

Article

# Exponential Families with External Parameters

Marco Favretti 

Dipartimento di Matematica Tullio Levi-Civita, Università degli Studi di Padova, 35131 Padova, Italy; favretti@math.unipd.it

**Abstract:** In this paper we introduce a class of statistical models consisting of exponential families depending on additional parameters, called external parameters. The main source for these statistical models resides in the Maximum Entropy framework where we have thermal parameters, corresponding to the natural parameters of an exponential family, and mechanical parameters, here called external parameters. In the first part we study the geometry of these models introducing a fibration of parameter space over external parameters. In the second part we investigate a class of evolution problems driven by a Fokker-Planck equation whose stationary distribution is an exponential family with external parameters. We discuss applications of these statistical models to thermodynamic length and isentropic evolution of thermodynamic systems and to a problem in the dynamic of quantitative traits in genetics.

**Keywords:** exponential family; Fisher metric; Maximum Entropy principle; Ehresmann connection; thermodynamic length; Fokker-Planck equation; generalized Pythagorean theorem



**Citation:** Favretti, M. Exponential Families with External Parameters. *Entropy* **2022**, *24*, 698. <https://doi.org/10.3390/e24050698>

Academic Editor: Antonio M. Scarfone

Received: 25 April 2022

Accepted: 12 May 2022

Published: 14 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

This work is a first attempt to study the geometrical properties and potential applications of a class of statistical models consisting of exponential families depending on additional parameters, called external parameters. The main source for these statistical models comes from the application of E.T. Jaynes Maximum Entropy framework [1] to thermodynamical systems, where we can identify in a natural way thermal parameters (corresponding to natural parameters in an exponential family) and mechanical parameters, here called external parameters. While the construction of equilibrium Statistical Mechanics from the Maximum Entropy principle is a well established domain of science, little attention is paid in the literature to the intrinsic geometrical structure of these statistical models. Given the widespread application of Maximum Entropy principle to disparate fields of science, it is reasonable to assume that a closer scrutiny of these models can pave the way to further applications outside statistical thermodynamics.

Here is the plan of the paper: in Section 2 we recall the definitions of regular statistical model and of exponential family. The main point is that we are dealing with a finite dimensional Riemannian manifold with respect to the Fisher metric. In Section 3 we introduce the exponential families with external parameters, we state the conditions that render them a regular statistical model and we compute the Fisher metric. The additional geometrical structure that we get with these exponential families is a fibration over the space of external parameters  $U$  in the sense that for every fixed  $u \in U$  the fiber is a standard exponential family. The notion of Ehresmann connection on a fibered bundle and of parallel transport is recalled in Section 4. In Section 5 we outline some applications of these parameterized exponential families: we give a formula for the thermodynamic length of a process described by a path in both natural and external parameters and we give conditions for the isentropic evolution of the system. Section 6 is motivated by a model problem in quantitative genetics (briefly recalled in Appendix A) where the dynamics of the system is given by a Fokker-Planck equation with gradient drift and the equilibrium or stationary distribution is an exponential family with external parameters. We recast the dynamic

approximation procedure exposed in [2–4] in the framework of exponential family with external parameters and we give a generalization of the ODE that drives the approximating dynamics. We think that the consideration from the present point of view of the problem exposed in [2–4] may shed light on some still poorly understood aspects of the model.

#### *Exponential Families in Statistical Thermodynamics*

To help locate the contribution of the present paper in the scientific literature we briefly review and compare some of the geometrical approaches to statistical mechanics that are most relevant for our argument. A line of research initiated by the influential papers of Wheinhold [5] and Ruppeiner [6] investigates the Riemannian metric structure on parameter space related to the Boltzmann-Gibbs canonical distribution. This Riemannian metric is the one defined by the Hessian matrix of the free energy  $\psi = \log Z$  (which coincides with the Fisher metric) with respect to the canonical parameters or by its inverse which is the Hessian of the entropy  $S$ , related to  $\psi$  by the Legendre transform. The Levi-Civita connection with respect to this metric allows to define the Riemannian curvature tensor and its sectional and scalar curvature. For a two-dimensional parameter space the divergence of the scalar curvature is a signal of the existence of a phase transition in the underlying physical system. This theory has been applied to Ising and Potts lattice system, to the ideal and Van der Waals gas and to black hole thermodynamics (see e.g., [7–10]). However in dimension greater of two the scalar curvature has a less stringent role and care must be taken in the interpretation of the results.

In this work we also start from the Boltzmann-Gibbs distribution but we stress the different role of natural or thermal parameters  $\theta$ , which occur linearly in (15), and external parameters  $u$  which may enter nonlinearly in the Boltzmann-Gibbs distribution. In particular we are interested in using the external parameters as control parameters on the evolution of the system. The related geometrical framework exposed in Section 4 adopts the connection and curvature associated to the Ehresmann connection on the fibration locally described by  $(\theta, u) \rightarrow u$ , which is fit for describing the isentropic evolution of the system or the dependence of the work control protocol on the global geometric structure i.e., the holonomy of the path of the external control space.

A second line of research relating information geometry and statistical thermodynamics concerns the notion of thermodynamic length (see [11–13]), which is important in the design of optimal driving protocols for the non-equilibrium evolution of (small) thermodynamic systems, see [14,15], both for classical and quantum descriptions. In this work (see Section 5.1) we investigate the notion of thermodynamic length using our geometric framework and we give a formula for for thermodynamic length that highlights the contribution of natural and external (controlled) parameters.

For the sake of completeness we cite the statistical models introduced by J. Naudits (see [16,17]) called generalized exponential families and  $q$ -exponential families by Amari-Ohara, [18]. In these models the exponential function is generalized by introducing the so-called  $q$ -deformed exponential. In practice one considers simultaneously two elements of an exponential family, the second one is called escort distribution. These deformed exponential families are useful for describing Tsallis thermostatics [19] which gives a more accurate description for thermodynamic systems where the extensivity of the classical definition of entropy notion is defied. However this highly debated topic is not relevant for the present work.

This paper is a first attempt to study the exponential families with external parameters using geometrical tools. Even if we were inspired by the Maximum Entropy formalism our result are completely general. In particular we investigated the case where the family (with respect to the the natural and external parameters) is a regular statistical models. This is only a first step in the analysis of these parameterized models; a further step would be in the direction of singular (in opposition to regular) statistical models (see [20]) a domain where there is nowadays an increasing attention in the information geometry community. A drawback of this work is that most of the results are presented in a coordinate-dependent

way and have a local character. We hope to resolve these issues in a subsequent work. Some of the results presented here were introduced in a less refined form in [21].

## 2. Statistical Models and Exponential Families

Before introducing their generalization in Section 3 below, we recall the definitions of regular statistical model and of exponential family (see [22–24]). Let  $(X, \mathcal{B}, dx)$  be a probability space where  $X$  may be a discrete or continuous set. We stipulate that in case of a discrete set the integrals over  $X$  with respect to the measure  $dx$  are substituted by sum symbols. Let

$$\mathcal{P}(X) = \{p : X \rightarrow [0, +\infty), p(x) \geq 0, \int_X p dx = 1\} \subset L_1(X)$$

be the infinite dimensional space of probability densities over  $X$ . Let  $Z \subset \mathbb{R}^d$  be the open set of the parameters,  $f : Z \rightarrow \mathcal{P}(X)$  be a given smooth map and consider the subset of  $\mathcal{P}(X)$

$$\mathcal{S} = \{p = f(z) : z \in Z\} \subset \mathcal{P}(X).$$

To avoid technicalities, we stipulate that the support of  $p$ , i.e., the set where  $p > 0$  is the same for all  $p \in \mathcal{S}$  and that it coincides with  $X$ . We now state the conditions under which  $\mathcal{S}$  is a regular  $d$ -dimensional statistical model (see [22,24,25]).

**Definition 1** (Regular statistical model).  $\mathcal{S}$  is a regular statistical model if the following conditions are satisfied:

1. (injectivity) the map  $f : Z \rightarrow \mathcal{S}, z \mapsto f(z) = p(z)$  is one to one,
2. (regularity) the  $d$  functions defined on  $X$

$$p_i(x; z) = \frac{\partial p}{\partial z_i}(x; z), \quad i = 1, \dots, d$$

are linearly independent as functions on  $X$  for every  $z \in Z$ .

A statistical model which is not regular is called singular (see [20] for a comprehensive discussion on singular models). If condition 1. hold the model is called identifiable, otherwise it is called unidentifiable. If condition 2. fail the main consequence is that the Fisher metric (22) is only positive semidefinite because condition (23) fail. Many statistical model e.g., Boltzmann machines, Bayes networks, hidden Markov models are singular. Note that for a regular statistical model the inverse  $\varphi$  of the map  $f$ ,  $\varphi(p) = z$  defines a global coordinate system for  $\mathcal{S}$ .

To check regularity condition 2. it is convenient to introduce the so called log-likelihood  $l = \ln p$  of  $p$  and the score base

$$l_i(x; z) = \frac{\partial l}{\partial z_i} = \frac{\partial \ln p}{\partial z_i} = \frac{1}{p} p_i(x; z). \tag{1}$$

Since  $l_i$  and  $p_i$  are proportional, the regularity condition 2. holds if and only if the elements of the score base are linearly independent on  $X$ .

### Exponential Family

Fundamental examples of statistical models are the exponential families. Let us introduce  $a$  observable functions  $h : X \rightarrow \mathbb{R}^a, h = (h_1, \dots, h_a)$  and suppose that the  $a + 1$  functions

$$h_1(x), \dots, h_a(x), 1, \tag{2}$$

are linearly independent as functions over  $X$ , where 1 denotes the constant function over  $X$ . Moreover, let  $k = k(x)$  be a function defined on  $X$  and let us introduce the *free energy*  $\psi : \Theta \subset \mathbb{R}^a \rightarrow \mathbb{R}$ ,  $\psi = \psi(\theta)$  as (here  $\theta \cdot h$  denotes the scalar product in  $\mathbb{R}^a$ )

$$\psi(\theta) = \ln \int_X e^{\theta \cdot h(x) + k(x)} dx \tag{3}$$

where the parameter space  $\Theta$  is the subset of  $\mathbb{R}^a$  where  $e^\psi(\theta) < +\infty$ . The  $a$  real numbers  $\theta$  are called *natural* parameters. It is known that the set  $\Theta$  is open and convex in  $\theta$  and that  $\psi$  is a convex function in the  $\theta$  variable (see [23,26]).

