

Article

Information Geometry Formalism for the Spatially Homogeneous Boltzmann Equation

Bertrand Lods¹ and Giovanni Pistone^{2,*}

¹ Department of Economics and Statistics, Università di Torino & Collegio Carlo Alberto, Corso Unione Sovietica, 218/bis, 10134 Torino, Italy; E-Mail: bertrand.lods@unito.it

² de Castro Statistics, Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy

* Author to whom correspondence should be addressed; E-Mail: giovanni.pistone@carloalberto.org; Tel.: +39-011-670-5033; Fax: +39-011-670-5082.

Academic Editor: Giorgio Kaniadakis

Received: 15 February 2015 / Accepted: 16 June 2015 / Published: 19 June 2015

Abstract: Information Geometry generalizes to infinite dimension by modeling the tangent space of the relevant manifold of probability densities with exponential Orlicz spaces. We review here several properties of the exponential manifold on a suitable set \mathcal{E} of mutually absolutely continuous densities. We study in particular the fine properties of the Kullback-Liebler divergence in this context. We also show that this setting is well-suited for the study of the spatially homogeneous Boltzmann equation if \mathcal{E} is a set of positive densities with finite relative entropy with respect to the Maxwell density. More precisely, we analyze the Boltzmann operator in the geometric setting from the point of its Maxwell's weak form as a composition of elementary operations in the exponential manifold, namely tensor product, conditioning, marginalization and we prove in a geometric way the basic facts, *i.e.*, the H-theorem. We also illustrate the robustness of our method by discussing, besides the Kullback-Leibler divergence, also the property of Hyvärinen divergence. This requires us to generalize our approach to Orlicz–Sobolev spaces to include derivatives.

Keywords: information geometry; Orlicz space; spatially homogeneous Boltzmann equation; Kullback-Leibler divergence; Hyvärinen divergence

1. Introduction

Information geometry (IG) has been essentially developed by S.-I. Amari, see the monograph by Amari and Nagaoka [1]. In his work, all previous geometric—essentially metric—descriptions of probabilistic and statistics concepts are extended in the direction of affine differential geometry, including the fundamental treatment of connections. A corresponding concept for abstract manifold, called statistical manifold, has been worked out by Lauritzen in [2]. Amari’s framework is today considered a case of Hessian geometry as it is described in the monograph by Shima [3].

Other versions of IG have been studied to deal with a non-parametric settings such as the Boltzmann equation as it is described in [4,5]. A very general set-up for information geometry is the following. Consider a one-dimensional family of positive densities with respect to a measure μ , $\theta \mapsto p_\theta$, and a random variable U . A classical statistical computation, possibly due to Ronald Fisher, is

$$\begin{aligned} & \frac{d}{d\theta} \int U(x)p(x; \theta) \mu(dx) \\ &= \int U(x) \left(\frac{d}{d\theta} p(x; \theta) \right) \mu(dx) \\ &= \int U(x) \left(\frac{d}{d\theta} \log p(x; \theta) \right) p(x; \theta) \mu(dx) \\ &= \mathbb{E}_\theta \left[(U - \mathbb{E}_\theta [U]) \frac{d}{d\theta} \log p_\theta \right] . \end{aligned}$$

The previous computation suggests the following geometric construction which is rigorous if the sample space is finite and can be forced to work in general under suitable assumptions. We use the differential geometry language, e.g., [6]. If Δ is the *probability simplex* on a given sample space (Ω, \mathcal{F}) , we define the *statistical bundle* of Δ to be

$$T\Delta = \{(\pi, u) \mid \pi \in \Delta, u \in \mathcal{L}^2(\pi), \mathbb{E}_\pi [u] = 0\} .$$

Given a one dimensional curve in Δ , $\theta \mapsto \pi_\theta$ we can define its *velocity* to be the curve

$$\theta \mapsto \left(\pi_\theta, \frac{D}{d\theta} \pi_\theta \right) \in T\Delta ,$$

where we define

$$\frac{D}{d\theta} \pi_\theta = \frac{d}{d\theta} \log (\pi_\theta) = \frac{\frac{d}{d\theta} \pi_\theta}{\pi_\theta} .$$

Each fiber $T_\pi \Delta = \mathcal{L}_0^2(\mu)$ has a scalar product and we have a parallel transport

$$\mathbb{U}_\pi^\nu : T_\pi \Delta \ni V \mapsto V - \mathbb{E}_\nu [V] \in T_\nu \Delta .$$

This structure provides an interesting framework to interpret the Fisher computation cited above. The basic case of a finite state space has been extended by Amari and coworkers to the case of a parametric set of strictly positive probability densities on a generic sample space. Following a suggestion by Dawid in [7,8], a particular nonparametric version of that theory was developed in a series of papers [9–19], where the set $\mathcal{P}_>$ of all strictly positive probability densities of a measure space is shown to be a Banach manifold (as it is defined in [20–22]) modeled on an Orlicz Banach space, see, e.g., [23, Chapter II].

In the present paper, Section 2 recalls the theory and our notation about the model Orlicz spaces. This material is included for convenience only and this part should be skipped by any reader aware of any of the papers [9–19] quoted above. The following Section 3 is mostly based on the same references and it is intended to introduce that manifold structure and to give a first example of application to the study of Kullback-Liebler divergence. The special features of statistical manifolds that contain the Maxwell density are discussed in Section 4. In this case we can define the Boltzmann–Gibbs entropy and study its gradient flow. The setting for the Boltzmann equation is discussed in Section 5 where we show that the equation can be derived from probabilistic operations performed on the statistical manifold. Our application to the study of the Kullback-Liebler divergence is generalized in Section 6 to the more delicate case of the Hyvärinen divergence. This requires in particular a generalization of the manifold structure to include differential operators and leads naturally to the introduction of Orlicz-Sobolev spaces.

We are aware that there are other approaches to non-parametric information geometry that are not based on the notion of the exponential family and that we do not consider here. We mention in particular [24] and [25].

2. Model Spaces

Given a σ -finite measure space, $(\Omega, \mathcal{F}, \mu)$, we denote by $\mathcal{P}_>$ the set of all densities that are positive μ -a.s, by \mathcal{P}_\geq the set of all densities, by \mathcal{P}_1 the set of measurable functions f with $\int f d\mu = 1$.

We introduce here the Orlicz spaces we shall mainly investigate in the sequel. We refer to [10] and [23, Chapter 2] for more details on the matter. We consider the Young function

$$\Phi: \mathbb{R} \ni x \mapsto \Phi(x) = \cosh x - 1$$

and, for any $p \in \mathcal{P}_>$, the Orlicz space $L^\Phi(p) = L^{\cosh-1}(p)$ is defined as follows: a real random variable U belongs to $L^\Phi(p)$ if

$$\mathbb{E}_p [\Phi(\alpha U)] < +\infty \quad \text{for some } \alpha > 0 .$$

The Orlicz space $L^\Phi(p)$ is a Banach space when endowed with the Luxemburg norm defined as

$$\|U\|_{\Phi,p} = \inf \{ \lambda > 0 \mid \mathbb{E}_p [\Phi(U/\lambda)] \leq 1 \} .$$

The conjugate function of $\Phi = \cosh - 1$ is

$$\Phi_*(y) = (\cosh - 1)_*(y) = \int_0^y \operatorname{arsinh} u \, du = y \operatorname{arsinh} y + 1 - \sqrt{1 + y^2}, \quad y \in \mathbb{R}$$

which satisfies the so-called Δ_2 -condition as

$$\Phi_*(\alpha y) = \int_0^{|\alpha y|} (|\alpha y| - u) \Phi_*''(u) du = \int_0^{|\alpha y|} \frac{|\alpha| |y| - u}{1 + u^2} du \leq \max(1, \alpha^2) \Phi_*(y) .$$

Since Φ_* is a Young function, for any $p \in \mathcal{P}_>$, one can define as above the associated Orlicz space $L^{\Phi_*}(p) = L^{(\cosh-1)_*}(p)$ and its corresponding Luxemburg norm $\|\cdot\|_{\Phi_*,p}$. Because the functions Φ and Φ_* form a Young pair, for each $U \in L^{\cosh-1}(p)$ and $V \in L^{(\cosh-1)_*}(p)$ we can deduce from Young’s inequality

$$xy \leq \Phi(x) + \Phi_*(y), \quad \forall x, y \in \mathbb{R}^2$$

that the following Holder’s inequality holds:

$$|\mathbb{E}_p [UV]| \leq 2 \|U\|_{\cosh^{-1}, p} \|V\|_{(\cosh^{-1})^*, p} < \infty .$$

Moreover, it is a classical result that the space $L^{(\cosh^{-1})^*}(p)$ is separable and its dual space is $L^{\cosh^{-1}}(p)$, the duality pairing being

$$(V, U) \in L^{(\cosh^{-1})^*}(p) \times L^{\cosh^{-1}}(p) \longmapsto \langle U, V \rangle_p := \mathbb{E}_p [UV] .$$

We recall the following continuous embedding result that we shall use repeatedly in the paper:

Theorem 1. *Given $p \in \mathcal{P}_>$, for any $1 < r < \infty$, the following embeddings*

$$L^\infty(p) \hookrightarrow L^{\cosh^{-1}}(p) \hookrightarrow L^r(p) \hookrightarrow L^{(\cosh^{-1})^*}(p) \hookrightarrow L^1(p)$$

are continuous.

From this result, we deduce easily the following useful Lemma

Lemma 1. *Given $p \in \mathcal{P}_>$ and $k \geq 1$. For any $u_1, \dots, u_k \in L^{\cosh^{-1}}(p)$ one has*

$$\prod_{i=1}^k u_i \in \bigcap_{1 < r < \infty} L^r(p) .$$

Proof. According to Theorem 1, $u_i \in L^r(p)$ for any $1 < r < \infty$ and any $i = 1, \dots, k$. The proof follows then simply from the repeated use of Holder inequality. \square

From now on we also define, for any $p \in \mathcal{P}_>$

$$B_p = L_0^\Phi(p) := \{u \in L^{\cosh^{-1}}(p) \mid \mathbb{E}_p [u] = 0\} .$$

In the same way, we set ${}^*B_p = L_0^{\Phi^*}(p) := \{u \in L^{(\cosh^{-1})^*}(p) \mid \mathbb{E}_p [u] = 0\}$.

2.1. Cumulant Generating Functional

Let $p \in \mathcal{P}_>$ be given. With the above notations one can define:

Definition 1. *The cumulant generating functional is the mapping*

$$K_p : u \in B_p \longmapsto \log \mathbb{E}_p [e^u] \in [0, +\infty] .$$

The following result [9,10] shows the properties of the exponential function as a superposition mapping [26].

Proposition 1. *Let $a \geq 1$ and $p \in \mathcal{P}_>$ be given.*

1. For any $n = 0, 1, \dots$ and $u \in L^\Phi(p)$:

$$\lambda_{a,n}(u) : (w_1, \dots, w_n) \mapsto \frac{w_1}{a} \dots \frac{w_n}{a} e^{\frac{u}{a}}$$

is a continuous, symmetric, n -multi-linear map from $(L^\Phi(p))^n$ to $L^a(p)$.

2. $v \mapsto \sum_{n=0}^\infty \frac{1}{n!} \left(\frac{v}{a}\right)^n$ is a power series from $L^\Phi(p)$ to $L^a(p)$, with radius of convergence larger than 1.

3. The superposition mapping, $v \mapsto e^{v/a}$, is an analytic function from the open unit ball of $L^\Phi(p)$ to $L^a(p)$.

The cumulant generating functional enjoys the following properties (see [9,10,14]):

Proposition 2.

1. $K_p(0) = 0$; otherwise, for each $u \neq 0$, $K_p(u) > 0$.

2. K_p is convex and lower semi-continuous, and its proper domain

$$\text{dom}(K_p) = \{u \in B_p \mid K_p(u) < \infty\}$$

is a convex set that contains the open unit ball of B_p .

3. K_p is infinitely Gâteaux-differentiable in the interior of its proper domain.

4. K_p is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of B_p .

Remark 1. One sees from the above property 2 that the interior of the proper domain of K_p is a non-empty open convex set. From now on we shall adopt the notation

$$\mathcal{S}_p = \text{Int}(\text{dom}(K_p)) .$$

Other properties of the functional K_p are described below, as they relate directly to the exponential manifold.

3. Exponential Manifold

The set of positive densities, $\mathcal{P}_>$, locally around a given $p \in \mathcal{P}_>$, is modeled by the subspace of centered random variables in the Orlicz space, $L^{\cosh^{-1}}(p)$. Hence, it is crucial to discuss the isomorphism of the model spaces for different p 's in order to show the existence of an atlas defining a Banach manifold.

Definition 2 (Statistical exponential manifold [10, Definition 20]). For $p \in \mathcal{P}_>$, the statistical exponential manifold at p is

$$\mathcal{E}(p) = \{e^{u-K_p(u)}p \mid u \in \mathcal{S}_p\} .$$

We also need the following definition of connection

Definition 3 (Connected densities). Densities $p, q \in \mathcal{P}_>$ are connected by an open exponential arc if there exists an open exponential family containing both, i.e., if for a neighborhood I of $[0, 1]$

$$\int_{\Omega} p^{1-t} q^t d\mu = \mathbb{E}_p \left[\left(\frac{q}{p} \right)^t \right] = \mathbb{E}_q \left[\left(\frac{p}{q} \right)^{1-t} \right] < +\infty, \quad t \in I$$

In such a case, one simply writes $p \smile q$.

Theorem 2 (Portmanteau theorem, see [10, Theorems 19 and 21] and [27, Theorem 4.7]). Let $p, q \in \mathcal{P}_>$. The following statements are equivalent:

1. $p \smile q$, (i.e., p and q are connected by an open exponential arc);
2. $q \in \mathcal{E}(p)$;
3. $\mathcal{E}(p) = \mathcal{E}(q)$;
4. $\log \frac{q}{p} \in L^\Phi(p) \cap L^\Phi(q)$;
5. $L^\Phi(p) = L^\Phi(q)$, (i.e., they both coincide as vector spaces and their norms are equivalent);
6. There exists $\varepsilon > 0$ such that

$$\frac{q}{p} \in L^{1+\varepsilon}(p) \quad \text{and} \quad \frac{p}{q} \in L^{1+\varepsilon}(q)$$

We can now define the charts and atlas of the exponential manifold as follows:

Definition 4 (Exponential manifold [9,10,14,15]). For each $p \in \mathcal{P}_>$, define the charts at p as:

$$s_p : q \in \mathcal{E}(p) \mapsto \log \left(\frac{q}{p} \right) - \mathbb{E}_p \left[\log \left(\frac{q}{p} \right) \right] \in \mathcal{S}_p \subset B_p,$$

with inverse

$$s_p^{-1} = e_p : u \in \mathcal{S}_p \mapsto e^{u-K_p(u)} p \in \mathcal{E}(p) \subset \mathcal{P}_>$$

The atlas, $\{s_p : \mathcal{S}_p | p \in \mathcal{P}_>\}$ is affine and defines the exponential (statistical) manifold $\mathcal{P}_>$.

We collect here various results from [9,10,14,15] about additional properties of K_p .

Proposition 3. Let $q = e^{u-K_p(u)} p \in \mathcal{E}(p)$ with $u \in \mathcal{S}_p$.

1. The first three derivatives of K_p on \mathcal{S}_p are:

$$\begin{aligned} dK_p(u)[v] &= \mathbb{E}_q [v] , \\ d^2 K_p(u)[v_1, v_2] &= \text{Cov}_q (v_1, v_2) , \\ d^3 K_p(u)[v_1, v_2, v_3] &= \text{Cov}_q (v_1, v_2, v_3) . \end{aligned}$$

2. The random variable, $\frac{q}{p} - 1$, belongs to *B_p and:

$$dK_p(u)[v] = \mathbb{E}_p \left[\left(\frac{q}{p} - 1 \right) v \right] .$$

In other words, the gradient of K_p at u is identified with an element of the predual space of B_p , viz. *B_p , denoted by $\nabla K_p(u) = e^{u-K_p(u)} - 1 = \frac{q}{p} - 1$.

3. The weak derivative of the map, $\mathcal{S}_p \ni u \mapsto \nabla K_p(u) \in {}^*B_p$, at u applied to $w \in B_p$ is given by:

$$d(\nabla K_p(u))[w] = \frac{q}{p} (w - \mathbb{E}_q [w])$$

and it is one-to-one at each point.

4. $q/p \in L^{(\cosh -1)*}(p)$.

5. B_q is defined by an orthogonality property:

$$B_q = L_0^{\cosh -1}(q) = \left\{ u \in L^{\cosh -1}(p) \mid \mathbb{E}_p \left[u \frac{q}{p} \right] = 0 \right\} .$$

On the basis of the above result, it appears natural to define the following *parallel transports*:

Definition 5.

