Lifting statistical manifolds

Katarzyna Grabowska∗
Faculty of Physics, University of Warsaw

Janusz Grabowski†
Institute of Mathematics, Polish Academy of Sciences

Marek Kus‡
Center for Theoretical Physics, Polish Academy of Sciences

Giuseppe Marmo§
Dipartimento di Fisica “Ettore Pancini”, Università “Federico II” di Napoli
and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli

Abstract

We consider some natural (functorial) lifts of geometric objects associated with statistical manifolds (metric tensor, dual connections, skewness tensor, etc.) to higher tangent bundles. It turns out that the lifted objects form again a statistical manifold structure, this time on the higher tangent bundles, with the only difference that the metric tensor is pseudo-Riemannian. This, however, is nowadays commonly accepted, if abstract statistical manifolds (not necessarily associated with the standard statistical models) are concerned. What is more, natural lifts of potentials (called also divergence or contrast functions) turn out to be again potentials, this time for the lifted statistical structures. We illustrate these geometric constructions with lifting statistical structures of some important statistical models.

1 Introduction

Since the pioneering work of Amari and Chentsov [1, 2, 3, 4, 5, 6, 12], information geometry has flourished greatly and became an object of intensive studies and various applications, especially in statistical decision rules and optimal inferences. This theory appeared to provide a link between different disciplines where statistical-probabilistic aspects play a relevant role. The search for such a link was clearly advocated by J. A. Wheeler [40]. A remarkable theorem by Chentsov [12], proved in a categorical setting, states that the

∗email:konicznu@fuw.edu.pl
†email: jagrab@impan.pl
‡email: marek.kus@cft.edu.pl
§email: marmo@na.infn.it
Fisher-Rao metric tensor is the only one respecting a monotonicity requirement under coarse grained transformations and, moreover, that the metric is invariant under the diffeomorphism group of the sample space.

Also quantum information geometry has been developed which is not strange, as the standard quantum mechanics is a probabilistic-statistical theory. The geometrical formulation of quantum mechanics has been used to show that the Fisher-Rao metric tensor can be obtained as the pull-back of the Fubini-Study metric tensor when the statistical manifold is embedded as an appropriate Lagrangian submanifold of the complex projective space of pure quantum states [19]. The possibility to obtain the Fisher-Rao metric from the Fubini-Study metric shows a deep connection between geometric information theory and the geometrical formulation of quantum mechanics (indeed, both theories admit a statistical-probabilistic description).

The geometrical objects introduced for statistics are Riemannian metrics, symmetric \((0,3)\) tensors and dual pairs of torsionless affine connections. Such geometric structures give rise to the concept of so called \textit{statistical manifold} which are purely geometric abstracts of the geometries of statistical models. Moreover the notion of ‘directed distances’ (called also distinguishability functions, divergence functions, or contrast functions), introduced as potential functions in the classical setting, serve as ‘generating objects’ of statistical manifolds. Classical potentials are related to relative entropies in the quantum setting. Thus the latter were used as potential functions for quantum metrics on the space of quantum states. It should be noticed, however, that potential functions appear also in the description of Kähler manifolds, Frobenius manifolds, and that the connection with statistical manifolds was considered in a paper by Y. Manin and his collaborators [10].

A natural problem is to find natural examples of statistical structures. An obvious idea is to build additional statistical models out of a given one. One possibility is lifting statistical structures to fibrations over statistical manifolds. In this paper we consider this lifting problem from the point of view of intrinsic differential geometry and functorial constructions, by lifting statistical structures to higher tangent bundles. We will show that this approach provides lifted potential functions for the lifted geometric objects, an interesting result providing an effective way of producing whole families of statistical structures.

The ‘naturality requirement’ for the construction provides metric tensors which are not Riemannian but only pseudo-Riemannian what is nowadays commonly accepted. As the metric is used mainly to introduce a notion of distinguishability among probability distributions or quantum states, this raises the question of a proper interpretation for probability distribution which have zero-distance even though they are distinct. This aspect, however, will not be considered in this paper and will be dealt with in the future.

2 Differential geometry of statistical models

Let us consider a paradigmatic experiment in which we perform a series of measurements obtaining some experimental distribution of them. Typically, upon our results we want to find the ‘true’ distribution, or rather identify a distribution (taken from a prescribed family), that best fits to the obtained data. It means that we are dealing with probability distributions \(p(\omega, x) d\omega\) depending on some parameters \(x\), identifying members of the considered family of distributions. Thus, to achieve the mentioned goals we need a parameterization of the family of distributions and possibly a kind of distance-like function that measures a relative distance between distributions, or more generally, allows
Information geometry provides a link between statistical models and differential (Riemannian) geometry. The idea is to parameterize the space of probability distributions on a measure space \( \Omega \), with a measure \( d\omega \) (a sample space). Let thus \( \mathcal{P}(\Omega) \) denote such a sample space. We parameterize it via a map \( M \ni m \mapsto p(\omega, x) d\omega \) from a differentiable manifold \( M \). That is, for each point \( x \) in the parameter space \( M \), we have a probability measure \( p(\cdot, x) \) on \( \Omega \).

The notions of a distance and distinguishability between distributions can be formulated \([3, 6]\), by introducing a two-point potential function \( F : M \times M \to \mathbb{R} \) \([1, 3]\), usually called a contrast function or a divergence. It is usually assumed that the potential function is non-negative, \( F(x, y) \geq 0 \), and vanishes exactly on the diagonal, i.e., \( F(x, y) = 0 \), if and only if \( x = y \). Let, \( \{\zeta^i\} \) be a coordinate system on the first manifold \( M \) and \( \{\xi^j\} \) on the second. If \( F \) is at least \( \mathcal{C}^3 \) the condition imposed on \( F \) imply \([3, 14]\),

\[
\frac{\partial F}{\partial \zeta^j} \bigg|_{\zeta = \xi} = \frac{\partial F}{\partial \xi^j} \bigg|_{\zeta = \xi} = 0.
\]

We can now construct a two- and three-tensors \( g \) (a metric) and \( T \) (the skewness tensor) as

\[
g_{jk} = \left. \frac{\partial^2 F}{\partial \zeta^j \partial \xi^k} \right|_{\zeta = \xi} = \left. \frac{\partial^2 F}{\partial \xi^j \partial \xi^k} \right|_{\zeta = \xi} = - \left. \frac{\partial^2 F}{\partial \xi^j \partial \eta^k \partial \xi^l} \right|_{\zeta = \xi},
\]

and

\[
T_{jkl} = \left. \frac{\partial^3 F}{\partial \zeta^j \partial \xi^k \partial \xi^l} \right|_{\zeta = \xi} = - \left. \frac{\partial^3 F}{\partial \xi^j \partial \xi^k \partial \xi^l} \right|_{\zeta = \xi}.
\]

For given tensors \( g \) and \( T \) we consider a one-parameter family of torsionless connections

\[
\Gamma_{jkl}^\alpha := \Gamma_{jkl}^g - \frac{\alpha}{2} T_{jkl},
\]

where \( \Gamma_{jkl}^g \) are the Christoffel symbols for the metric tensor \( g \) (the Levi-Civita connection). On Riemannian manifold there is a concept of dual connections, namely \( \nabla \) and \( \nabla^\ast \) are dual with respect to \( g \) if for all vector fields \( X, Y, Z \) on \( M \)

\[
Z \left( g(X, Y) \right) = g \left( \nabla_Z X, Y \right) + g \left( X, \nabla_Z^\ast Y \right).
\]

One can check that \( \nabla^\alpha \) and \( \nabla^{-\alpha} \) are dual to each other, i.e. \( (\nabla^\alpha)^\ast = \nabla^{-\alpha} \). The statistical model is called self-dual if \( \nabla^\alpha = \nabla^{-\alpha} \). The self-duality implies \( T = 0 \).