The following subset of the infinite dimensional space  $\mathcal{P}(X)$

$$\mathcal{E} = \{ p(x; \theta) = e^{\theta \cdot h(x) - \psi(\theta) + k(x)}, \theta \in \Theta \} \subset \mathcal{P}(X). \tag{4}$$

is called exponential family. We show that  $\mathcal{E}$  is an  $a$ -dimensional regular statistical model. For  $p \in \mathcal{E}$  we have

$$l = \ln p = \theta \cdot h(x) - \psi(\theta) + k(x) \tag{5}$$

therefore the injectivity condition 1. above holds if and only if for all  $\theta, \theta' \in \Theta$

$$(\theta - \theta') \cdot h(x) + 1(\psi(\theta') - \psi(\theta)) = 0 \quad \forall x \in X \quad \Rightarrow \quad \theta = \theta' \tag{6}$$

holds and this is true by the independence condition (2) above. To check regularity condition 2 above, we compute the elements of the score base. They are (here we use the shorthand notation  $\partial_i f = \partial f / \partial z_i$  and  $\langle f \rangle = \int f p dx$ , moreover summation over repeated indices is understood)

$$l_\alpha = h_\alpha - \partial_\alpha \psi = h_\alpha - \langle h_\alpha \rangle, \quad \alpha = 1, \dots, a. \tag{7}$$

The last equality  $\partial_\alpha \psi = \langle h_\alpha \rangle$  holds if we assume that the integrability condition  $\langle h_\alpha \rangle$  is satisfied for every  $\alpha$ . It is not restrictive to assume that  $\langle h_\alpha \rangle = 0$  therefore the regularity condition 2. holds if and only if the  $d$  functions  $h_\alpha$  are linearly independent over  $X$ , which again follows from (2).

One can show (see [22,27]) that every smooth diffeomorphism  $\theta \mapsto m(\theta)$  give an equivalent parameterization of the elements of the exponential family. In this sense  $\mathcal{E}$  has the structure of a smooth manifold, called statistical manifold. Another coordinate system for  $\mathcal{E}$  (we will denote it with  $p = p(x; \eta)$ ) is provided by the so called *expectation* parameters  $\eta \in E \subset \mathbb{R}^a$  defined by (here  $(\partial_\theta \psi)_i = \partial \psi / \partial \theta_i$ )

$$\eta = \partial_\theta \psi(\theta) = \int_X h(x) p(x; \theta) dx.$$

Since  $\psi$  is a convex function, the gradient map  $\theta \mapsto \partial_\theta \psi(\theta)$  is globally invertible with inverse  $\theta = \hat{\theta}(\eta)$  which is also a gradient map  $\hat{\theta}(\eta) = \partial_\eta \varphi(\eta)$ , where

$$\varphi(\eta) = \hat{\theta}(\eta) \cdot \eta - \psi(\hat{\theta}(\eta)) \tag{8}$$

is the Legendre transform of  $\psi$  (see [22]).

### 3. Exponential Families Depending on External Control Parameters

These statistical models are introduced by supposing that the observables  $h$  that defines an exponential family depend on so-called *external* parameters  $u \in U \subset \mathbb{R}^b$ , which are to be distinguished from the natural parameters  $\theta$ . These generalized exponential families arise naturally when one applies the Maximum Entropy formalism to equilibrium Statistical Mechanics, that we briefly recall here (see E.T. Jaynes books [1,28]).

It is well known that when the information consist of the average values of some random variables  $h_\alpha$  describing observables of interest for the system, the maximum

entropy probability densities are exponential families. Indeed, if we introduce the Shannon entropy functional for a probability density  $p \in \mathcal{P}(X)$

$$H(p) = - \int_X p \ln p dx, \tag{9}$$

then the probability density that maximize  $H$  on the set of probability densities that satisfy the constraints  $\langle h \rangle = \int_X h p dx = c \in \mathbb{R}^a$  has the form of an exponential family of the form in (4) with  $k = 0$ . If the observables of interest for the system  $h = h(x, u)$  depend on extra parameters, the exponential family inherits naturally a dependence on the external parameters, see (15) below. Typical examples of external parameters are the magnetic or electric field applied to the system or the length of a polymer chain (see [12,29]). Also, for a quantum system confined in an infinite square well potential, the discrete energy levels  $h_i$  depends on the width  $L$  of the well. Another typical example of a thermodynamic system subject to an external parameter is an ideal gas in a container of variable volume  $V$ ; however in this case the parameter  $V$  affects the state space  $X = X(V)$  and not the observables  $h$  therefore this important system it is not described by a generalized exponential family (see [21] for a discussion of this point).

An important difference between the natural parameters  $\theta$  and the external ones  $u$  is that the former are the Lagrange multipliers associated to the constraints when one solves the constrained extremization problem for  $H$  using Lagrange multipliers method, while the latter are parameters in the problem formulation that can be controlled by an agent external to the system under consideration. This difference is displayed when we consider the variation of  $\langle h \rangle$  for  $p = p(x; \theta, u)$ . If we suppose, as we will always do, that we can exchange the order of integration and differentiation with respect to a parameter, we have

$$d\langle h \rangle = \left( \int_X h \partial_\theta p dx \right) d\theta + \left( \int_X p \partial_u h dx \right) du = dQ + \langle \partial_u h \rangle \tag{10}$$

where  $dQ$  has the meaning of generalized heat exchanged and  $\langle \partial_u h \rangle$  of generalized work exchanged (see [28]). Moreover, while the value of the external parameters  $u$  is controlled and can be varied by an agent external to the system, the value of the natural parameters  $\theta$  can be varied only by putting the system in contact with an heath bath at a prescribed value of the inverse temperature  $\theta$  (see again [28]).

The Kullback-Leibler divergence, also called relative entropy (see [27]) is defined for  $p, q \in \mathcal{P}(X)$  and  $q > 0$  as

$$D(p|q) = \int_X p(x) \log \frac{p(x)}{q(x)} dx.$$

It is well known that the probability density  $\hat{p}$  that minimize  $D$  on the set of probability densities that satisfy the constraints  $\langle h \rangle = \int_X h p dx = c \in \mathbb{R}^a$  has the form of an exponential family as in (4) with  $q = e^K > 0$

$$\hat{p}(x; \theta) = e^{\theta \cdot h(x) - \psi(\theta) + k(x)}. \tag{11}$$

The probability distribution  $\hat{p}$  is the distribution that gives the minimum information gain when one wants to update the current statistical description of the system given by  $q$  using the new available information  $\langle h \rangle = c$ . We will refer in the sequel to this as the minimum Relative Entropy principle. The parameters  $\theta$  of  $\hat{p}(x; \theta)$  in (4) are uniquely determined as  $\theta = \hat{\theta}(c)$  by the constraint conditions

$$\langle h \rangle = \partial_\theta \psi(\theta) = c \tag{12}$$

since the gradient map  $\theta \mapsto \partial_\theta \psi$  is invertible. Note that for  $\theta = 0$  we have  $p(x; 0) = q(x)$  therefore the case  $\hat{\theta}(c) = 0$  corresponds uniquely to the constraint value  $c = \int_X h q dx$  meaning that the constraints do not represent a new piece of information on the system. We will use this fact in the following.

Having exposed the motivations for considering these probability distributions, in the sequel we will investigate the geometrical properties of exponential families with external parameters or controlled exponential families for short.

### 3.1. Exponential Families with External Parameters

Let  $U \subset \mathbb{R}^b$  be the external parameter space and consider the  $a$  observables

$$h_\alpha : X \times U \longrightarrow \mathbb{R}.$$

Let  $k(x)$  be a function on  $X$  and define the free energy  $\psi : Z \subset \mathbb{R}^d, \psi = \psi(\theta, u), d = a + b$

$$\psi(\theta, u) = \ln \int_X e^{\theta \cdot h(x, u) + k(x)} dx \tag{13}$$

where the parameter space  $Z$  is the subset of  $\mathbb{R}^d$  where  $e^\psi < +\infty$ . We suppose that (i)  $Z$  is open and we introduce the map

$$\pi : Z \longrightarrow U, \quad \pi(\theta, u) = u. \tag{14}$$

We consider the following subset of the infinite dimensional space  $\mathcal{P}(X)$

$$\mathcal{F} = \{ p(x; \theta, u) = e^{\theta \cdot h(x, u) - \psi(\theta, u) + k(x)}, (\theta, u) \in Z \} \subset \mathcal{P}(X) \tag{15}$$

and we suppose that

(ii) for every fixed  $u \in \pi(Z)$  the set  $\mathcal{E}(u) \subset \mathcal{F}$

$$\mathcal{E}(u) = \{ p(x; \theta) = e^{\theta \cdot h(x, u) - \psi(\theta, u) + k(x)}, (\theta, u) \in \pi^{-1}(u) \} \subset \mathcal{P}(X) \tag{16}$$

is an exponential family. As a consequence  $\pi^{-1}(u)$  is a convex subset in  $\theta$  and  $h_\alpha(x, u), 1$  are  $a + 1$  functions linearly independent over  $X$ .

A natural question is to ask if the set  $\mathcal{F}$  can be seen as a foliated manifold whose leaves are the statistical manifolds  $\mathcal{E}(u)$ . Note however that if  $\theta = 0$  is allowed (that is  $\int_X e^k dx < +\infty$ ) we have for  $\theta = 0$  in (13)  $\psi(0) = \psi(0, u)$  and  $p(x; 0, u) = e^{k(x) - \psi(0)}$  for every  $u \in \pi(Z)$  therefore

$$e^{k(x) - \psi(0)} \in \mathcal{E}(u) \cap \mathcal{E}(u') \quad \forall u, u' \in \pi(Z).$$

So the statistical manifold leaves are not disjoint.