1. The exponential transport ${}^e\mathbb{U}_p^q: B_p \rightarrow B_q$ is computed as

$${}^e\mathbb{U}_p^q u = u - \mathbb{E}_q [u] \quad u \in B_p .$$

2. The mixture transport ${}^m\mathbb{U}_q^p: {}^*B_q \rightarrow {}^*B_p$ is computed as

$${}^m\mathbb{U}_q^p v = \frac{p}{q} v \quad v \in {}^*B_q .$$

One has the following properties

Proposition 4. Let $p, q \in \mathcal{P}_>$ be given. Then

1. ${}^e\mathbb{U}_p^q$ is an isomorphism of B_p onto B_q .
2. ${}^m\mathbb{U}_q^p$ is an isomorphism of *B_q onto *B_p .
3. The mixture transport ${}^m\mathbb{U}_q^p$ and the exponential transport ${}^e\mathbb{U}_p^q$ are dual of each other: if $u \in B_q$ and $v \in {}^*B_p$ then

$$\langle u, {}^m\mathbb{U}_p^q v \rangle_q = \langle {}^e\mathbb{U}_q^p u, v \rangle_p .$$

4. ${}^m\mathbb{U}_p^q$ and ${}^e\mathbb{U}_p^q$ together transport the duality pairing: if $u \in B_p$ and $v \in {}^*B_p$, then

$$\langle u, v \rangle_p = \langle {}^e\mathbb{U}_p^q u, {}^m\mathbb{U}_p^q v \rangle_q .$$

We reproduce here a scheme of how the affine manifold works. The domains of the charts centered at p and q respectively are either disjoint or equal if $p \sim q$:

$$\begin{array}{ccccccc} \mathcal{E}(p) & \xrightarrow{s_p} & \mathcal{S}_p & \xrightarrow{I} & B_p & \xrightarrow{I} & L^{\cosh^{-1}}(p) \\ \parallel & & \downarrow s_q \circ s_p^{-1} & & \downarrow d(s_q \circ s_p^{-1}) & & \parallel \\ \mathcal{E}(q) & \xrightarrow{s_q} & \mathcal{S}_q & \xrightarrow{I} & B_q & \xrightarrow{I} & L^{\cosh^{-1}}(q) \end{array}$$

Our discussion of the tangent bundle of the exponential manifold is based on the concept of the velocity of a curve as in [20, Section 3.3] and it is mainly intended to underline its statistical interpretation, which is obtained by identifying curves with one-parameter statistical models. For a statistical model $p(t)$, $t \in I$, the random variable, $\dot{p}(t)/p(t)$ (which corresponds to the Fisher score), has zero expectation with respect to $p(t)$, and its meaning in the exponential manifold is velocity; see [28] on exponential families. More precisely, let $p(\cdot): I \rightarrow \mathcal{E}(p)$, I the open real interval containing zero. In the chart centered at p , the curve is $u(\cdot): I \rightarrow B_p$, where $p(t) = e^{u(t) - K_p(u(t))} p$.

Definition 6 (Velocity field of a curve and tangent bundle).

1. Assume $t \mapsto u(t) = s_p(p(t))$ is differentiable with derivative $\dot{u}(t)$. Define:

$$\frac{D}{dt}p(t) = {}^e\mathbb{U}_p^{p(t)}\dot{u}(t) = \dot{u}(t) - \mathbb{E}_{p(t)}[\dot{u}(t)] = \frac{d}{dt}(u(t) - K_p(u(t))) = \frac{d}{dt} \log \left(\frac{p(t)}{p} \right) = \frac{\frac{d}{dt}p(t)}{p(t)}. \tag{1}$$

Note that Dp does not depend on the chart s_p and that the derivative of $t \mapsto p(t)$ in the last term of the equation is computed in $L^{\Phi^*}(p)$. The curve $t \mapsto (p(t), Dp(t))$ is the velocity field of the curve.

2. On the set $\{(p, v) | p \in \mathcal{P}_>, v \in B_p\}$, the charts:

$$s_p: \{(q, w) | q \in \mathcal{E}(p), w \in B_q\} \ni (q, w) \mapsto (s_p(q), {}^e\mathbb{U}_q^p w) \in \mathcal{S}_p \times B_p \subset B_p \times B_p \tag{2}$$

define the tangent bundle, $T\mathcal{P}_>$.

Remark 2. Let $E: \mathcal{E}(p) \rightarrow \mathbb{R}$ be a C^1 function. Then, $E_p = E \circ e_p: \mathcal{S}_p \rightarrow \mathbb{R}$ is differentiable and:

$$\frac{d}{dt}E(p(t)) = \frac{d}{dt}E_p(u(t)) = dE_p(u(t))\dot{u}(t) = dE_p(u(t)) {}^e\mathbb{U}_{p(t)}^p Dp(t)$$

3.1. Pretangent Bundle

Let M be a density and $\mathcal{E}(M)$ its associated exponential manifold. Here M is generic, later it will be the Maxwell distribution. All densities are assumed to be in $\mathcal{E}(M)$.

Definition 7 (Pretangent bundle ${}^*T\mathcal{E}(p)$). The set

$${}^*T\mathcal{E}(M) = \{(q, V) | q \in \mathcal{E}(M), V \in {}^*B_q\}$$

together with the charts:

$${}^*s_p: {}^*T\mathcal{E}(M) \ni (q, V) \mapsto (s_p(q), {}^m\mathbb{U}_q^p V)$$

is the pretangent bundle, ${}^*T\mathcal{E}(M)$.

Let F be a vector field of the pretangent bundle, $F: \mathcal{E}(M) \rightarrow {}^*T\mathcal{E}(M)$. In the chart centered at p , the vector field is expressed by

$$F_p(u) = {}^m\mathbb{U}_{e_p(u)}^p F \circ e_p(u) = \frac{e_p(u)}{p} F \circ e_p(u) \in {}^*B_p, \quad u \in \mathcal{S}_M$$

If F_p is of class C^1 with derivative $dF_p(u) \in L(B_p, {}^*B_p)$, for each differentiable curve $t \mapsto p(t) = e^{U(t)-K_p(U(t))} \cdot p$ we have $Dp(t) = \frac{d}{dt} \log p(t) = \dot{U}(t) - \mathbb{E}_{p(t)} [\dot{U}(t)]$ and also

$$\frac{d}{dt} F_p(U(t)) = dF_p(U(t))[\dot{u}(t)] = dF_p(u(t)) [{}^e\mathbb{U}_{p(t)}^p Dp(t)] \in {}^*B_p$$

Definition 8 (Covariant derivative in ${}^*T\mathcal{E}(M)$). *Let F be a vector field of class C^1 of the pretangent bundle ${}^*T\mathcal{E}(M)$, and let G be a continuous vector field in the tangent bundle $T\mathcal{E}(M)$. The covariant derivative is the vector field $D_G F$ of ${}^*T\mathcal{E}(M)$ defined at each $q \in \mathcal{E}(M)$ by*

$$(D_G F)(q) = dF_q(u)|_{u=0} [w], \quad w = G(q)$$

In the definition above the covariant derivative is computed in the mobile frame because its value at q is computed using the expression in the chart centered at q . In a fixed frame centered at p we write $s_p(q) = w$ so that $e_q(u) = e_p(u - \mathbb{E}_p[u] + w)$, and compare the two expressions of F as follows.

$$\begin{aligned} F_q(u) &= {}^m\mathbb{U}_{e_q(u)}^q F \circ e_q(u) \\ &= {}^m\mathbb{U}_{e_q(u)}^q \quad {}^m\mathbb{U}_{e_p(u - \mathbb{E}_p[u] + w)}^{e_p} \quad {}^m\mathbb{U}_{e_p(u - \mathbb{E}_p[u] + w)}^p F \circ e_p(u - \mathbb{E}_p[u] + w) \\ &= {}^m\mathbb{U}_p^q \quad {}^m\mathbb{U}_{e_p(u - \mathbb{E}_p[u] + w)}^p F \circ e_p(u - \mathbb{E}_p[u] + w) \\ &= {}^m\mathbb{U}_p^q F_p(u - \mathbb{E}_p[u] + w). \end{aligned}$$

Derivation in the direction $v \in B_q$ gives

$$dF_q(u)[v] = {}^m\mathbb{U}_p^q dF_p(u - \mathbb{E}_p[u] + w)[v - \mathbb{E}_p[v]],$$

hence $dF_q(u) = {}^m\mathbb{U}_p^q dF_p({}^e\mathbb{U}_q^p u + w) {}^e\mathbb{U}_q^p$ and at $u = 0$, we have $dF_q(0) = {}^m\mathbb{U}_p^q dF_p(w) {}^e\mathbb{U}_q^p$. It follows that the covariant derivative in the fixed frame at p is

$$(D_G F)(q) = {}^m\mathbb{U}_p^q dF_p(s_p(q)) {}^e\mathbb{U}_q^p G(q).$$

The tangent and pretangent bundle can be coupled to produce the vector bundle of order 2 defined by

$$({}^*T \times T)\mathcal{E}(M) = \{(p, v, w) | p \in \mathcal{E}(M), v \in B_p, w \in {}^*B_p\}$$

with charts

$$(q, v, w) \mapsto (s_p(q), {}^e\mathbb{U}_q^p v, {}^m\mathbb{U}_q^p w)$$

and the duality coupling:

$$(q, v, w) \mapsto \langle {}^e\mathbb{U}_q^p v, {}^m\mathbb{U}_q^p w \rangle_q$$

Proposition 5 (Covariant derivative of the duality coupling). *Let F be a vector field of ${}^*T\mathcal{E}(M)$, and let G, X be vector fields of $T\mathcal{E}(M)$, F, G of class C^1 and X continuous.*

$$D_X \langle F, G \rangle = \langle D_X F, G \rangle + \langle F, D_X G \rangle$$

Proof. Consider the real function $\mathcal{E} \ni q \mapsto \langle F, G \rangle (q) = \mathbb{E}_q [F(q)G(q)]$ in the chart centered at any $p \in \mathcal{E} (M)$:

$$\begin{aligned} \mathcal{S}_p \ni u \mapsto \mathbb{E}_{e_p(u)} [F(e_p(u))G(e_p(u))] &= \\ \mathbb{E}_p \left[{}^m\mathbb{U}_{e_p(u)}^p F \circ e_p(u) \ {}^e\mathbb{U}_{e_p(u)}^p G \circ e_p(u) \right] &= \mathbb{E}_p [F_p(u)G_p(u)], \end{aligned}$$

and compute its derivative at 0 in the direction $X(p)$. \square

We refer to [18,19] for further details on the geometric structure, namely the Hilbert bundle, the tangent mapping of an homeomorphism, the Riemannian Hessian. We now turn to a basic example.

3.2. Kullback-Leibler Divergence

The Kullback-Leibler divergence [29] on the exponential manifold \mathcal{E} is the mapping

$$\mathbf{D} : (q_1, q_2) \in \mathcal{E} \times \mathcal{E} \mapsto \mathbf{D} (q_1 \| q_2) = \mathbb{E}_{q_1} \left[\log \left(\frac{q_1}{q_2} \right) \right].$$

Notice that, if $q_2 = e^{u-K_{q_1}(u)} \cdot q_1 \in \mathcal{E} (q_1)$, $u \in \mathcal{S}_{q_1}$, then $K_{q_1}(u) = \mathbb{E}_{q_1} [\log (q_1/q_2)]$ is the expression in the chart centered at q_1 of the marginal Kullback-Leibler divergence $q_2 \mapsto \mathbf{D} (q_1 \| q_2)$. Therefore, the Kullback-Leibler divergence is non-negative valued and zero if and only if $q_2 = q_1$ because of Item (4) in Theorem 2. Its expression in the chart centered at a generic $p \in \mathcal{E}$ is

$$D_p : (u_1, u_2) \in \mathcal{S}_p \times \mathcal{S}_p \mapsto K_p(u_2) - K_p(u_1) - dK_p(u_1)[u_2 - u_1],$$

which is the Bregman divergence [30] of the convex function $K_p : \mathcal{S}_p \rightarrow \mathbb{R}$.

It follows from Item (4) in Proposition 2 that it is C^∞ jointly in both variables and, moreover, analytic with

$$D_p(u_1, u_2) = \sum_{n \geq 2} \frac{1}{n!} d^n K_p(u_1)[(u_1 - u_2)^{\otimes n}], \quad \|u_1 - u_2\|_{\Phi,p} < 1.$$

This regularity result is to be compared with what is available when the restriction, $q_1 \smile q_2$, is removed, *i.e.*, the semi-continuity [31, Section 9.4].

The partial derivative of D_p in the first variable, that is the derivative of $u_1 \mapsto D_p(u_1, u_2)$, in the direction $v \in B_p$ is

$$d(u_1 \mapsto D_p(u_1, u_2))[v] = -d^2 K_p(u_1)[u_2 - u_1, v] = -\text{Cov}_{q_1} \left(\log \frac{q_2}{q_1}, v \right),$$

with $q_i = e_p(u_i)$, $i = 1, 2$. If $v = {}^e\mathbb{U}_{q_1}^p w$, we have $-\text{Cov}_{q_1} \left(\log \frac{q_2}{q_1}, v \right) = \text{Cov}_{q_1} \left(\log \frac{q_1}{q_2}, w \right)$, so that we can compute both the covariant derivative of the partial functional $q \mapsto \mathbf{D} (q \| q_2)$ and its gradient as

$$\begin{aligned} D_w(q \mapsto \mathbf{D} (q \| q_2)) &= \text{Cov}_q \left(\log \frac{q}{q_2}, w \right) = \mathbb{E}_q \left[\left(\log \frac{q}{q_2} - \mathbf{D} (q \| q_2) \right) w \right], \\ \nabla(q \mapsto \mathbf{D} (q \| q_2)) &= \log \frac{q}{q_2} - \mathbf{D} (q \| q_2). \end{aligned}$$

The negative gradient flow is

$$\frac{d}{dt} \log \frac{q(t)}{q_2} = -\log \frac{q(t)}{q_2} + \mathbf{D}(q(t) \| q_2).$$

As $\frac{d}{dt} \left(e^t \log \frac{q(t)}{q_2} \right) = e^t \mathbf{D}(q(t) \| q_2)$, for each t the random variable

$$e^t \log \frac{q(t)}{q_2} - \log \frac{q(0)}{q_2} = e^t \log \left(\frac{q(t)}{q_2} \left(\frac{q_2}{q(0)} \right)^{e^{-t}} \right) = e^t \log \left(q(t) \frac{q_2^{e^{-t}-1}}{q(0)^{e^{-t}}} \right)$$

is constant, so that $q(t) \propto q(0)e^{-t} q_2^{1-e^{-t}}$. It is the exponential arc of $q(0) \smile q_2$ in an exponential time scale.

The partial derivative of D_p in the second variable, that is the derivative of $u_2 \mapsto D_p(u_1, u_2)$, in the direction $v \in B_p$ is

$$d(u_2 \mapsto D_p(u_1, u_2))v = dK_p(u_2)[v] - dK_p(u_1)[v] = \mathbb{E}_{q_2}[v] - \mathbb{E}_{q_1}[v],$$

with $q_i = e_p(u_i), i = 1, 2$. If $v = {}^c\mathbb{U}_{q_2}^p w$, we have $\mathbb{E}_{q_2}[v] - \mathbb{E}_{q_1}[v] = \mathbb{E}_{q_2}[w] - \mathbb{E}_{q_1}[w]$, so that we can compute both the covariant derivative of the partial functional $q \mapsto \mathbf{D}(q_1 \| q)$ and its gradient as

$$D_w(q \mapsto \mathbf{D}(q_1 \| q)) = \mathbb{E}_q[w] - \mathbb{E}_{q_1}[w] = \mathbb{E}_q \left[\left(1 - \frac{q_1}{q} \right) w \right],$$

$$\nabla(q \mapsto \mathbf{D}(q_1 \| q)) = 1 - \frac{q_1}{q}.$$

The negative gradient flow is

$$\frac{\dot{q}(t)}{q(t)} = - \left(1 - \frac{q_1}{q(t)} \right) = \frac{q_1 - q(t)}{q(t)},$$

whose solution starting at q_0 is $q(t) = q_1 + (q_0 - q_1)e^{-t}$. It is a mixture model in an exponential time scale.

4. Gaussian Space

In this Section the sample space is \mathbb{R}^n , M denotes the standard n -dimensional Gaussian density (we denoted it M because of James Clerk Maxwell (1831–1879))

$$M(x) = \frac{1}{(2\pi)^{n/2}} \exp \left(-\frac{|x|^2}{2} \right), \quad x \in \mathbb{R}^n \tag{3}$$

and \mathcal{E} is the exponential manifold containing M . We recall that the Orlicz space $L^{\cosh^{-1}}(M)$ is defined with the Young function $\Phi := x \mapsto \cosh x - 1$. The following propositions depend on the specific properties of the Gaussian density M . They do not hold in general.

Proposition 6.

1. The Orlicz space $L^{\cosh^{-1}}(M)$ contains all polynomials with degree up to 2.
2. The Orlicz space $L^{(\cosh^{-1})^*}(M)$ contains all polynomials.

Proof. 1. If f is a polynomial of degree $d \leq 2$ then

$$\int_{\mathbb{R}^n} e^{\alpha f(x)} M(x) dx = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{\alpha f(x) - \frac{1}{2}|x|^2} dx$$

and the latter is finite for all α such that $\alpha \text{Hess } f - I$ is negative definite.