The most prominent example of a statistical model is that proposed by Rao \([38]\). The manifold \( M \) is equipped with the metric (the Fisher-Rao metric),

\[
g_{jk} = \int_\Omega p(\omega, \zeta) \left( \frac{\partial \log (p(\omega, \zeta))}{\partial \zeta^j} \right) \left( \frac{\partial \log (p(\omega, \zeta))}{\partial \zeta^k} \right) d\omega.
\]

and the corresponding skewness tensor

\[
T_{jkl} = \int_\Omega p(\omega, \zeta) \left( \frac{\partial \log (p(\omega, \zeta))}{\partial \zeta^j} \right) \left( \frac{\partial \log (p(\omega, \zeta))}{\partial \zeta^k} \right) \left( \frac{\partial \log (p(\omega, \zeta))}{\partial \xi^l} \right) d\omega.
\]

As a contrast function defining the Fisher-Rao metric we can take the so called Kullback-Leibler divergence (Shannon relative entropy)

\[
F(x, y) = D_{KL}(x, y) = \int_\Omega p(\omega; x) \log \frac{p(\omega; x)}{p(\omega; y)} d\omega.
\]
The geometrical structures of statistical models were an inspiration for the concept of a general statistical manifold, a priori not necessarily associated with a statistical model. It can be analyzed in terms of (pseudo)-Riemannian geometry without any particular connection to statistics. However, every statistical manifold has a statistical model, as it was demonstrated in [30]. In the following section we shall review the definition of a statistical manifold as a geometric construction.

3 Statistical manifolds

The original definition of a statistical manifold by Lauritzen [29] is the following (see also [11, 27, 32, 34, 43]).

**Definition 3.1.** A statistical manifold is a Riemannian manifold \((M, g)\) with a symmetric covariant 3-tensor \(T\).

**Remark 3.2.** It is obvious that any submanifold \(N\) of a statistical manifold is a statistical manifold itself with the tensors \(g_N = g|N\) and \(T_N = T|N\).

All geometric statistical models carry such a structure. The tensor \(T\) is called the skewness tensor (in statistical models – Amari-Chensov tensor) or a cubic form (cubic tensor). Recently, also pseudo-Riemannian metrics have been admitted in the definition of statistical manifolds, as all main features of statistical manifolds may be carried over to this more general framework. We will use this more general definition. Note however, that Remark 3.2 is no longer true in the more general situation.

In our considerations all geometric objects (manifolds, functions, vector fields, metrics, etc.) will be smooth and all connections will be affine connections.

There are several equivalent definitions of a statistical manifolds in the literature. Depending on the problem, one or another may be more suitable. A particular role is played by affine connections and their duals with respect to the metric. We mentioned it already in previous section about statistical models, see (1).

**Definition 3.3.** Let \(\nabla\) be a connection on a pseudo-Riemannian manifold \((M, g)\). The dual connection (with respect to \(g\)), called also the \(g\)-conjugate connection, is the connection \(\nabla^*\) defined by

\[
g(\nabla^*_X Y, Z) = X g(Y, Z) - g(Y, \nabla_X Z)\,.
\]

Note that it is a true duality, as \((\nabla^*)^* = \nabla\) and that the Levi-Civita connection \(\nabla^g\) for \((M, g)\) is the only self-dual connection which is torsion-free.

We will say that two geometric structures on a pseudo-Riemannian manifold are equivalent if one of them canonically determines the other and vice versa. The following is essentially due to Lauritzen [29].

**Theorem 3.4.** Let \((M, g)\) be a pseudo-Riemannian manifold. The following geometric structures on \((M, g)\) are equivalent.

1. a symmetric \((0,3)\)-tensor \(T\), i.e. \((M, g, T)\) is a statistical manifold;

2. a torsion-free connection \(\nabla\) such that \(\nabla g\) is symmetric (\((g, \nabla)\) are Codazzi coupled);
3. a torsion free connection $\nabla$ such that
$$\nabla^g = \frac{1}{2}(\nabla + \nabla^\ast);$$

4. a pair $(\nabla, \nabla^\ast)$ of dual torsion-free connections (dualistic structure).

The tensor $T$ and connections $\nabla, \nabla^\ast$ are related by
$$g(\nabla_X Y, Z) = g(\nabla_X^0 Y, Z) - \frac{1}{2}T(X, Y, Z)$$
$$g(\nabla^\ast_X Y, Z) = g(\nabla^\ast_X^0 Y, Z) + \frac{1}{2}T(X, Y, Z).$$

Remark 3.5. It is clear that $(M, g, \nabla)$ is a statistical manifold if and only if $(M, g, \nabla^\ast)$ is a statistical manifold, so called the dual statistical manifold.

Actually, the dual pair $(\nabla, \nabla^\ast)$ can be extended to a one-parametr family of connections,
$$g(\nabla^\alpha_X Y, Z) = g(\nabla^\alpha_X^0 Y, Z) - \frac{\alpha}{2}T(X, Y, Z).$$

Here, $\alpha \in \mathbb{R}$ and $\nabla^{-\alpha} = (\nabla^\alpha)^\ast$. Clearly $\nabla^1 = \nabla$ (called the exponential connection), $\nabla^{-1} = \nabla^\ast$ (called the mixture connection), and $\nabla^0 = \nabla^g$ is the metric (Levi-Civita) connection. The tensor $T$ can be obtained from the connection via formula
$$T(X, Y, Z) = g(\nabla^\ast_X Y, Z) - g(\nabla_X Y, Z) = (\nabla_X g)(Y, Z)$$
and is sometimes also called the cubic form of $(M, g, \nabla)$.

Recently, statistical manifolds admitting torsion are considered as well [26, 28, 33].

Definition 3.6. Let $g$ be a pseudo-Riemannian metric on a manifold $M$ and $\nabla$ be a connection on $M$. We call the pair $(M, g, \nabla)$ a statistical manifold admitting torsion (SMAT) if
$$(\nabla_X g)(Y, Z) - (\nabla_Y g)(X, Z) = -g(\text{Tor}^\nabla(X, Y), Z).$$

In this case we call the pair $(g, \nabla)$ torsion coupled.

Proposition 3.7. A pair $(g, \nabla)$ is torsion coupled if and only if $\nabla^\ast$ is torsion-free.

In consequence, if $\nabla, \nabla^\ast$ are torsion coupled with $g$, then $(M, g, \nabla, \nabla^\ast)$ is a statistical manifold.

A very useful method of constructing statistical manifolds is that by means of contrast functions [9, 13, 14, 15, 17, 18]. They are also called potentials, divergences, yokes etc.

