Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Frobenius Statistical manifolds, Geometric invariants & Hidden symmetries

by

Noémie Combe, Philippe Combe, and Hanna Nencka

Preprint no.: 3 2021

Frobenius Statistical manifolds, Geometric invariants & Hidden symmetries

Noemie Combe, Philippe Combe, and Hanna Nencka*

Max Planck Institute for Maths in the Sciences, Inselstrasse 22 ,04103 Leipzig, Germany
noemie.combe@mis.mpg.de

Abstract. In this paper, we explicitly prove that statistical manifolds have a Frobenius manifold structure. This latter object, at the interplay of beautiful interactions between topology and quantum field theory, raises natural questions, concerning the existence of Gromov–Witten invariants for statistical manifolds. We prove that an analog of Gromov–Witten invariants for statistical manifolds (GWS) exists, and that it plays an important role in the learning process. These new invariants have a geometric interpretation concerning intersection points of para-holomorphic curves. In addition, we unravel the hidden symmetries of statistical manifolds. It decomposes into a pair of totally geodesic submanifolds, containing a pair of flat connections. We prove that the pair of pseudo-Riemannian submanifolds are symmetric to each other with respect to Pierce mirror.

Keywords: Statistical manifold · Frobenius manifold · Gromov–Witten invariants · Paracomplex geometry
Mathematics Subject Classification 53B99, 62B10, 60D99, 53D45

1 Introduction

Statistical manifolds have been for more than 60 years a domain of great interest, in information theory and machine learning [2–4, 7–10] and in decision theory [9]. Statistical manifolds are given by a family of probability distributions, indexed by some real parameter and carry an affine connection structure, compatible with the Fisher–Rao metric. Recently, in [11], a new perspective on this object has been given. It was shown that statistical manifolds have a structure of F-manifolds (see [13]). We prove explicitly that they have the structure of a *Frobenius manifold*. It is the *fourth type* of Frobenius manifolds, known nowadays:

Theorem A *Statistical manifolds have the structure of Frobenius manifolds.*

Moreover, we investigate geometric properties, and show that statistical manifolds split into a pair of totally geodesic submanifolds, containing a pair of flat

* This research was supported by the Max Planck Society's Minerva grant. The authors express their gratitude towards MPI MiS for excellent working conditions.

connections. We prove that the pair of pseudo-Riemannian submanifolds are symmetric to each other with respect to Pierce mirror i.e.:

Theorem B *The manifold of probability distributions is a paracomplex manifold. It decomposes into a pair of totally geodesic submanifolds, containing a pair of flat connections. Moreover, the pair of pseudo-Riemannian submanifolds are symmetric to each other with respect to Pierce mirror.*

The notion of Frobenius manifolds (resp. F-manifolds) is the fruit of fifty years of remarkable interaction between topology and quantum physics. This relation involves the most advanced and sophisticated ideas on each side, and lead to Topological Quantum Field Theory. The three other classes include *quantum cohomology* (topological sigma-models), *unfolding spaces of singularities* (Saito's theory, Landau-Ginzburg models), and *Barannikov–Kontsevich construction* starting with the Dolbeault complex of a Calabi–Yau manifold and conjecturally producing the B-side of the mirror conjecture in arbitrary dimension (see [14]).

Moreover, Frobenius manifolds being at the interplay of beautiful interactions between topology and quantum field theory, raises natural questions concerning the role and existence of Gromov–Witten invariants, for statistical manifolds. We prove that an analog of Gromov–Witten invariants for statistical manifolds exists, and that it plays an important role in the learning process. These new invariants have a geometric interpretation concerning the intersection of paraholomorphic curves.

In particular, we show that:

Theorem C *Consider a statistical manifold S. Then, the Gromov–Witten invariants for statistical manifolds (GWS) determine the learning process.*

2 Statistical manifolds

Let $(\Omega, \mathcal{F})$ be a measure space, where $\mathcal{F}$ denotes the σ-algebra of elements of Ω. We consider a family of parametric probabilities $\mathfrak{S}$ on the measure space $(\Omega, \mathcal{F})$. We ask all parametric probabilities of $\mathfrak{S}$ to be *absolutely continuous* w.r to a σ-finite measure λ i.e. a measure $P \in (\mathfrak{S}, \mathcal{F})$ is absolutely continuous w.r to λ if for every measurable set $A \subset \mathcal{F}$:

$$\lambda(A) = 0 \Rightarrow P(A) = 0, \quad \forall A \subset \mathcal{F},$$

which we denote by $P \ll \lambda$. Note that $P \ll \lambda$ does not imply that $\lambda \ll P$.

If λ is positive and σ-finite there exists a measurable function ρ called density of the measure P w.r to the measure λ, denoted

$$\rho = \frac{dP}{d\lambda}, \qquad P(A) = \int_A \rho d\lambda, \ \forall A \subset \mathcal{F}.$$

We denote by S the associated family of probability densities of the parametric probabilities. We limit ourselves to the case where S is a smooth topological manifold.

$$S = \left\{ \rho_\theta \in L^1(\Omega, \lambda),\ \theta = \{\theta_1, \dots \theta_n\};\ \rho > 0\ \lambda - a.e.,\ \int_\Omega \rho d\lambda = 1 \right\}.$$

This generates the space of probability measures absolutely continuous with respect to the measure λ, i.e. $P_\theta(A) = \int_A \rho_\theta d\lambda$ where $A \subset \mathcal{F}$.