2. The result comes from the fact that all polynomials belong to $L^2(M)$ and one has $L^2(M) \subset L^{(\cosh-1)^*}(M)$. \square

4.1. Boltzmann–Gibbs Entropy

While the Kullback-Leibler divergence $\mathbf{D}(q_1 || q_2)$ of Section 3.2 is defined and finite if the densities q_1 and q_2 belong to the same exponential manifold, the *Boltzmann–Gibbs entropy* (BG-entropy in the sequel)

$$H(q) = -\mathbb{E}_q [\log(q)]$$

could be either non defined or infinite, precisely $-\infty$, everywhere on some exponential manifolds, or finite everywhere on other exponential manifolds.

Proposition 7. Assume $p \smile q$, Then:

1. For each $a \geq 1$, $\log p \in L^a(p)$ if, and only if, $\log q \in L^a(q)$.
2. $\log p \in L^{\cosh-1}(p)$ if, and only if, $\log q \in L^{\cosh-1}(q)$.

Proof. If p, q belong to the same exponential manifold, we can write $q = e^{u - K_p(u)} \cdot p$ and, from Item (4) in Theorem 2, we obtain $\log q - \log p = u - K_p(u) \in L^{\cosh-1}(p) = L^{\cosh-1}(q)$, so that $\log q \in L^a(q)$ if, and only if, $\log p \in L^a(p)$, $a \geq 1$, and $\log q \in L^{\cosh-1}(q)$ if, and only if, $\log p \in L^{\cosh-1}(p)$. \square

In order to obtain a smooth function, we study the BG-entropy $H(q)$ on all manifolds \mathcal{E} , such that for at least one, and, hence for all, $p \in \mathcal{E}$, it holds $\log(p) \in L^{\cosh-1}(p)$. In such an exponential manifold we can write

$$\mathbf{D}(q || p) = -H(q) - \mathbb{E}_q [\log p] ,$$

so that $H(q) \leq -\mathbb{E}_q [\log p]$.

For example, it is the case when the reference measure is finite and p is constant. Another notable example is the Gaussian case, where the sample space is \mathbb{R}^n endowed with the Lebesgue measure and $p(x) \propto \exp(-\frac{|x|^2}{2})$. In such case $\int \cosh(\alpha|x|^2) \exp(-|x|^2/2) dx < +\infty$ if $0 < \alpha < 1/2$.

We investigate here the main properties of the BG-entropy in this context. First, one has

Proposition 8. The BG-entropy is a smooth real function on the exponential manifold \mathcal{E} . Namely, if $p \in \mathcal{E}$ then

$$H_p : u \in \mathcal{S}_p \mapsto H \circ e_p(u)$$

is a C^∞ real function. Moreover, its derivative in the direction v equals

$$dH_p(u)[v] = -\text{Cov}_q(u + \log p, v) ,$$

where $q = e_p(u)$.

Proof. As $-\log q = -u + K_p(u) - \log p = -(u + \log p + H(p)) + K_p(u) + H(p) \in L^{\cosh^{-1}}(p)$, with $-(u + \log p + H(p)) \in B_p$, the representation of the BG-entropy in the chart centered at p is

$$\begin{aligned} H_p(u) &= H \circ e_p(u) \\ &= -\mathbb{E}_{e_p(u)} [u + \log p + H(p)] + K_p(u) + H(p) \\ &= -dK_p(u) [u + \log p + H(p)] + K_p(u) + H(p) \\ &= (K_p(u) - d_u K_p(u)) - dK_p(u)[\log p + H(p)] + H(p), \end{aligned}$$

hence, $u \mapsto H_p(u)$ is a C^∞ real function. Notice that $K_p(u) - d_u K_p(u) \leq 0$, hence $H(q) \leq \mathbb{E}_q [\log p]$, as we already know. The derivative of H_p in the direction v equals

$$dH_p(u)[v] = -d^2 K_p(u)[(u + \log p + H(p)), v] + d_v K_p(u) = -\text{Cov}_q(u + \log p, v) .$$

□

Notice that, for $u = 0$ and $q = e_p(u)$, we have $\mathbb{E}_q [v] = \mathbb{E}_p [v] = 0$, hence

$$dH_p(0)[v] = -\text{Cov}_p(\log p, v) = -\mathbb{E}_p [\log(p) v] .$$

Proposition 9. The gradient field ∇H over \mathcal{E} can be identified, at each p , with random variable $\nabla H(p) \in B_p \subset {}^*B_q$,

$$\nabla H(p) = -(\log(p) + H(p)).$$

Proof. The covariant derivative $D_G H$ at $p \in \mathcal{E}$ with respect to the vector field G defined on \mathcal{E} with $G(p) \in B_p$ and $p \in \mathcal{E}$ is

$$D_G H(p) = -\mathbb{E}_p [\log(p) G(p)] = -\langle \log(p) + H(p), G(p) \rangle_p .$$

The gradient field ∇H over \mathcal{E} , is then defined by $D_G H(p) = \langle \nabla H(p), G(p) \rangle_p$. This justifies the identification with the random variable $\nabla H(p) = -(\log(p) + H(p)) \in B_p \subset {}^*B_q$. □

Remark 3. The equation $\nabla H(p) = 0$ implies $\log p = -H(p)$, hence p has to be constant and this requires it is the finite reference measure μ .

We refer to [19] for more details on the BG-entropy and in particular on the evolution of H on C^1 curve in \mathcal{E} of the type $I \ni t \mapsto f_t$.

5. Boltzmann Equation

We consider a space-homogeneous Boltzmann operator as it is defined, for example, in [5] and [32]. We retell the basic story in order to introduce our notations and the IG background. Orlicz spaces as a setting for Boltzmann’s equation have been recently proposed by [33], while the use of exponential statistical manifolds has been suggested in [18, Example 11] and sketched in [19, Section 4.4]. We start with an improvement of the latter, a few repetitions being justified by consistency between this presentation and [32, Sections 1.3, 4.5–6], compare also Proposition 10 below.

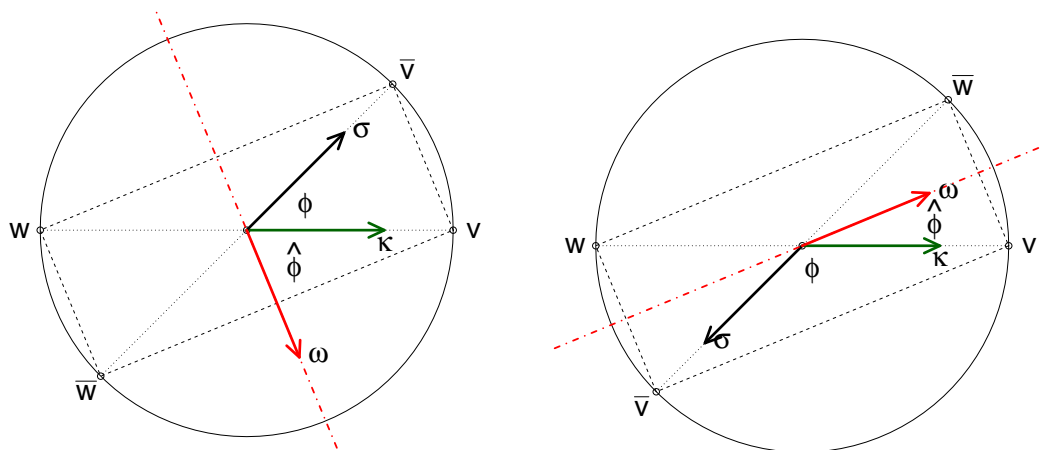


Figure 1. Elastic collision for two opposite σ 's corresponding to the two labeling of the velocities after collision. Velocities v and w are before collision, \bar{v} and \bar{w} , after collision. The unit vectors are $\kappa = \widehat{v-w}$, $\sigma = \widehat{\bar{v}-\bar{w}}$, $\omega = \widehat{v-\bar{v}} = \widehat{\kappa-\sigma}$. The red dotted line represents the space generated by $\pm\omega$. The angles are given by $\cos \phi = \kappa \cdot \omega$, $\phi \in [0, \pi]$. As $\omega \cdot (\sigma + \kappa) = 0$, the angles are related by $\hat{\phi} = (\pi - \phi)/2$.

5.1. Collision Kinematics

We review our notations, see our Figure 1, cf. [32, Figure 1]. We denote by $v, w \in \mathbb{R}^3$ the velocities before collision, while the velocities after collision are denoted by $\bar{v}, \bar{w} \in \mathbb{R}^3$. The quadruple $(v, w, \bar{v}, \bar{w}) \in (\mathbb{R}^3)^4$, is assumed to satisfy the conservation laws

$$F_1(v, w, \bar{v}, \bar{w}) = v + w - (\bar{v} + \bar{w}) = 0, \tag{4}$$

$$F_2(v, w, \bar{v}, \bar{w}) = |v|^2 + |w|^2 - (|\bar{v}|^2 + |\bar{w}|^2) = 0, \tag{5}$$

which define an algebraic variety \mathcal{M} that we expect to have dimension $12 - (3 + 1) = 8$. The Jacobian matrix of the four defining Equations (4) and (5) is

$$\begin{array}{l}
 (4)_1 \\
 (4)_2 \\
 (4)_3 \\
 (5)
 \end{array}
 \begin{bmatrix}
 v_1 & v_2 & v_3 & w_1 & w_2 & w_3 & \bar{v}_1 & \bar{v}_2 & \bar{v}_3 & \bar{w}_1 & \bar{w}_2 & \bar{w}_3 \\
 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 \\
 2v_1 & 2v_2 & 2v_3 & 2w_1 & 2w_2 & 2w_3 & -2\bar{v}_1 & -2\bar{v}_2 & -2\bar{v}_3 & -2\bar{w}_1 & -2\bar{w}_2 & -2\bar{w}_3
 \end{bmatrix}
 \tag{6}$$

The Jacobian matrix in Equation (6) has in general position full rank, and rank 3 if $v = w = \bar{v} = \bar{w}$. We denote by \mathcal{M}_* the 8 dimensional manifold $\mathcal{M} \setminus \{v = w = \bar{v} = \bar{w}\}$. In the sequel, for $v \neq w$, we set

$$\widehat{v-w} = \frac{v-w}{|v-w|}.$$

From (4) and (5) it follows the conservation of both the scalar product, $v \cdot w = \bar{v} \cdot \bar{w}$ and of the norm of the difference, $|v-w| = |\bar{v}-\bar{w}|$, so that all the vectors of the quadruple lie on a circle with center $z = (v+w)/2 = (\bar{v}+\bar{w})/2$ and are the four vertexes of a rectangle. If $v = w$ then $\bar{v} = \bar{w}$, and also $v = w = \bar{v} = \bar{w}$ as the circle collapse to one point, hence we have $\mathcal{M}_* = \mathcal{M} \setminus \{v = w\} = \mathcal{M} \setminus \{\bar{v} = \bar{w}\}$.

There are various explicit and interesting parametrizations of \mathcal{M}_* available.

An elementary parametrization consists of any algebraic solution of Equations (4) and (5) with respect to any of the free 8 coordinates. Other parametrizations are used in the literature, see classical references on the Boltzmann equation, e.g., [5].

A first parametrization is

$$(u, v, \sigma) \in (\mathbb{R}^3 \times \mathbb{R}^3)_* \times \mathbb{S}^2 \mapsto (u, v, \hat{A}_\sigma(u, v)) \in \mathcal{M}_* \subset (\mathbb{R}^3)^4, \tag{7}$$

where $\mathbb{S}^2 = \{\sigma \in \mathbb{R}^3 \mid |\sigma| = 1\}$ and the collision transformation $\hat{A}_\sigma: (v, w) \mapsto (v, w) = (v_\sigma, w_\sigma)$ is:

$$\hat{A}_\sigma: \begin{cases} v_\sigma = \frac{v+w}{2} + \frac{|v-w|}{2}\sigma \\ w_\sigma = \frac{v+w}{2} - \frac{|v-w|}{2}\sigma \end{cases}. \tag{8}$$

Viceversa, on \mathcal{M}_* the collision transformation depends on the unit vector $\sigma = \widehat{v-w} \in \mathbb{S}^2$, while the other terms depend on the collision invariants, as $|v-w|^2 = 2(|v|^2 + |w|^2) - |v+w|^2$. In conclusion, the transformation in Equation (7) is 1-to-1 from $(\mathbb{R}^3 \times \mathbb{R}^3)_* \times \mathbb{S}^2$ to \mathcal{M}_* , where $(\mathbb{R}^3 \times \mathbb{R}^3)_* = \{(u, v) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid v \neq w\}$.

A second parametrization of \mathcal{M}_* is obtained using the common space of two parallel sides of the velocity's rectangle, $\text{Span}(v - \bar{v}) = \text{Span}(\bar{w} - w)$, so that $v - \bar{v} = \bar{w} - w = \Pi(v - w)$, where Π is the orthogonal projection on the subspace. Vice versa, given any Π in the set $\Pi(1)$ of projections of rank 1, the mapping

$$A_\Pi = \begin{bmatrix} (I - \Pi) & \Pi \\ \Pi & (I - \Pi) \end{bmatrix}, \quad \begin{cases} v_\Pi = v - \Pi(v - w) = (I - \Pi)v + \Pi w \\ w_\Pi = w + \Pi(v - w) = \Pi v + (I - \Pi)w \end{cases}. \tag{9}$$

The components in the direction of the image of Π are exchanged, $\Pi v_\Pi = \Pi w$ and $\Pi w_\Pi = \Pi v$, while the orthogonal components are conserved. If ω is any of the two unit vectors such that $\Pi = \omega \otimes \omega'$, the matrix $A_\Pi = A_{\omega \otimes \omega'}$ does not depend on the direction of $\omega \in \mathbb{S}^2$. Notice that $A'_\Pi = A_\Pi$ and $A_\Pi A'_\Pi = I_6$, that is A_Π is an orthogonal symmetric matrix. This parametrization uses the set $\Pi(1)$ of projection matrices of rank 1,

$$(\mathbb{R}^3 \times \mathbb{R}^3)_* \times \Pi(1) \ni (u, v, \Pi) \mapsto (u, v, A_\Pi(u, v)) \in \mathcal{M}_* \in (\mathbb{R}^3)^4. \tag{10}$$

The σ -parametrization (7) and the Π -parametrization (10) are related as follows. Given the unit vector $\kappa = \widehat{v-w}$, the parameters Π and σ in Equations (9) and (8) are in 1-to-1 relation as

$$\sigma = (I - 2\Pi)\kappa, \quad \Pi = \widehat{\kappa - \sigma} \otimes \widehat{\kappa - \sigma}. \tag{11}$$

The transition map from the parametrization (7) to the parametrization (10) is

$$(v, w, \sigma) \mapsto (v, w, \widehat{v-w - \sigma} \otimes \widehat{v-w - \sigma}),$$

while the inverse transition is

$$(v, w, \Pi) \mapsto (v, w, (I - 2\Pi)\widehat{v-w}).$$

5.2. Uniform Probabilities on \mathbb{S}^2 and on $\Pi(1)$

Let μ be the uniform probability on \mathbb{S}^2 , computed, for example, in polar coordinates by

$$\int_{\mathbb{S}^2} f(\sigma) \mu(d\sigma) = \frac{1}{4\pi} \int_0^\pi \sin \varphi d\varphi \int_0^{2\pi} f(\sin \varphi \cos \theta u_1 + \sin \varphi \sin \theta u_2 + \cos \varphi u_3) d\theta, \tag{12}$$

where u_1, u_2, u_3 is any orthonormal basis of \mathbb{R}^3 that is $U = [u_1 u_2 u_3] \in \text{SO}(3)$. In such a way, $U: \mathbb{S}^2 \mapsto \mathbb{S}^2$ and the right hand side of Equation (12) does not depend on U .