A second natural question to ask is if  $\mathcal{F}$  can be given the structure of a regular statistical model. To this we need to check conditions 1. and 2. in the Definition 1 above. Concerning injectivity condition 1. for the map  $z \mapsto f(z)$  we have that  $f(0, u) = e^{k - \psi(0)}$  for all  $u \in U$  so injectivity condition 1. may fail for controlled exponential families at  $\theta = 0$ . However, if we recall the statistical mechanics interpretations of controlled exponential families made in Section 3 and in particular in (12), we can consider the point of singularity  $\theta = 0$  outside the domain of application of the statistical model (see however [20] for a discussion of this point). If we assume  $\theta \neq 0$ , due to the possibly nonlinear dependence of  $h(x, u)$  on  $u$ , condition (6) to assess injectivity for a controlled exponential family becomes

$$p_z = \theta \cdot h(x, u) - \psi(\theta, u) = \theta' \cdot h(x, u') - \psi(\theta', u') = p_{z'} \quad \forall x \in X \Rightarrow \theta = \theta', u = u'. \tag{17}$$

Condition (17) seems hard to satisfy even if we assume hypothesis (ii) as the following example shows. Suppose that the observables  $h$  depends linearly on  $u$

$$h_\alpha(x, u) = A_{\alpha k}(x)u_k \tag{18}$$

and (see (ii)) suppose that the  $d + 1$  functions in (18)  $h_\alpha(x, u), 1$  are linearly independent over  $X$  for every fixed  $u$ . Note that the elements of  $\mathcal{F}$  in (15) depend on  $\theta, u$  through the

scalar quantity  $\theta \cdot Au$ . To prove injectivity of the map  $z \mapsto p_z, z = (\theta, u)$  we need to prove that if  $z \neq z'$

$$\theta \cdot A(x)u \neq \theta' \cdot A(x)u'$$

as functions on  $X$ . But this is not true if for example  $\theta' = \lambda\theta$  and  $u' = u/\lambda$  for  $\lambda \neq 0$ . So the model (18) is singular. This should not be a surprise because elements of the family  $\mathcal{F}$  are not characterized by the observables  $A_{\alpha k}(x)$  but by the linear space spanned by the  $A_{\alpha k}(x)$ . Indeed, if we set  $\theta = B\theta'$  and  $u = Cu'$  where  $B$  and  $C$  are nonsingular square matrices, then

$$\theta \cdot Au = \theta^T Au = (B\theta')^T ACu' = \theta'^T B^T ACu' = \theta'^T A'u' = \theta' \cdot A'u'$$

hence the family  $\mathcal{F}$  is equally described by  $A' = B^T AC$  with respect to the parameters  $(\theta', u')$ . Another lesson we can draw from this example is that for an exponential family linearly dependent in the external parameters, the distinction between natural and external ones is lost, as their role can be interchanged.

All that said, we stipulate that

**Definition 2.**  $\mathcal{F}$  in (15) is an exponential family depending on the parameters  $z = (\theta, u)$  if (i) for every fixed  $u$  the set  $\mathcal{E}(u)$  is an exponential family and (ii)  $\mathcal{F}$  is a regular  $d = a + b$  statistical model for a suitable choice of the open parameter set  $Z \subset \mathbb{R}^a \times U$ .

In the case of an exponential family (15) depending on natural and external parameters in addition to  $a$  natural parameters score base vectors

$$l_\alpha = \frac{\partial \ln p}{\partial \theta_\alpha} = h_\alpha - \partial_\alpha \psi = h_\alpha - \langle h_\alpha \rangle \tag{19}$$

we have  $b$  external parameters score base vectors

$$l_k = \frac{\partial \ln p}{\partial u_k} = \theta_\alpha \partial_k h_\alpha - \partial_k \psi = \theta_\alpha (\partial_k h_\alpha - \langle \partial_k h_\alpha \rangle) = \theta_\alpha L_{\alpha k}. \tag{20}$$

Note that  $\langle l_\alpha \rangle = 0$  and  $\langle l_k \rangle = 0$  because  $\langle L_{\alpha k} \rangle = 0$ . Moreover, one can always assume that  $\langle h_\alpha \rangle = 0$  and  $\langle \partial_k h_\alpha \rangle = 0$  therefore the regularity condition 2. above holds if and only if the  $a + b$  functions

$$h_\alpha(x, u), \quad \theta_\alpha \frac{\partial h_\alpha}{\partial u_k}(x, u) \tag{21}$$

are linearly independent over  $X$ .

### 3.2. Fisher Metric for an Exponential Family with External Parameters

Regular statistical models can be endowed with a Riemannian metric defined on their parameter space  $Z$ . This is called Fisher metric [30] and it has the form

$$g_{ij}(z) = \langle l_i l_j \rangle = \int_X \frac{\partial l}{\partial z_i} \frac{\partial l}{\partial z_j} p dx. \tag{22}$$

The Fisher matrix is symmetric and positive definite therefore it defines a Riemannian metric on  $\mathcal{Z}$  (see [24], p. 24). In fact we have

$$g_{ij} v_i v_j = \langle l_i l_j v_i v_j \rangle = \langle (l_i v_i)^2 \rangle = 0 \iff l_i v_i = 0 \iff v_i = 0 \quad \forall i \tag{23}$$

since the score vectors  $l_i$  are linearly independent over  $X$ . Note also (see [24]) that  $g$  is invariant with respect to change of coordinates in the state space  $X$  and covariant (as an order 2 tensor) with respect to change of coordinates in the parameter space  $\mathcal{Z}$ .

The elements of the Fisher matrix (22) relative to an exponential family with external parameters (15) can be detailed as follows: using (19)

$$g_{\alpha\beta} = \langle l_\alpha l_\beta \rangle = \langle (\langle h_\alpha \rangle - h_\alpha)(\langle h_\beta \rangle - h_\beta) \rangle = \text{cov}(h_\alpha, h_\beta); \tag{24}$$

we also have from (20)

$$g_{\alpha k} = \langle l_\alpha l_k \rangle = \langle (\langle h_\alpha \rangle - h_\alpha)\theta_\beta(\langle \partial_k h_\beta \rangle - \partial_k h_\beta) \rangle = \theta_\beta \text{cov}(h_\alpha, \partial_k h_\beta) \tag{25}$$

and

$$g_{km} = \langle l_k l_m \rangle = \langle \theta_\alpha(\langle \partial_k h_\alpha \rangle - \partial_k h_\alpha)\theta_\beta(\langle \partial_m h_\beta \rangle - \partial_m h_\beta) \rangle = \theta_\alpha \theta_\beta \text{cov}(\partial_k h_\alpha, \partial_m h_\beta). \tag{26}$$

It is useful to set

$$\mathcal{A}_{\alpha\beta} = g_{\alpha\beta}, \quad \mathcal{M}_{\alpha k} = g_{\alpha k}, \quad \mathcal{B}_{km} = g_{km}$$

and introduce a block representation of the symmetric  $(a + b)$ -dimensional Fisher matrix  $g$  as

$$g(z) = \begin{pmatrix} \mathcal{A} & \mathcal{M} \\ \mathcal{M}^T & \mathcal{B} \end{pmatrix}. \tag{27}$$

We now give the expression of the Fisher metric coefficients using the free entropy function  $\psi$  in (13), which is also called the moment generating function because its derivative with respect to the  $\theta$  parameters give the different moments of the random variables  $h$ . We thus have the well know relation

$$\partial_\beta \partial_\alpha \psi = \text{cov}(h_\alpha, h_\beta) = g_{\alpha\beta}. \tag{28}$$

By direct computation on (13) we have also

$$\partial_k \partial_\alpha \psi = \theta_\beta \text{cov}(h_\alpha, \partial_k h_\beta) + \langle \partial_k h_\alpha \rangle$$

hence

$$g_{\alpha k} = \partial_k \partial_\alpha \psi - \langle \partial_k h_\alpha \rangle. \tag{29}$$

Moreover we have

$$\partial_k \partial_m \psi = \theta_\alpha \theta_\beta \text{cov}(\partial_k h_\alpha, \partial_m h_\beta) + \theta_\alpha \langle \partial_k \partial_m h_\alpha \rangle \tag{30}$$

hence

$$g_{km} = \partial_k \partial_m \psi - \theta_\alpha \langle \partial_k \partial_m h_\alpha \rangle. \tag{31}$$

We see that, unlike the case of natural parameters  $\theta$ , second order derivatives of the free entropy  $\psi$  with respect to mixed or external parameters do not coincides with the elements of Fisher matrix.

**Example 1.** As a toy model, we introduce the following example of a controlled exponential family. Let  $X = [0, +\infty)$  and  $U = [0, +\infty)$  and consider the two observables where  $x \in X, u \in U$

$$h_1(x) = \ln x, \quad h_2(x, u) = \ln(x + u). \tag{32}$$

For this example we set  $k(x) = -\ln x$ . We check that we have an integrable free energy function

$$\begin{aligned} e^\psi &= \int_X e^{\theta \cdot h + k} dx = \int_0^{+\infty} e^{(\theta_1 - 1) \ln x + \theta_2 \ln(x + u)} dx \\ &= \int_0^{+\infty} x^{\theta_1 - 1} (x + u)^{\theta_2} dx = u^{\theta_1 + \theta_2} \Gamma(\theta_1) \frac{\Gamma(-\theta_2 - \theta_1)}{\Gamma(-\theta_2)} \end{aligned}$$

which is finite if  $\theta_1 > 0, u > 0$  and  $\theta_2 + \theta_1 < 0$ . Here  $\Gamma(z)$  is the Gamma function defined as

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

Note that since  $e^k(x) = 1/x$  is non integrable over  $X$ ,  $\theta_1 = \theta_2 = 0$  is a non feasible value. By inspection  $h_1, h_2, 1$  are linearly independent over  $X$  for every fixed  $u$ , the map

$$(\theta_1, \theta_2, u) \mapsto \theta_1 \ln(x) + \theta_2 \ln(x + u) - \ln(x)$$

is injective. From the likelihood

$$l = \theta_1 \ln(x) + \theta_2 \ln(x + u) - \ln(x) - \psi(\theta, u),$$

the elements of the score base are

$$l_1 = \ln(x) - \partial_1 \psi, \quad l_2 = \ln(x + u) - \partial_2 \psi, \quad l_u = \theta_2 \left( \frac{1}{x + u} - \partial_u \psi \right)$$

which are linearly independent over  $X$ . So the statistical model defined by (32) is a 2 + 1 dimensional controlled exponential family. Note that the probability density

$$p(x; \theta_1, \theta_2, u) = e^{\theta_1 h_1 + \theta_2 h_2 - \psi + k} = \frac{x^{\theta_1 - 1} (x + u)^{\theta_2}}{Z(\theta_1, \theta_2, u)}$$

is known as a (possible formulation of a) compound Gamma distribution; moreover, for  $u = 1$ , this is the Beta distribution of second kind [31].