Let $M$ be a manifold and $X$ be a vector field on $M$. Denote with $^1X$ (resp., $^2X$) the vector field on $M \times M$ which is $X$ on the left factor (resp., on the right factor). In coordinates, if $(z^i)$ are coordinates on $M$ and $(x^i)$, $(y^i)$ denote the same coordinates on the first and the second factor, we have
$$^1(f(z)\partial_{z^i})(x, y) = f(x)\partial_{x^i}, \quad ^2(f(z)\partial_{z^i})(x, y) = f(y)\partial_{y^i}.$$ Note that the vector fields $^1X$ and $^2Y$ commute for all $X, Y$. It will be convenient to use the notation
$$^1X_1 \cdots ^1X_k \cdot ^2Y_1 \cdots ^2Y_l(F) =: F[X_1 \cdots X_k|Y_1 \cdots Y_l].$$
Remark 3.8. For a function $F$ on a manifold $N$ of dimension $n$ with vanishing $k$-th jet at $p \in N$, and for any vector fields $X_1, \ldots, X_{k+1}$ on $N$, the expression $X_1 \cdots X_{k+1}(F)(p)$ depends only on $X_i(p)$, $i = 1, \ldots, k+1$, and

$$g^F(X_1, \ldots, X_{k+1}) = (-1)^k (X_1 \cdots X_{k+1})(F)(p)$$

(3)

defines uniquely a symmetric $(0, k+1)$-tensor on $N$ (cf. [20, Lemma 5.1]). We say that $g^F$ is induced by the function $F$.

The following proposition will be useful while discussing local forms of contrast functions and other geometric objects on statistical manifold.

Proposition 3.9. If $u^i$ are local coordinates in a neighbourhood of $p \in N$, $u^i(p) = 0$, then $F : N \to \mathbb{R}$ has the $k$-th jet at $p$ vanishing if and only if $F$ can be locally written as

$$F(u) = F_{\alpha_1, \ldots, \alpha_n}(u)(u^i)^{\alpha_i} \cdots (u^n)^{\alpha_n}, \quad \alpha_j \in \mathbb{N} \cup \{0\}, \quad \sum_{j=1}^n \alpha_j = k + 1$$

(4)

for some functions $F_{\alpha_1, \ldots, \alpha_n}$ defined in a neighbourhood of $p$. Moreover, for the $(0, k+1)$-tensor (3) we have

$$g^F(\partial_{u^i_1}, \ldots, \partial_{u^i_{k+1}})(p) = (\beta_1)! \cdots (\beta_n)! F_{\beta_1, \ldots, \beta_n}(p),$$

where $\beta_j$ is the number of those $i_i$ which equal $j$. The converse is obvious.

Proof. Indeed, it is well known (cf. also the proof of the next proposition) that if $F(p) = 0$, then there exist functions $F_i$ such that $F(u) = F_i(u)u^i$. We get (3.9) by induction. The rest follows easily by differentiating (4).

Denote the diagonal submanifold in $M \times M$, i.e. $\{(x, x) \mid x \in M\} \subset M \times M$, with $\Delta_M$.

Definition 3.10. A function $F : M \times M \to \mathbb{R}$ we call a contrast function (potential function, divergence function) on $M$ if the first jet of $F$ vanishes on $\Delta_M$, i.e. $F|_{\Delta_M} = 0$ and $dF|_{\Delta_M} = 0$, and $g^F$ is a pseudo-Riemannian metric.

Remark 3.11. Note that, in consequence of the fact that we admit pseudo-Riemannian metrics in the definition of a statistical manifold, our concept of a contrast function is more general that the ‘classical’ one and accepts $g^F$ to be only pseudo-Riemannian.

Remark 3.12. According to Remark 3.8 the pseudo-Riemannian metric $g^F$ reads

$$g^F(X, Y)(x) = -F[X|Y](x, x).$$

Denote $F^*(x, y) = F(y, x)$. It is easy to see that $F$ is a contrast function if and only if $F^*$ is a contrast function and that $g^F = g^{F^*}$. Of course, if $F$ is a contrast function and $F \geq 0$, then $g$ is Riemannian.

Remark 3.13. Actually, any statistical manifold is induced by a contrast function [34]. It is interesting that contrast functions on $M$ define also symplectic structures on $M \times M$ [8].
Remark 3.14. In [20] (see also [21]) we discovered that contrast functions on $M \times M$ and the way of inducing statistical structures out of them uses actually only the canonical structure of a Lie groupoid on $M \times M \rightrightarrows M$, called the pair groupoid. This observation was a motivation for developing a concept of contrast functions on Lie groupoids $\mathcal{G} \rightrightarrows M$ inducing statistical structures on the Lie algebroid $\text{Lie}(\mathcal{G})$. In the case of the pair groupoid $\text{Lie}(M \times M) = T M$.

Let $M$ be a manifold of dimension $n$, $p \in M$ and $(x^i)$ be local coordinates on $M$ in a neighbourhood of $p \in M$, $x^i(p) = 0$. Let $(y^i)$ be the same coordinates on another copy of $M$, so that $(x^i, y^i)$ are local coordinates in a neighbourhood of $(p, p) \in M \times M$.

**Theorem 3.15.** (Local characterization of contrast functions) Let $M$ be a manifold, $p \in M$ and $(x^i, y^i)$ be the local coordinates in a neighbourhood of $(p, p)$ as above. Then a function $F : M \times M$ is a contrast function on $M$ with the induced pseudo-Riemannian metric $g^F$ which locally reads

$$g^F = g^F_{ij}(x) \, dx^i \otimes dx^j$$

if and only if, in a neighbourhood of $(p, p) = (0, 0)$,

$$F(x, y) = \frac{1}{2} (x^i - y^i)(x^j - y^j) \, h_{ij}(x, y),$$

where $h_{ij} = h_{ji}$ and $[h_{ij}(0, 0)]$ is an invertible matrix. In such a case $g^F_{ij}(x) = h_{ij}(x, x)$ for $x$ in a neighbourhood of $0$.

**Proof.** Suppose $F$ is a contrast function on $M$ and consider new coordinates in a neighbourhood of $(p, p) \in M \times M$:

$$u^i = x^i - y^i \quad v^i = x^i + y^i.$$ 

The diagonal submanifold $\Delta_M$ is defined locally by $z^i = 0$, $i = 1, \ldots, n$ and $F(0, v) = 0$. Moreover, as $dF(0, v) = 0$, we have

$$\left(\frac{\partial F}{\partial u^i}\right)(0, v) = 0 \quad \text{for} \quad j = 1, \ldots, n.$$

For fixed $(u, v)$ and $t \in \mathbb{R}$ put $G(t) = F(tu, v)$. We have

$$F(u, v) = G(1) - G(0) = \int_0^1 G'(s) \, ds = \int_0^1 \frac{\partial F}{\partial u^i}(su, v) \, u^i \, ds = \left(\int_0^1 \frac{\partial F}{\partial u^i}(su, v) \, ds\right) u^i.$$