As the mapping $\omega \in \mathbb{S}^2 \mapsto \Pi = \omega \otimes \omega \in \Pi(1)$ is a 2-covering, we define the image ν of μ by the equation

$$\int_{\Pi(1)} g(\Pi) \nu(d\Pi) = \int_{\mathbb{S}^2} g(\omega \otimes \omega) \mu(d\omega) = 2 \int_{\{\sigma \in \mathbb{S}^2 | \kappa \cdot \sigma > 0\}} g(\omega \otimes \omega) \mu(d\omega), \tag{13}$$

where $\kappa \in \mathbb{S}^2$ is any unit vector used to split \mathbb{S}^2 in two parts, $\{\sigma | \kappa \cdot \sigma > 0\}$ and $\{\sigma | \kappa \cdot \sigma < 0\}$. Equation (13) defines a probability ν on $\Pi(1)$ such that we have the invariance

$$\int_{\Pi(1)} g(U\Pi U') \nu(d\Pi) = \int_{\Pi(1)} g(\Pi) d\nu(\Pi), \quad U \in \text{SO}(3).$$

Let us compute the image $T_\kappa^* \mu$ of the uniform measure μ under the action of the transformation $T_\kappa: \sigma \mapsto \omega = \widehat{\kappa - \sigma}$. If in Equation (12) we take $u_3 = \kappa$, that is an orthonormal basis (u_1, u_2, κ) , then, for ϕ, θ such that $\sigma = \sin \phi \cos \theta u_1 + \sin \phi \sin \theta u_2 + \cos \phi \kappa$ and $\hat{\phi} = (\pi - \phi)/2$ (see Figure 1) we have

$$\begin{aligned} \omega = \widehat{\kappa - \sigma} &= \sin \hat{\phi} \cos \theta u_1 + \sin \hat{\phi} \sin \theta u_2 + \cos \hat{\phi} \kappa \\ &= \sin \left(\frac{\pi - \phi}{2} \right) \cos \theta u_1 + \sin \left(\frac{\pi - \phi}{2} \right) \sin \theta u_2 + \cos \left(\frac{\pi - \phi}{2} \right) \kappa \\ &= \cos \left(\frac{\phi}{2} \right) \cos \theta u_1 + \cos \left(\frac{\phi}{2} \right) \sin \theta u_2 + \sin \left(\frac{\phi}{2} \right) \kappa \end{aligned} \tag{14}$$

and for all integrable $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ one has

$$\begin{aligned} \int_{\mathbb{S}^2} f(\omega) T_\kappa^*(d\omega) &= \int_{\mathbb{S}^2} f(\widehat{\kappa - \sigma}) \mu(d\sigma) \\ &= \frac{1}{4\pi} \int_0^\pi \sin \phi d\phi \int_0^{2\pi} d\theta f \left(\cos \left(\frac{\phi}{2} \right) \cos \theta u_1 + \cos \left(\frac{\phi}{2} \right) \sin \theta u_2 + \sin \left(\frac{\phi}{2} \right) \kappa \right) \\ &= \frac{1}{4\pi} \int_0^{\pi/2} 2 \sin(2\phi) d\phi \int_0^{2\pi} d\theta f(\cos \phi \cos \theta u_1 + \cos \phi \sin \theta u_2 + \sin \phi \kappa) \\ &= \frac{1}{4\pi} \int_0^{\pi/2} 4 \cos \phi \sin(\phi) d\phi \int_0^{2\pi} d\theta f(\cos \phi \cos \theta u_1 + \cos \phi \sin \theta u_2 + \sin \phi \kappa) \\ &= \int_{\{\sigma \in \mathbb{S}^2 | \kappa \cdot \omega \geq 0\}} f(\omega) 4(\kappa \cdot \omega) \mu(d\omega), \end{aligned} \tag{15}$$

compare [32, Section 4.5].

In particular, for a symmetric function, $f(\omega) = f(-\omega)$, we have

$$\int_{\mathbb{S}^2} f(\omega) T_\kappa^*(d\omega) = \int_{\mathbb{S}^2} f(\widehat{\kappa - \sigma}) \mu(d\sigma) = \int_{\mathbb{S}^2} f(\omega) 2 |\kappa \cdot \omega| \mu(d\omega). \tag{16}$$

It follows, for each integrable $g: \Pi(1) \rightarrow \mathbb{R}$, that

$$\int_{\mathbb{S}^2} g(\widehat{(\kappa - \sigma)} \otimes \widehat{(\kappa - \sigma)}) \mu(d\sigma) = \int_{\mathbb{S}^2} g(\omega \otimes \omega) 2|\kappa \cdot \omega| \mu(d\omega) \tag{17}$$

Notice that if we integrate Equation (15) with respect to κ we obtain

$$\iint_{\mathbb{S}^2 \times \mathbb{S}^2} f(\widehat{(\kappa - \sigma)}) \mu(d\sigma)\mu(d\kappa) = \int_{\mathbb{S}^2} f(\sigma) \mu(d\sigma), \tag{18}$$

because

$$\int_{\{\kappa \in \mathbb{S}^2 | \kappa \cdot \sigma \geq 0\}} \kappa \cdot \sigma \mu(d\kappa) = \frac{1}{4}. \tag{19}$$

5.3. Conditioning on the Collision Invariants

Given a function $g: \mathbb{R}^3 \times \mathbb{R}^3$, Equation (8) shows that the function

$$\hat{g}: (u, v) \mapsto \int_{\mathbb{S}^2} g(v_\sigma, w_\sigma) \mu(d\sigma) = \int_{\mathbb{S}^2} g(\hat{A}_\sigma(v, w)) \mu(d\sigma) \tag{20}$$

depends on the collision invariants only. This, in turn, implies that \hat{g} is the conditional expectation of g with respect of the collision invariants under any probability distribution on $\mathbb{R}^3 \times \mathbb{R}^3$ such that the collision invariants and the unit vector of σ are independent, the unit vector σ being uniformly distributed. See below a more precise statement in the case of the Gaussian distribution.

On the sample space (\mathbb{R}^3, dv) , let M be the standard normal density defined in (3) (the Maxwell density). As, for all $\Pi \in \Pi(1)$, $A_\Pi A'_\Pi = A_\Pi A_\Pi = I_6$, in particular $|\det A_\Pi| = 1$, we have for each $(V, W) \sim M \otimes M$, (i.e., V, W are i.i.d. with distribution M) that

$$A_\Pi(V, W) = (V_\omega, W_\omega) \sim (V, W). \tag{21}$$

Under the same distributions, the random variables $\frac{V+W}{\sqrt{2}}$, $\frac{|V-W|}{\sqrt{2}}$, $\widehat{V-W}$, are independent, with distributions given by

$$\frac{V+W}{\sqrt{2}} \sim N(0_3, I_3), \quad \frac{|V-W|^2}{2} \sim \chi^2(3), \quad \widehat{V-W} \sim \mu, \tag{22}$$

respectively. Hence, given any $S \sim \mu$ such that $\frac{V+W}{\sqrt{2}}$, $\frac{|V-W|}{\sqrt{2}}$, S , are independent, we get

$$\hat{A}_S(V, W) \sim (V, W). \tag{23}$$

This equality of distribution generalizes the equality of random variables $\hat{A}_{\widehat{V-W}}(V, W) = (V, W)$. We state the results obtained above as follows.

The image distribution of $M \otimes M \otimes U$ induced on $(\mathbb{R}^3)^4$ by the parametrization in Equation (7) is supported by the manifold \mathcal{M}_* . Such a distribution has the property that the projections on both the first two and the last two components are $M \otimes M$. The joint distribution is not Gaussian; in fact the support \mathcal{M}_* is not a linear subspace. We will call this distribution the *normal collision distribution*.

The second parametrization in Equation (9) shows that the variety \mathcal{M} contains the bundle of linear spaces

$$(v, w) \mapsto (v, w, (I - \Pi)v + \Pi w, \Pi v + (I - \Pi)w), \quad \pi \in \Pi(1).$$

The distribution of

$$\mathcal{M} \ni (v, w, \bar{v}, \bar{w}) \mapsto \Pi = \widehat{v - \bar{v}} \otimes \widehat{v - \bar{w}} \in \Pi(1) \tag{24}$$

under the normal collision distribution is obtained from Equation (11). In fact Π is the projector on the subspace generated by $\kappa - \sigma$ where $(v, w, \sigma) \mapsto \kappa = \widehat{v - w}$ is uniformly distributed and independent from σ . Hence, Equation (18) shows that $(v, w, \sigma) \mapsto \widehat{\kappa - \sigma}$ is uniformly distributed on \mathbb{S}^2 so that the distribution in Equation (24) is the ν measure defined in Equation (13). Conditionally to Π , the normal collision distribution is Gaussian with covariance

$$\begin{bmatrix} I & 0 & I - \Pi & \Pi \\ 0 & I & \Pi & I - \Pi \\ I - \Pi & \Pi & I & 0 \\ \Pi & I - \Pi & 0 & I \end{bmatrix} .$$

We can give the previous remarks a more probabilistic form as follows.

Proposition 10 (Conditioning). *Let M be the density of the standard normal $N(0_3, I_3)$ and $g: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ be an integrable function. It holds the following*

1. *If $(V, W) \sim M \otimes M$ and $\Pi \in \Pi(1)$, then*

$$\mathbb{E} (g(V, W) | V + W, |V|^2 + |W|^2) = \mathbb{E} (g(A_\Pi(V, W)) | V + W, |V|^2 + |W|^2)$$

2. *If $(V, W) \sim M \otimes M$, then*

$$\int_{\mathbb{S}^2} g(\hat{A}_\sigma(V, W)) \mu(d\sigma) = \mathbb{E} (g(V, W) | V + W, |V|^2 + |W|^2) .$$

3. *If $(V, W) \sim M \otimes M$, then*

$$2 \int_{\mathbb{S}^2} g(A_\Pi(V, W)) \left| \Pi(\widehat{V - W}) \right| \nu(d\Pi) = \mathbb{E} (g(V, W) | V + W, |V|^2 + |W|^2) .$$

4. *Assume $(V, W) \sim F$, $F \in \mathcal{E}(M \otimes M)$, with $F(v, w) = f(v, w)M(v)M(w)$. Then*

$$\left(\int_{\mathbb{S}^2} f \circ \hat{A}_\sigma \mu(d\sigma) \right) \cdot M \otimes M \in \mathcal{E}(M \otimes M) ,$$

and

$$\begin{aligned} & \mathbb{E}_{f \cdot M \otimes M} (g(V, W) | V + W, |V|^2 + |W|^2) \\ &= \frac{\int_{\mathbb{S}^2} g(\hat{A}_\sigma(V, W)) f(\hat{A}_\sigma(V, W)) \mu(d\sigma)}{\int_{\mathbb{S}^2} f(\hat{A}_\sigma(V, W)) \mu(d\sigma)} \\ &= \frac{\int_{\mathbb{S}^2} g(A_\Pi(V, W)) f(A_\Pi(V, W)) \left| \Pi(\widehat{V - W}) \right| \nu(d\Pi)}{\int_{\mathbb{S}^2} f(A_\Pi(V, W)) \left| \Pi(\widehat{V - W}) \right| \nu(d\Pi)} . \end{aligned}$$

Proof.

1. We use $V + W = V_{\Pi} + W_{\Pi}$, $|V|^2 + |W|^2 = |V_{\Pi}|^2 + |W_{\Pi}|^2$, $(V, W) \sim (V_{\Pi}, W_{\Pi})$. For all bounded $h_1: \mathbb{R}^3 \rightarrow \mathbb{R}$ and $h_2: \mathbb{R} \rightarrow \mathbb{R}$, we have

$$\begin{aligned} & \mathbb{E} \left(\mathbb{E} \left(g \left(A_{\Pi}(V, W) \right) \middle| V + W, |V|^2 + |W|^2 \right) h_1(V + W) h_2(|V|^2 + |W|^2) \right) \\ &= \mathbb{E} \left(g(V_{\Pi}, W_{\Pi}) h_1(V + W) h_2(|V|^2 + |W|^2) \right) \\ &= \mathbb{E} \left(g(V_{\omega}, W_{\omega}) h_1(V_{\Pi} + W_{\Pi}) h_2(|V_{\omega}|^2 + |W_{\omega}|^2) \right) \\ &= \mathbb{E} \left(g(V, W) h_1(V + W) h_2(|V|^2 + |W|^2) \right) \\ &= \mathbb{E} \left(\mathbb{E} \left(g(V, W) \middle| V + W, |V|^2 + |W|^2 \right) h_1(V + W) h_2(|V|^2 + |W|^2) \right). \end{aligned}$$

2. For a generic integrable $h: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ we have

$$\begin{aligned} & \mathbb{E} (h(V, W)) \\ &= \mathbb{E} \left(h \left(\frac{V + W}{2} + \frac{|V - W|}{2} \widehat{V - W}, \frac{V + W}{2} - \frac{|V - W|}{2} \widehat{V - W} \right) \right) \\ &= \int_{\mathbb{S}^2} \mathbb{E} \left(h \left(\frac{V + W}{2} + \frac{|V - W|}{2} \sigma, \frac{V + W}{2} - \frac{|V - W|}{2} \sigma \right) \right) \mu(d\sigma) \\ &= \int_{\mathbb{S}^2} \mathbb{E} \left(h(\hat{A}_{\sigma}(V, W)) \right) \mu(d\sigma), \end{aligned}$$

because $\widehat{V - W} \sim \mu$, and $(V + W), |V - W|, \widehat{V - W}$ are independent.

The random variable

$$\int_{\mathbb{S}^2} g \circ \hat{A}_{\sigma}(V, W) \mu(d\sigma) = \int_{\mathbb{S}^2} g \left(\frac{V + W}{2} + \frac{|V - W|}{2} \sigma, \frac{V + W}{2} - \frac{|V - W|}{2} \sigma \right) \mu(d\sigma)$$

is a function of the collision invariants, *i.e.*, it is of the form $\tilde{g}(V + W, |V|^2 + |W|^2)$. For all bounded $h_1: \mathbb{R}^3 \rightarrow \mathbb{R}$ and $h_2: \mathbb{R} \rightarrow \mathbb{R}$, we apply the previous computation to $h = gh_1h_2$ to get

$$\begin{aligned} & \mathbb{E} \left(\left(\int_{\mathbb{S}^2} g \circ \hat{A}_{\sigma}(V, W) \mu(d\sigma) \right) h_1(V + W) h_2(|V|^2 + |W|^2) \right) \\ &= \int_{\mathbb{S}^2} \mathbb{E} \left(g \circ \hat{A}_{\sigma}(V, W) h_1(V + W) h_2(|V|^2 + |W|^2) \right) \mu(d\sigma) \\ &= \int_{\mathbb{S}^2} \mathbb{E} \left(g \circ \hat{A}_{\sigma}(V, W) h_1(V_{\sigma} + W_{\sigma}) h_2(|V_{\sigma}|^2 + |W_{\sigma}|^2) \right) \mu(d\sigma) \\ &= \mathbb{E} (g(V, W) h_1(V + W) h_2(|V|^2 + |W|^2)). \end{aligned}$$

3. We use Item 2 and the equality $\hat{A}_{\sigma}(v, w) = A_{\Pi}(v, w)$ when $\Pi = \widehat{\kappa - \sigma} \otimes \widehat{\kappa - \sigma}$ and $\kappa = \widehat{v - w}$ to write

$$\int_{\mathbb{S}^2} g \left(A_{\widehat{V - W - \sigma} \otimes \widehat{V - W - \sigma}}(V, W) \right) \mu(d\sigma) = \mathbb{E} (g(V, W) |V + W, |V|^2 + |W|^2)$$

where, for given vectors $u, v \in \mathbb{R}^3$, $u, v \neq 0$ we simply denote $u \otimes v = \widehat{u} \otimes \widehat{v}$. From Equation (17), the left-end-side can be rewritten as an integral with respect to $\omega \in \mathbb{S}^2$,

$$\int_{\mathbb{S}^2} g \left(A_{\omega \otimes \omega}(V, W) \right) 2 \left| \widehat{V - W} \cdot \omega \right| \mu(d\omega) = \int_{\mathbb{S}^2} g \left(A_{\widehat{V - W - \sigma} \otimes \widehat{V - W - \sigma}}(V, W) \right) \mu(d\sigma).$$

If $\Pi = \omega \otimes \omega$, then $|\kappa \cdot \omega| = |\omega \otimes \omega \kappa| = |\Pi \kappa|$. Using that together with the definition of the measure ν on $\Pi(1)$ in Equation (13), we have the result:

$$\begin{aligned} & \int_{\mathbb{S}^2} g(A_{\omega \otimes \omega}(V, W)) 2 \left| \widehat{V - W} \cdot \omega \right| \mu(d\omega) \\ &= \int_{\mathbb{S}^2} g(A_{\omega \otimes \omega}(V, W)) 2 \left| (\omega \otimes \omega) \widehat{V - W} \right| \mu(d\omega) \\ &= 2 \int_{\Pi(1)} g(A_{\Pi}(V, W)) \left| \Pi(\widehat{V - W}) \right| \nu(d\Pi). \end{aligned}$$

4. We use Theorem 2. If $F \in \mathcal{E}(M \otimes M)$, then

$$F = e^{U - K_{M \otimes M}(U)} M \otimes M, \quad U \in \mathcal{S}_{M \otimes M},$$

and there exists a neighborhood I of $[0, 1]$, where the one dimensional exponential family

$$F_t = e^{tU - K_0(tM)} M \otimes M, \quad t \in I,$$

exists. The random variable

$$\hat{f} = \int_{\mathbb{S}^2} f \circ \hat{A}_\sigma \mu(d\sigma) \tag{25}$$

is a positive probability density with respect to $M \otimes M$ because it is the conditional expectation in $M \otimes M$ of the positive density f .

In order to show that $\mathbb{E}_{M \otimes M} [(\hat{f})^t] < +\infty$ for $t \in I$, it is enough to consider the convex cases, $t < 0$ and $t > 1$, because otherwise $\mathbb{E}_{M \otimes M} [(\hat{f})^t] \leq 1$. We have

$$\int_{\mathbb{S}^2} f \circ \hat{A}_\sigma \mu(d\sigma) = \int_{\mathbb{S}^2} e^{U \circ \hat{A}_\sigma - K_{M \otimes M}(U)} \mu(d\sigma),$$

so that in the convex cases:

$$\begin{aligned} & \mathbb{E}_{M \otimes M} \left[\left(\int_{\mathbb{S}^2} f \circ \hat{A}_\sigma \mu(d\sigma) \right)^t \right] \\ &= \mathbb{E}_{M \otimes M} \left[\left(\int_{\mathbb{S}^2} e^{U \circ \hat{A}_\sigma - K_{M \otimes M}(U)} \mu(d\sigma) \right)^t \right] \\ &\leq \mathbb{E}_{M \otimes M} \left[\int_{\mathbb{S}^2} e^{tU \circ \hat{A}_\sigma - tK_{M \otimes M}(U)} \mu(d\sigma) \right] \\ &= \mathbb{E}_{M \otimes M} [e^{tU - tK_{M \otimes M}(U)}] \\ &= e^{K_{M \otimes M}(tU) - tK_{M \otimes M}(U)} < +\infty, \quad t \in I \setminus [0, 1]. \end{aligned}$$

To conclude, use Bayes' formula for conditional expectation,

$$\mathbb{E}_{f.M \otimes M} (A|\mathcal{B}) = \frac{\mathbb{E}_{M \otimes M} (Af|\mathcal{B})}{\mathbb{E}_{M \otimes M} (f|\mathcal{B})}.$$

to the expressions of conditional expectation in Item 2 and Item 3 above.