We now compute the Fisher matrix elements for this example. Let us introduce the Polygamma function  $\Phi_m$  for  $m \in \mathbb{N}$

$$\Phi_m(z) = \frac{d^m}{dz^m} \Phi_0(z), \quad \Phi_0(z) = \frac{d}{dz} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}.$$

We have

$$\psi(\theta_1, \theta_2, u) = (\theta_1 + \theta_2) \ln u + \ln \frac{\Gamma(\theta - 1) \Gamma(-\theta_1 - \theta_2)}{\Gamma(-\theta_2)}$$

and from relation (28) above we have

$$\begin{aligned} g_{11} &= \partial_1 \partial_1 \psi = \Phi_1(\theta_1) + \Phi_1(-\theta_1 - \theta_2) \\ g_{12} &= \partial_1 \partial_2 \psi = \Phi_1(-\theta_1 - \theta_2) \\ g_{22} &= \partial_2 \partial_2 \psi = -\Phi_1(-\theta_2) + \Phi_1(-\theta_1 - \theta_2) \end{aligned}$$

so the  $\mathcal{A}$  block of  $g$  depends only on  $\theta$ . Moreover from (29) and (31) we have

$$\begin{aligned} g_{1u} &= \partial_u \partial_1 \psi - \langle \partial_u h_1 \rangle = \partial_u \partial_1 \psi = \frac{1}{u} \\ g_{2u} &= \partial_u \partial_2 \psi - \langle \partial_u h_2 \rangle = \frac{1}{u} + \frac{\theta_1 + \theta_2}{\theta_2 u} = -\frac{\theta_1}{u \theta_2} \\ g_{uu} &= \partial_u \partial_u \psi - \theta_\alpha \langle \partial_u \partial_u h_\alpha \rangle = \frac{(\theta_1 + \theta_2) \theta_1}{u^2 (\theta_2 - 1)}. \end{aligned}$$

#### 4. A Synopsis of Ehresmann Connections

On a smooth fibration  $\pi : M \rightarrow N$ , where  $M, N$  are smooth manifolds, with  $\dim M = m, \dim N = n$ , the set  $VM = \ker T\pi$  of the vectors that project onto the null space of  $TN$  is an integrable subbundle of  $TM$  called the vertical bundle.

An Ehresmann connection (see e.g., [32]) on  $\pi : M \rightarrow N$  is the assignment of a distribution  $HM$  transversal to  $VM$ , so that  $HM \oplus VM = TM$ . The elements of  $HM$  are the

horizontal vectors; since  $T\pi$  restricted to  $HM$  is an isomorphism, it has a fiberwise defined inverse, the horizontal lift:  $\text{hor} : T_{\pi(z)}N \rightarrow T_zM$ ,  $\text{hor}(X) \in H_zM$ . Let  $X = X^h + X^v$  be the splitting of a vector in  $T_zM$  into its horizontal and vertical component. The projection on  $VM$  with respect to the horizontal subspace defines the vector-valued connection one-form

$$\omega : TM \rightarrow VM, \quad \omega(z)(X) = X^v, \tag{33}$$

whose kernel is the horizontal distribution. The assignment of an horizontal distribution, of an horizontal lift operator or of a connection one-form are equivalent ways to define a connection on  $\pi : M \rightarrow N$ . The curvature of the connection is the  $VM$ -valued two-form defined as

$$\Omega(X, Y) = -\omega([X^h, Y^h]) \tag{34}$$

which shows that the curvature measures the failure of the horizontal distribution to be integrable. Moreover, the curvature relates the Lie brackets of vector fields  $X, Y$  on the base manifold  $N$  with the Lie bracket of their horizontal lifts through the formula

$$\Omega(\text{hor}X, \text{hor}Y) = [\text{hor}X, \text{hor}Y]_M - \text{hor}[X, Y]_N. \tag{35}$$

Again, we find that if the curvature is vanishing the horizontal distribution, spanned by vectors of the type  $\text{hor}X$ , is involutive hence integrable. Next we give the local expressions of a connection in a fibered chart. Let  $z = (x, y)$  be a fibered chart on  $U \subset M$ ,  $\pi(x, y) = y$ . Then the vertical space is,  $\alpha = 1, \dots, a = \dim M - \dim N$ ,

$$V_zU = \ker T_z\pi = \text{span}\left\{\frac{\partial}{\partial x^\alpha}\right\} \tag{36}$$

and the connection one-form  $\omega$  is:

$$\omega = \omega^\alpha \otimes \frac{\partial}{\partial x^\alpha}, \quad \omega^\alpha = dx^\alpha + A_l^\alpha(z)dy^l. \tag{37}$$

The  $A_l^\alpha(z)$  are the connection's coefficients. The horizontal vectors have the coordinate expression

$$X \in HM \Leftrightarrow \omega(X) = 0, \Leftrightarrow X^l = X^l\left(\frac{\partial}{\partial y^l} - A_l^\alpha \frac{\partial}{\partial x^\alpha}\right)$$

while the horizontal lift of a base vector  $U = U^l \frac{\partial}{\partial y^l} \in T_{\pi(z)}N$  has the form

$$(\text{hor } U)^l(z) = U^l\left(\frac{\partial}{\partial y^l} - A_l^\alpha \frac{\partial}{\partial x^\alpha}\right). \tag{38}$$

We now specialize the above relations to the important case where the horizontal distribution  $H_zM$  is defined to be the  $g$ -orthogonal of  $V_zM$  with respect to a Riemannian metric  $g$  on  $M$ . Referring to a block representation of the metric  $g$  in the coordinates  $(x, y)$  like the above one (27) for  $(\theta, u)$  we ask that every  $X^h \in H_zM$ ,  $X^h = (-A_l^\alpha U^l, U^l)$  be orthogonal to all  $X^v \in V_zM$ ,  $X^v = (W, 0)$ . As a consequence

$$g(X^v, X^h) = W \cdot (-AAU + MU) = 0 \quad \forall W \Leftrightarrow A = \mathcal{A}^{-1}\mathcal{M}. \tag{39}$$

The connection one-form (37) becomes from (39)

$$\omega^\alpha = dx^\alpha + A_l^\alpha(z)dy^l, \quad \text{where } A_l^\alpha = (\mathcal{A}^{-1}\mathcal{M})_l^\alpha \tag{40}$$

and it is called mechanical connection in the control theory for mechanical systems, where  $g$  is the kinetic energy of a mechanical system. In the orthogonal splitting case the metric  $g$  has the simpler form by (39)

$$g(X, Y) = g(X^v + X^h, Y^v + Y^h) = g(X^v, Y^v) + g(X^h, Y^h)$$

Since  $X^v = (\omega(X), 0)$  and using again the block representation (27) of  $g$  we have

$$g(X^v, Y^v) = \mathcal{A}_{\alpha\beta}(z)\omega^\alpha(X)\omega^\beta(Y)$$

and

$$g(X^h, Y^h) = (-AU, U)^T g(-AV, V) = U^T \mathcal{K} V$$

where  $\mathcal{K} = \mathcal{B} - \mathcal{M}^T \mathcal{A}^{-1} \mathcal{M} = \mathcal{K}^T$  hence

$$g(z) dz \otimes dz = \mathcal{A}(z)\omega(\cdot) \otimes \omega(\cdot) + \mathcal{K}(z) dy \otimes dy. \tag{41}$$

and

$$g = \begin{pmatrix} \mathcal{A} & 0 \\ 0 & \mathcal{K} \end{pmatrix}. \tag{42}$$

*Parallel Transport Equation*

Let  $\gamma : [0, T] \rightarrow N$  be a smooth path in the base manifold and let  $z_0 \in \pi^{-1}(\gamma(0))$ . The parallel transport equation is the following ODE for the horizontal lift vector field

$$\dot{z} = \frac{dz}{dt} = \text{hor}(\dot{\gamma}), \quad z(0) = z_0 \tag{43}$$

with local expression

$$\dot{x}_\alpha = -A_\alpha^l(x, \gamma)\dot{\gamma}_l, \quad \dot{y}_l = \dot{\gamma}_l. \tag{44}$$

The connection is called complete if the parallel transport equation has a solution defined on the whole  $[0, T]$ . If in (41) we have  $\mathcal{K} = \mathcal{K}(y)$  then the metric  $g$  is called bundle-like metric. The main geometric consequence is that if we introduce the Riemannian manifold  $(N, \mathcal{K})$  then the horizontal lift is an isometry and the solution  $z(t)$  of the parallel transport equation is a curve that projects over  $\gamma$  of the same length.

**5. Some Applications of Exponential Families with External Parameters**

In this Section we apply the geometric framework of the previous Section 4 to the fibration  $\pi : Z \rightarrow U, \pi(z, u) = u$  introduced in (14). We can also consider the inverse  $\varphi$  of the map  $z \mapsto f(z) = p_z$  and introduce the fibration

$$\mathcal{F} \xrightarrow{\varphi} Z \xrightarrow{\pi} U, \quad \tilde{\pi} = \pi \circ \varphi.$$

Since  $\tilde{\pi}^{-1}(u) = \mathcal{E}(u)$ , fibers of  $\tilde{\pi}$  are exponential families for every fixed value of the external parameters. One can show that the orthogonal splitting of  $TZ$  induces and orthogonal splitting of  $T\mathcal{F}$  with respect to the Fisher metric (see [21]).