Denote

$$h_i(u, v) = \left(\int_0^1 \frac{\partial F}{\partial u^i}(su, v) \, ds\right),$$

so that $F(u, v) = h_i(u, v) u^i$. Since $\left(\frac{\partial F}{\partial u^i}\right)(0, v) = h_i(0, v) = 0$, repeating the previous calculation, we get $h_i(u, v) = \frac{1}{2} h_{ij}(u, v) u^j$, so finally

$$F(u, v) = \frac{1}{2} u^i u^j h_{ij}(u, v)$$

for some functions $h_{ij} = h_{ji}$ defined in a neighbourhood of $(p, p)$. It is now easy to see that

$$g^F_{ij}(0) = g^F(\partial_{x^i}, \partial_{x^j})(x) = -\frac{\partial^2 F}{\partial x^i \partial x^j}(x, x) = h_{ij}(x, x).$$
Conversely, if $F$ has the local form (5) with $h_{ij} = h_{ji}$, then clearly $F(x, x) = 0$. Moreover,
\[
\frac{\partial F}{\partial x^i}(x, x) = \left[ (x^j - y^j) h_{ij} \right](x, x) = 0.
\]
Similarly, $\frac{\partial F}{\partial x^i}(x, x) = 0$ so $dF(x, x) = 0$. Finally, we show, exactly like in (6) that
the induced tensor $g^F$ satisfies $g^F_{ij}(x) = h_{ij}(x, x)$. Hence, $g^F$ is non-degenerate, so a pseudo-Riemannian metric, that finishes the proof.

\[\square\]

**Theorem 3.16.** In coordinates $(x^i, y^j)$ on $M \times M$ as above, any contrast function $F$ in a neighbourhood of $(p, p)$ can be written as
\[
F(x, y) = \frac{1}{2} (x^i - y^i)(x^j - y^j) g^F_{ij}(0) + x^i x^j y^k \gamma_{ijk}(x, y) + y^i y^j x^k \theta_{ijk}(x, y)
\]
\[\text{where } \gamma_{ijk} \text{ and } \theta_{ijk} \text{ are symmetric with respect to the first two indices. Moreover, } t_{ijk} = \gamma_{ijk} - \theta_{ijk} \text{ and } a_{ijk}, b_{ijk} \text{ are totally symmetric.}
\]

**Proof.** Since $h_{ij} - h_{ij}(p, p)$ vanishes at $(p, p)$, using original coordinates $(x^i, y^j)$, we can write (cf. (5))
\[
h_{ij}(x, y) = g^F_{ij}(0, 0) + 2 \alpha_{ijk}(x, y) x^k + 2 \beta_{ijk}(x, y) y^k,
\]
so that (cf. (5))
\[
F(x, y) = (x^i - y^i)(x^j - y^j) \left( \frac{1}{2} g^F_{ij}(0, 0) + \alpha_{ijk}(x, y) x^k + \beta_{ijk}(x, y) y^k \right).
\]
Since $h_{ij} = h_{ji}$, we have $\alpha_{ijk} = \alpha_{jik}$ and $\beta_{ijk} = \beta_{jik}$ and finally
\[
F(x, y) = \frac{1}{2} (x^i - y^i)(x^j - y^j) g^F_{ij}(0) + x^i x^j y^k \gamma_{ijk}(x, y) + y^i y^j x^k \theta_{ijk}(x, y)
\]
\[\text{where } \gamma_{ijk} = \beta_{ijk} - (\alpha_{ikj} + \alpha_{jki}), \quad \theta_{ijk} = \alpha_{ijk} - (\beta_{ikj} + \beta_{jki})\]
and
\[a_{ijk} = \alpha_{ijk} + (\text{cycl}), \quad b_{ijk} = \beta_{ijk} + (\text{cycl}),\]
where ‘(cycl)’ denotes the cyclic permutations of $(ijk)$ and (7) follows. It is obvious that
$\gamma_{ijk}, \theta_{ijk}$ are symmetric with respect to the first two indices and that $a_{ijk}, b_{ijk}$ are totally symmetric. Moreover,
\[
t_{ijk} = \theta_{ijk} - \gamma_{ijk} = (\alpha_{ikj} + \alpha_{jki} + \alpha_{ijk}) - (\beta_{ikj} + \beta_{ikj} + \beta_{jki})
\]
is totally symmetric.

\[\square\]

**Remark 3.17.** As follows from the proof, the functions $\gamma_{ijk}, \theta_{ijk}, a_{ijk}, b_{ijk}$ in (7) cannot be arbitrary.
It is easy to see now that
\[
F^*(x, y) = \frac{1}{2} (x^i - y^i)(x^j - y^j) g^F_{ij}(0) + x^i x^j y^k \cdot \theta_{ijk}(y, x) + y^i y^j x^k \cdot \gamma_{ijk}(y, x) \\
+ x^i x^j x^k \cdot b_{ijk}(y, x) + y^i y^j y^k \cdot a_{ijk}(y, x),
\]
so that \((F - F^*)(x, y)\) reads
\[
(F - F^*)(x, y) = x^i x^j y^k \cdot (\gamma_{ijk}(x, y) - \theta_{ijk}(y, x)) + y^i y^j x^k \cdot (\theta_{ijk}(x, y) - b_{ijk}(y, x)) \\
+ y^i y^j y^k \cdot (b_{ijk}(x, y) - a_{ijk}(y, x)).
\]

We have already discussed the connection between a contrast function and a metric on statistical manifold. Now we pass to the rest of the structure. Let \(F\) be a contrast function on a manifold \(M\) and \(g^F\) be the pseudo-Riemannian metric induced by \(F\). The induced connection \(\nabla^F\) and the skewness tensor \(T^F\) are determined by the formulae
\[
g^F(\nabla^F_X Y, Z)(x) = -F[XY]Z(x), \quad T^F(X, Y, Z)(x) = (F - F^*)[X|YZ](x).
\]
Note that the formula for \(T^F\) makes sense, since the second jets of \(F - F^*\) vanish on the diagonal (cf. Remark 3.21). The connection dual to \(\nabla^F\) can be obtained from
\[
g^F((\nabla^F)^*_X Y, Z)(x) = -F[Z|XY](x, x) = -F^*[XY|Z](x, x).
\]
It is easy to see that \((\nabla^F)^* = \nabla^{F^*}\).\]

**Theorem 3.18.** The connections \(\nabla^F, (\nabla^F)^* = \nabla^{F^*}\) are torsion-free and the pair \((g^F, \nabla)\) is Codazzi coupled.

**Corollary 3.19.** If \(F\) is a contrast function on \(M\), then \((M, g^F, T^F), (M, g^F, \nabla^F)\) and \((M, g^F, \nabla^F, (\nabla^F)^*)\) are different forms of the same statistical manifold.

Writing \(\nabla^F, (\nabla^F)^* = \nabla^{F^*}\), and \(T^F\) in local coordinates of Theorem 3.15, we get
\[
\nabla^F_{\partial_x^i} \partial_x^j = (\Gamma^F)^i_{ij}(x) \partial_x^i, \quad (\nabla^F)^*_x \partial_x^j = (\Gamma^F)^j_{ij}(x) \partial_x^i, \quad T^F(\partial_x^i, \partial_x^j, \partial_x^k) = T_{ijk}(x),
\]
and using the local form \(\Box\) of \(F\), the local form \(\Box\) of \(F^*\), and the local form \(\Box\) of \((F - F^*)\), we get the following.

