Gaussian and binomial distributions. It is useful to introduce the concept of entropy as a means to define distances. In information theory, Shannon entropy [12] represents the information content of a message or, from the receiver's point of view, the uncertainty about the message the sender produced prior to its reception. It is defined as

*Email address: calmet@physics.unc.edu
†Email address: calmet@ira.uka.de

$$-\sum_i p(i)\log p(i), \qquad (2)$$

where $p(i)$ is the probability of receiving the message i. The unit used is the bit. The relative entropy can be used to define a "distance" between two distributions $p(i)$ and $g(i)$. The Kullback-Leibler [13] distance or relative entropy is defined as

$$D(g\|p) = \sum_i g(i)\log \frac{g(i)}{p(i)}, \qquad (3)$$

where $p(i)$ is the real distribution and $g(i)$ is an assumed distribution. Clearly the Kullback-Leibler relative entropy is not a distance in the usual sense: it satisfies the positive definiteness axiom, but not the symmetry or the triangle inequality axioms. It is nevertheless useful to think of the relative entropy as a distance between distributions.

The Kullback-Leibler distance is relevant to discrete sets. It can be generalized to the case of continuous sets. For our purposes, a probability distribution over some field (or set) X is a distribution $p: X \in \mathbb{R}$, such that
  (1) $\int_X d^4x\, p(x) = 1$,
  (2) for any finite subset $S \subset X$, $\int_S d^4x\, p(x) > 0$.
We shall consider families of distributions, and parametrize them by a set of continuous parameters θ^i that take values in some open interval $M \subseteq \mathbb{R}^4$. We use the notation p_θ to denote members of the family. For any fixed θ, $p_\theta: x \mapsto p_\theta(x)$ is a mapping from X to $\mathbb{R}$. We shall consider the extension of the family of distributions $F = \{p_\theta | \theta \in M\}$, to a manifold $\mathcal{M}$ such that the points $p \in \mathcal{M}$ are in one to one correspondence with the distributions $p \in F$. The parameters θ of F can thus be used as coordinates on $\mathcal{M}$.

The Kullback number is the generalization of the Kullback-Leibler distance for continuous sets. It is defined as

$$I(g_\theta\|p_\theta) = \int d^4x\, g_\theta(x)\log \frac{g_\theta(x)}{p_\theta(x)}. \qquad (4)$$

Let us now study the case of an infinitesimal difference between $q_\theta(x) = p_{\theta+\epsilon v}(x)$ and $p_\theta(x)$:

$$I(p_{\theta+\epsilon v}\|p_\theta) = \int d^4x\, p_{\theta+\epsilon v}(x)\log \frac{p_{\theta+\epsilon v}(x)}{p_\theta(x)}. \qquad (5)$$

Expanding in ϵ and keeping θ and v fixed one finds

$$I(p_{\theta+\epsilon v}\|p_\theta) = I(p+\epsilon\|p)|_{\epsilon=0} + \epsilon I'(\epsilon)|_{\epsilon=0}$$
$$+ \frac{1}{2}\epsilon^2 I''(\epsilon)|_{\epsilon=0} + O(\epsilon^3). \qquad (6)$$

One finds $I(0) = I'(0) = 0$ and

$$I''(0) = v^\mu \left[\int_X d^4x\, p_\theta(x)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial \theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial \theta^\nu}\right) \right] v^\nu. \qquad (7)$$

We can identify the Fisher information metric [14] on a manifold of probability distributions as

$$g_{\mu\nu} = \int_X d^4x\, p_\theta(x)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial \theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial \theta^\nu}\right). \qquad (8)$$

It has been shown that this matrix is a metric on a manifold of probability distributions (see, e.g., [11]) Corcuera and Giummolè [4] have shown that the Fisher information metric is invariant under reparametrization of the sample space X and that it is covariant under reparametrizations of the manifold, i.e., the parameter space (see, e.g., [15] for a review.)

III. DYNAMICS OF FISHER INFORMATION METRIC

Given a distribution $p(\theta)$, the Fisher information metric can be calculated using Eq. (8). Here we wish to approach this issue from a different point of view and derive the Fisher information metric from a dynamical perspective. We introduce a functional $J[g^{\mu\nu}(\theta^i)]$. To constrain the functional dependence on $g^{\mu\nu}(\theta^i)$, we impose that this functional is invariant under general transformations of the coordinates θ^i. As already mentioned, the Fisher information metric is covariant under reparametrizations of the manifold, i.e., it is covariant under general coordinate transformations $\theta^i \rightarrow \theta^{i'}$. It thus seems natural to posit that the functional $J[g_{\mu\nu}]$, describing the dynamics of the metric, is invariant under general coordinate transformations $\theta^i \rightarrow \theta^{i'}$. We shall first introduce a few basic quantities which are well known from differential geometry and the general theory of relativity.