*5.1. Thermodynamic Length*

Let  $t \mapsto z(t) = (\theta(t), u(t)) \in Z, t \in [0, T]$  be a path in parameter space. Define the time-dependent relative entropy along the path as  $D(t) = D(p(z(t))|p(z(0)))$  and compute the Taylor expansion of  $D(t)$  at  $t = 0$ . A direct computation shows that  $D(0) = 0, D'(0) = 0$  hence

$$D(dt) = D(0) + D'(0)dt + \frac{1}{2}D''(0)dt^2 + \mathcal{O}(dt^3) = \frac{1}{2}D''(0)dt^2 + \mathcal{O}(dt^3) = \frac{1}{2}\| \dot{z} \|_g^2 + \mathcal{O}(dt^3)$$

where  $\|\dot{z}\|_g^2$  is the scalar product with respect to the Fisher metric in  $(Z, g)$ . It holds that

$$\begin{aligned} \|\dot{z}\|_g^2 &= \dot{z} \cdot g(z)\dot{z} \langle l_i l_j \rangle \dot{z}_i \dot{z}_j = \langle l_i l_j \dot{z}_i \dot{z}_j \rangle \\ &= \int_X p \partial_i \ln p \partial_j \ln p \dot{z}_i \dot{z}_j dx = \int_X p \left( \frac{1}{p} \partial_i p \frac{1}{p} \partial_j p \right) \dot{z}_i \dot{z}_j dx \\ &= \int_X \frac{1}{p} \left( \frac{dp}{dt}(z(t)) \right)^2 dx. \end{aligned}$$

The quantity  $2D(dt) = \|\dot{z}\|_g^2$  can be related to the entropy change rate  $d\sigma/dt$  of the heat bath and to the total system entropy production rate  $dF/dt$  in a non quasi-static evolution of the system by the formula (see [14])

$$\|\dot{z}\|_g^2 = \left\langle \frac{d\sigma}{dt} \right\rangle - \left\langle \frac{dF}{dt} \right\rangle \geq 0.$$

Therefore  $\|\dot{z}\|_g^2$  is a measure of the system entropy production rate  $d\sigma_{sys}/dt$  in a non-quasi static evolution of the system. When integrated along the finite time evolution protocol  $z(t)$ , the quantity

$$2C = \int_0^T \|\dot{z}\|_g^2 dt = \int_0^T \left[ \left\langle \frac{d\sigma}{dt} \right\rangle - \left\langle \frac{dF}{dt} \right\rangle \right] dt$$

is called action of the path and can be interpreted as the thermodynamic cost (loss in the entropy transfer due to the system entropy production) associated to the protocol therefore it is a measure of the dissipated (non available) work. The quantity (see [11,12,15])

$$\mathcal{L}(z) = \int_0^T \|\dot{z}\|_g(t) dt$$

is called the thermodynamic length of the path  $z(\cdot)$ . By the Cauchy-Schwartz inequality one obtains the inequality (see [14])

$$T2C \geq (\mathcal{L}(z))^2$$

showing that the thermodynamic length (TL) gives a lower bound on the dissipated work in a non quasi-static evolution of the system [11,15]. The above relation is used when studying the controlled evolution of classical and quantum small thermodynamic systems, e.g., molecular motors (see [15]).

Using the representation (41) of the scalar product with respect to the Fisher metric  $g$  we have the interesting formula for the TL of a controlled exponential family

$$\mathcal{L}(z) = \int_0^T \|\dot{z}\|_g(t) dt = \int_0^T \left( \mathcal{A}_{\alpha\beta}(z) \omega^\alpha(\dot{z}) \omega^\beta(\dot{z}) + \mathcal{K}_{lm}(z) \dot{u}^l \dot{u}^m \right)^{\frac{1}{2}} dt.$$

In particular, if the path  $z$  is the horizontal lift of a path  $u$  in the external parameter space then  $\dot{z} = \text{hor}(\dot{u})$  and  $\omega(\dot{z}) = 0$ . If moreover the metric  $g$  is bundle-like with respect to the fibration  $\pi$  we have  $\mathcal{K}(z) = \mathcal{K}(\pi(z))$  and the thermodynamic length can be expressed as

$$\mathcal{L}(z) = \mathcal{L}(u) = \int_0^T \sqrt{\mathcal{K}(u) \dot{u} \cdot \dot{u}} dt$$

showing that TL depends solely on the external parameters evolution  $u = u(t)$ .

### 5.2. Isentropic Evolution Driven by External Parameters

We have recalled in Section 3 that the elements of a controlled exponential family  $\mathcal{F}$  where  $q = e^K \in \mathcal{P}$  are the solution of the constrained minimization problem for the relative entropy  $D(p|q)$  of the form (11)

$$\hat{p}(x; c, u) = e^{\hat{\theta} \cdot h(x, u) - \psi(\hat{\theta}, u) + k(x)}$$

where  $\hat{\theta} = \hat{\theta}(c, u)$  is uniquely determined by inverting the gradient map  $\partial_\theta \psi(\theta, u) = c$ . We have that

$$D(c, u) = D(\hat{p}|q) = \int_X \hat{p} \ln \frac{\hat{p}}{q} dx = \hat{\theta} \cdot \langle h \rangle - \psi = \hat{\theta} \cdot \partial_\theta \psi - \psi = -S(c, u)$$

where  $S(c, u) = \psi - \theta \cdot \partial_\theta \psi$  is the entropy of the statistical system when the information on the system is described by the constraint  $\langle h \rangle = c$ . In the following we consider  $D(c, u)$  as a function of  $(\theta, u)$  knowing that  $\theta$  is in a one-to-one correspondence with  $c$ .

Let us compute the differential of  $D(\theta, u)$  corresponding to a infinitesimal variation of the parameters  $z = (\theta, u)$ . We have

$$dD(\theta, u) = D(\hat{p}|q) = \partial_\theta D d\theta + \partial_u D du \tag{45}$$

More in detail, using (28) we obtain

$$\partial_\beta D = \partial_\beta (\theta_\alpha \partial_\alpha \psi - \psi) = \partial_\beta \psi + \theta_\alpha \partial_\beta \partial_\alpha \psi - \partial_\beta \psi = \theta_\alpha g_{\alpha\beta}$$

and using (29) we obtain

$$\partial_k D = \partial_k (\theta_\alpha \partial_\alpha \psi - \psi) = \theta_\alpha \partial_k \partial_\alpha \psi - \partial_k \psi = \theta_\alpha (g_{\alpha k} + \langle \partial_k h_\alpha \rangle) - \theta_\alpha \langle \partial_k h_\alpha \rangle = \theta_\alpha g_{\alpha k}$$

so collecting the results and using (40) we have

$$dD(z) = \theta_\alpha (g_{\alpha\beta} d\theta^\beta + g_{\alpha k} du_k) = \theta_\alpha (\mathcal{A}_{\alpha\beta} d\theta^\beta + \mathcal{M}_{\alpha k} du_k) = \theta_\alpha \mathcal{A}_{\alpha\beta} \omega^\beta$$

and the following proposition holds

**Proposition 1.** (1) *The variation of entropy for an infinitesimal change in the parameters  $z = (\theta, u)$  can be expressed using the g-orthogonal Ehresmann connection  $\omega$  on  $\pi : Z \rightarrow U$*

$$-dS = dD = \theta_\alpha \mathcal{A}_{\alpha\beta} \omega^\beta = \theta_\alpha \partial_\alpha \partial_\beta \psi \omega^\beta \tag{46}$$

(2) *the change in entropy for the system along a given path  $z = z(t)$ ,  $t \in [0, T]$  in parameter space is given by*

$$\Delta S(z) = \int_0^T dS(\dot{z}) dt = - \int_0^T \theta_\alpha(t) \mathcal{A}_{\alpha\beta}(z(t)) \omega^\beta(\dot{z}) dt$$

(3) *since  $\omega(\dot{z}) = 0$  for an horizontal path, the horizontal lift  $\dot{z} = \text{hor}\dot{\gamma}$  of a path  $\gamma$  in the external parameter space  $U$  gives an isentropic ( $\Delta S = 0$ ) evolution of the system.*

Note that the horizontal lift do not represent all the possible isentropic evolution of the system. These are characterized by the weaker (with respect to  $\omega(\dot{z}) = 0$ ) condition  $\theta \cdot \mathcal{A}\omega(\dot{z}) = 0$ . Let us investigate this condition using the general relation (10) that we can now write as

$$d\langle h_\alpha \rangle = d(\partial_\alpha \psi) = \partial_\alpha \partial_\beta \psi d\theta^\beta + \partial_k \partial_\alpha \psi du_k = \mathcal{A}_{\alpha\beta} \omega^\beta + \langle \partial_k h_\alpha \rangle du_k = dQ_\alpha + dW_\alpha.$$

If we want to gain insight into the above relation using a thermodynamic analogy, then  $h_\alpha$  is the  $\alpha$ -type energy,  $dQ_\alpha = \mathcal{A}_{\alpha\beta} \omega^\beta$  is the  $\alpha$ -type heat exchanged and  $dW_\alpha$  the  $\alpha$ -type

work exchanged. If we interpret the natural parameters  $\theta$  as the  $\alpha$ -type inverse temperature  $\theta_\alpha = 1/T_\alpha$  then (46) display as

$$-dS = \theta_\alpha \mathcal{A}_{\alpha\beta} \omega^\beta = \sum_\alpha \frac{dQ_\alpha}{T_\alpha}.$$

Therefore an horizontal path corresponds to the condition  $dQ_\alpha = 0$  for all  $\alpha$  and certainly it represents an isentropic evolution of the system, but we can have an isentropic evolution even if  $dQ_\alpha \neq 0$  if the heat fluxes divided by their temperatures have a zero sum. As a final remark, note that in the exponential family we have the scalar product  $\theta \cdot h$  hence the inverse temperature vector  $\theta \in \mathbb{R}^a$  should be seen as an element of the dual space of the  $h \in \mathbb{R}^a$  vector space and not as a point in a local coordinate chart. See [33] on this point.