□

Remark 4. In the last Item, we compute a conditional expectation of a density f , that is

$$\mathbb{E} (f(V, W)|V + W, |V|^2 + |W|^2) = \hat{f}(V + W, |V|^2 + |W|^2).$$

The random variable $\hat{f}: \mathbb{R}^3 \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is the density of the image of F $dvdw$ with respect to the image of $M(v)M(w) dvdw$.

5.4. Interactions

We introduce here the crucial role played by microscopic interaction in the definition of the Boltzmann collision operator. In the physics literature, such interaction are referred to as the kinetic collision kernel and takes into account the intermolecular forces suffered by particles during a collision [5]. Before defining formally what we mean by interaction, we first observe that, if M is the Maxwell density on \mathbb{R}^3 and $f, g \smile M$ then

$$f \otimes g \smile M \otimes M$$

where $M \otimes M$ is the standard normal density on \mathbb{R}^6 and $f \otimes g$ is a density on \mathbb{R}^6 . Indeed, one has

$$\begin{aligned} f(v) &= e^{U(v)-K_M(U)} M(v), \quad U \in \mathcal{S}_M, \\ g(w) &= e^{V(w)-K_M(V)} M(w), \quad V \in \mathcal{S}_M. \end{aligned}$$

It follows that the product density has the form

$$f(v)g(w) = e^{U(v)+V(w)-K_M(U)-K_M(V)} M(v)M(w), \quad U \oplus V \in \mathcal{S}_M \otimes \mathcal{S}_M \subset \mathcal{S}_{M \otimes M},$$

which implies $f \otimes g \smile M \otimes M$, with $K_{M \otimes M}(U \oplus V) = K_M(U) + K_M(V)$.

Definition 9. With the previous notations, we say that $b: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^+$ is an interaction on $\mathcal{E}(M) \otimes \mathcal{E}(M)$, if $\mathbb{E}_{f \otimes g}[b] < +\infty$, so that

$$\frac{b}{\mathbb{E}_{f \otimes g}[b]} f \otimes g: (v, w) \in \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \frac{b(v, w)}{\mathbb{E}_{f \otimes g}[b]} f(v)g(w)$$

is a density.

Sometimes, we make the abuse of notation by writing $\mathbb{E}_{b, f \otimes g}[\cdot]$, where the obvious normalization is not written down. As can be seen, here interactions indicate only a class of suitable weight functions b for which $b(v, w)f(v)g(w)$ is still (up to normalization) a density. This is in accordance with the usual role played by the kinetic collision kernel (see [5]).

According to the Portmanteau Theorem 2, it holds $\frac{b}{\mathbb{E}_{f \otimes g}[b]} \cdot f \otimes g \smile M \otimes M$ if and only if $\frac{b}{\mathbb{E}_{f \otimes g}[b]} \cdot f \otimes g \smile f \otimes g$, which, in turn, is equivalent to $\mathbb{E}_{f \otimes g}[b^{1+\epsilon}], \mathbb{E}_{b, f \otimes g}[b^{-1-\epsilon}] < \infty$ for some $\epsilon > 0$.

Proposition 11. Let $b: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^+$ be such that for some real $A \in \mathbb{R}$ and positive $B, C, \lambda \in \mathbb{R}_{>}$, it holds

$$C|u - v|^\lambda \leq b(u, v) \leq A + B|u - v|^2.$$

Then the b is an interaction on $\mathcal{E}(M) \times \mathcal{E}(M)$ and for all $f, g \smile M$ the following holds.

1. $\frac{b}{\mathbb{E}_{f \otimes g}[b]} \cdot f \otimes g \smile M \otimes M$.
2. Assume moreover that the interaction b is a function of the invariants only

$$b(v, w) = \hat{b}(v + w, |v|^2 + |w|^2) .$$

It follows

$$\mathbb{E}_{M \otimes M} (b \cdot f \otimes g | V + W, |V|^2 + |W|^2) = b(V, W) \int_{\mathbb{S}^2} f \otimes g \circ \hat{A}_\sigma(V, W) \mu(d\sigma) .$$

and moreover a sufficiency relation holds, i.e., for all integrable $F: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$, it holds

$$\mathbb{E}_{b \cdot f \otimes g} (F(V, W) | V + W, |V|^2 + |W|^2) = \mathbb{E}_{f \otimes g} (F(V, W) | V + W, |V|^2 + |W|^2) .$$

Proof. 1. We can assume $\mathbb{E}_{f \otimes g} [b] = 1$. As $b(u, v) \leq A + B|u - v|^2 \leq A + 2B|(u, v)|^2$, we have $b \in L^\Phi(M \otimes M) = L^\Phi(f \otimes g)$, and hence $b \in L^{1+\epsilon}(f \otimes g)$ for all $\epsilon > 0$. For the second inequality we use the Hardy-Littlewood-Sobolev inequality [34, Theorem 4.3]. We have

$$\mathbb{E}_{b \cdot f \otimes g} [b^{-1-\epsilon}] = \iint \frac{f(u)g(v)}{b(u, v)^\epsilon} dudv \leq C^{-1} \iint \frac{f(u)g(v)}{|u - v|^{\epsilon\lambda}} dudv .$$

If $\frac{1}{\beta} + \frac{\epsilon\lambda}{3} + \frac{1}{\alpha} = 2$, by the H-L-S inequality the last integral is bounded by a constant times $\|f\|_\alpha \|g\|_\beta$. From $f, g \smile M$ we get that $\|f\|_\alpha \|g\|_\beta$ is finite for α, β in a right neighborhood of 1. There exists $\epsilon = \frac{3}{\lambda} \left(2 - \frac{1}{\alpha} - \frac{1}{\beta}\right) > 0$ satisfying all conditions.

2. It is a special case of the Conditioning Theorem 10. \square

Let us discuss the differentiability of the operations we have just introduced.

Proposition 12.

1. The product mapping $\mathcal{E}(M) \ni f \mapsto f \otimes f$ is a differentiable map into $\mathcal{E}(M \otimes M)$ with tangent mapping given for any vector field $X \in T\mathcal{E}(M)$ by

$$T_f \mathcal{E}(M) \ni X_f \mapsto X_f \oplus X_f \in T_{f \otimes f} \mathcal{E}(M \otimes M) .$$

2. Let b be an interaction on $\mathcal{E}(M) \otimes \mathcal{E}(M)$. If the mapping

$$f \mapsto \frac{b}{\mathbb{E}_{f \otimes f}[b]} \cdot f \otimes f$$

is defined on $\mathcal{E}(M)$ with values in $\mathcal{E}(M \otimes M)$, then it is differentiable with tangent mapping given for all vector field $X \in T\mathcal{E}(M)$ by

$$X_f \mapsto X_f \oplus X_f - \mathbb{E}_{b \cdot f \otimes f} [X_f \oplus X_f] .$$

Proof. 1. We have already proved that $f \smile M$ implies $f \otimes f \smile M \otimes M$ and that the mapping in the charts centered at M and $M \otimes M$ respectively is represented as $\mathcal{S}_M \ni U \mapsto U + U \in \mathcal{S}_{M \otimes M}$. The differential of the linear map is again $T_M \mathcal{E}(M) \ni V \mapsto V \oplus V \in T_{M \otimes M} \mathcal{E}(M \otimes M)$. The transport commutes with the \oplus operation,

$$\mathbb{U}_{M \otimes M}^{f \otimes f} (V \oplus V) = \left(\mathbb{U}_M^f V\right) \oplus \left(\mathbb{U}_M^f V\right) ,$$

and the result follows.

2. Let U be the coordinate of f at M . By assumption, we have

$$\frac{b}{\mathbb{E}_{f \otimes f}[b]} \cdot f \otimes f \smile \frac{b}{\mathbb{E}_{M \otimes M}[b]} \cdot M \otimes M$$

and

$$\frac{\frac{b}{\mathbb{E}_{f \otimes f}[b]} \cdot f \otimes f}{\frac{b}{\mathbb{E}_{M \otimes M}[b]} \cdot M \otimes M} = \frac{\mathbb{E}_{M \otimes M}[b]}{\mathbb{E}_{f \otimes f}[b]} \exp(U \oplus U - 2K_M(U)) ,$$

so that the coordinate of $\frac{b}{\mathbb{E}_{f \otimes f}[b]} \cdot f \otimes f$ in the chart centered at $\frac{b}{\mathbb{E}_{M \otimes M}[b]} \cdot M \otimes M$ is

$$U \oplus U - \mathbb{E}_{b \cdot M \otimes M}[U \oplus U] = \mathbb{U}_{M \otimes M}^{b \cdot M \otimes M}(U \oplus U).$$

The expression is linear and so is the expression of the tangent map

$$V \mapsto V \oplus V - \mathbb{E}_{b \cdot M \otimes M}[V \oplus V] .$$

In conclusion, for each vector field X of $T\mathcal{E}(M)$, at f we have $V = \mathbb{U}_f^M X_f$ and $V \oplus V = \mathbb{U}_{f \oplus f}^{M \oplus M}(X_f \oplus X_f)$, hence the action on $T_f\mathcal{E}(M)$ is as stated.

□

Assume $f \smile M$ and let b be an interaction on $f \otimes f$ which depends on the invariants only and such that $b \cdot f \otimes f \smile M \otimes M$. For each random variable $g \in L^{\cosh^{-1}}(M) = L^{\cosh^{-1}}(f)$, define $\bar{g} = \frac{1}{2} g \oplus g$, which belongs to $L^{\cosh^{-1}}(M \otimes M) = L^{\cosh^{-1}}(f \otimes f)$. Define the operator

$$A: L^{\cosh^{-1}}(M) \rightarrow L^{\cosh^{-1}}(M \otimes M) ,$$

by

$$Ag(v, w) = \int_{\mathbb{S}^2} \bar{g} \left(\hat{A}_\sigma(v, w) \right) \mu(d\sigma) - \bar{g}(V, W) .$$

As constant random variables are in the kernel of the operator A , we assume $\mathbb{E}_f[g] = 0$.

5.5. Maxwell-Boltzmann and Boltzmann Operator

As explained by C. Villani [5, Section I.2.3], Maxwell obtained a weak form of Boltzmann operator before Boltzmann himself. We rephrase in geometric-probabilistic language such a Maxwell's weak form by expanding and rigorously proving what was hinted to in [19].

Let $b: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^+$ be an interaction on $\mathcal{E}(M) \times \mathcal{E}(M)$ which depends on the invariants only and such that $b \cdot f \otimes f \smile M \otimes M$ if $f \smile M$, cf. Proposition 11. We shall call such a b a *proper interaction*.

For each random variable $g \in L^{\cosh^{-1}}(M) = L^{\cosh^{-1}}(f)$, define $\bar{g} = \frac{1}{2} g \oplus g$ by

$$\bar{g}(v, w) = \frac{1}{2} (g(v) + g(w))$$

which it is easily shown to belong to $L^{\cosh^{-1}}(M \otimes M) = L^{\cosh^{-1}}(f \otimes f)$. The mapping $g \mapsto \bar{g}$ is a version of the conditional expectation $\mathbb{E}_{M \otimes M}(g|\mathcal{S})$, where \mathcal{S} is the σ -algebra generated by symmetric random variables.

We define the operator

$$A: L^{\cosh^{-1}}(M) \rightarrow L^{\cosh^{-1}}(M \otimes M) ,$$

by

$$Ag(v, w) = \int_{\mathbb{S}^2} \bar{g} \left(\hat{A}_\sigma(v, w) \right) \mu(d\sigma) - \bar{g}(V, W) ,$$

which is a version of

$$Ag = \mathbb{E}_{M \otimes M} (\mathbb{E}_{M \otimes M} (g|\mathcal{S})|\mathcal{I}) - \mathbb{E}_{M \otimes M} (g|\mathcal{S}) = \mathbb{E}_{M \otimes M} (g|\mathcal{I}) - \mathbb{E}_{M \otimes M} (g|\mathcal{S}) ,$$

where $\mathcal{I} \subset \mathcal{S}$ is the σ -algebra generated by the collision invariants $(v, w) \mapsto (v + w, |v|^2 + |w|^2)$.

As constant random variables are in the kernel of the operator A , we assume $\mathbb{E}_f [g] = 0$. Analogously, as the kernel of the operator contains all symmetric random variables, we could always assume that g is anti-symmetric.

The nonlinear operator $f \mapsto \mathbb{E}_{b, f \otimes f} [Ag]$ is the Maxwell's weak form of the Boltzmann operator, g being a test function.

Proposition 13. *Given a proper interaction b and a density $f \in \mathcal{E} (M)$, the linear map*

$$L_0^{\cosh^{-1}}(M) \ni g \mapsto \mathbb{E}_{b, f \otimes f} [Ag]$$

is continuous.

1. It can be represented in the duality $L^{(\cosh^{-1})^*}(f) \times L^{\cosh^{-1}}(f)$ by

$$\begin{aligned} \mathbb{E}_{b, f \otimes f} [Ag] &= \mathbb{E}_{M \otimes M} \left[b \left(\mathbb{E}_{M \otimes M} \left(\frac{f}{M} \otimes \frac{f}{M} \middle| \mathcal{I} \right) - \frac{f}{M} \otimes \frac{f}{M} \right) g \right] \\ &= \langle Q(f)/f, g \rangle_f , \end{aligned}$$

where Q is the Boltzmann operator with interaction b ,

$$Q(f)(v) = \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} \left(f \otimes f \left(\hat{A}_\sigma(v, w) \right) - f(v)f(w) \right) b(v, w) \mu(d\sigma) dw$$

2. Especially, if $f = e^{U-K_M(U)} M$ and we take $g = \log \left(\frac{f}{M} \right)$, then

$$\mathbb{E}_{b, f \otimes f} \left[A \log \left(\frac{f}{M} \right) \right] = \langle Q(f)/f, U \rangle_f .$$

Proof. The continuity follows from the Portmanteau theorem and the continuity of the conditional expectation. Item 1 follows from the projection properties of the conditional expectation and from general properties of Orlicz spaces. Item 2 is a special case of the previous one. \square

It follows from the previous theorem and from the discussion in [19, Proposition 10] that $f \mapsto Q(f)/f$ is a vector field in the cotangent bundle ${}^*T\mathcal{E} (M)$ and its flow

$$\frac{D}{dt} f_t = \frac{\dot{f}_t}{f_t} = \frac{Q(f_t)}{f_t}$$

is equivalent to the standard Boltzmann equation $\dot{f}_t = Q(f_t)$. We do not discuss in this paper the implications of that presentation of the Boltzmann equation to the existence properties. We turn our attention to the comparison of the Boltzmann field to the gradient field of the entropy.

5.6. Entropy Generation

If $t \mapsto p(t)$ is a curve in $\mathcal{E}(M)$, the entropy of $p(t)$ is defined for all t and the variation of the entropy along the curve is computed as

$$\begin{aligned} \frac{d}{dt}H(p_t) &= - \int \frac{d}{dt} (p(v; t) \log p(v; t)) \, dv \\ &= - \int (\log p(v; t) - 1) \frac{d}{dt} p(v; t) \, dv \\ &= - \int \log p(v; t) \frac{d}{dt} p(v; t) \, dv . \end{aligned}$$

In our setup, the computation takes the following form. Let $t \mapsto p_t$ be a differentiable curve in $\mathcal{E}(M)$ with velocity $t \mapsto Dp_t \in {}^*B_{p_t}$. As the gradient ∇H is given at p by $(\nabla H)(p) = -(\log p + H(p)) \in B_p$, we have

$$\frac{d}{dt}H(p(t)) = - \langle \log p(t) + H(p(t)), Dp(t) \rangle_{p(t)} .$$

In particular, if $Dp(t) = \dot{p}(v; t)/p(v; t)$ we recover the previous computation:

$$\begin{aligned} \frac{d}{dt}H(p(t)) &= - \int (\log p(t) + H(p(t))) \frac{\dot{p}(v; t)}{p(v; t)} p(v; t) \, dv \\ &= - \int \log p(v; t) \frac{d}{dt} p(v; t) \, dv . \end{aligned}$$

Assume now that $(p, t) \mapsto \gamma(t; p)$ is the flow of a vector field $F: \mathcal{E}(M) \rightarrow {}^*T\mathcal{E}(M)$. Then

$$\begin{aligned} \frac{d}{dt}H(\gamma(t; p)) &= - \langle \log \gamma(t; p) + H(\gamma(t; p)), D\gamma(t; p) \rangle_{\gamma(t; p)} \\ &= - \langle \log \gamma(t; p) + H(\gamma(t; p)), F(\gamma(t; p)) \rangle_{\gamma(t; p)} , \end{aligned}$$

that is, for each $p \in \mathcal{E}(M)$, the entropy production at p along the vector field F is

$$\begin{aligned} (\nabla_F H)(p) &= - \langle \log p + H(p), F(p) \rangle_p \\ &= - \int (\log p(v) + H(p)) F(v, p) p(x) \, dx \\ &= - \int \log p(v) F(v, p) p(v) \, dv . \end{aligned}$$

In particular, if the vector field F is the Boltzmann vector field, $F(f) = Q(f)/f$, we have that for each $f \in \mathcal{E}(M)$ the Boltzmann's entropy production is

$$\begin{aligned} \mathcal{D}(f) &= (\nabla_{Q(f)/f} H)(f) \\ &= - \langle \log f + H(f), Q(f)/f \rangle_f \\ &= -\mathbb{E}_{b, f \otimes f} [A \log f] \\ &= -\frac{1}{2} \mathbb{E}_{b, f \otimes f} [A (\log f \ominus \log f)] , \end{aligned}$$

where

$$(\log f \ominus \log f)(v, w) = \log f(v) - \log f(w) = \log \frac{f(v)}{f(w)} .$$

We recover a well-known formula for the entropy production of the Boltzmann operator [5]. We now proceed to compute the covariant derivative of the entropy production.