We construct its tangent space as follows. Let $u \in L^2(\Omega, P_\theta)$ be a tangent vector to S at the point ρ_θ.

$$T_\theta = \left\{ u \in L^2(\Omega, P_\theta); \mathbb{E}_{P_\theta}[u] = 0, u = \sum_{i=1}^{d} u^i \partial_i \ell_\theta \right\},$$

where $\mathbb{E}_{P_\theta}[u]$ is the expectation value, w.r. to the probability P_θ.

Remark 1. The elements of T_θ generate the family of signed measures with bounded variations (i.e. signed measures whose total variation $\|\mu\| = |\mu|(X)$ is bounded, vanishing only on an ideal $\mathcal{I}$ of the σ-algebra $\mathcal{F}$) which are absolutely continuous with respect to P_θ and such that $\int_\Omega u dP_\theta = 0$. This forms a real affine space.

In 1945, Rao [17] introduced the Riemannian metric on a statistical manifold, using the Fischer information matrix. The statistical manifold forms a (pseudo)-Riemannian manifold.

In the basis, where $\{\partial_i \ell_\theta\}, \{i = 1, \dots, n\}$ where $\ell_\theta = \ln \rho_\theta$, the Fisher metric are just the *covariance matrix* of the score vector. Citing results of [5] (p89) we can in particular state that:

$$g_{i,j}(\theta) = \mathbb{E}_{P_\theta}[\partial_i \ell_\theta \partial_j \ell_\theta]$$
$$g^{i,j}(\theta) = \mathbb{E}_{P_\theta}[a^i_\theta a^j_\theta],$$

where $\{a^i\}$ form a dual basis to $\{\partial_j \ell_\theta\}$:

$$a^i_\theta(\partial_j \ell_\theta) = \mathbb{E}_{P_\theta}[a^i_\theta \partial_j \ell_\theta] = \delta^i_j$$

with

$$\mathbb{E}_{P_\theta}[a^i_\theta] = 0.$$

3 Frobenius manifolds & statistical manifolds

A Frobenius manifold is a manifold M endowed with an affine flat structure[1], a compatible metric g, and an even symmetric rank 3 tensor t. Define a symmetric

[1] Here the affine flat structure is equivalently described as complete atlas whose transition functions are affine linear. Since the statistical manifolds are (pseudo)Riemannian manifolds this condition is fulfilled.

bilinear multiplication on the tangent bundle:

$$\circ : TM \otimes TM \to TM.$$

M endowed with these structures is called *pre-Frobenius.* A pre-Frobenius manifold is Frobenius if it verifies the following associativity and potentiality properties (see [14]):

– *Associativity:* for any (flat) local tangent fields u, v, w, we have: $t(u, v, w) = g(u \circ v, w) = g(u, v \circ w)$. The metric is invariant with respect to the multiplication.
– *Potentiality:* A admits locally everywhere locally a potential function Φ such that, for any local tangent fields ∂_i we have $t(\partial_a, \partial_b, \partial_c) = \partial_a, \partial_b, \partial_c \Phi$.

We now prove, in this section, explicitly that the statistical manifold is Frobenius. Let (S, g, t) be a statistical manifold equipped with the Fischer-Rao Riemannian metric g and a 3-covariant tensor field t called the *skewness tensor*. It is a covariant tensor of rank 3 which is fully symmetric:

$$t : TS \times TS \times TS \to \mathbb{R},$$

given by

$$t|_{\rho_\theta}(u, v, w) = \mathbb{E}_{P_\theta}[u_\theta v_\theta w_\theta].$$

In other words, in the score coordinates, we have:

$$t_{ijk}(\theta) = \mathbb{E}_{P_\theta}[\partial_i \ell_\theta \partial_j \ell_\theta \partial_k \ell_\theta].$$

Denote the mixed tensor by $\bar{t} = t.g^{-1}$. It is bilinear map $\bar{t} : TS \times TS \to TS$, given componentwise by:

$$\bar{t}^k_{ij} = g^{km} t_{ijm}, \tag{1}$$

where $g^{km} = \mathbb{E}_{P_\theta}[a^k_\theta a^m_\theta]$. **NB:** This is written using Einstein's convention.

Remark 2. The Einstein convention will be used throughout this paper, whenever needed.

We have:

$$\bar{t}^k_{ij} = \bar{t}|_{\rho_\theta}(\partial_i \ell_\theta, \partial_j \ell_\theta, a^k) = \mathbb{E}_{P_\theta}[\partial_i \ell_\theta \partial_j \ell_\theta a^k_\theta].$$

As for the connection, it is given by:

$$\overset{\alpha}{\nabla}_X Y = \overset{0}{\nabla}_X Y + \frac{\alpha}{2} \bar{t}(X, Y), \quad \alpha \in \mathbb{R}, X, Y \in T_\rho S$$

where $\overset{\alpha}{\nabla}_X Y$ denotes the α-covariant derivative.

Remark 3. Whenever we have a pre-Frobenius manifold (S, g, t) we call the connection $\overset{\alpha}{\nabla}$ the *structure connection.*

In fact $\overset{\alpha}{\nabla}$ is the unique torsion free connection satisfying:

$$\overset{\alpha}{\nabla} g = \alpha t,$$

i.e.

$$\overset{\alpha}{\nabla}_X g(Y, Z) = \alpha t(X, Y, Z).$$

Proposition 1. *The tensor $\bar{t} : TS \times TS \to TS$ allows to define a multiplication $\circ$ on TS, such that for all $u, v, \in T_{\rho_\theta} S$, we have:*

$$u \circ v = \bar{t}(u, v).$$

Proof. By construction, in local coordinates, for any $u, v, \in T_{\rho_\theta} S$, we have $u = \partial_i \ell_\theta$ and $v = \partial_j \ell_\theta$. In particular, $\partial_i \ell_\theta \otimes \partial_j \ell_\theta = \bar{t}_{ij}^k \partial_k \ell_\theta$, which by calculation turns to be $\mathbb{E}_{P_\theta}[\partial_i \ell_\theta \partial_j \ell_\theta a_\theta^k]$.