A. Definitions

Differential geometry is the tool we shall be using, just as in general relativity; the concepts of differential geometry we need are nicely introduced in [2]. Let us define a few quantities. The affine connection is defined as

$$\Gamma^\sigma_{\lambda\nu} = \frac{1}{2}g^{\nu\sigma}\left(\frac{\partial g_{\mu\nu}}{\partial \theta^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial \theta^\mu} - \frac{\partial g_{\mu\lambda}}{\partial \theta^\nu}\right). \qquad (9)$$

The curvature tensor is given by

$$R^\lambda_{\mu\nu\kappa} = \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial \theta^\kappa} - \frac{\partial \Gamma^\lambda_{\mu\kappa}}{\partial \theta^\nu} + \Gamma^\eta_{\mu\nu}\Gamma^\lambda_{\kappa\eta} - \Gamma^\eta_{\mu\kappa}\Gamma^\lambda_{\nu\eta}. \qquad (10)$$

It is the only tensor that can be constructed from the metric tensor $g_{\mu\nu}$ and its first and second derivatives. We shall also need the Ricci tensor, which reads

$$R_{\mu\kappa} = R^\lambda_{\mu\lambda\kappa}, \qquad (11)$$

and the curvature scalar is given by

$$R = g^{\mu\kappa}R_{\mu\kappa}. \qquad (12)$$

B. Dynamics

We now have all the necessary tools to introduce the invariant functional which describes the dynamics of Fisher information metric. Because of [4], it seems natural to posit that the functional $J[g_{\mu\nu}]$, describing the dynamics of the

metric, is invariant under general coordinate transformations[1] $\theta^i \to \theta^{i'}$. We shall start from the functional:

$$J[g_{\mu\nu}] = \frac{-1}{16\pi} \int \sqrt{g(\theta)} R(\theta) d^4\theta, \qquad (13)$$

where $g = \det g^{\mu\nu}$ and $R(\theta)$ is the curvature scalar. This functional is a scalar and it is invariant under general coordinate transformations. Note that $R(\theta)$ is a scalar and is thus invariant under general coordinate transformations, $d^4\theta$ transforms as

$$d^4\theta' \to \left| \frac{\partial\theta'}{\partial\theta} \right| d^4\theta, \qquad (14)$$

and $\sqrt{g(\theta)}$ transforms as

$$\sqrt{g(\theta')} \to \left| \frac{\partial\theta}{\partial\theta'} \right| \sqrt{g(\theta)}. \qquad (15)$$

The functional is a scalar and thus invariant under general coordinate transformations.

We assume the least action principle and consider small virtual fluctuations of the metric. The variation of $J[g_{\mu\nu}]$ with respect to $g_{\mu\nu}$ leads to

$$\delta J[g_{\mu\nu}] = \frac{1}{16\pi} \int \sqrt{g} \left[R^{\mu\nu}(\theta) - \frac{1}{2} g^{\mu\nu}(\theta) R(\theta) \right] \delta g_{\mu\nu} d^4\theta, \qquad (16)$$

where $R^{\mu\nu}(\theta)$ is the Ricci tensor introduced in Eq. (11) and $R(\theta)$ is the curvature scalar [see Eq. (12)]. The requirement that $J[g_{\mu\nu}]$ be invariant with respect to variation of the metric $g_{\mu\nu}$, $\delta g_{\mu\nu}$ leads to the following partial differential equations (the Euler-Lagrange equations):

$$R^{\mu\nu}(\theta) - \frac{1}{2} g^{\mu\nu}(\theta) R(\theta) = 0. \qquad (17)$$

Contracting Eq. (17) with $g_{\mu\nu}$, one finds $R = 0$; the partial differential equations (the Euler-Lagrange equations) are thus

$$R^{\mu\nu}(\theta) = 0, \qquad (18)$$

which correspond to the equations obtained by Einstein in his theory of general relativity [1]. If the statistical system under consideration is invariant under reparametrization of the manifold, the dynamics of the Fisher information metric is governed by these partial differential equations.