Proposition 14. Let X be a vector field of $T\mathcal{E}(M)$ and let F be a vector field of ${}^*T\mathcal{E}(M)$.

1. The Hessian of the entropy (in the exponential connection) is

$$D_X \nabla H(f) = -X.$$

2. The covariant derivative of the entropy production along F is

$$D_X \mathcal{D} = -\langle X, F \rangle + \langle \nabla H, D_X F \rangle.$$

Proof. We note that the entropy production along a vector field F , $\mathcal{D} = (\nabla_F H)(p) = \langle \nabla H, F \rangle$, is a function of the duality coupling of $T\mathcal{E}(M) \times {}^*T\mathcal{E}(M)$, so that we can apply Proposition 5 to compute its covariant derivative along X as

$$D_X \mathcal{D} = \langle D_X \nabla H, F \rangle + \langle \nabla H, D_X F \rangle.$$

The first term at p is

$$\langle D_X \nabla H, F \rangle(p) = \langle D_X \nabla H(p), F(p) \rangle_p.$$

Let us compute $D_X \nabla H(p)$, which is the Hessian of the entropy in the exponential connection. First, we compute the expression of $\nabla H(q) = -(\log q + H(q)) \in B_q$, $q \in \mathcal{E}(M)$ in the chart centered at p . We have

$$-\log q = K_p(U) - U - \log p, \quad U \in \mathcal{S}_p, q = e_p(U),$$

and

$$H_p(U) = -\mathbb{E}_q[\log q] = K_p(U) - dK_p(U)[U] + dK_p(U)[\nabla H(p)] + H(p)$$

so that

$$(\nabla H) \circ e_p(U) = -U + dK_p(U)[U] + \nabla H(p) - dK_p(U)[\nabla H(p)]$$

and, finally, the expression of ∇H in the chart centered at p is

$$(\nabla H)_p(U) = {}^e\mathbb{U}_{e_p(U)}^p(\nabla H) \circ e_p(U) = \nabla H(p) - U.$$

Note that this function is affine, and its derivative in the direction $X(p)$ is $d(\nabla H)_p(U)[X(p)] = -X(p)$. It follows that $D_X \nabla H = -X$. \square

The application of this computation to the Boltzmann field, *i.e.*, $F(f) = Q(f)/f$ requires the existence of the covariant derivative of the Boltzmann operator. We leave this discussion as a research plan.

6. Weighted Orlicz–Sobolev Model Space

We show in this section that the Information Geometry formalism described in Sections 2 and 3 is robust enough to allow to take into account differential operators (e.g., the classical Laplacian). This yields naturally to the introduction of weighted Orlicz–Sobolev spaces. While the case “without derivative” studied in Section 3 was well-suited for the study of the fine properties of the Kullback–Leibler divergence, we illustrate in Section 6.3 our use of Orlicz–Sobolev spaces with the fine study of the Hyvärinen divergence. This is a special type of divergence between densities that involves an L^2 -distance between gradients of densities [35] which has multiple applications. In particular, it is related with the so called Fisher information as it is defined for example in [32] (p. 49), which has deep connections with Boltzmann equation, see [36]. However the name Fisher information should not be used in a statistics context where it rather refers to the expression in coordinates of the metric of statistical models considered as pseudo-Riemannian manifolds, e.g., [1].

We introduce the Orlicz–Sobolev spaces with weight M , Maxwell density on \mathbb{R}^n ,

$$W_{\cosh -1}^1(M) = \{f \in L^{\cosh -1}(M) \mid \partial_j f \in L^{\cosh -1}(M), j = 1, \dots, n\}, \tag{26}$$

$$W_{(\cosh -1)_*}^1(M) = \{f \in L^{(\cosh -1)_*}(M) \mid \partial_j f \in L^{(\cosh -1)_*}(M), j = 1, \dots, n\}, \tag{27}$$

where ∂_j is the derivative in the sense of distributions. They are both Banach spaces, see [23, Section 10]. (The classical Adams’s treatise [37, Chapter 8] has Orlicz–Sobolev spaces, but does not consider the case of a weight. The product functions $(u, x) \mapsto (\cosh -1)(u)M(x)$ and $(u, x) \mapsto (\cosh -1)_*(u)M(x)$ are ϕ -functions according the Musielak’s definition.) The norm on $W_{\cosh -1}^1(M)$ is

$$\|f\|_{W_{\cosh -1}^1(M)} = \|f\|_{L^{\cosh -1}(M)} + \sum_{j=1}^n \|\partial_j f\|_{L^{\cosh -1}(M)}, \tag{28}$$

and similarly for $W_{(\cosh -1)_*}^1(M)$. One begins with a first technical result in order to relate such spaces with statistical exponential families:

Proposition 15. *Given $u \in \mathcal{S}_M \cap W_{\cosh -1}^1(M)$ and $f \in W_{\cosh -1}^1(M)$, one has*

$$f e^{u-K_M(u)} \in W_{(\cosh -1)_*}^1(M).$$

Proof. For simplicity, set $G = e^{u-K_M(u)}$. One knows from the Portmanteau Theorem 2 that $G M \in \mathcal{E}(M)$ and therefore, there exists $\varepsilon > 0$ such that $G \in L^{1+\varepsilon}(M)$. Let us prove that $fG \in L^{(\cosh -1)_*}(M)$. First of all, since $L^{1+\varepsilon}(M) \subset L^{(\cosh -1)_*}(M)$ and $f \in L^{\cosh -1}(M)$ one has $fG \in L^1(M)$. Moreover, for any $x \in \mathbb{R}^n$, according to classical Young’s inequality

$$f(x)G(x) \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|G(x)|^q \quad \forall p > 1, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Since Φ_* is increasing and convex

$$\Phi_*(f(x)G(x)) \leq \frac{1}{p}\Phi_*(|f(x)|^p) + \frac{1}{q}\Phi_*(G(x)^q) \quad x \in \mathbb{R}^n.$$

Now, since $f \in L^\Phi(M)$, one has $|f|^p \in L^{(\cosh-1)^*}(M)$ for all $p > 1$, i.e., $\Phi_*(|f|^p) \in L^1(M)$ for all $p > 1$. Choosing then $1 < q < 1 + \varepsilon$ one has $G^q \in L^{\frac{1+\varepsilon}{q}}(M) \subset L^{(\cosh-1)^*}(M)$ so that $\Phi_*(G^q) \in L^1(M)$. This proves that $\Phi_*(fG) \in L^1(M)$, i.e.,

$$fG \in L^{(\cosh-1)^*}(M).$$

In the same way, since $f \in W^1_{\cosh-1}(M)$ one also has

$$G\partial_j f \in L^{(\cosh-1)^*}(M) \quad \forall j = 1, \dots, n.$$

Moreover, $u \in W^1_{\cosh-1}(M)$ so that, for any $j = 1, \dots, n$, $G\partial_j u \in L^r(M)$ for any $r > 1$ and therefore $\Phi_*(|G\partial_j u|^p) \in L^1(M)$ for any $p > 1$. Repeating the above argument we get therefore

$$fG\partial_j u \in L^{(\cosh-1)^*}(M) \quad \forall j = 1, \dots, n.$$

Since $\partial_j(fG) = G\partial_j f + Gf\partial_j u$ a.e., one gets $\partial_j(fG) \in L^{(\cosh-1)^*}(M)$ for any $j = 1, \dots, n$ which proves the result. \square

Remark 5. As a particular case of the above Proposition, if $u \in \mathcal{S}_M \cap W^1_{\cosh-1}(M)$ then

$$e^{u-K_M(u)} \in W^1_{(\cosh-1)^*}(M) \quad \text{with} \quad \nabla e^{u-K_M(u)} = \nabla u e^{u-K_M(u)}.$$

The Orlicz–Sobolev spaces $W^1_{\cosh-1}(M)$ and $W^1_{(\cosh-1)^*}(M)$, as defined in Equations (26) and (27) respectively, are instances of Gaussian spaces of random variables and they inherit from the corresponding Orlicz spaces a duality form. In this duality the adjoint of the partial derivative has a special form coming from the form of the weight M see, e.g., [38, Chapter V]. We have the following:

Proposition 16.

1. Let $f \in C^\infty_0(\mathbb{R}^n)$ and $g \in W^1_{\cosh-1}(M)$. Then

$$\langle f, \partial_j g \rangle_M = \langle X_j f - \partial_j f, g \rangle_M \tag{29}$$

where X_j is the multiplication operator by the j -th coordinate x_j .

2. If $f \in W^1_{(\cosh-1)^*}(M)$, then $X_j f \in L^{(\cosh-1)^*}(M)$. More precisely, there exists $C > 0$ such that

$$\|X_j f\|_{L^{(\cosh-1)^*}(M)} \leq C \|f\|_{W^1_{(\cosh-1)^*}(M)} \quad \forall f \in W^1_{(\cosh-1)^*}(M). \tag{30}$$

3. If $f \in W^1_{(\cosh-1)^*}(M)$ and $g \in W^1_{\cosh-1}(M)$, then (29) holds.

Proof. 1. As $fM \in C^\infty_0(\mathbb{R}^n)$, we have by definition of distributional derivative,

$$\begin{aligned} \langle f, \partial_j g \rangle_M &= \int_{\mathbb{R}^n} f(x) \partial_j g(x) M(x) dx \\ &= - \int_{\mathbb{R}^n} g(x) \partial_j (f(x) M(x)) dx \\ &= \int_{\mathbb{R}^n} (x_j f(x) - \partial_j f(x)) M(x) dx \\ &= \langle X_j f - \partial_j f, g \rangle_M. \end{aligned}$$

2. Let us observe first that, according to Holder’s inequality

$$\mathbb{E}_M [|X_j f|] \leq 2 \|X_j\|_{L^{\cosh-1}(M)} \|f\|_{L^{(\cosh-1)^*(M)}} < \infty,$$

i.e., $X_j f \in L^1(M)$. Since $\Phi_* = (\cosh-1)_*$ enjoys the so-called Δ_2 -condition, to prove the stronger result $X_j f \in L^{(\cosh-1)^*(M)}$, it is enough to show that $\mathbb{E}_M [\Phi_*(X_j f)] < \infty$. First of all, using the tensorization property of the Gaussian measure, i.e., the fact that $M(x) = M_1(x_1) \dots M_1(x_n)$ for any $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ where M_1 stands for the one-dimensional standard Gaussian, we claim that it is enough to prove the result for $n = 1$. Indeed, given $f \in W^1_{(\cosh-1)^*(M)}$ and $x_j \in \mathbb{R}$ ($j = 1, \dots, n$), any $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ can be identified with $x = (x_j, \underline{x})$ with $\underline{x} \in \mathbb{R}^{n-1}$ and $x_j f(x) = x_j F_{\underline{x}}(x_j)$ where $F_{\underline{x}}(y) = f(y, \underline{x})$ for any $\underline{x} \in \mathbb{R}^{n-1}$, $y \in \mathbb{R}$. We also set $M_{n-1}(\underline{x}) = M(x)/M_1(x_j)$. Then, for a.e. $\underline{x} \in \mathbb{R}^{n-1}$, $F_{\underline{x}} \in W^1_{(\cosh-1)^*(M_1)}$ with

$$\int_{\mathbb{R}^{n-1}} M_{n-1}(\underline{x}) d\underline{x} \int_{\mathbb{R}} \Phi_*(F_{\underline{x}}(y)) M_1(y) dy = \int_{\mathbb{R}^n} \Phi_*(f(x)) M(x) dx$$

and

$$\int_{\mathbb{R}^{n-1}} M(\underline{x}) d\underline{x} \int_{\mathbb{R}} M_1(y) \Phi_*(F'_{\underline{x}}(y)) dy = \int_{\mathbb{R}^n} M(x) \Phi_*(\partial_j f(x)) dx$$

where F' denotes the distributional derivative of $F = F(y)$. In particular, if there exists $C > 0$ such that

$$\int_{\mathbb{R}} \Phi_*(yF(y)) M_1(y) dy \leq C \int_{\mathbb{R}} (\Phi_*(F(y)) + \Phi_*(F'(y))) M_1(y) dy \quad \forall F \in W^1_{(\cosh-1)^*(M_1)} \tag{31}$$

we get the desired result.

Let us then prove (31) and fix $F \in W^1_{(\cosh-1)^*(M_1)}$. From $\Phi_*(y) = \int_0^{|y|} \operatorname{arsinh} u \, du$ together with the evenness of arsinh we obtain

$$\Phi_*(yF(y)) = \int_0^{|F(y)|} |y| \operatorname{arsinh}(|y|v) \, dv = \int_0^{|F(y)|} y \operatorname{arsinh}(yv) \, dv.$$

Write for simplicity

$$G(y) := \int_0^{|F(y)|} \operatorname{arsinh}(yv) \, dv$$

one has

$$\begin{aligned} \int_{\mathbb{R}} \Phi_*(yF(y)) M_1(y) \, dy &= \int_{\mathbb{R}} y M_1(y) G(y) \, dy \\ &= - \int_{\mathbb{R}} G(y) M'_1(y) \, dy \\ &= \int_{\mathbb{R}} M_1(y) G'(y) \, dy. \end{aligned}$$

Now, the derivative of G exists because of the assumption $F \in W^1_{(\cosh-1)^*(M_1)}$ (that is, $|F| \in W^1_{(\cosh-1)^*(M_1)}$ and its derivative is given by the derivation of a composite function) and it is computed as

$$G'(y) = \operatorname{arsinh}(y|F(y)|) \frac{d}{dy} |F(y)| + \int_0^{|F(y)|} \frac{v}{\sqrt{1+y^2v^2}} \, dv.$$

Using Young’s inequality with $\Phi = \cosh - 1$ and $\Phi_* = (\cosh - 1)_*$ we get

$$G'(y) \leq \Phi\left(\operatorname{arsinh}(y|F(y)|)\right) + \Phi_*\left(\frac{d}{dy}|F(y)|\right) + \frac{\sqrt{1+y^2|F(y)|^2}-1}{y^2}.$$

All the terms in the right-hand side of the above inequality are integrable with respect to the measure $M_1(y) dy$ over \mathbb{R} . Indeed the first term is bounded as

$$\Phi(\operatorname{arsinh}(y|F(y)|)) = \sqrt{1+y^2|F(y)|^2} \leq \sqrt{2}(1 \vee |yF(y)|)$$

and $y \mapsto yF(y) \in L^1(\mathbb{R}, M_1(y)dy)$. The second term is integrable by assumption. The only concern is then the last term. For any $r > 0$,

$$\int_{|y|>r} M_1(y) \frac{\sqrt{1+y^2|F(y)|^2}-1}{y^2} dy \leq \frac{1}{r} \int_{\mathbb{R}} |F(y)| M_1(y) dy \leq \frac{1}{r} \int_{\mathbb{R}} \Phi_*(F(y)) M_1(y) dy$$

while,

$$\int_{|y|<r} M_1(y) \frac{\sqrt{1+y^2|F(y)|^2}-1}{y^2} dy \leq \int_{-r}^r M_1(y) \frac{|F(y)|}{|y|} dy \leq \frac{1}{\sqrt{2\pi}} \int_{-r}^r \frac{|F(y)|}{|y|} dy.$$

Now, splitting the integral into the two integrals \int_0^r and \int_{-r}^0 , one can use the one-dimensional Hardy inequality in Orlicz–Sobolev space [39] to get that there exists $C > 0$ such that

$$\int_{-r}^r \frac{|F(y)|}{|y|} dy \leq C \int_{-r}^r \Phi_*(F'(y)) dy.$$

This achieves to prove (31).