**Theorem 3.20.** The local coefficients of \(\nabla^F\) and \(T^F\) read
\[
(\Gamma^F)^i_{ij}(x) = g^F \left( \nabla^F_{\partial_x^i} \partial_x^j, \partial_x^k \right) (x) \cdot (g^F)^{kl}(x) = -[\partial_x^i, \partial_x^j, \partial_x^k] F(x, x)] \cdot (g^F)^{kl}(x)
\]
\[
= -\gamma_{ijk}(x, x) \cdot (g^F)^{kl}(x);
\]
\[
(\Gamma^{F^*})^i_{ij}(x) = g^F \left( \nabla^{F^*}_{\partial_x^i} \partial_x^j, \partial_x^k \right) (x) \cdot (g^F)^{kl}(x) = -[\partial_x^i, \partial_x^j, \partial_x^k] F^*(x, x)] \cdot (g^F)^{kl}(x)
\]
\[
= -\theta_{ijk}(x, x) \cdot (g^F)^{kl}(x);
\]
\[
T^F_{ijk}(x) = \partial_x^i \partial_x^j \partial_x^k (F - F^*)(x, x) = \theta_{kji}(x, x) - \gamma_{kji}(x, x) = t_{ijk}(x, x).
\]

It is easy to see that \(\nabla^F\) is symmetric (torsion-free) and \(T^F\) is totally symmetric.

**Remark 3.21.** Statistical manifolds admitting torsion are induced from the so called pre-contrast functions \(\Box, \Box, \Box\), but we will not consider pre-contrast functions in this paper.
4 Lifting geometrical structures to the tangent bundle

Let $M$ be a manifold and

$$\mathcal{T}(M) = \bigoplus_{p,q=0}^{\infty} \mathcal{T}^q_p(M)$$

be the algebra of tensor fields on $M$, elements of $\mathcal{T}^q_p(M)$ being the $q$-contravariant and $p$-covariant tensor fields on $M$. There is a canonical injective homomorphism of the algebra $\mathcal{T}(M)$ into the algebra of $\mathcal{T}(TM)$ tensor fields on the tangent bundle $TM$,

$$\text{Vert} : \mathcal{T}(M) \rightarrow \mathcal{T}(TM), \quad K \mapsto K^v,$$

The tensor $K^v$ is called the vertical lift of $K$. Since Vert is an algebra homomorphism, $(K \otimes S)^v = K^v \otimes S^v$, it is enough to define the vertical lifts of functions, 1-forms, and vector fields. In local coordinates $(x^i)$ on $M$ and the adapted coordinates $(x^i, \dot{x}^j)$ on $TM$ the lifts read

$$f^v(x, \dot{x}) = f(x),$$

$$\left( f_i(x) dx^i \right)^v = f_i(x) dx^i,$$

$$\left( X_i(x) \partial_{x^i} \right)^v = X_i(x) \partial_{\dot{x}^i}.$$

There is another lift of tensor fields to the tangent bundle $[22, 41, 42]$

$$d_T : \mathcal{T}(M) \rightarrow \mathcal{T}(TM), \quad K \mapsto K^c,$$

called the tangent or complete lift. In this case, $d_T$ is a Vert-derivation, i.e.

$$(K \otimes S)^c = K^c \otimes S^v + K^v \otimes S^c. \quad (10)$$

Again, due to (10), it is enough to define the vertical lifts of functions, 1-forms, and vector fields:

$$f^c(x, \dot{x}) = \frac{\partial f}{\partial x^i}(x) \dot{x}^i,$$

$$\left( f_i(x) dx^i \right)^c = \frac{\partial f}{\partial x^i}(x) dx^i + f_i(x) d\dot{x}^i,$$

$$\left( X_i(x) \partial_{x^i} \right)^c = \frac{\partial X}{\partial x^j}(x) \dot{x}^j \partial_{\dot{x}^i} + X_i(x) \partial_{x^i}.$$

In particular, the complete lift of a 2-covariant tensor $g = g_{ij} \, dx^i \otimes dx^j$ is

$$(g_{ij} \, dx^i \otimes dx^j)^c = \frac{\partial g_{ij}}{\partial x^k} \, \dot{x}^k \, dx^i \otimes dx^j + g_{ij} \left( d\dot{x}^i \otimes dx^j + dx^i \otimes d\dot{x}^j \right). \quad (11)$$

If $g$ is symmetric (anti-symmetric), then $g^c$ is symmetric (anti-symmetric). If $g$ is non-degenerate, then $g^c$ is non-degenerate. However, if $g$ is a Riemannian metric, then $g^c$ is never Riemannian, but pseudo-Riemannian of index 0.

Proposition 4.1. ([41]) If $K \in \mathcal{T}^q_p(M)$, then

$$K^c(X_1^c, \ldots, X_p^c) = (K(X_1, \ldots, X_p))^c,$$

for any vector fields $X_1, \ldots, X_p$ on $M$. 10
Remark 4.2. Note that the complete lift of vector fields can be generalized a complete
lift of sections of a Lie algebroid \( \tau : E \to M \) into the space of vector fields on \( E \). It is a homomorphism of Lie brackets, \([X,Y]_E = [X^e, Y^e]\). Here, the bracket on the left hand side is the Lie algebroid bracket of sections \( X, Y \in \text{Sec}(E) \), while the bracket on the right hand side is the Lie bracket of vector fields. This lift completely characterizes the Lie algebroid.

There exist not only complete lifts of tensor fields, but also complete lifts of affine connections.

\[ \text{Theorem 4.3. } (41) \] Let \( \nabla \) be a connection on a manifold \( M \). Then there exists a unique connection \( \nabla^c \) on \( \mathcal{T}M \) such that
\[ \nabla^c_{X^e}(Y^e) = (\nabla_X Y)^c \]
for any vector fields \( X, Y \) on \( M \). Moreover, for any tensor field \( K \) on \( M \), we have
\[ \nabla^c_{X^e}(K^e) = (\nabla_X K)^c \quad \text{and} \quad \nabla^c(K^e) = (\nabla K)^c. \]

If \( \text{Tor}^\nabla \) and \( \text{R}^\nabla \) are the torsion and the curvature of \( \nabla \), then
\[ \text{Tor}^\nabla^c = (\text{Tor}^\nabla)^c \quad \text{and} \quad \text{R}^\nabla^c = (\text{R}^\nabla)^c. \]

\[ \text{Corollary 4.4. If } \nabla^g \text{ is the metric (Levi-Civita) connection for a pseudo-Riemannian metric } g, \text{ then } (\nabla^g)^c \text{ is the metric connection for the pseudo-Riemannian metric } g^c. \]

Let \( \Gamma^i_{jk} \) be the connection components for \( \nabla \) with respect to a local coordinate system \((x^i)\). Then the coordinate components of \( \nabla^c \) with respect to the induced coordinate system \((x^i, \dot{x}^j)\) are
\[ \tilde{\Gamma}^i_{jk} = \Gamma^i_{jk}, \quad \tilde{\Gamma}^i_{jk} = \Gamma^i_{jk}, \quad \tilde{\Gamma}^i_{jk} = \Gamma^i_{jk}, \quad \tilde{\Gamma}^i_{jk} = \frac{\partial \Gamma^i_{jk}}{\partial x^l} \dot{x}^l, \quad \tilde{\Gamma}^i_{jk} = 0, \quad \tilde{\Gamma}^i_{jk} = 0, \quad \tilde{\Gamma}^i_{jk} = 0. \]

Here the indices with bars refer to \((\dot{x}^i)\).