Lemma 1. *For any local tangent fields $u, v, w \in T_{\rho_\theta} S$ the associativity property holds:*

$$g(u \circ v, w) = g(u, v \circ w).$$

Proof. Let us start with the left hand side of the equation. Suppose that $u = \partial_i \ell_\theta$, $v = \partial_j \ell_\theta$, $w = \partial_l \ell_\theta$. By previous calculations: $\partial_i \ell_\theta \circ \partial_j \ell_\theta = \bar{t}_{ij}^k \partial_k \ell_\theta$. Insert this result into $g(u \circ v, w)$, which gives us $g(\partial_i \ell_\theta \circ \partial_j \ell_\theta, \partial_l \ell_\theta)$, and leads to $g(\bar{t}_{ij}^k \partial_k \ell_\theta, \partial_l \ell_\theta)$. By some calculations and formula (1) it turns out to be equal to $t(u, v, w)$.

Consider the right hand side. Let $g(u, v \circ w) = g(\partial_i \ell_\theta, \partial_j \ell_\theta \circ \partial_l \ell_\theta)$. Mimicking the previous approach, we show that this is equivalent to $g(\partial_i \ell_\theta, \bar{t}_{jl}^k \partial_l \ell_\theta)$, which is equal to $t(u, v, w)$.

Proposition 2. *The manifold S verifies the potentiality condition.*

Proof. Consider the skewness tensor t, which is given by $t = \bar{t} \cdot g$. Then, in local coordinates we have that: $t_{ijm} = \bar{t}_{ij}^k \cdot g_{km}$. The metric g_{km} is also given by $g_{km} = \partial_k \partial_m \Phi$, where Φ is a local potential function. Therefore, $t_{ijm} = \bar{t}_{ij}^k \partial_k \partial_m \Phi$, and hence $t_{ijm} = \partial_i \partial_j \partial_m \Phi$. The skewness tensor everywhere, locally, admits such a potential function. So, the pre-Frobenius manifold S is potential.

Theorem A *The statistical manifold (S, g, t) is a Frobenius manifold for $\alpha = \pm 1$.*

Proof. The statistical manifold S comes equipped with a Riemannian metric g, and a skew symmetric tensor t. We have proved that the associativity conditions and the potentiality are fulfilled. Moreover, we have a pencil of connections depending on an even parameter α and defined by:

$$\overset{\alpha}{\nabla}_X Y = \overset{0}{\nabla}_X Y + \frac{\alpha}{2}(X \circ Y), \quad \alpha \in \mathbb{R}, X, Y \in T_\rho S$$

where $\overset{\alpha}{\nabla}_X Y$ denotes the α-covariant derivative. We call the connection $\overset{\alpha}{\nabla}$ the *structure connection* of the pre-Frobenius manifold (S, g, t). Now it remains to apply the Theorem 1.5 from [14] (p.20), stating that the triplet (S, g, t) is Frobenius if and only if the *structure connection* $\overset{\alpha}{\nabla}$ is flat. By direct computation, we show that for $\alpha = \pm 1$, the the structure connection is flat. Therefore, the conclusion is straightforward.

4 Hidden geometry of statistical manifolds

Let $\mathcal{W}$ be a linear space of signed measures with bounded variations, vanishing on an ideal $\mathcal{I}$ of the σ-algebra $\mathcal{F}$. Let $\mathcal{C}$ be a cone in $\mathcal{W}$ of (strictly) positive measures on the space $(X, \mathcal{F})$, vanishing only on an ideal $\mathcal{I}$ of the σ–algebra $\mathcal{F}$.

From [11], it is known that:

Theorem 1. *The positive cone $\mathcal{C}$, defined above, is a paracomplex (Vinberg) cone*[2] *.*

According to Vinberg 's classification theorem there exist five types of cones. As it was shown in [11], the cone of positive measures of bounded variations belongs to the fifth class, that means the cone defined over the algebra of paracomplex numbers. A Vinberg n-cone is in bijection with a (semi-simple) Jordan n-algebra. Properties of these algebras imply that for a pair of idempotents, there exists a Pierce decomposition, and Pierce mirror [15].

We will prove the following theorem:

Theorem B *The manifold of probability distributions is a paracomplex manifold. It decomposes into a pair of totally geodesic submanifolds, containing a pair of flat connections. Moreover, the pair of pseudo-Riemannian submanifolds are symmetric to each other with respect to Pierce mirror.*

In order to prove the theorem, we introduce a certain number of intermediary results, which are listed below.

Lemma 2 (Duality lemma). *Let $H \subset \mathcal{W}$ be a hyperplane, given by the constraints:*

$$\langle 1, \mu \rangle = 1, \;\; \text{where } \langle f, \mu \rangle = \int_X f d\mu,$$

with $\mu \in \mathcal{W}$, and being a section of the cone $\mathcal{C}$. Then, H is the dual space of S, defined above.