Another interesting case is the one of infinitesimal transformations $\theta = \theta + \epsilon(\theta)$. The functional $J[g_{\mu\nu}]$ being a scalar is also invariant under this class of transformation. This transformation of θ implies the following variation for the metric:

$$\delta g_{\mu\nu}(\theta) = -g_{\mu\lambda}(\theta) \frac{\partial\epsilon^\lambda(\theta)}{\partial\theta^\nu} - g_{\lambda\nu}(\theta) \frac{\partial\epsilon^\lambda(\theta)}{\partial\theta^\mu} - \frac{\partial g_{\mu\nu}(\theta)}{\partial\theta^\lambda} \epsilon^\lambda(\theta). \qquad (19)$$

The variation of the functional J with respect to this transformation leads to the contracted Bianchi identity

$$\left[R^\lambda_{\ \nu} - \frac{1}{2} \delta^\lambda_{\ \nu} R \right]_{;\nu} = 0, \qquad (20)$$

where the semicolon stands for the covariant derivative which is defined by

$$V_{\mu;\lambda} = \frac{\partial V_\mu}{\partial\theta^\lambda} - \Gamma^\lambda_{\mu\nu} V^\nu. \qquad (21)$$

It is possible to introduce more dynamics in the model by introducing another source (the partial differential equations being nonlinear, the metric is a source for itself already) for the metric in the form of a symmetric tensor $T^{\mu\nu}$ which transforms as a contravariant tensor [i.e., $T^{\mu\nu'} \to (\partial\theta^\mu/\partial\theta^\rho) \times (\partial\theta^\nu/\partial\theta^\sigma) T^{\rho\sigma}$]. The modified functional then becomes

$$J[g_{\mu\nu}] = \frac{-1}{16\pi} \int \sqrt{g(\theta)} R(\theta) d^4\theta + \frac{1}{2} \int \sqrt{g(\theta)} T^{\mu\nu} g_{\mu\nu} d^4\theta. \qquad (22)$$

This functional leads to the partial differential equations

$$R^{\mu\nu}(\theta) - \frac{1}{2} g^{\mu\nu}(\theta) R(\theta) + 8\pi T^{\mu\nu}(\theta) = 0. \qquad (23)$$

In the general theory of relativity the case $T^{\mu\nu}(\theta) = 0$ corresponds to empty space. In the statistical system case, the tensor $T^{\mu\nu}(\theta)$ can be used to implement constraints on the statistical system under consideration. Notice that symmetries, in our case general coordinate invariance, can be broken by generating terms in $T^{\mu\nu}$ that are not transforming as second rank tensors. Imposing symmetries on a system and breaking these symmetries usually leads to relations between the parameters of the model. This can be very useful for applications.

Obviously the functional we have chosen is the simplest case possible. We have chosen to build the curvature tensor using only the metric and its first and second derivatives. One could obviously obtain more complicated models by taking higher derivatives into account. One could also introduce further dynamics in the model. As an example one can add a scalar ϕ in the theory and consider the so-called Brans-Dicke model [16]

$$J[g_{\mu\nu}] = \frac{-1}{16\pi} \int \sqrt{g(\theta)} \phi(\theta) R(\theta) d^4\theta + [\text{dynamics for } \phi(\theta)]. \qquad (24)$$

This functional is also a scalar, i.e., it is invariant under general coordinate transformations, but it leads to a different set of partial differential equations.

---

[1]This situation arises in Einstein's theory of general relativity.

C. Constraints for the probability distributions

If one inserts the expression for the Fisher information metric (8) in the partial differential equations (18) obtained by postulating that the functional is invariant under general coordinate transformations (note that $R_{\mu\nu}$ and R are functions of $g_{\mu\nu}$), one finds a lengthy and complicated expression. It is a differential equation for the distribution $p_\theta(x)$. It is best to first find solutions to the partial differential equation (17) or (18) to obtain the metric constrained by the symmetries of the problem and then to deduce the constraints for the distribution $p_\theta(x)$. Finding an exact solution to the partial differential equation (17) or (18) is in general very difficult; however, when the problem under consideration has enough symmetries, certain solutions are known.

Obviously, a trivial diagonal metric with constant entries of the type $(1, \ldots, 1)$ is a solution of the partial differential equations (17). One obtains the following constraints for probability distributions:

$$\int_X d^4x\, p_\theta(x)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\mu}\right) = 1,$$

$$\int_X d^4x\, p_\theta(x)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\nu}\right) = 0 \quad \text{if } \mu \neq \nu.$$

$$(25)$$

This constraint is trivial and for example Gaussian distributions satisfy it.

Another example is that of a four-dimensional problem which is spherically symmetric in three coordinates (isotropic) say θ^1, θ^2, and θ^3 and with a metric that is independent of the fourth coordinate θ^0. In that case a solution was found by Schwarzschild [17] (see also, e.g., [2]). Using spherical coordinates for the three coordinates θ^1, θ^2, and θ^3, the metric is of the form $(1-2\alpha/r, (1-2\alpha/r)^{-1}, r^2, r^2\sin^2\theta)$ for a four-dimensional problem denoting the coordinates by (τ, r, θ, ϕ), and where α is an integration constant. We thus obtain the following constraints for a probability distribution:

$$\int_X d^4x \frac{1}{p_\theta(x)}\left(\frac{\partial p_\theta(x)}{\partial\tau}\right)^2 = 1 - \frac{2\alpha}{r}, \qquad (26)$$