5 Higher lifts

The tangent lifts described in the previous sections have been generalized to lifts from \( M \) to higher tangent bundles \( \mathcal{T}^r M \) by Morimoto [35]. Let us recall that \( \mathcal{T}^r M \) is the bundle of \( r \)-jets of smooth curves
\[ \gamma : \mathbb{R} \to M, \quad \mathbb{R} \ni t \mapsto \gamma(t) \in M \]
at \( t = 0 \). Such jets are classes \([\gamma]_r\) of curves \( \gamma \) which at \( 0 \in \mathbb{R} \) are tangent of the order \( r \), i.e. have the same position, velocities, acceleration, etc. More precisely, the two curves \( \gamma_1 \) and \( \gamma_2 \) are equivalent if for all smooth functions \( f \)
\[ \frac{d^\lambda (f \circ \gamma_1)}{d t^\lambda}(0) = \frac{d^\lambda (f \circ \gamma_2)}{d t^\lambda}(0), \; \lambda = 0, 1, \ldots, r. \]
Coordinates \((x^i)\) on \(M\) induce coordinates \((x^i_\lambda)\) on \(\mathcal{T}^*M\), where \(\lambda = 0, 1, \ldots, r\). When considering jets up to order 2 one frequently uses the notation according to which coordinates \((x^i, \dot{x}^j, \ddot{x}^k)\) of \([\gamma]_2\) are defined by derivatives with respect to \(t\), namely
\[
x^i(\gamma(t)) = x^i + \dot{x}^i t + \frac{1}{2} \ddot{x}^i t^2 + o(t^2).
\]
Coordinates \((x^i_\lambda)\) are however defined using the following convention. For \([\gamma]_r\) the values of \(x^i_\lambda([\gamma]_r)\) are such that
\[
x^i(\gamma(t)) = x^i_0 + x^i_1 t + x^i_2 t^2 + \cdots + x^i_r t^r + o(t^r).
\]
This definition of coordinates is more convenient for the lifting procedures (see e.g. Proposition 5.3).

Let us fix \(r \in \mathbb{N}\) for the rest of this section.

**Definition 5.1** \((\text{35})\). Let \(f \in C^\infty(M)\) and \(\lambda\) be a non-negative integer not bigger than \(r\). Then, the \(\lambda\)-lift of \(f\) is the function \(L^\lambda f = f^{(\lambda)}\) on \(\mathcal{T}^*M\) defined by
\[
f^{(\lambda)}([\gamma]_r) = \frac{1}{\lambda!} \left[ \frac{d^\lambda (f \circ \gamma)}{dt^\lambda} \right]_{t=0},
\]
for \([\gamma]_r \in \mathcal{T}^*M\), where \(\gamma : \mathbb{R} \to M\) is a smooth curve. We put by convention \(f^{(\lambda)} = 0\) for \(\lambda < 0\).

**Example 5.2.** We have
\[
\begin{align*}
f^{(0)}(x_0, x_1, x_2) &= f(x_0); \\
f^{(1)}(x_0, x_1, x_2) &= \frac{\partial f}{\partial x^i}(x_0) x^i_1; \\
f^{(2)}(x_0, x_1, x_2) &= \frac{1}{2} \frac{\partial^2 f}{\partial x^i \partial x^j}(x_0) x^i_1 x^j_1 + \frac{\partial f}{\partial x^k}(x_0) x^k_2.
\end{align*}
\]

Lifting \(f \mapsto f^{(\lambda)}\) is linear and the following Leibniz rule
\[
(f \cdot g)^{(\lambda)} = \sum_{\mu=0}^{\lambda} f^{(\mu)} \cdot g^{(\lambda-\mu)}
\]
is satisfied for all \(f, g \in C^\infty(M)\). It is easy to see the following.

**Proposition 5.3.** For local coordinates \((x^i)\) on \(M\) we have \((x^i)^{(\lambda)} = x^i_\lambda\).

The \(\lambda\)-lifts of one-forms \(\omega \in \Omega^1(M)\) and vector fields \(X \in \mathfrak{X}(M)\) are defined as follows.

**Theorem 5.4.**
- There exists a unique \(\mathbb{R}\)-linear lift \(L_\lambda : \Omega^1(M) \to \Omega^1(\mathcal{T}^*M)\) such that
\[
L_\lambda(f \cdot dg) = (f \cdot dg)^{(\lambda)} := \sum_{\mu=0}^{\lambda} f^{(\mu)} dg^{(\lambda-\mu)}.
\]
In particular, \((dx^i)^{(\lambda)} = dx^i_\lambda\).
• There exists a unique \( \mathbb{R} \)-linear lift \( L_\lambda : \mathfrak{X}(M) \rightarrow \mathfrak{X}(\mathcal{T}^r M) \) such that for \( L_\lambda (X) = X^{(\lambda)} \) we have

\[
X^{(\lambda)} f(\mu) = (X f)^{(\lambda+\mu-r)}
\]

for any smooth function \( f \) on \( M \). In particular, \((\partial_{x^i} f)^{(\lambda)} = \partial_{x^{i-\lambda}} f\).

The lifts \( f^{(r)} \), \( \omega^{(r)} \), and \( X^{(r)} \) will be called complete lifts to \( \mathcal{T}^r M \) and denoted also \( f^{(c)} \), \( \omega^{(c)} \), and \( X^{(c)} \). For \( r = 1 \) they coincide with the complete lifts from the previous section, while \( 0 \)-lifts coincide with vertical lifts. If the vector field \( X \in \mathfrak{X}(M) \) induces a one-parameter group of transformations \( \psi_t \), then \( X^c \in \mathfrak{X}(\mathcal{T}^r M) \) induces the one-parameter group of transformations \( \mathcal{T}^r \psi_t \) on \( \mathcal{T}^r M \).

Example 5.5. For \( r = 2 \) and a vector field \( X = X_i(x) \frac{\partial}{\partial x^i} \) we have in coordinates \((x^i, x^j, x^k) \) on \( \mathcal{T}^2 M \)

\[
X^{(0)} = X^i(x_0) \frac{\partial}{\partial x^i},
\]

\[
X^{(1)} = X^i(x_0) \frac{\partial}{\partial x^i} + \frac{\partial X^k}{\partial x^j}(x_0) x^j \frac{\partial}{\partial x^k},
\]

\[
X^{(2)} = X^i(x_0) \frac{\partial}{\partial x^i} + \frac{\partial X^k}{\partial x^j}(x_0) x^j \frac{\partial}{\partial x^k} + \left( \frac{1}{2} \frac{\partial^2 X^l}{\partial x^m \partial x^n}(x_0) x^m x^n + \frac{\partial X^l}{\partial x^p}(x_0) x^p \right) \frac{\partial}{\partial x^l}.
\]

For a one-form \( \alpha = \alpha_i(x) \, dx^i \) we have in turn

\[
\alpha^{(0)} = \alpha_i(x_0) \, dx^i;
\]

\[
\alpha^{(1)} = \frac{\partial \alpha_i}{\partial x^j}(x_0) x^j \, dx^i + \alpha_k(x_0) \, dx^k;
\]

\[
\alpha^{(2)} = \left( \frac{1}{2} \frac{\partial \alpha_i}{\partial x^k \partial x^j}(x_0) x^k x^j + \frac{\partial \alpha_i}{\partial x^l}(x_0) x^l \right) \, dx^i + \frac{\partial \alpha_m}{\partial x^n}(x_0) x^n \, dx^m + \alpha_p(x_0) \, dx^p.
\]