[2] A cone $V \subset R$ is a non–empty subset, closed with respect to addition and multiplication by positive reals. A convex cone V in a vector space R with an inner product has a dual cone $V^* = \{a \in R : \forall b \in V, \langle a, b \rangle > 0\}$. The cone is self-dual when $V = V^*$. It is homogeneous when to any points $a, b \in V$ there is a real linear transformation $T : V \to V$ that restricts to a bijection $V \to V$ and satisfies $T(a) = b$. Moreover, the closure of V should not contain a real linear subspace of positive dimension.

Proof. This follows from the properties of the dual space, obtained using the Radon–Nikodym derivatives $\frac{d\nu}{d\mu}$ of the measure ν w.r.t. the measure μ.

We observe the following:

Lemma 3. *Let S be the space of probability distributions defined above and P_θ a point on S. The tangent space T_θ at this point P_θ can be identified to a module over an algebra.*

Hence, the following proposition is true:

Proposition 3 (Paracomplex spaces). *The space S of probability distributions and its dual H are paracomplex spaces.*

The manifold over paracomplex algebra have the following property:

Theorem 2 (Paracomplex manifold). *The space of probability distributions S and its dual H are paracomplex manifolds.*

Lemma 4. *The manifold of probability distributions is decomposed into a pair of totally geodesic submanifolds*[3]*.*

Čentsov [9] has shown that the totally geodesic submanifolds turn out to be the exponential families of probability distributions with local parametrization. Therefore, we can obtain the following:

Lemma 5. *The manifold of probability distributions has a pair of flat connections.*

This follows from the calculation in [7, 8].

Proposition 4. *The space of probability distributions S and H are decomposed into pseudo-Riemannian submanifolds which are symmetric to each other wrt Pierce mirror.*

In order to prove these statements, we recall the tools from paracomplex geometry, in the following subsections.

5 Proof of theorem B

We start by introducing the tools to prove the Lemma 3 and Proposition 3.

[3] A submanifold N of a Riemannian manifold (M, g) is called totally geodesic if any geodesic on the submanifold N with its induced Riemannian metric g is also a geodesic on the Riemannian manifold (M, g).

5.1 Modules over spin factor algberas

Let $\mathfrak{A}$ be the unital and bi-dimensional algebra over paracomplex numbers $\mathfrak{C}$, generated by 1 and ε where $\varepsilon^2 = 1$, verifying the following relations:

$$e_i \cdot e_j = \sum_k C_{ij}^k e_k \quad \text{with} \quad C_{ij}^k = C_{ji}^k.$$

The algebra $\mathfrak{A}$ is known as the spin factor algebra or the algebra of paracomplex numbers.

The structure constants C_{ij}^k are:

$$C_{11}^1 = C_{12}^2 = C_{22}^1 = 1,$$

the other structure constants are null.

Let us change the basis such that the new generators are defined by:

$$e_- = \frac{1-\varepsilon}{2}, \quad e_+ = \frac{1+\varepsilon}{2}.$$

These generators have the following relations:

$$e_- \circ e_- = e_-, \; e_+ \circ e_+ = e_+, e_- \circ e_+ = 0,$$

$$e_- + e_+ = 1, \; e_- - e_+ = \varepsilon.$$

we call this new basis a canonical basis. Notice that this new basis highlights the existence of a pair of idempotents i.e. $e_-^2 = e_-$ and $e_+^2 = e_+$.

Remark This semi-simple algebra is isomorphic to $\mathbb{R} \oplus \mathbb{R}$. As a set, it can be identified to $\mathbb{R}^2$ but NOT as an algebra.

We construct the module over the spin factor algebra. To go back to the Amari–Čentsov statistical manifolds, we use the Norden–Shirokov method [16, 19]. Let $\mathfrak{A}$ be the spin factor algebra. Let us construct an m-module over the spin factor algebra $\mathfrak{M}^m(\mathfrak{A})$. The affine representation of the algebra A, or free module AE^m, admits a real interpretation in the real linear space E^{2m}([18], section 2.1.2).

Let E^{2m} be a $2m$-dimensional real linear space. *A paracomplex structure* on E^{2m} is an endomorphism $\mathfrak{K} : E^{2m} \to E^{2m}$ such that $\mathfrak{K}^2 = I$. The eigenspaces E_+^m, E_-^m of $\mathfrak{K}$ with eigenvalues $1, -1$ respectively, have the same dimension.

The pair $(E^{2m}, \mathfrak{K})$ will be called a *paracomplex vector space*. We define *the paracomplexification of* E^{2m} as $E_{\mathfrak{C}}^{2m} = E^{2m} \otimes_{\mathbb{R}} \mathfrak{C}$ and we extend $\mathfrak{K}$ to a $\mathfrak{C}$-linear endomorphism $\mathfrak{K}$ of $E_{\mathfrak{C}}^{2m}$.

Lemma 6. *Let $E_{\mathfrak{C}}^{2m} = E^{2m} \otimes_{\mathbb{R}} \mathfrak{C}$ be endowed with an involutive $\mathfrak{C}$-linear endomorphism $\mathfrak{K}$ of $E_{\mathfrak{C}}^{2m}$. Then, the space $E_{\mathfrak{C}}^{2m}$ is decomposed into the direct sum of a pair of m-dimensional subspaces E_+^m and E_-^m such that:*

$$E_{\mathfrak{C}}^{2m} = E_+^m \oplus E_-^m,$$

verifying:

$$E_+^m = \{v \in E_{\mathfrak{C}}^{2m} \,|\, \mathfrak{K}v = \varepsilon v\} = \{v + \varepsilon\mathfrak{K}v \,|\, v \in E_{\mathfrak{C}}^{2m}\},$$
$$E_-^m = \{v \in E_{\mathfrak{C}}^{2m} | \, \mathfrak{K}v = -\varepsilon v\} = \{v - \varepsilon\mathfrak{K}v \,|\, v \in E_{\mathfrak{C}}^{2m}\}.$$

Proof This statement and its proof can be found in the literature. For example, see [18].