### 6. Information Geometry of Gradient Systems

In this Section we consider a class of evolution problems described by a Fokker-Planck type equation (FPE) on a regular connected domain  $X \subset \mathbb{R}^n$  which is open and bounded. We write FPE as in [34] ( $i, j = 1, \dots, n$ , repeated indices are summed)

$$\partial_t p = -\partial_i(D_i p) + \partial_i \partial_j(D_{ij} p) = -\nabla \cdot S \tag{47}$$

where  $D_i(x)$  is the drift field,  $D_{ij}(x)$  is the symmetric diffusion matrix,  $\nabla \cdot$  denotes divergence and  $S$  is the probability current

$$S_i(p) = D_i p - \partial_j(D_{ij} p), \quad i = 1, \dots, n.$$

To ensure that a solution  $p(x, t)$  is normalized to one for all  $t \geq 0$  we need to ask

$$0 = \frac{d}{dt} \int_X p dx = \int_X \partial_t p dx = - \int_X \nabla \cdot S dx = - \int_{\partial X} S \cdot \nu d\sigma$$

that is  $S \cdot \nu = 0$  on  $\partial X$ . We restrict to the case that the diffusion matrix is diagonal  $D_{ij} = d_i(x)\delta_{ij}$  and positive definite ( $d_i > 0$ ) and therefore we rewrite  $S$  as

$$S_i(p) = D_i p - \partial_i(d_i p) = p[D_i - d_i \partial_i \ln(p d_i)]. \tag{48}$$

Moreover we suppose that the drift field is of the form

$$D_i(x) = d_i(x)\partial_i \phi(x) \tag{49}$$

where  $\phi$  is a function defined on  $X$ . A stationary solution  $p_\infty$  of FPE is obtained if we have  $S_i(p_\infty) = 0$  for all  $i$  that is from (48)

$$d_i[\partial_i(\phi - \ln(p_\infty d_i))] = 0. \tag{50}$$

One can show ([34], Chapter 6) that in this setting the stationary solution to FPE (47) is unique. We can rewrite the FPE using  $p_\infty$  from (48) and (49) as follows

$$\begin{aligned} \partial_t p &= -\partial_i S_i = \partial_i [p d_i \partial_i (-\phi + \ln(p d_i))] \\ &= \partial_i [p d_i \partial_i (-\ln(p_\infty d_i) + \ln(p d_i))] = \partial_i [p d_i \partial_i (\ln \frac{p}{p_\infty})] \end{aligned}$$

or in compact notation as

$$\partial_t p = \nabla \cdot \left( p D \nabla \ln \frac{p}{p_\infty} \right) \tag{51}$$

where  $D$  is the diagonal diffusion matrix. The trend to the equilibrium can be studied by computing the “distance in entropy” between a solution  $p$  of (47) and  $p_\infty$ . Setting  $\lambda(p) = \ln(p/p_\infty)$  we have from (51)

$$\frac{d}{dt}D(p|p_\infty) = \frac{d}{dt} \int_X p \ln \frac{p}{p_\infty} dx = \int_X (\partial_t p) \lambda(p) dx = \int_X \nabla \cdot (pD\nabla\lambda) \lambda dx$$

and using the relation  $\lambda \nabla \cdot X = \nabla \cdot (\lambda X) - X \cdot \nabla \lambda$  where  $X = pD\nabla\lambda$  we get

$$\begin{aligned} \frac{d}{dt}D(p|p_\infty) &= \int_X [\nabla \cdot (\lambda pD\nabla\lambda) - pD\nabla\lambda \cdot \nabla\lambda] dx \\ &= \int_{\partial X} \lambda pD\nabla\lambda \cdot \nu d\sigma - \int_X pD\nabla\lambda \cdot \nabla\lambda dx \\ &= - \int_X pD\nabla \ln \frac{p}{p_\infty} \cdot \nabla \ln \frac{p}{p_\infty} dx < 0 \end{aligned} \tag{52}$$

because  $S \cdot \nu = pD\nabla\lambda \cdot \nu = 0$  on  $\partial X$ . So the distance in entropy tends to zero independently of the initial conditions. One can show that for the FPE we have (Csiszar-Kullback-Pinsker inequality, see [35], Chapter 9)

$$D(p|p_\infty) \geq \frac{1}{2} \|p - p_\infty\|_{L_1}.$$

If we have a constant diffusion matrix  $D_{ij} = d\delta_{ij}$ ,  $d > 0$ , the above inequality (52) can be rewritten as

$$\frac{d}{dt}D(p|p_\infty) = -d \int_X p \left(\nabla \ln \frac{p}{p_\infty}\right)^2 dx = -d \mathcal{R}(p|p_\infty) \tag{53}$$

where  $\mathcal{R}(p|p_\infty)$  is called relative Fisher information (see [35], Chapter 9 or [36]).

A probability density  $p_\infty$  satisfies a logarithmic Sobolev inequality (LSI) with positive constant  $\sigma > 0$  if

$$2\sigma D(p|p_\infty) \leq \mathcal{R}(p|p_\infty) \quad \forall p \in \mathcal{P}(X).$$

If  $p_\infty$  satisfies a Logarithmic Sobolev inequality we can prove the exponential speed of convergence to equilibrium in entropy; indeed we have  $-\mathcal{R}(p|p_\infty) \leq 2\sigma D(p|p_\infty)$  and by substitution in (53) we obtain

$$\frac{d}{dt}D(p|p_\infty) \leq -2d\sigma D(p|p_\infty) \Rightarrow D(p|p_\infty) \leq D(p_0|p_\infty)e^{-2d\sigma t}.$$

A sufficient condition for LSI is the following one (see [35], Chapter 9):

(LSI condition) Let  $V$  be a  $C^2$  function on  $X$  with  $\int_X e^{-V} dx = 1$  and  $Hess(V) \geq \sigma \mathbb{I}$  for some  $\sigma > 0$ . Then  $e^{-V}$  satisfies LSI with positive constant  $\sigma$ .

### 6.1. A Dynamic Approximation Problem

This section is motivated by a problem in quantitative genetics which has been dealt with in a series of papers [2–4]. See also Appendix A for a brief account. Here we introduce a slightly simplified version of the original model problem which has the advantage of a greater generality. Let us consider FPE (51) with a gradient drift field of the form (49) with  $x = (x_1, \dots, x_n)$  and

$$d_i(x) = d(x_i), \quad \phi(x) = \theta \cdot h(x), \quad D_{ij} = d(x_i)\delta_{ij}, \quad D_i(x) = d(x_i)\partial_i(\theta \cdot h(x)). \tag{54}$$

In this case one can prove that the stationary solution satisfying (50) of FPE (51) has the form of an exponential family of the form

$$p_\infty(x) = p_\theta = e^{\theta \cdot h(x) - \psi + k(x)} \quad \text{where} \quad k(x) = - \sum_i \ln d(x_i). \tag{55}$$

We are free to set the value of the natural parameters and we set  $\theta = \theta_1$  where  $\theta_1$  is a feasible value.

The explicit solution of FPE (51) is difficult to study and one could be content with the study of the time evolution of the average values of the observables  $h_\alpha$  that is the functions

$$t \mapsto \langle h_\alpha \rangle_{p(t)} = \int_X h_\alpha(x) p(x, t) dx$$

along the unknown solution of FPE. With this aim, it is natural to consider the following:

(Approximation problem): to find the time evolution of the natural parameters  $\theta = \theta(t)$  such that the density

$$p_\theta(x, t) = e^{\theta(t) \cdot h(x) - \psi(\theta(t)) - k(x)} \tag{56}$$

has the same average values of the unknown solution of FPE i.e.,

$$\langle h_\alpha \rangle_{p(t)} = \langle h_\alpha \rangle_{p_{\theta(t)}}. \tag{57}$$

This strategy (called Dynamic Maximum Entropy method in [3]) seems reasonable because the exponential density (55) is the maximum entropy distribution which satisfy the constraints of the form  $\langle h \rangle = c$  therefore it contains exactly the required amount of information needed to satisfy the average values constraint. In the following we will investigate the interplay between the following three densities: (1) the unknown solution  $p$  of FPE, (2) the approximating exponential density  $p_\theta$  and (3) the exponential equilibrium density  $p_\infty = p_{\theta_1}$ .

Triangular Relation

To start with note that for (56) (dropping the explicit time dependence in  $\theta$  and  $p$ )

$$(b) \quad D(p|p_\theta) = \int_X p \ln \frac{p}{p_\theta} dx = -H(p) - \theta \cdot \langle h \rangle_p + \psi - \langle k \rangle_p \tag{58}$$

therefore the condition (57) can be rewritten as

$$\partial_\theta D(p|p_\theta) = -\langle h \rangle_p + \partial_\theta \psi(\theta) = \langle h \rangle_\theta - \langle h \rangle_p = 0. \tag{59}$$

Note that the equation  $\partial_\theta \psi(\theta) = \langle h \rangle_p$  has a unique solution  $\theta = \theta(t)$  for all  $t \geq 0$ .

Next, let us compute the distance in entropy between the solution of FPE and its stationary solution (55) with  $\theta = \theta_1$

$$(a) \quad D(p|p_{\theta_1}) = -H(p) - \theta_1 \cdot \langle h \rangle_p + \psi(\theta_1) - \langle k \rangle_p \tag{60}$$

and the distance between  $p$  and  $p_\theta$  (here  $\theta = \theta(t)$  is the value of the approximating solution satisfying (59))

$$(c) \quad D(p_\theta|p_{\theta_1}) = (\theta - \theta_1) \cdot \langle h \rangle_\theta + \psi(\theta_1) - \psi(\theta) = (\theta - \theta_1) \cdot \partial_\theta \psi(\theta) + \psi(\theta_1) - \psi(\theta) \tag{61}$$

which coincides with the Bregman divergence (see [27]) of the convex function  $\psi$

$$\mathcal{D}_\psi(\theta_1, \theta) = \psi(\theta_1) - \psi(\theta) - (\theta_1 - \theta) \cdot \partial_\theta \psi(\theta) = D(p_\theta|p_{\theta_1}).$$

Collecting the above results (58) (60) (61) and summing the right hand sides we obtain the triangular relation (see [27], Theorem 1.2 or [22], Theorem 3.7)

$$D(p|p_\theta) + D(p_\theta|p_{\theta_1}) = D(p|p_{\theta_1}) + (\theta - \theta_1) \cdot (\langle h \rangle_\theta - \langle h \rangle_p). \tag{62}$$

It follows that

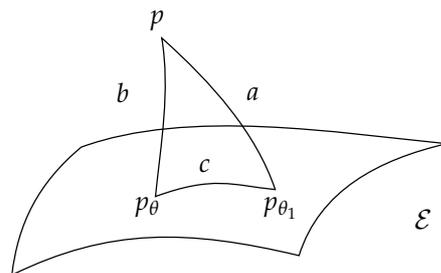
**Proposition 2.** The function  $\theta(t)$  satisfies condition (57) of the Approximation problem if and only if the following relation (called generalized Pythagorean theorem in [27]) holds (see Figure 1)

$$b + c = D(p|p_\theta) + D(p_\theta|p_{\theta_1}) = D(p|p_{\theta_1}) = a \tag{63}$$

meaning that  $p_{\theta(t)}$  is the geodesic projection of  $p$  on the exponential family (flat submanifold)  $\mathcal{E}$  satisfying to

$$D(p|p_\theta) = \min\{D(p|p_{\theta'}) : p_{\theta'} \in \mathcal{E}\}$$

that is  $p_{\theta(t)}$  is the best approximation of  $p$  on  $\mathcal{E}$  with respect to the information gain.