$$\int_X d^4x \frac{1}{p_\theta(x)}\left(\frac{\partial p_\theta(x)}{\partial r}\right)^2 = \left(1 - \frac{2\alpha}{r}\right)^{-1},$$

$$\int_X d^4x \frac{1}{p_\theta(x)}\left(\frac{1}{r}\frac{\partial p_\theta(x)}{\partial\theta}\right)^2 = r^2,$$

$$\int_X d^4x \frac{1}{p_\theta(x)}\left(\frac{1}{r\sin\theta}\frac{\partial p_\theta(x)}{\partial\phi}\right)^2 = r^2\sin^2\theta,$$

$$\int_X d^4x\, p_\theta(x)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\nu}\right) = 0 \quad \text{if } \mu \neq \nu.$$

$$(27)$$

This example illustrates how to obtain nontrivial constraints on the probability distributions.

IV. CONCLUSIONS

In this work we have presented a method to generate probability distributions that correspond to metrics obeying partial differential equations generated by extremizing a functional $J[g^{\mu\nu}(\theta^i)]$, where $g^{\mu\nu}(\theta^i)$ is the Fisher metric. We have postulated that this functional of the dynamical variable $g^{\mu\nu}(\theta^i)$ be stationary with respect to small variations of these variables. Our approach enables a dynamical approach to the Fisher information metric. It allows us to impose symmetries on a statistical system in a systematic way. We have presented different models and some solutions to these partial differential equations. There is a very nice analogy between the Fisher information metric and Einstein's theory of general relativity. We have argued that since the Fisher information metric is covariant under reparametrizations of the manifold, i.e., it is covariant under general coordinate transformations $\theta^i \rightarrow \theta^{i'}$, it is natural to posit that the functional $J[g_{\mu\nu}]$, describing the dynamics of the metric, is invariant under general coordinate transformations $\theta^i \rightarrow \theta^{i'}$. This led us to the functional that determines the dynamics of our models.

We foresee several application domains such as reasoning or quantum computing. There is an additional application that is under investigation and is a classical one for stochastic methodologies: classification in insurance [18]. We expect to refine the classification process through symmetry considerations. The Fisher information metric is a standard tool to analyze data. We expect that our work will help to classify data and verify whether reparametrization symmetries are present in the data. Our work will also have implications for thermodynamics, e.g., for the problem leading to Landau's diamagnetism where Fisher's information plays an important role [19] or for the canonical ensemble entropy for the one-dimensional quantum harmonic oscillator which has also been shown to be related to Fisher's information [20].

[1] A. Einstein, Ann. Phys. **49**, 769 (1916).

[2] S. Weinberg, *Gravitation and Cosmology* (John Wiley, New York, 1972).

[3] B. R. Frieden, *Physics from Fisher Information: A Unification* (Cambridge University Press, Cambridge, U.K., 1998).

[4] J. M. Corcuera and F. Giummolè, Ann. Inst. Stat. Math. **50**, 433 (1998).

[5] S. Yahikozawa, Phys. Rev. E **69**, 026122 (2004).

[6] J. Calmet and A. Daemi, in *Proceedings of the 17th European Meeting on Cybernetics and Systems Research EMCSR 2004*, edited by R. Trappl (Austrian Society for Cybernetic Studies, Vienna, 2004), Vol. 2, pp. 547–551.

[7] J. Calmet and X. Calmet, WSEAS Trans. Circuits Syst., **10**, 2267 (2004).

[8] P. Marriott and M. Salmon, http://www.business.city.ac.uk/ferc//wpapers/ms2.pdf

[9] R. L. M. Peeters and B. Hanzon, Automatica **35**, 1059 (1999).

[10] E. R. Caianiello, Riv. Nuovo Cimento **15** (4) (1992).

[11] C. C. Rodriguez, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic, Dordrect, 1989), pp. 415–422.

[12] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).

[13] S. Kullback, *Information Theory and Statistics* (John Wiley, New York, 1959).

[14] R. A. Fisher, Philos. Trans. R. Soc. London, Ser. A **222**, 309 (1922).

[15] D. A. Wagenaar, http://www.its.caltech.edu/pinelab/wagenaar/infogeom.pdf

[16] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).

[17] K. Schwarzschild, Sitzungsber. K. Preuss. Akad. Wiss. **1916**, 189; **1916**, 424.

[18] C. Hipp and M. Taksar, Insur. Math. Econ. **26**, 185 (2000).

[19] S. Curilef, F. Pennini, and A. Plastino, Phys. Rev. B **71**, 024420 (2005).

[20] F. Pennini and A. Plastino, Phys. Rev. E **69**, 057101 (2004).