- Recall that (29) holds for any $f \in C_0^\infty(\mathbb{R}^n)$ and any $g \in L^{\cosh-1}(M)$. Since $\Phi_* = (\cosh - 1)_*$ enjoys the Δ_2 -condition, it is a well-known fact that $C_0^\infty(\mathbb{R}^n)$ is dense in $W_{(\cosh-1)_*}^1(M)$ (for the norm $\|\cdot\|_{W_{(\cosh-1)_*}^1}$). Therefore, approximating any $f \in W_{(\cosh-1)_*}^1(M)$ by a sequence of $C_0^\infty(\mathbb{R}^n)$ functions, we deduce the result from point 1.

□

Remark 6. In the second Item of the above Proposition, notice that a priori $X_j f$ belongs to $L^{(\cosh-1)_*}(M)$ but not to $W_{(\cosh-1)_*}^1(M)$, $j = 1, \dots, n$. For this to be true, one would require that $\partial_k f \in W_{(\cosh-1)_*}^1(M)$ for any $k = 1, \dots, n$.

$$\|X_j f\|_{L^{(\cosh-1)_*}(M)} \leq \delta + C_\delta \|f\|_{W_{(\cosh-1)_*}^1(M)} \quad \forall f \in W_{(\cosh-1)_*}^1(M).$$

6.1. Stein and Laplace Operators.

Following the language of [38, Chapter V], Item 3 of the above Proposition 16 can be reformulated saying that

$$\langle f, \partial_j g \rangle_M = \langle \delta_j f, g \rangle_M \quad \forall j = 1, \dots, n$$

where

$$\delta_j f = X_j f - \partial_j f.$$

This allows to define the *Stein operator* δ on $L^{(\cosh -1)*}(M)$ as

$$\delta : f \in \text{Dom}(\delta) \subset L^{(\cosh -1)*}(M) \mapsto \delta f = (\delta_j f)_{j=1, \dots, n} \in (L^{(\cosh -1)*}(M))^n.$$

where the the domain $\text{Dom}(\delta)$ of δ is exactly $W^1_{(\cosh -1)*}(M)$ according to point 2. of the above Proposition. Notice that, since Φ_* enjoys the Δ_2 -condition, $\text{Dom}(\delta)$ is dense in $L^{(\cosh -1)*}(M)$. One deduces then easily that δ is a closed and densely defined operator in $L^{(\cosh -1)*}(M)$.

One sees that

$$\begin{aligned} \langle f, \text{div } \mathbf{g} \rangle_M &:= \sum_{j=1}^n \langle f, \partial_j g_j \rangle_M = \sum_{j=1}^n \langle \delta_j f, g_j \rangle_M =: \langle \delta f, \mathbf{g} \rangle_M \\ \forall f \in W^1_{\cosh -1}(M); \mathbf{g} &= (g_j)_{j=1, \dots, n} \in (W^1_{(\cosh -1)*}(M))^n \end{aligned} \quad (32)$$

where $\text{div } \mathbf{g} = \sum_{j=1}^n \partial_j g_j$ is the divergence of \mathbf{g} . This allows to define the adjoint operator δ^* as follows, see [40]

$$\delta^* : \text{Dom}(\delta^*) \subset (L^{\cosh -1}(M))^n \rightarrow L^{\cosh -1}(M)$$

with

$$\text{Dom}(\delta^*) = \left\{ \mathbf{g} = (g_j)_{j=1, \dots, n} \in (L^{\cosh -1}(M))^n \mid (\exists c > 0)(\forall f \in \text{Dom}(\delta)) \mid \langle \mathbf{g}, \delta f \rangle_M \mid \leq c \|f\|_{L^{(\cosh -1)*}(M)} \right\}$$

and

$$\langle \mathbf{g}, \delta f \rangle_M = \langle \delta^* \mathbf{g}, f \rangle_M \quad \forall \mathbf{g} \in \text{Dom}(\delta^*), \forall f \in \text{Dom}(\delta).$$

One sees from (32) that

$$(W^1_{\cosh -1}(M))^n \subset \text{Dom}(\delta^*) \quad \text{and} \quad \delta^* \mathbf{g} = \nabla \cdot \mathbf{g} \quad \forall \mathbf{g} \in (W^1_{\cosh -1}(M))^n.$$

$$\begin{aligned} \left| \sum_{j=1}^n \langle g_j, \delta_j f \rangle_M \right| &\leq c \|f\|_{V_*} \quad \forall f \in W^1_{(\cosh -1)*}(M). \\ \left| \langle g_j, X_j f \rangle_M \right| &\leq 2 \|g_j\|_{L^{\cosh -1}(M)} \|X_j f\|_{L^{(\cosh -1)*}(M)} \end{aligned}$$

and one sees that there exists $C > 0$ such that, for any $j = 1, \dots, n$ and any $f \in W^1_{(\cosh -1)*}(M)$, it holds

$$\begin{aligned} \left| \langle g_j, \partial_j f \rangle_M \right| &\leq C \|f\|_{V_*}. \\ \left| \langle \partial_j g_j, f \rangle_M \right| &\leq C \|f\|_{V_*} \quad \forall f \in W^1_{(\cosh -1)*}(M) \end{aligned}$$

Remark 7. Notice that, since $\Phi = \cosh -1$ does not satisfies the Δ_2 -condition, it is not clear whether $\text{Dom}(\delta^*)$ is dense in $(L^{\cosh -1}(M))^n$ or not. However, from the general theory of adjoint operators and since δ is a closed densely defined operator in $L^{(\cosh -1)*}(M)$, the domain $\text{Dom}(\delta^*)$ is dense in $(L^{\cosh -1}(M))^n$ endowed with the weak- \star topology. Moreover, δ^* is a closed operator from $(L^{\cosh -1}(M))^n$ to $L^{\cosh -1}(M)$ (see [40, Chapter 2] for details).

We also define the gradient operator

$$\nabla : \text{Dom}(\nabla) \subset L^{\cosh^{-1}}(M) \rightarrow (L^{\cosh^{-1}}(M))^n$$

by

$$\text{Dom}(\nabla) = W^1_{\cosh^{-1}}(M) \quad \text{and} \quad \nabla f = (\partial_j f)_{j=1,\dots,n} \quad f \in W^1_{\cosh^{-1}}(M).$$

One sees that, if $f \in \text{Dom}(\nabla^2)$, i.e., if $f \in W^1_{\cosh^{-1}}(M)$ is such that $\nabla f \in (W^1_{\cosh^{-1}}(M))^n$, then $\nabla f \in \text{Dom}(\delta^*)$ and

$$\delta^* \nabla f = \sum_{j=1}^n \partial^2_{jj} f = \Delta f.$$

corresponds to the Laplace operator.

From now, set $W^{-1}_{\cosh^{-1}}(M)$ as the dual space of $W_{(\cosh^{-1})_*}(M)$:

$$W^{-1}_{\cosh^{-1}}(M) := (W_{(\cosh^{-1})_*}(M))^*,$$

i.e., the set of all continuous linear forms continuous $F : W_{(\cosh^{-1})_*}(M) \rightarrow \mathbb{R}$ and let $\langle \cdot, \cdot \rangle_{-1,1}$ denotes the duality pairing:

$$\langle F, u \rangle_{-1,1} = F(u) \quad \forall u \in W_{(\cosh^{-1})_*}(M); F \in W^{-1}_{\cosh^{-1}}(M).$$

For any $f \in \text{Dom}(\nabla)$, let

$$L_f : g \in W_{(\cosh^{-1})_*}(M) \mapsto \langle \nabla f, \delta g \rangle_M \in \mathbb{R}.$$

One easily checks that L_f is continuous and defines therefore an element of $W^{-1}_{\cosh^{-1}}(M)$. Clearly, the operator $f \in \text{Dom}(\nabla) = W^1_{\cosh^{-1}}(M) \mapsto L_f \in W^{-1}_{\cosh^{-1}}(M)$ is also linear. Using the identification $\Delta = \delta^* \nabla$, we denote it by

$$\begin{aligned} \Delta : W^1_{\cosh^{-1}}(M) &\longrightarrow W^{-1}_{\cosh^{-1}}(M) \\ f &\longmapsto \Delta f := L_f \end{aligned}$$

defined as

$$\langle \Delta f, u \rangle_{-1,1} = \langle \nabla f, \delta u \rangle_M, \quad f \in W^1_{\cosh^{-1}}(M), u \in W_{(\cosh^{-1})_*}(M).$$

Notice that, with this definition, Δf is an element of $W^{-1}_{\cosh^{-1}}(M)$ whereas, from the above observation, $\Delta f \in L^{\cosh^{-1}}(M)$. Also the domains of both operators are different. Of course, if $f \in \text{Dom}(\nabla^2)$ then Δf is actually an element of $L^{\cosh^{-1}}(M)$ and it coincides with Δf , namely, in such a case

$$\langle \Delta f, u \rangle_{-1,1} = \langle \Delta f, u \rangle_M \quad \forall u \in W_{(\cosh^{-1})_*}(M).$$

Remark 8. Given $f \in W^1_{\cosh^{-1}}(M)$ and $u \in C^\infty_0(\mathbb{R}^n)$ one has

$$\langle \Delta f, u \rangle_{-1,1} = \langle \nabla f, \delta u \rangle_M = \sum_{j=1}^n \langle \partial_j f, \delta_j u \rangle_M$$

but, since $\delta_j u \in W^1_{(\cosh-1)_*}(M)$, one can use (29) again to get

$$\langle \Delta f, u \rangle_{-1,1} = \sum_{j=1}^n \langle f, \delta_j^2 u \rangle_M.$$

One readily computes, for any $j = 1, \dots, n$, $\delta_j^2 u = (x_j^2 - 1)u - 2x_j \partial_j u + \partial_j^2 u$, so that

$$\langle \Delta f, u \rangle_{-1,1} = \langle f, (|\mathbf{X}|^2 - 1)u - 2\mathbf{X} \cdot \nabla u + \Delta u \rangle_M \quad \forall u \in C_0^\infty(\mathbb{R}^n).$$

Remark 9. Since the constant function $1 \in W^1_{(\cosh-1)_*}(M)$ one checks easily that

$$\langle \Delta f, 1 \rangle_{-1,1} = \langle \mathbf{X}, \nabla f \rangle_M = \mathbb{E}_M[\mathbf{X} \cdot \nabla f] \quad \forall f \in W^1_{\cosh-1}(M).$$

where we notice that $\mathbf{X} \cdot \nabla f \in L^1(M)$ for any $f \in W^1_{\cosh-1}(M)$. Actually, since $\mathbf{X} \in (W^1_{(\cosh-1)_*}(M))^n$ one can use (29) again to get,

$$\langle \mathbf{X}, \nabla f \rangle_M = \sum_{j=1}^n \langle \delta_j x_j, f \rangle_M = \sum_{j=1}^n \langle x_j^2 - 1, f \rangle_M$$

i.e.,

$$\langle \Delta f, 1 \rangle_{-1,1} = \langle |\mathbf{X}|^2 - n, f \rangle_M = \int_{\mathbb{R}^n} f(x) \Delta M(x) dx \quad \forall f \in W^1_{\cosh-1}(M).$$

Of course, if $f \in W^1_{\cosh-1}(M)$ and $\nabla f \in (W^1_{\cosh-1}(M))^n$ one actually gets $\Delta f = \Delta f \in L^{\cosh-1}(M)$ and

$$\mathbb{E}_M[\Delta f] = \mathbb{E}_M[\mathbf{X} \cdot \nabla f].$$

For technical purposes, we finally state the following Lemma

Lemma 2. Given $w_1, w_2 \in W^1_{\cosh-1}(M)$, $v \in \mathcal{S}_M \cap W^1_{\cosh-1}(M)$ and $g = e_M(v)$ one has

$$\langle \Delta w_2, w_1 e^{v-K_M(v)} \rangle_{-1,1} = -\mathbb{E}_g[\nabla w_1 \cdot \nabla w_2] + \mathbb{E}_g[w_1 (\mathbf{X} - \nabla v) \cdot \nabla w_2]. \tag{33}$$

Proof. For simplicity, set $h = w_1 e^{v-K_M(v)}$. One knows from Proposition 15 that $h \in W^1_{(\cosh-1)_*}(M)$ so that, by definition

$$\langle \Delta w_2, h \rangle_{-1,1} = \langle \nabla w_2, \delta h \rangle_M.$$

Moreover, one checks easily that

$$\delta h = \mathbf{X}h - \nabla h = \mathbf{X}h - \nabla w_1 e^{v-K_M(v)} - h \nabla v$$

so that

$$\langle \Delta w_2, h \rangle_{-1,1} = \langle \nabla w_2, (\mathbf{X} - \nabla v)h \rangle_M - \langle \nabla w_2, \nabla w_1 e^{v-K_M(v)} \rangle_M$$

which gives the result. \square

6.2. Exponential Family Based on Orlicz–Sobolev Spaces with Gaussian Weights

If we restrict the exponential family $\mathcal{E}(M)$ to M -centered random variable $W_{\cosh^{-1}}^1(M)$, that is in

$$W_M = W_{\cosh^{-1}}^1(M) \cap B_M = \{U \in W_{\cosh^{-1}}^1(M) \mid \mathbb{E}_M[U] = 0\},$$

we obtain the following non parametric exponential family

$$\mathcal{E}_1(M) = \{e^{U-K_M(U)} \cdot M \mid U \in W_{\cosh^{-1}}^1(M) \cap \mathcal{S}_M\}.$$

Because of the continuous embedding $W_{\cosh^{-1}}^1(M) \hookrightarrow L^{\cosh^{-1}}(M)$ the set $W_{\cosh^{-1}}^1(M) \cap \mathcal{S}_M$ is open in W_M and the cumulant functional $K_M : W_{\cosh^{-1}}^1(M) \cap \mathcal{S}_M \rightarrow \mathbb{R}$ is convex and differentiable.

In a similar way, we can define

$${}^*W_M = W_{(\cosh^{-1})_*}^1(M) \cap {}^*B_M = \{f \in W_{(\cosh^{-1})_*}^1(M) \mid \mathbb{E}_M[f] = 0\}$$

so that for each $f \in \mathcal{E}_1(M)$ we have $\frac{f}{M} - 1 \in {}^*W_M$, see Remark 5.

Every feature of the exponential manifold carries over to this case. In particular, we can define the spaces

$$W_f = W_{\cosh^{-1}}^1(M) \cap B_M = \{U \in W_{\cosh^{-1}}^1(M) \mid \mathbb{E}_f[U] = 0\}, \quad f \in \mathcal{E}_1(M),$$

to be models for the tangent spaces of $\mathcal{E}_1(M)$. Note that the transport acts on these spaces

$$\mathbb{U}_f^g : W_f \ni U \mapsto U - \mathbb{E}_g[U] \in W_g,$$

so that we can define the tangent bundle to be

$$T\mathcal{E}_1(M) = \{(g, V) \mid g \in \mathcal{E}_1(M), V \in W_g\}$$

and take as charts the restrictions of the charts defined on $T\mathcal{E}(M)$.

As a first example of application, note that the gradient of the BG-entropy

$$\nabla H(f) = -\log f - H(f)$$

is a vector field on $\mathcal{E}_1(M)$, which implies the solvability in $\mathcal{E}_1(M)$ of the gradient flow equation. Our concern here is to set up a framework for the study of evolution equation in $\mathcal{E}_1(M)$. Following sections are devoted to discuss a special functional and its gradient.

6.3. Hyvärinen Divergence

We begin with the following general properties of $W_M = \{U \in W_{\cosh^{-1}}^1(M) \mid \mathbb{E}_M[U] = 0\}$

Proposition 17. *Let $f, g \in \mathcal{E}_1(M)$, with $f = e_M(u), g = e_M(v), u, v \in W_{\cosh^{-1}}^1(M) \cap \mathcal{S}_M$. The following hold*

1. $\mathbb{E}_M[\nabla u] = \text{Cov}_M(u, \mathbf{X})$.

2. $\mathbb{E}_g [\mathbf{X} - \nabla v] = 0.$
3. $\mathbb{E}_g [\nabla u] = \text{Cov}_g (u, \mathbf{X} - \nabla v).$
4. $\mathbb{E}_g [\nabla u - \nabla v] = \text{Cov}_g (u - v, \mathbf{X} - \nabla v).$

Proof. In all the sequel, we set $F = e^{u-K_M(u)} = \frac{f}{M}$ and $G = e^{v-K_M(v)} = \frac{g}{M}$. Recall from Proposition 15 that $F, G \in W^1_{(\cosh-1)_*}(M).$

1. One has $\mathbb{E}_M [\nabla u] = \langle 1, \partial_j u \rangle_M$ where $1 \in W^1_{(\cosh-1)_*}(M)$ is the constant function. Then, from (29), $\mathbb{E}_M [\nabla u] = \langle \delta 1, u \rangle_M = \text{Cov}_M (u, \mathbf{X})$ since $\delta 1 = \mathbf{X}$.
2. As above, one has

$$\mathbb{E}_g [\nabla v] = \int \nabla v e_M(v) dx = \int \nabla v G M dx$$

and, from Proposition 15, $\nabla v G = \nabla G$ so that

$$\mathbb{E}_g [\nabla v] = \mathbb{E}_M [\nabla G].$$

Recall that $G \in W^1_{(\cosh-1)_*}(M)$ and that $\nabla G = \mathbf{X}G - \delta G$ so that

$$\mathbb{E}_g [\nabla v] = \langle \mathbf{X}G, 1 \rangle_M - \langle \delta G, 1 \rangle_M$$

where $1 \in W^1_{\cosh-1}(M)$ is the constant function equal to 1. Applying again (29) we get $\langle \delta G, 1 \rangle_M = \langle G, \nabla 1 \rangle_M = 0$ and

$$\mathbb{E}_g [\nabla v] = \langle \mathbf{X}G, 1 \rangle_M = \mathbb{E}_g [\mathbf{X}].$$

3. Observe first that $\mathbb{E}_g [\nabla u] = \langle G, \nabla u \rangle_M$. Using again (29), since $u \in W^1_{\cosh-1}(M)$ and $G \in W^1_{(\cosh-1)_*}(M)$ we have

$$\mathbb{E}_g [\nabla u] = \langle \delta G, u \rangle_M = \langle \mathbf{X}G - \nabla G, u \rangle_M.$$

Since $GM = g$ and $\nabla G = \nabla v G$ (see Proposition 15) we get $\langle \mathbf{X}G - \nabla G, u \rangle_M = \mathbb{E}_g [(\mathbf{X} - \nabla v) u]$ which gives the result.