Remark 4. From this, it follows that any hypersurface $x^i = 0$ of $E^{2m}_{\mathfrak{C}}$ is decomposed into two hypersurfaces $x^i_+ = 0$ and $x^i_- = 0$ respectively in E^m_+ and E^m_-. Similarly, the coordinates of any point of the space E^{2m} can be given by

$$x^i = x^i_+ e_+ + x^i_- e_-.$$

Furthermore, consider a vector $\mathbf{X}(x^i)$ at the coordinate x^i in the space $E^{2m}_{\mathfrak{C}}$. Then, due to the splitting we have that $\mathbf{X}(x^i) = \mathbf{X}(x^i_+) \oplus \mathbf{X}(x^i_-)$. Therefore, the vector $\mathbf{X}$ splits into $\mathbf{X} = \mathbf{X}_+ \oplus \mathbf{X}_-$ and $\mathbf{X}^2$ is given by $\sum_i x^i_+ x^i_-$.

Proof (of Lemma 3). The tangent space is identified with the space of signed measures with bounded variations, being absolutely continuous with respect to P_θ and such that $\mu(\Omega) = \int_\Omega u dP_\theta = 0$, with $u \in L^2(\Omega, P_\theta)$. Such a measure can be split into two positive measures μ^+ and μ^- such that: $\mu = \mu^+ - \mu^-$ and $|\mu| = \mu^+ + \mu^- < \infty$. This forms a real affine space. Using the Norden–Shirokov construction in [16, 19], leads to the existence of a module over an algebra.

Proposition 5 (*The rank 2* **Lemma, [11]**). *Consider an affine, symmetric space over a Jordan algebra. There exists exactly two affine and flat connections on this space if and only if the algebra is of rank 2, and generated by* $\{1, \varepsilon\}$ *with* $\varepsilon^2 = 1$ *or* -1.

Proof (of Proposition 3). Let $\mathcal{C} \subset \mathcal{W}$ be the cone of positive measures. Let $H \cap \mathcal{C} \neq \emptyset$. By Theorem 1, $\mathcal{C}$ is a cone over $\mathfrak{C}$. So, H inherits the paracomplex structure and by duality, S is therefore also a paracomplex space. We have shown in Lemma 3 that the tangent space to S is identified to a module over an algebra. Hence, the tangent space to H is also a module over an algebra.

5.2 Paracomplex manifold

In this subsection, we expose paracomplex manifolds and their properties. This will be used to prove theorem 2. Let $y = f(x)$ be a (analytic) function, whose domain and range belong to a commutative algebra (i.e. $C^h_{jk} = C^h_{kj}$). We put $x = \sum_i x_i e_i$, $y = \sum_i y_i e_i$. From the generalized Cauchy–Riemann we have the following:

$$\sum_h \frac{\partial y_i}{\partial x_h} C^h_{jk} = \sum_h \frac{\partial y_h}{\partial x_i} C^j_{hk},$$

where C^h_{jk} are the constant structures.

A *paracomplex manifold* is a real manifold M endowed with a paracomplex structure $\mathfrak{K}$ that admits an atlas of paraholomorphic coordinates (which are functions with values in the algebra $\mathfrak{C} = \mathbb{R} + \varepsilon\mathbb{R}$ defined above), such that the transition functions are paraholomorphic.

Explicitly, this means the existence of local coordinates (z_+^α, z_-^α), $\alpha = 1\ldots, m$ such that paracomplex decomposition of the local tangent fields is of the form

$$T^+M = span\left\{\frac{\partial}{\partial z_+^\alpha}, \alpha = 1, ..., m\right\},$$

$$T^-M = span\left\{\frac{\partial}{\partial z_-^\alpha}, \alpha = 1, ..., m\right\}.$$

Such coordinates are called *adapted coordinates* for the paracomplex structure $\mathfrak{K}$.

By abuse of notation, we write ∂_z instead of $\frac{\partial}{\partial z^\alpha}$.

We associate with any adapted coordinate system (z_+^α, z_-^α) a paraholomorphic coordinate system z^α by

$$z^\alpha = \frac{z_+^\alpha + z_-^\alpha}{2} + \varepsilon\frac{z_+^\alpha - z_-^\alpha}{2}, \alpha = 1, ..., m.$$

We define the paracomplex tangent bundle as the $\mathbb{R}$-tensor product $T^{\mathfrak{C}}M = TM \otimes \mathfrak{C}$ and we extend the endomorphism $\mathfrak{K}$ to a $\mathfrak{C}$-linear endomorphism of $T^{\mathfrak{C}}M$. For any $p \in M$, we have the following decomposition of $T_p^{\mathfrak{C}}M$:

$$T_p^{\mathfrak{C}}M = T_p^{1,0}M \oplus T_p^{0,1}M$$

where

$$T_p^{1,0}M = \{v \in T_p^{\mathfrak{C}}M | \mathfrak{K}v = \varepsilon v\} = \{v + \varepsilon\mathfrak{K}v | v \in E^{2m}\},$$
$$T_p^{0,1}M = \{v \in T_p^{\mathfrak{C}}M | \mathfrak{K}v = -\varepsilon v\} = \{v - \varepsilon\mathfrak{K}v | v \in E^{2m}\}$$

are the eigenspaces of $\mathfrak{K}$ with eigenvalues $\pm\varepsilon$. The following paracomplex vectors

$$\frac{\partial}{\partial z_+^\alpha} = \frac{1}{2}\left(\frac{\partial}{\partial x^\alpha} + \varepsilon\frac{\partial}{\partial y^\alpha}\right), \quad \frac{\partial}{\partial z_-^\alpha} = \frac{1}{2}\left(\frac{\partial}{\partial x^\alpha} - \varepsilon\frac{\partial}{\partial y^\alpha}\right)$$

form a basis of the spaces $T_p^{1,0}M$ and $T_p^{0,1}M$.