Theorem 5.6. ([33]) If \( X, Y \) are vector fields on \( M \) and \( f \in C^\infty(M) \), then

\[
(f \cdot X)^{(\lambda)} = \sum_{\mu=0}^{\lambda} f^{(\mu)} X^{(\lambda-\mu)}.
\]

In particular,

\[
\left( \sum_i a_i \partial_{x^i} \right)^{(\lambda)} = \sum_i \sum_{\nu=r-\lambda}^r a_i^{(\nu+\lambda-r)} \partial_{x^i}.
\]

Moreover,

\[
[X^{(\lambda)}, Y^{(\mu)}] = [X, Y]^{(\lambda+\mu-r)}.
\]

Finally, we apply the generalized Leibniz rule for the lifts of tensor product:

\[
(T \otimes S)^{(\lambda)} = \sum_{\mu=0}^{\lambda} T^{(\mu)} \otimes S^{(\lambda-\mu)}
\]

to obtain the lifts of \( q \)-contravariant tensor fields

\[
(X_1 \otimes \ldots \otimes X_q)^{(\lambda)} = \sum_{\mu_1 + \ldots + \mu_q = \lambda} (X_1)^{\mu_1} \otimes \ldots \otimes (X_q)^{\mu_q}
\]
and $p$-covariant tensor fields

$$(\alpha_1 \otimes \ldots \otimes \alpha_p)^{(\lambda)} = \sum_{\mu_1 + \ldots + \mu_p = \lambda} (\alpha_1)^{(\mu_1)} \otimes \ldots \otimes (\alpha_p)^{(\mu_p)}.$$ 

The lifts of symmetric tensors are symmetric, the lifts of antisymmetric tensors are antisymmetric. Actually, we can obtain in this way the lifts of arbitrary tensor fields:

$$L_\lambda : \mathcal{T}_p^q(M) \to \mathcal{T}_p^q(T^r M),$$

so

$$L_\lambda : \mathcal{T}(M) \to \mathcal{T}(T^r M).$$

The lifts $K^{(c)}$ we will call complete lifts of $K$ and denote with $K^{(c)}$.

**Proposition 5.7.** ([35]) If $K \in \mathcal{T}_p^q(M)$, then

$$K^{(\lambda)}(X^{(\mu_1)}_1, \ldots, X^{(\mu_p)}_p) = (K(X_1, \ldots, X_p))^{(\lambda + \mu - r \cdot p)},$$

where $\mu = \sum_i \mu_i$, for any vector fields $X_1, \ldots, X_p$ on $M$. In particular,

$$K^{(c)}(X^c_1, \ldots, X^c_p) = (K(X_1, \ldots, X_p))^c.$$ 

Similarly to the case of the tangent bundle, there are higher lifts of affine connections.

**Theorem 5.8.** ([54, 41]) Let $\nabla$ be a connection on a manifold $M$. There is a unique affine connection $\nabla^{(c)}$ on $T^r M$ which satisfies the condition

$$\nabla^{(c)}_{X^{(\lambda)}} K^{(\mu)} = (\nabla_X K)^{(\lambda + \mu - r)}$$

for $\lambda, \mu = 0, \ldots, r$, any vector fields $X$ on $M$, and any tensor field $K$ on $M$. We have

$$\nabla^{(c)} K^{(\mu)} = (\nabla K)^{(\mu)}.$$ 

In particular,

$$\nabla^{(c)}_{X^{(c)}} K^{(c)} = (\nabla_X K)^c \quad \text{and} \quad \nabla^{(c)} K^{(c)} = (\nabla K)^c.$$ 

6 \ Lifts of statistical manifolds

Let $(M, g, T)$ be a statistical manifold. The obvious lift of this structure to $T^r M$ is $(T^r M, g^c, T^c)$. As $T$ is a symmetric (0,3) tensor, then $T^c$ is a symmetric (0,3)-tensor, so that $(T^r M, g^c, T^c)$ is a statistical manifold. We will call it the $r$-lift of the statistical manifold $(M, g, T)$. Note however, that even if $g$ is a Riemannian metric, the metric $g^c$ is only pseudo-Riemannian.

Let us check the other ways of defining a statistical manifold structure.

**Theorem 6.1.** For $(M, g)$ being a pseudo-Riemannian manifold, $\nabla$ being a connection on $M$, and $\nabla^*$ being the dual connection, let $g^c$, $\nabla^c$, $(\nabla^*)^c$ be the corresponding complete lifts to the bundle $T^r M$.

1. If $g$ is a pseudo-Riemannian metric on $M$, then $g^c$ is a pseudo-Riemannian metric on $T^r M$. 

14
2. If $\nabla$ is a connection on $M$ and $\text{Tor}^{\nabla}$, $R^{\nabla}$ are the torsion and the curvature of $\nabla$, then

$$\text{Tor}^{\nabla^c} = (\text{Tor}^{\nabla})^c \quad \text{and} \quad R^{\nabla^c} = (R^{\nabla})^c.$$  

In particular, $\nabla^c$ is torsion-free (flat) if $\nabla$ is torsion-free (flat).

3. The complete lift of the dual connection $\nabla^*$ is the connection dual to $\nabla^c$, $(\nabla^*)^c = (\nabla^c)^*$.

4. If $\nabla^g$ is the metric connection for $g$, then $(\nabla^g)^c$ is the metric connection for $g^c$, $\nabla^{g^c} = (\nabla^g)^c$.

5. If $\nabla$ is torsion-free and $\nabla^g = \frac{1}{2}(\nabla + \nabla^*)$, then $\nabla^c$ is torsion-free and

$$\nabla^{g^c} = \frac{1}{2}(\nabla^c + (\nabla^c)^*).$$

6. If $(\nabla, \nabla^*)$ is a dualistic structure on $M$, then $(\nabla^c, (\nabla^c)^*)$ is a dualistic structure on $\mathcal{T}^*M$.

7. If $\nabla$ is torsion-free and Codazzi coupled with $g$, then $\nabla^c$ is torsion-free and Codazzi coupled with $g^c$.

**Remark 6.2.** The above theorem expresses the functoriality of our construction.

**Lemma 6.3.** Let $M$ be a manifold of dimension $n$ and $\pi : \mathcal{T}^*M \to M$ be the canonical projection. If vector fields $X_1, \ldots, X_n$ generate the vector bundle $\mathcal{T}M$ in a neighbourhood of $x \in M$ and $\pi(p) = x$ for some $p \in \mathcal{T}^*M$, then $X_i^{(\lambda)}$, $i = 1, \ldots, n$, $\lambda = 0, \ldots, r$, generate the vector bundle $\mathcal{T}\mathcal{T}^*M$ in a neighbourhood of $p$ in $\mathcal{T}^*M$.