4. Arguing as above one sees that $\mathbb{E}_g [\nabla v] = \mathbb{E}_g [(\mathbf{X} - \nabla v) v]$ and therefore the conclusion follows from the previous item.

□

Remark 10.

1. The equations in Proposition 17 could be written without reference to the score (chart) mapping $s_M: f \mapsto u$ by writing $\nabla u = \nabla \log \left(\frac{f}{M}\right) = \nabla \log f + \mathbf{X}$ to get

(a) $\mathbb{E}_M [\nabla \log \left(\frac{f}{M}\right)] = \text{Cov}_M \left(\log \left(\frac{f}{M}\right), \mathbf{X}\right).$

(b) $\mathbb{E}_g [\nabla \log g] = 0.$

(c) $\mathbb{E}_g [\nabla \log \left(\frac{f}{M}\right)] = -\text{Cov}_g \left(\log \left(\frac{f}{M}\right), \nabla \log g\right).$

$$(d) \mathbb{E}_g \left[\nabla \log \left(\frac{f}{g} \right) \right] = - \text{Cov}_g \left(\log \left(\frac{f}{g} \right), \nabla \log \left(\frac{g}{M} \right) \right).$$

However, we feel that explicit reference to the chart clarifies the geometric picture.

2. The mapping $f \mapsto \nabla \log f$ is a vector field in $T\mathcal{E}_1(M)$ with flow given by the translations:

$$\frac{d}{dt} \log f_t = \nabla \log f_i, \quad f(\mathbf{x}, t) = f(\mathbf{x} + \mathbf{1}t).$$

3. The KL-divergence $(f, g) \mapsto \mathbf{D}(f \| g)$ has expression $(u, v) \mapsto dK_M(v)[u - v] - K_M(u) + K_M(v)$ in the chart centered at M , with partial derivative with respect to v in the direction w given by $\text{Cov}_g(u - v, w)$. If the direction is $w = \mathbf{X} - \nabla v = -\nabla \log g$, we have that $\mathbb{E}_g[\nabla v - \nabla u]$ is the derivative of the KL-divergence along the vector field of translations.

We introduce here the Hyvärinen divergence between two elements of $\mathcal{E}_1(M)$:

Definition 10 (Hyvärinen divergence). For each $f, g \in \mathcal{E}_1(M)$ the Hyvärinen divergence is the quantity

$$\text{DH}(g|f) = \mathbb{E}_g \left[|\nabla \log f - \nabla \log g|^2 \right].$$

The expression in the chart centered at M is

$$\text{DH}_M(v|u) := \text{DH}(e_M(v)|e_M(u)) = \mathbb{E}_M \left[|\nabla u - \nabla v|^2 e^{v - K_M(v)} \right],$$

where $f = e_M(u)$, $g = e_M(v)$.

Remark 11.

1. The mapping $\mathcal{E}_1(M) \ni f \mapsto \nabla \log f = f^{-1} \nabla f$ is, in statistical terms, an estimating function or a pivot, because $\mathbb{E}_f[\nabla \log f] = 0$, $f \in \mathcal{E}_1(M)$. This means, it is a random variable whose value is zero in the mean if f is correct. If g is correct, then the expected value is $\mathbb{E}_g[\nabla \log f] = \mathbb{E}_g \left[\nabla \log \left(\frac{f}{g} \right) \right] = - \text{Cov}_g \left(\log \left(\frac{f}{g} \right), \nabla \log \left(\frac{g}{M} \right) \right)$. The second moment of $\nabla \log \left(\frac{f}{g} \right)$ was used by Hyvärinen as a measure of deviation from f to g [35].
2. Hyvärinen work has been used to discuss proper scoring rules in [41].
3. The same notion is known in Physics under the name of relative Fisher information, e.g., see [32].

In the following we denote the gradient of a function defined on the exponential manifold $\mathcal{E}_1(M)$, which is a random variable, by ∂ .

Proposition 18.

1. The Hyvärinen divergence is finite and infinitely differentiable everywhere in both variables.
2. $\partial(f \mapsto \text{DH}(g|f)) = -2 \nabla \log g \cdot \nabla \log \frac{f}{g} - 2 \Delta \log \frac{f}{g}$
3. $\partial(g \mapsto \text{DH}(f|g)) = 2 \nabla \log g \cdot \nabla \log \frac{f}{g} + 2 \Delta \log \frac{f}{g} + \text{DH}(f|g)$.

Proof. 1. For each $w \in V$ the gradient ∇w is in $(L^\Phi(g))^n = (L^\Phi(M))^n$ for all $g \in \mathcal{E}_1(M)$, hence it is g -square integrable for all $g \in \mathcal{E}_1(M)$. Moreover, the squared norm function $\mathcal{S}_M \times (B_M)^n \ni (v, \mathbf{w}) \mapsto \mathbb{E}_{e_M(v)} [|\mathbf{w}|^2]$ is ∞ -differentiable because it is the moment functional,

$$\mathbb{E}_g [|\mathbf{w}|^2] = \sum_{j=1}^n (d^2 K_M(v)[w_j, w_j] + (dK_M(v)[w_j])^2).$$

We can compose this function with the linear function

$$V_M \cap \mathcal{S}_M \times V_M \ni (v, u) \mapsto (v, \mathbf{w}) = (v, \nabla(u - v)) \in \mathcal{S}_M \times B_M^n.$$

2. Let $g = e_M(v), f = e_M(u), u, v \in \mathcal{S}_M \cap V$ be given. For any $w \in V$, we compute first the directional derivative:

$$d(u \mapsto \text{DH}_M(v||u)) [w] = 2\mathbb{E}_M [\nabla w \cdot (\nabla u - \nabla v)e^{v-K_M(v)}] = 2\mathbb{E}_g [\nabla w \cdot (\nabla u - \nabla v)]$$

where we notice that all the terms are well defined whenever $u, v \in \mathcal{S}_M \cap V, w \in V$. Using now (33) with $w_2 = u - v$ and $w_1 = w$ we get that

$$\begin{aligned} & d(u \mapsto \text{DH}_M(v||u)) [w] \\ &= 2\mathbb{E}_g [w(\mathbf{X} - \nabla v) \cdot \nabla(u - v)] - 2 \langle \Delta(u - v), we^{v-K_M(v)} \rangle_{-1,1} \\ &= 2\mathbb{E}_M [we^{v-K_M(v)} (\mathbf{X} - \nabla v) \cdot \nabla(u - v)] - 2 \langle \Delta(u - v), we^{v-K_M(v)} \rangle_{-1,1}. \end{aligned}$$

Since this is true for any $w \in V$ we get

$$d(u \mapsto \text{DH}_M(v||u)) = 2(\mathbf{X} - \nabla v) \cdot \nabla(u - v) - 2\Delta(u - v)$$

where of course, $\Delta(u - v)$ is meant in $W_{\cosh^{-1}}^{-1}(M)$ (notice that $w \in W_{\cosh^{-1}}^1(M) \subset W_{(\cosh^{-1})_*}^1(M)$). The formula for the partial gradient in absolute variables follows from

$$(\mathbf{X} - \nabla v) \cdot \nabla(u - v) = -\nabla \log g \cdot \nabla \log \frac{f}{g}$$

for, with a slight abuse of notations, we identify $\Delta \log \frac{f}{g}$ to $\Delta(u - v)$.

3. As above, let $g = e_M(v), f = e_M(u), u, v \in \mathcal{S}_M \cap V$ be given. For any $w \in V$, we compute first the directional derivative. One gets now

$$\begin{aligned} & d(v \mapsto \text{DH}_M(u||v)) [w] \\ &= -2\mathbb{E}_M [\nabla w \cdot (\nabla u - \nabla v)e^{v-K_M(v)}] + \mathbb{E}_M [(w - dK_M(v)[w]) |\nabla u - \nabla v|^2 e^{v-K_M(v)}] \end{aligned}$$

One recognizes in the first term $-d(u \mapsto \text{DH}_M(v||u)) [w]$ while the second term is given by

$$\begin{aligned} \mathbb{E}_M [(w - dK_M(v)[w]) |\nabla u - \nabla v|^2 e^{v-K_M(v)}] &= \mathbb{E}_g [(w - \mathbb{E}_g [w]) |\nabla u - \nabla v|^2] \\ &= \text{Cov}_g (w, |\nabla u - \nabla v|^2). \end{aligned}$$

As in the previous item, this gives the result.

□

As well-documented, the Hyvärinen divergence is a powerful tool for the study of general diffusion processes. We have just shown that the Information Geometry formalism and the exponential manifold approach are robust enough to allow for a generalization in Orlicz–Sobolev spaces. We believe then that, as Boltzmann equation can be studied through the exponential manifold formalism in $L^{\cosh^{-1}}(M)$, general diffusion processes can be investigated in $W_{\cosh^{-1}}^1(M)$ with the formalism discussed in the present section. This is a plan for future work.

7. Conclusions and Discussion

We have shown that well known geometric feature of problems in Statistical Physics can be turned into precise formal results via a careful consideration of the relevant functional analysis.

In particular the notion of flow in a Banach manifold modeled on Orlicz spaces can be used to clarify arguments based on the evolution of the classical Boltzmann–Gibbs entropy in the vector field associated to the Boltzmann equation.

In the last section we have shown how to construct a similar theory in the case the generalized entropy under consideration is the so-called Fisher functional or Hyvärinen divergence. Such a generalized entropy is particularly well-suited for the study of general diffusion problems and the results presented in Section 6 can be seen as the first outcome of an ongoing joint research program.

Acknowledgments

Both B. Lods and G. Pistone acknowledge support from the de Castro Statistics Initiative, Collegio Carlo Alberto, Moncalieri, Italy. G. Pistone is a member of INdAM/GNAMPA.

Author Contributions

This work has been conceived and written by both authors. Bertrand Lods' field of research is the study of Boltzmann equation while Giovanni Pistone's field of research is Information Geometry. Both authors have read and approved the final manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; Volume 191, Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA; Oxford University Press: Oxford, UK, 2000.
2. Amari, S.I.; Barndorff-Nielsen, O.E.; Kass, R.E.; Lauritzen, S.L.; Rao, C.R. *Differential Geometry in Statistical Inference*; Institute of Mathematical Statistics Lecture Notes—Monograph Series, 10; Institute of Mathematical Statistics: Hayward, CA, USA, 1987.
3. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Hackensack, NJ, USA, 2007.

4. Cercignani, C. *The Boltzmann Equation and Its Applications*; Volume 67, Applied Mathematical Sciences; Springer: New York, NY, USA, 1988.
5. Villani, C. A review of mathematical topics in collisional kinetic theory. In *Handbook of Mathematical Fluid Dynamics*, Volume I; North-Holland: Amsterdam, The Netherlands, 2002.
6. Do Carmo, M.P. *Riemannian Geometry*; Mathematics: Theory & Applications; Birkhäuser: Boston, MA, USA, 1992.
7. Dawid, A.P. Further comments on: “Some comments on a paper by Bradley Efron”. *Ann. Stat.* **1977**, *5*, 1249.
8. Efron, B. Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Stat.* **1975**, *3*, 1189–1242.
9. Cena, A. Geometric structures on the non-parametric statistical manifold. Ph.D. Thesis, Dottorato in Matematica, Università di Milano, Milano, Italy, 2002.
10. Cena, A.; Pistone, G. Exponential statistical manifold. *Ann. Inst. Stat. Math.* **2007**, *59*, 27–56.
11. Gibilisco, P.; Isola, T. Connections on statistical manifolds of density operators by geometry of noncommutative L^p -spaces. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1999**, *2*, 169–178.
12. Gibilisco, P.; Pistone, G. Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1999**, *2*, doi: 10.1142/S0219025799000096
13. Imperato, D. Exponential models and Fisher information. Geometry and applications. Ph.D. Thesis, DIMAT Politecnico di Torino, Torino, Italy, 2008.
14. Pistone, G.; Sempi, C. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **1995**, *23*, 1543–1561.
15. Pistone, G.; Rogantin, M.P. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **1999**, *5*, 721–760.
16. Pistone, G. κ -exponential models from the geometrical viewpoint. *Europ. Phys. J. B* **2009**, *71*, 29–37.
17. Pistone, G. Algebraic varieties vs differentiable manifolds in statistical models. In *Algebraic and Geometric Methods in Statistics*; Gibilisco, P., Riccomagno, E., Rogantin, M.P., Wynn, H.P., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 341–365.
18. Pistone, G. Nonparametric information geometry. In *Geometric Science of Information*; Volume 8085, Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; pp. 5–36.
19. Pistone, G. Examples of the application of nonparametric information geometry to statistical physics. *Entropy* **2013**, *15*, 4042–4065.
20. Abraham, R.; Marsden, J.E.; Ratiu, T. *Manifolds, Tensor Analysis, and Applications*, 2nd ed.; Volume 75, Applied Mathematical Sciences; Springer: New York, NY, USA, 1988.
21. Bourbaki, N. *Variétés différentielles et analytiques. Fascicule de résultats / Paragraphes 1 à 7*; Number XXXIII in *Éléments de mathématiques*; Hermann: Paris, France, 1971.
22. Lang, S. *Differential and Riemannian Manifolds*, 3rd ed.; Volume 160, Graduate Texts in Mathematics; Springer: New York, NY, USA, 1995.

23. Musielak, J. *Orlicz Spaces and Modular Spaces*; Volume 1034, Lecture Notes in Mathematics; Springer: Berlin, Germany, 1983.
24. Newton, N.J. An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **2012**, *263*, 1661–1681.
25. Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L. Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **2014**, *162*, 327–364.
26. Appell, J.; Zabrejko, P.P. *Nonlinear Superposition Operators*; Volume 95, Cambridge Tracts in Mathematics; Cambridge University Press: Cambridge, UK, 1990.
27. Santacroce, M.; Siri, P.; Trivellato, B. New Results on Mixture and Exponential Models by Orlicz Spaces. *Bernoulli* **2015**, doi: 10.3150/15-BEJ698.
28. Brown, L.D. *Fundamentals of statistical exponential families with applications in statistical decision theory*; Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9; Institute of Mathematical Statistics: Hayward, CA, USA, 1986.
29. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
30. Brègman, L.M. A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Ž. Vyčisl. Mat. i Mat. Fiz.* **1967**, *7*, 620–631.
31. Ambrosio, L.; Gigli, N.; Savaré, G. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed.; Lectures in Mathematics ETH Zürich; Birkhäuser Verlag: Basel, Switzerland, 2008.
32. Villani, C. Entropy production and convergence to equilibrium. In *Entropy Methods for the Boltzmann Equation*; Volume 1916, Lecture Notes in Mathematics; Springer: Berlin, Germany, 2008; pp. 1–70.
33. Majewski, W.A.; Labuschagne, L.E. On applications of Orlicz spaces to statistical physics. *Ann. Henri Poincaré* **2014**, *15*, 1197–1221.
34. Lieb, E.H.; Loss, M. *Analysis*, 2nd ed.; Volume 14, Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2001.
35. Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **2005**, *6*, 695–709.
36. Toscani, G.; Villani, C. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Commun. Math. Phys.* **1999**, *203*, 667–706.
37. Adams, R.A.; Fournier, J.J.F. *Sobolev Spaces*, 2nd ed.; Volume 140, Pure and Applied Mathematics (Amsterdam); Elsevier/Academic Press: Amsterdam, The Netherlands, 2003.
38. Malliavin, P. *Integration and Probability*; Volume 157, Graduate Texts in Mathematics; Springer: New York, NY, USA, 1995.
39. Cianchi, A. Some results in the theory of Orlicz spaces and applications to variational problems. In *Nonlinear Analysis, Function Spaces and Applications 6*, Proceedings of the Spring School held in Prague, Czech, May 31–June 6 1998; pp. 50–92.
40. Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*; Universitext; Springer: New York, NY, USA, 2011.

41. Parry, M.; Dawid, A.P.; Lauritzen, S. Proper local scoring rules. *Ann. Stat.* **2012**, *40*, 561–592.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).