Proof (of Theorem 2). We use the above subsection on paracomplex manifolds. Now, consider an affine and symmetric space over a Jordan algebra A and consider S as above. From [16, 2, 3, 7, 8], it is shown that there exist 2 connections. Applying Proposition 5, we have that the algebra is of rank 2, and generated by $\{1, \varepsilon\}$ with $\varepsilon^2 = 1$ or -1. These flat affine connections are constructed from a field of objects, having the components:

$$\Gamma_{jk}^i = \Gamma_{jk}^{i\alpha} e_\alpha \in A.$$

Suppose that $\mathbf{v}^i = \mathbf{v}^{(i,\alpha)} e_\alpha$ are quantities from the algebra corresponding to a tangent vector $\mathbf{v}$ to S. Then, from the following condition

$$d\mathbf{v}^i + \Gamma_{jk}^i \mathbf{v}^j d\mathbf{x}^k = 0,$$

we can define an affine connection equipped with the following components:

$$\Gamma^{(i,\alpha)}_{(j,\beta)(k,\gamma)} = \Gamma^{is}_{jk} C^{\delta}_{s\beta} C^{\alpha}_{\delta\gamma},$$

where the $C^{\alpha}_{\beta\gamma}$ are structure constants of algebra A, with respect to the local adapted coordinates $x^{(\alpha,i)}$. Now, these objects are indexed by the number of generators of the algebra A. There exist 2 connections. So, it implies that $s \in \{1,2\}$ and A is of rank 2, generated by $\{1, \varepsilon \,|\, \varepsilon^2 = \pm 1\}$. Since, we cannot be in a complex framework, we have an algebra of paracomplex numbers. Using the duality Lemma 2 between S and H, it follows that H is also a paracomplex manifold.

Proof (of Proposition 4). By theorem 2 we know that S is a paracomplex manifold. Therefore, it can be decomposed into a pair of submanifolds. The structure is ruled by the Jordan algebra $\mathfrak{A}$ of paracomplex numbers, which has a pair of idempotents. Using [15], we can provide a Pierce decomposition and thus a Pierce mirror. The totally geodesic submanifolds of the paracomplex manifold are symmetric to each other wrt the Pierce mirror.

6 Learning & Gromov–Witten invariants for statistical manifolds

6.1 Comparison with the Partial Differential Equations approach

A complementary, approach to the Frobenius problem can be given through the Witten–Dijkgraaf–Verlinde–Verlinde (WDVV) highly non-linear Partial Differential Equations system:

$$\forall a, b, c, d : \sum_{ef} \Phi_{abe} g^{ef} \Phi_{fcd} = (-1)^{a(b+c)} \sum_{ef} \Phi_{bce} g^{ef} \Phi_{fad}.$$

Geometrically, the (WDVV) PDE system express a flatness condition [20].

Proposition 6. *For $\alpha = \pm 1$, the (WDVV) PDE system are always (uniquely) integrable over (S, g, t).*

Proof. The WDVV equations are always integrable if and only if the curvature is null. In the context of (S, g, t) the curvature tensor of the covariant derivative is null for $\alpha = \pm 1$. Therefore, in this context the WDVV equation is always integrable (uniquely).

6.2 Gromov–Witten invariants for statistical manifolds

We introduce Gromov–Witten invariants, in short (GW), for statistical manifolds. In symplectic geometry and algebraic geometry, the (GW) are rational numbers that count (pseudo)holomorphic curves under some conditions, on the (symplectic) manifold, considered as a kind of generalisation of the phase state.

Since statistical manifolds are Frobenius, it is a natural question to ask if one can define analogous (GW) invariants on it. We show that there exist analogous invariants, called (GWS). Notice that these invariants have naturally a different meaning from the classical Gromov–Witten invariants, because we do not work on moduli spaces of genus $\mathfrak{g}$ curves with marked points. However, (GWS) still encode deep geometric aspects of statistical manifolds, and tell whether the learning process has succeeded or not.

Let us consider the (formal) Frobenius manifold (H, g). We denote k a (super)commutative $\mathbb{Q}$-algebra. Let H be a k-module of finite rank and $g : H \otimes H \to k$ an even symmetric pairing (which is non degenerate). We denote H^* the dual to H. The structure of the Formal Frobenius manifold on (H, g) is given by an even potential $\mathbf{\Phi} \in k[[H^*]]$:

$$\mathbf{\Phi} = \sum_{n \geq 3} \frac{1}{n!} Y_n,$$

where $Y_n \in (H^*)^{\otimes n}$ can also be considered as an even symmetric map $H^{\otimes n} \to k$. This system of *Abstract Correlation Functions* in (H, g) is a system of (symmetric, even) polynomials. The (GW) invariants appear in these poly-linear maps.