**Figure 1.** Triangular relation between  $a = D(p|p_{\theta_1})$ ,  $b = D(p|p_\theta)$  and  $c = D(p_\theta|p_{\theta_1})$ .  $\mathcal{E}$  is the exponential family submanifold.

Note that this relation is exact and does not need the hypothesis that  $\theta$  be close to  $\theta_1$ . Relation (63) characterizes  $\theta(t)$  from a geometrical point of view.

We now take the time derivative of (63) with a double aim: on the one hand to find a differential relation (ODE) for  $\theta$  and on the other hand to find an upper bound for the distance in entropy  $b = D(p|p_\theta)$  knowing that  $a = D(p|p_{\theta_1})$  tends to 0, possibly with exponential speed. Note that taking the time derivative of the relation (63)  $b + c = a$  is equivalent to taking the time derivative of the relation (59), since the two are equivalent conditions on  $\theta(t)$ . We have from (52)

$$\dot{a} = \frac{d}{dt}D(p|p_{\theta_1}) = - \int_X p D \nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla \ln \frac{p}{p_{\theta_1}} dx < 0. \tag{64}$$

Moreover

$$\dot{b} = \frac{d}{dt}D(p|p_\theta) = \frac{d}{dt} \int_X p \ln \frac{p}{p_\theta} dx = \int_X [\dot{p} \ln \frac{p}{p_\theta} + \dot{p} - \frac{p}{p_\theta} \dot{p}_\theta] dx = \int_X [\dot{p} \ln \frac{p}{p_\theta} - \frac{p}{p_\theta} \dot{p}_\theta] dx.$$

Since

$$\frac{\dot{p}_\theta}{p_\theta} = \frac{d}{dt} \ln p_\theta = \dot{\theta} \cdot h - \partial_\theta \psi \cdot \dot{\theta} = \dot{\theta} \cdot (h - \langle h \rangle_\theta)$$

we have from (59)

$$\int_X p \frac{\dot{p}_\theta}{p_\theta} dx = \int_X \dot{\theta} (h - \langle h \rangle_\theta) dx = 0$$

and recalling that  $p$  is a solution of FPE (51) we obtain

$$\dot{b} = \frac{d}{dt}D(p|p_\theta) = \int_X \nabla \cdot \left( pD\nabla \ln \frac{p}{p_{\theta_1}} \right) \ln \frac{p}{p_\theta} dx = - \int_X pD\nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla \ln \frac{p}{p_\theta} dx \quad (65)$$

since we can get rid of the boundary term as done above.

The side  $c = D(p_\theta|p_{\theta_1})$  does not contain the solution  $p$  of FPE therefore its time derivative can be computed as, see (45)

$$\begin{aligned} \dot{c} &= \frac{d}{dt}D(p_\theta|p_{\theta_1}) = \int_X \dot{p}_\theta \ln \frac{p_\theta}{p_{\theta_1}} dx = \int_X p_\theta \frac{d}{dt}(\theta \cdot h - \psi(\theta) + k) \ln \frac{p_\theta}{p_{\theta_1}} dx \\ &= \int_X p_\theta \dot{\theta} (h - \langle h \rangle_\theta) [(\theta - \theta_1) \cdot h + \psi(\theta_1) - \psi(\theta)] dx \\ &= \langle h_\alpha h_\beta - \langle h_\alpha \rangle \langle h_\beta \rangle \rangle \dot{\theta}_\alpha (\theta - \theta_1)_\beta = \text{cov}_\theta(h_\alpha, h_\beta) \dot{\theta}_\alpha (\theta - \theta_1)_\beta \\ &= g_{\alpha\beta}(\theta) \dot{\theta}_\alpha (\theta - \theta_1)_\beta. \end{aligned} \quad (66)$$

On the other hand,  $\dot{c}$  can also be computed from (64) and (65) as

$$\begin{aligned} \dot{c} = \dot{a} - \dot{b} &= \int_X pD\nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla (\ln \frac{p}{p_\theta} - \ln \frac{p}{p_{\theta_1}}) dx \\ &= \int_X pD\nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla \ln \frac{p_{\theta_1}}{p_\theta} dx \\ &= - \int_X pD\nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla \ln \frac{p_\theta}{p_{\theta_1}} dx = V(p, \theta, \theta_1). \end{aligned} \quad (67)$$

By equating (66) and (67) we obtain an ODE for the evolution of  $\theta(t)$

$$g_{\alpha\beta}(\theta) \dot{\theta}_\alpha (\theta - \theta_1)_\beta = V(p, \theta, \theta_1)$$

which depends on the unknown solution  $p$  of FPE. In the paper [3] the following approximation is made: if we substitute  $p$  with  $p_\theta$  in  $V(p, \theta, \theta_1)$  in (67) we get

$$\begin{aligned} V(p_\theta, \theta, \theta_1) &= - \int_X p_\theta D\nabla \ln \frac{p_\theta}{p_{\theta_1}} \cdot \nabla \ln \frac{p_\theta}{p_{\theta_1}} dx \\ &= - \int_X p_\theta D\nabla (h \cdot (\theta - \theta_1)) \cdot \nabla (h \cdot (\theta - \theta_1)) dx \\ &= - \left( \int_X p_\theta d_i \partial_i h_\alpha \partial_i h_\beta dx \right) (\theta - \theta_1)_\alpha (\theta - \theta_1)_\beta \\ &= - \langle B_{\alpha\beta} \rangle_\theta (\theta - \theta_1)_\alpha (\theta - \theta_1)_\beta \end{aligned} \quad (68)$$

where we have introduced the the symmetric matrix

$$B_{\alpha\beta} = (D\nabla h \cdot h)_{\alpha\beta} = d_i \partial_i h_\alpha \partial_i h_\beta.$$

By comparing (66) and (68) we obtain a closed form ODE for  $\theta$  since  $\theta - \theta_1 \neq 0$

$$\mathcal{A}_{\alpha\beta} \dot{\theta}_\alpha = g_{\alpha\beta}(\theta) \dot{\theta}_\alpha = \langle B_{\alpha\beta} \rangle_\theta (\theta_1 - \theta)_\alpha. \quad (69)$$

which is equation (5.2) in [3] or equation (12) in [4]. It can be given normal form since  $g_{\alpha\beta}$  is invertible. In these paper the above equation is solved numerically and it is shown that it gives very good (sometimes surprisingly good) estimates of  $\langle h \rangle_p$  using  $\langle h \rangle_\theta$  even if  $\theta$  is far from  $\theta_1$ .

Using the information geometry tools we have shown that the above triangular relation (62) holds independently from the assumption that  $\theta$  be close to  $\theta_1$  (called quasi equilibrium approximation in [2]). Moreover, is is evident from inspection of (65) that the

substitution of  $p$  with  $p_\theta$  renders  $\dot{b} = 0$  therefore  $\dot{a} = \dot{c}$ . Note that if  $D_{ij} = d\delta_{ij}$  and we substitute  $p$  with  $p_\theta$  in the above formula (67) we get

$$\dot{c} = \frac{d}{dt}D(p_\theta|p_{\theta_1}) = -d\mathcal{R}(p_\theta|p_{\theta_1}).$$

Hence, if  $p_{\theta_1}$  satisfies a LSI, we have exponential speed of convergence of  $p_\theta$  to equilibrium distribution  $p_{\theta_1}$ , which explains the good behavior of the approximation.

6.2. A Dynamic Approximation Problem with External Parameters

We now suppose that the drift field (54)  $D_i(x) = d(x_i)\partial_i(\theta \cdot h(x))$  which defines the FPE depends on external parameters because  $h = h(x, u)$ . We consider the same dynamic approximation problem of Section 6.1 with the extra degrees of freedoms given by the external parameters  $u$ . The approximation condition (57) now reads

$$\langle h \rangle_p = \langle h \rangle_{p_z}$$

where  $z = (\theta, u)$ . We take the time derivative of the above relation to find an ODE for  $z$ . We have

$$\begin{aligned} \frac{d}{dt}\langle h \rangle_{p_z} &= \frac{d}{dt} \int_X h p_z dx = \langle \partial_u h \rangle_z \dot{u} + \int_X h p_z \frac{d}{dt} \ln p_z dx \\ &= \langle \partial_u h \rangle_z + \int_X h p_z \frac{d}{dt} (\theta \cdot h - \psi + k) dx \\ &= \langle \partial_u h \rangle_z \dot{u} + \int_X p_z [h(h - \langle h \rangle) \dot{\theta} + \theta \cdot h (\partial_u h - \langle \partial_u h \rangle) \dot{u}] dx \\ &= \langle \partial_u h \rangle_z \dot{u} + \omega(\dot{z}) \end{aligned} \tag{70}$$

and

$$\begin{aligned} \frac{d}{dt}\langle h \rangle_p &= \frac{d}{dt} \int_X h p dx = \langle \partial_u h \rangle_p \dot{u} + \int_X h \partial_t p dx \\ &= \langle \partial_u h \rangle_p \dot{u} + \int_X h \nabla \cdot (p D \nabla \ln \frac{p}{p_{\theta_1}}) dx \\ &= \langle \partial_u h \rangle_p \dot{u} - \int_X p D \nabla \ln \frac{p}{p_{\theta_1}} \cdot \nabla h dx \end{aligned}$$

since the boundary term is vanishing. If we substitute  $p$  with  $p_z$  in the last line we get

$$\begin{aligned} \frac{d}{dt}\langle h \rangle_p &= \langle \partial_u h \rangle_p \dot{u} - \int_X p D \nabla \ln \frac{p_\theta}{p_{\theta_1}} \cdot \nabla h dx \\ &= \langle \partial_u h \rangle_z \dot{u} - \langle B \rangle_z (\theta - \theta_1) \end{aligned}$$

and therefore we have the ODE for  $z$  which is a direct generalization of (69)

$$\omega(\dot{z}) = \mathcal{A}\dot{\theta} + \mathcal{M}\dot{u} = \langle B \rangle_z (\theta_1 - \theta). \tag{71}$$

Note that in this case we have considerably more freedom because we have a system of  $a$  ODEs for the  $d = a + b$  variables  $z = (\theta, u)$  therefore we can assign the evolution  $u(t)$  of the external parameters to control the evolution  $\theta(t)$ .