We go back to statistical manifolds. Let us consider the discrete case of the exponential family formula:

$$\sum_{\omega \in \Omega} \exp\{-\sum \beta^j X_j(\omega)\} = \sum_{\omega \in \Omega} \sum_{m \geq 1} \frac{1}{m!} \left\{ -\sum_j \beta^j X_j(\omega) \right\}^{\otimes m}, \tag{2}$$

where $\beta = (\beta_0, ..., \beta_n) \in \mathbb{R}^{n+1}$ is a canonical affine parametrisation, $X_j(\omega)$ are directional co-vectors, belonging to a finite cardinality $n+1$ list $\mathcal{X}_n$ of random variables. These co-vectors represent necessary and sufficient statistics of the exponential family. We have $X_0(\omega) \equiv 1$, and $X_1(\omega), \ldots, X_n(\omega)$ are linearly independent co-vectors. The family in (2) describes an analytical n-dimensional hypersurface in the statistical manifold. It can be uniquely determined by $n+1$ points in general position.

Definition 1. *Let k be the field of real numbers. Let S be the statistical manifold. The Gromov–Witten invariants for statistical manifolds (GWS) are given by the poly-linear maps:*

$$\tilde{Y}_n : S^{\otimes n} \to k.$$

One can also write them as follows:

$$\tilde{Y}_n \in \left(-\sum_j \beta^j X_j(\omega) \right)^{\otimes n}.$$

These invariants appear as part of the potential function $\tilde{\mathbf{\Phi}}$ which is a Kullback–Liebler entropy function.

One can write the relative entropy function:

$$\tilde{\mathbf{\Phi}} = ln \sum_{\omega \in \Omega} \exp\left(-\sum_j \beta^j X_j(\omega)\right). \tag{3}$$

Therefore, we state the following:

Proposition 7. *The entropy function $\tilde{\mathbf{\Phi}}$ of the statistical manifold relies on the (GWS).*

Proof. Indeed, since $\tilde{\mathbf{\Phi}}$, in formula (3) relies on the poly-linear maps $\tilde{Y}_n \in \Big(-\sum_j \beta^j X_j(\omega)\Big)^{\otimes n}$, defining the (GWS), the statement follows.

6.3 Learning with statistical Gromov–Witten invariants

Consider the tangent fiber bundle over S, the space of probability distributions, with Lie group G. We denote it by (TS, S, π, G, F), where TS is the total space of the bundle $\pi : TS \to S$ is a continuous surjective map and F the fiber. Recall that for any point ρ on S, the tangent space at ρ is isomorphic to the space of bounded, signed measures vanishing on an ideal I of the $\sigma-$algebra. The Lie group G acts (freely and transitively) on the fibers by $f \overset{h}{\mapsto} f + h$, where h is a parallel transport, and f an element of the total space (see [11] for details).

Remark 5. Consider the (local) fibre bundle $\pi^{-1}(\rho) \cong \{\rho\} \times F$. Then F can be identified to a module over the algebra of paracomplex numbers $\mathfrak{C}$ (see [11] for details). By a certain change of basis, this rank 2 algebra generated by $\{e_1, e_2\}$, can always be written as $\langle 1, \varepsilon | \, \varepsilon^2 = 1 \rangle$.

We call a *canonical basis* for this paracomplex algebra, the one given by: $\{e_+, e_-\}$, where $e_\pm = \frac{1}{2}(1 \pm \varepsilon)$. Moreover, any vector $X = \{x^i\}$ in the module over the algebra is written as $\{x^{ia} e_a\}$, where $a \in \{1, 2\}$.

Lemma 7. *Consider the fiber bundle (TS, S, π, G, F). Consider a path γ being a geodesic in S. Consider its fiber F_γ. Then, the fiber contains two connected compontents: (γ^+, γ^-), lying respectively in totally geodesic submanifolds E^+ and E^-.*

Proof. Consider the fiber above γ. Since for any point of S, its the tangent space is identified to module over paracomplex numbers. This space is decomposed into a pair of subspaces (i.e. eigenspaces with eigenvalues $\pm\varepsilon$) (see Section 5.2). The geodesic curve in S is a path such that $\gamma = (\gamma^i(t)) : t \in [0, 1] \to S$. In local coordinates, the fiber budle is given by $\{\gamma^{ia} e_a\}$, and $a \in \{1, 2\}$. Therefore, the fiber over γ has two components (γ^+, γ^-). Taking the canonical basis for $\{e_1, e_2\}$, implies that (γ^+, γ^-) lie respectively in the subspaces E^+ and E^-. These submanifolds are totally geodesic in virtue of Lemma 4.

We define a learning process through the Ackley–Hilton–Sejnowski method [1], which consists in minimising the Kullback–Leibler divergence. By Propositions 2 and 3 in [12], we can restate it geometrically, as follows:

Proposition 8. *The learning process consists in determining if there exist intersections of the paraholomorphic curve γ^+ with the orthogonal projection of γ^- in the subspace E^+.*

In particular, a learning process succeeds whenever the distance between a geodesic γ^+ and the projected one in E^+ shrinks to become as small as possible.