**Funding:** This research received no external funding.

**Acknowledgments:** The author thanks the anonymous referees for valuable comments and suggestions on this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. An Approximation Problem in Quantitative Genetics

We give a brief account of a classical problem in dynamics of quantitative traits whose approach to equilibrium described by a Fokker-Planck equation has been investigated in [2–4]. See also [37] for a gentle introduction to dynamics of populations. We consider a polygenic trait located in  $n$  biallelic loci  $(A, a)$ . If we consider a sufficiently large population, the frequencies of genotype  $AA$  at locus  $i$  are described by  $n$  independent random variables  $x = (x_1, \dots, x_n)$ ,  $x_i \in [0, 1]$ . The dynamics of these allele frequencies can be described by a diffusion process under the action of stochastic forces which represent the effect of directional selection, dominance, mutation and random drift. This diffusion process is described by a linear Fokker-Planck equation whose equilibrium distribution is known from a long time ( see [38,39])

$$p(x) = e^{\theta \cdot h(x,u) - \psi(\theta,u) + k(x)}, \quad x = (x_1, \dots, x_n) \in X = [0, 1]^n$$

Below we show that it can be seen as a generalized exponential distribution with *natural* parameters  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$  and *external* parameters  $u = (v, w) \in \mathbb{R}_+^{2n}$ . Setting  $d(x_i) = x_i(1 - x_i)$ , we introduce the following four observables:

$$h_1(x, v) = - \sum_{i=1}^n v_i d'(x_i) = - \sum_{i=1}^n v_i (2x_i - 1)$$

$h_1$  is the directional selection and  $v = (v_1, \dots, v_n)$  are external parameters describing the effects on loci;

$$h_2(x, w) = \sum_{i=1}^n w_i 2d(x_i) = - \sum_{i=1}^n w_i 2x_i(1 - x_i)$$

$h_2$  is the dominance,  $w = (w_1, \dots, w_n)$  are external parameters describing the effects on loci;

$$h_3(x) = \sum_{i=1}^n \ln x_i \quad \text{and} \quad h_4(x) = \sum_{i=1}^n \ln(1 - x_i)$$

$h_3$  describes forward and backward mutations and

$$k(x) = - \sum_{i=1}^n \ln x_i(1 - x_i)$$

is the (non integrable) neutral distribution of allele frequencies in absence of selection and mutation. Since the random variables  $x_i$  are independent, the free energy can be factorized using Fubini theorem

$$e^\psi = \int_X e^{\theta \cdot h + k} dx = \prod_{i=1}^n \int_0^1 e^{-\theta_1 v_i (2x_i - 1) + \theta_2 w_i 2x_i(1 - x_i)} x_i^{\theta_3 - 1} (1 - x_i)^{\theta_4 - 1} dx_i.$$

We have  $e^\psi < +\infty$  if  $\theta_3, \theta_4 > 0$  so  $\theta = 0$  is a non feasible value. Moreover  $p(x)$  is integrable but tends to  $+\infty$  for  $x$  tending to  $\partial X$  if  $1 > \theta_3, \theta_4 > 0$  while  $p = 0$  on  $\partial X$  if  $\theta_3, \theta_4 > 1$ . Since the observables  $h_1$  and  $h_2$  have linear dependence on the external parameters  $v$  and  $w$ , this statistical model fails the injectivity condition 1. therefore it is not identifiable. The Fokker-Planck equation can be given the form (51) where  $D = d(x_i)\delta_{ij} = x_i(1 - x_i)$  is diagonal, but  $D = 0$  on  $\partial X$  therefore the inequalities like (52) that govern the trend to equilibrium are not strict ones due to the degeneracy of the diffusion matrix  $D$  on  $\partial X$ . This is a major source of difficulties in the analysis of this model as reported in [3]. A final remark is that the LSI sufficient condition  $Hess(V) \geq \sigma \mathbb{1}$  becomes in this case ( $V = \theta \cdot h - \psi + k$ )

$$\partial_i \partial_j V = \partial_i \partial_j (\theta \cdot h - k) = 4\theta_2 w_j \delta_{ij} - \left( \frac{\theta_4}{(1 - x_j)^2} - \frac{\theta_3}{x_j^2} \right) \delta_{ij}.$$

Therefore  $Hess(V)$  is not bounded below and the LSI condition fails. In the above cited papers the dynamic approximation procedure exposed in Section 6 is introduced to compute the evolution of the averages of the observables  $h$  without solving the FPE which is a computationally hard problem.

In this paper we have shown that this model problem is partly amenable to the controlled exponential family framework with some insight, however remain peculiar difficulties that prevent from a complete analysis of the model. We refer the interested reader to the specific literature, see [3].

## References

- Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: New York, NY, USA, 2003.
- De Vladar, H.P.; Barton, N.H. The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *J. R. Soc. Interface* **2011**, *8*, 720–739. [[CrossRef](#)] [[PubMed](#)]
- Bodova, K.; Haskovec, J.; Markowich, P. Well posedness and maximum entropy approximation for the dynamics of quantitative traits. *Phys. D Nonlinear Phenom.* **2018**, *376*, 108–120. [[CrossRef](#)]
- Bodova, K.; Szep, E.; Barton, N.H. Dynamic maximum entropy provides accurate approximation of structured population dynamics. *PLoS Comput. Biol.* **2021**, *17*, e1009661.
- Weinhold, F. Metric geometry of equilibrium thermodynamics. *J. Chem. Phys.* **1975**, *63*, 2479–2483. [[CrossRef](#)]
- Ruppeiner, G. Thermodynamics: A Riemannian geometric model. *Phys. Rev. A* **1979**, *20*, 1608. [[CrossRef](#)]
- Janke, W.; Johnston, D.A.; Kenna, R. Information geometry and phase transitions. *Phys. A Stat. Mech. Its Appl.* **2004**, *336*, 181–186. [[CrossRef](#)]
- Brody, D.; Rivier, N. Geometrical aspects of statistical mechanics. *Phys. Rev. E* **1995**, *51*, 1006. [[CrossRef](#)]
- Prokopenko, M.; Lizier, J.T.; Obst, O.; Wang, X.R. Relating Fisher information to order parameters. *Phys. Rev. E* **2011**, *84*, 041116. [[CrossRef](#)]
- Brody, D.C.; Ritz, A. Information geometry of finite Ising models. *J. Geom. Phys.* **2003**, *47*, 207–220. [[CrossRef](#)]
- Salamon, P.; Berry, R.S. Thermodynamic length and dissipated availability. *Phys. Rev. Lett.* **1983**, *51*, 1127–1130. [[CrossRef](#)]
- Crooks, G.E. Fisher information and statistical mechanics. *Tech. Rep.* **2011**, 1–3.
- Crooks, G.E. Measuring thermodynamic length. *Phys. Rev. Lett.* **2007**, *99*, 100602. [[CrossRef](#)] [[PubMed](#)]
- Ito, S. Stochastic thermodynamic interpretation of information geometry. *Phys. Rev. Lett.* **2018**, *121*, 030605. [[CrossRef](#)] [[PubMed](#)]
- Abiuso, P.; Miller, H.J.; Perarnau-Llobet, M.; Scandi, M. Geometric optimisation of quantum thermodynamic processes. *Entropy* **2020**, *22*, 1076. [[CrossRef](#)] [[PubMed](#)]
- Naudts, J. The q-exponential family in statistical physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405–413. [[CrossRef](#)]
- Naudts, J. Generalised exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149. [[CrossRef](#)]
- Amari, S.I.; Ohara, A. Geometry of q-exponential family of probability distributions. *Entropy* **2011**, *13*, 1170–1185. [[CrossRef](#)]
- Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
- Watanabe, S. *Algebraic Geometry and Statistical Learning Theory*; No. 25; Cambridge University Press: Cambridge, UK, 2009.
- Favretti, M. Geometry and Control of Thermodynamic Systems described by Generalized Exponential Families. *J. Geom. Phys.* **2022**, *176*, 104497. [[CrossRef](#)]
- Amari, S.; Nagaoka, H. *Methods of Information Geometry*; AMS: Premstätten, Austria; Oxford University Press: Oxford, UK, 2000.
- Barndorff-Nielsen, O. *Information and Exponential Families: In Statistical Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
- Calin, O.; Udriste, C. *Geometric Modeling in Probability and Statistics*; Springer: Berlin, Germany, 2014.
- Murray, M.K.; Rice, J.W. *Differential Geometry and Statistics*; CRC Press: Boca Raton, FL, USA, 1993; Volume 48.
- Souriau, J.-M. *Structure of Dynamical Systems: A Symplectic View of Physics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 149.
- Amari, S. *Information Geometry and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
- Jaynes, E.T. Brandeis Lectures. In *Papers on Probability, Statistics and Statistical Physics*; Rosenkrantz, E.D., Ed.; Springer: Dordrecht, The Netherlands, 1989; pp. 39–76
- Rubinstein, M.; Ralph, H.C. *Polymer Physics*; Oxford University Press: New York, NY, USA, 2003; Volume 23.
- Fisher, R.A. On the mathematical foundations of theoretical statistics. In *Philosophical Transactions of the Royal Society of London; Series A, Containing Papers of a Mathematical or Physical Character*; Royal Society of London: London, UK, 1922.
- Dubey, S.D. Compound gamma, beta and F distributions. *Metrika* **1970**, *16*, 27–31. [[CrossRef](#)]
- Marsden, J.E.; Montgomery, R.; Ratiu, T.S. *Reduction, Symmetry, and Phases in Mechanics*; American Mathematical Soc.: Providence, RI, USA, 1990; Volume 436.
- Souriau, J.M. Mécanique statistique, groupes de Lie et cosmologie. *GeoM. Symplectique et Phys. Math.* **1974**, 237.
- Risken, H. *The Fokker-Planck Equation. Methods of Solutions and Applications*; Springer: Berlin/Heidelberg, Germany, 1984.
- Villani, C. *Topics in Optimal Transportation*; American Mathematical Soc.: Providence, RI, USA, 2021; Volume 58.
- Yamano, T. Phase space gradient of dissipated work and information: A role of relative Fisher information. *J. Math. Phys.* **2013**, *54*, 113301. [[CrossRef](#)]

37. Rice, S. *Evolutionary Theory: Mathematical and Conceptual Foundations*; Sinauer Associates, Inc. Publishers: Sunderland, MA, USA, 2004.
38. Wright, S. Evolution in Mendelian populations. *Genetics* **1931**, *16*, 97. [[CrossRef](#)] [[PubMed](#)]
39. Kimura, M. Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. USA* **1955**, *41*, 144. [[CrossRef](#)]