More formally, as was depicted in [5] (sec. 3) let us denote by Υ the set of (centered) random variables over $(\Omega, \mathcal{F}, P_\theta)$ which admit an expansion in terms of the scores under the following form:

$$\Upsilon_P = \{X \in \mathbb{R}^\Omega \,|\, X - \mathbb{E}_P[X] = g^{-1}(\mathbb{E}_P[X d\ell]), d\ell\}.$$

By direct calculation, one finds that the log-likelihood $\ell = ln\rho$ of the usual (parametric) families of probability distributions belongs to Υ_p as well as the difference $\ell - \ell^*$ of log-likelihood of two probabilities of the same family. Being given a family of probability distributions such that $\ell \in \Upsilon_P$ for any P, let $\mathcal{U}_P$, let us denote P^* the set such that $\ell - \ell^* \in \Upsilon_p$. Then, for any $P^* \in \mathcal{U}_p$, we define $K(P, P^*) = \mathbb{E}_P[\ell - \ell^*]$.

Theorem 3. *Let (S, g, t) be statistical manifold. Then, the (GWS) determine the learning process.*

Proof. Since $K(P, P^*) = \mathbb{E}_P[\ell - \ell^*]$, this implies that $K(P, P^*)$ is minimised whenever there is a successful learning process. The learning process is by definition given by *deformation* of a pair of geodesics, defined respectively in the pair of totally geodesic manifolds E^+, E^-. Therefore, the (GWS), which arise in the $\tilde{Y}_n$ in the potential function $\tilde{\mathbf{\Phi}}$, which is directly related to the relative entropy function $K(P, P^*)$. Therefore, the (GWS) determine the learning process.

Similarly as in the classical (GW) case, the (GWS) count intersection numbers of certain para-holomorphic curves. In fact, we have the following statement:

Corollary 1. *Let (TS, S, π, G, F) be the fiber bundle above. Then, the (GWS) determine the number of intersection of the projected γ^- geodesic onto E^+, with the $\gamma^+ \subset E^+$ geodesic.*

References

1. Ackley, D.; Hilton G.; Sejnowski, T. *Learning algorithm for Boltzmann machine, Cognitive science.* **9**, 147–169, (1985).
2. Amari, S. *Differential-geometrical methods in statistics.* Lect. Notes Stat., **28**, Berlin: Springer-Verl., (1985).
3. Amari, S. *Information Geometry* in *Geometry and nature.* Eds. J.-P. Bourguignon, H. Nencka. Providence (R.I.): Amer. Math. Soc., Contemp. Math., **203**, 81–95 (1997).
4. Ay, N.; Jost, J.; Vân Lê, H.; Schwachhöfer, L. *Information geometry.* A Series of Modern Surveys in Mathematics, **64**, Springer, (2017).

5. Burdet, G.; Combe, Ph.; Nencka, H. *Statistical manifolds, self-parallel curves and learning processes.* Seminar on Stochastic Analysis, Random Fields and Applications (Ascona, 1996), 87–99, Progr. Probab., 45, Birkhäuser, Basel, (1999).
6. Burdet, G.; Combe, Ph.; Nencka, H. *Statistical manifolds: the natural affine-metric structure of probability theory.* Mathematical physics and stochastic analysis (Lisbon, 1998), 138–164, World Sci. Publ., River Edge, NJ, (2000).
7. Burdet, G.; Nencka, H. *Equation of self-parallel curve deviation on statistical manifolds.* Methods Funct. Anal. Topology 3 , **1**, 46–50 (1997).
8. Burdet, G.; Nencka, H.; Perrin, M. *An example of dynamical behaviour of the relative entropy*, in *Geometry and nature* Eds. J.-P. Bourguignon, H. Nencka. Providence (R.I.), Contemp. Math., **203**, Amer. Math. Soc., Providence, RI, 97–104, (1997).
9. Čensov, N. *Statistical decision rules and optimal inference.* Trans. of Math. monographs, **53**, Ed. Amer. Math. Soc., (1982).
10. F.M. Ciaglia, J. Jost, L. Schwachhöfer. *Differential Geometric Aspects of Parametric Estimation Theory for States on Finite-Dimensional C^*-Algebras.* Entropy, **22**, 1332, (2020).
11. Combe, N., C.; Manin, Yu., I. *F-manifolds and geometry information.* Bulletin London Maths Society, **52**, Issue 5 October, 777–792, (2020).
12. Combe, Ph.; Nencka, H. *Information Geometry and Learning in Formal Neural networks.* Eds. J.-P. Bourguignon, H. Nencka. Providence (R.I.): Amer. Math. Soc., Contemp. Math., **203**, 105–116 (1997).
13. Hertling, C.; Manin, Yu., I. *Weak Frobenius manifolds.* Int. Math. Res. Notices, **6**, 277–286, (1999).
14. Manin, Yu., I. *Frobenius manifolds, Quantum cohomology, and Moduli spaces.* Amer. Math. Soc., Colloquium Publications, **47**, (1999).
15. K. McCrimmon. *A taste of Jordan algebras.* Universitext. Springer-Verlag, New York, (2004).
16. A. P. Norden. *On intrinsic geometry of the second kind on hypersurfaces of an affine space.* (Russian) Izv. Vysš. Učebn. Zaved. Matematika **4** (5), 172–183 (1958).
17. Rao, C. R. *Information and the Accuracy Attainable in the Estimation of Statistical Parameters.* C.R. Bull. Calcutta Math. Soc., **37**, 81–89 (1945).
18. B. Rozenfeld. *Geometry of Lie groups* Springer-Science Business Media Dordrecht, (1997).
19. A. P. Shirokov. *Spaces over algebras and their applications.* Geometry, 7. J. Math. Sci. (New York) 108, **2**, 232–248 (2002).
20. Sikorski, R. *Differential modules.* Colloquium Mathematicum, Fasc. 1, **24**, 45–79, (1971).