**Proof.** That the vector fields $X_1, \ldots, X_n$ generate $\mathcal{T}M$ around $x \in M$ is equivalent to the fact that $X_1(x), \ldots, X_n(x)$ span $\mathcal{T}_xM$. Hence, there is a system of $(x^1, \ldots, x^n)$ of local coordinates around $x$ such that $X_i(x) = \partial_{x^i}$. In a neighbourhood of $x$ we can write $X_i = f_i^j \partial_{x^j}$, $i, j = 1, \ldots, n$, where $f_i^j(x) = 1$ and $f_i^j(x) = 0$ for $i \neq j$. According to \[14\],

$$X_i^{(\lambda)} = \left( \sum_j f_i^j \partial_{x^j} \right)^{(\lambda)} = \sum_{\nu=-\lambda}^{r} (f_i^j)^{(\nu+\lambda-r)} \partial_{x^\nu}.$$  

Let us first put $\lambda = r$. Then we have

$$X_i^{(r)}(p) = \sum_j (f_i^j)^{(0)}(p) \partial_{x^j} + \text{(higher)} = \sum_j f_i^j(x) \partial_{x^j} + \text{(higher)} = \partial_{x^i} + \text{(higher)}.$$  

The text (higher) means a vector field which is a combination of $\partial_{x^\nu}$ with $\nu > 0$. Generally, for $\lambda < r$,

$$X_i^{(\lambda)}(p) = \sum_{\nu=-\lambda}^{r} (f_i^j)^{\nu+\lambda-r} \partial_{x^\nu} = \sum_j f_i^j(x) \partial_{x^j} + \text{(higher)} = \partial_{x^i} + \text{(higher)},$$  

where (higher) means a vector field which is a combination of $\partial_{x^\nu}$ with $\nu > r - \lambda$. Now it is clear that $X_i^{(\lambda)}(p)$, $i = 1, \ldots, n$, $\lambda = 0, \ldots, r$, span $\mathcal{T}_pM$.  

\[ \square \]
Proof of Theorem 6.1.

1. Using the diagonalization algorithm for symmetric matrices, we can find local linearly independent vector fields $X_1, \ldots, X_n$ on $M$, $n = \dim(M)$, such that $g(X_i, X_i) = \pm 1$. According to (15),

$$g^c(X_i^{(\lambda)}, X_i^{(r-\lambda)}) = (g(X_i, X_i))^{(0)} = \pm 1$$

for $i = 1, \ldots, n$. But $X_i^{(\lambda)}$, $i = 1, \ldots, n$, $\lambda = 0, \ldots, r$, locally generate $TT^r M$, so $g^c$ is non-degenerated.

2. Using (16) and (15) we get

$$\nabla_{\alpha x_i}^c (\partial^i_{\alpha x_i}) = \nabla_{\alpha x_i}^c (\partial^i_{\alpha x_i}) = (\nabla_{\alpha x_i} \partial_{\alpha x_i})^{(\lambda + \mu - r)} = (\nabla_{\alpha x_i} \partial_{\alpha x_i})^{(\lambda + \mu - r)},$$

so that

$$\text{Tor}^c \left( \partial_{\alpha x_i}, \partial_{\alpha x_i} \right) = \nabla_{\alpha x_i}^c (\partial_{\alpha x_i}) - \nabla_{\alpha x_i}^c (\partial_{\alpha x_i}) = (\nabla_{\alpha x_i} \partial_{\alpha x_i} - \nabla_{\alpha x_i} \partial_{\alpha x_i})^{(\lambda + \mu - r)}$$

$$= \left[ \text{Tor}^c (\partial_{\alpha x_i}, \partial_{\alpha x_i}) \right]^{(\lambda + \mu - r)} = (\text{Tor}^c (\partial_{\alpha x_i}, \partial_{\alpha x_i}))^{(\lambda + \mu - r)} = \left( \text{Tor}^c \left( \partial_{\alpha x_i}, \partial_{\alpha x_i} \right) \right)^{\lambda + \mu - r}.$$

Similarly,

$$R^c \left( \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i} \right) = \nabla_{\alpha x_i}^c (\partial_{\alpha x_i}) - \nabla_{\alpha x_i}^c (\partial_{\alpha x_i}) = (\nabla_{\alpha x_i} \partial_{\alpha x_i} - \nabla_{\alpha x_i} \partial_{\alpha x_i})^{(\lambda + \mu - r)}$$

$$= \left[ \text{Tor}^c (\partial_{\alpha x_i}, \partial_{\alpha x_i}), \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i} \right]^{(\lambda + \mu - r)} = (\text{Tor}^c (\partial_{\alpha x_i}, \partial_{\alpha x_i}))^{(\lambda + \mu - r)} = \left( \text{Tor}^c \left( \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i} \right) \right)^{\lambda + \mu - r}.$$

3. In view of (13) and (18) we have

$$g^c \left( (\nabla)^c_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i} \right) = g^c \left( (\nabla_{\alpha x_i} ^c \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i}) \right)^{(\lambda + \mu - r)}$$

$$= g \left( \nabla_{\alpha x_i}^c \partial_{\alpha x_i}, (\partial_{\alpha x_i}) \right)^{(\lambda + \mu - r)}$$

$$= \left( \nabla_{\alpha x_i} ^c \partial_{\alpha x_i}, (\partial_{\alpha x_i}) \right)^{(\lambda + \mu - r)} = \left( \nabla_{\alpha x_i} ^c \partial_{\alpha x_i}, (\partial_{\alpha x_i}) \right)^{(\lambda + \mu - r)}$$

$$= g^c \left( (\nabla)^c_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i}, \partial_{\alpha x_i} \right).$$

4. Metric connections are characterized as torsion-free and self-dual. From 2) and 3) it follows that these properties are respected when going to the complete lifts.

5. and 6. are trivial in view of previous points.

6. If $\nabla g$ is symmetric, then according to (17)

$$\nabla^c g^c = (\nabla g)^c$$

is symmetric.
Note that the complete \( r \)-lifts of a statistical manifold are independent on the choice of description: in terms of the \((0,3)\) tensor or in terms of the connection and its dual. Actually, \((\nabla^\alpha)^c = (\nabla^c)^\alpha\). Indeed, if

\[
g(\nabla^\alpha_X Y, Z) = g(\nabla^\sigma_X Y, Z) - \frac{\alpha}{2} T(X, Y, Z),
\]

then

\[
g^c \left( (\nabla^\alpha)_X^\lambda Y^{(\mu)}, Z^{(\nu)} \right) = g^c \left( (\nabla_X Y)^{(\lambda+\mu-r)}, Z^{(\nu)} \right) = [g(\nabla_X Y, Z)]^{(\lambda+\mu+r-2r)}
\]

Example 6.4. The tangent lift of the Fisher-Rao metric is (cf. (11))

\[
\left( \int_\Omega \left[ \frac{\partial^2 \log p(\xi, x)}{\partial x^i \partial x^j} \frac{\partial \log p(\xi, x)}{\partial x^i} + \frac{\partial \log p(\xi, x)}{\partial x^j} \frac{\partial^2 \log p(\xi, x)}{\partial x^k \partial x^j} \right] p(\xi, x) \, d\xi \right)^{1/2} \, d^i \, d^j
\]

\[
+ \left( \int_\Omega \left[ \frac{\partial \log p(\xi, x)}{\partial x^i} \frac{\partial \log p(\xi, x)}{\partial x^j} \frac{\partial \log p(\xi, x)}{\partial x^k} \right] p(\xi, x) \, d\xi \right)^{1/2} \, d^i \, d^j
\]

\[
+ \left( \int_\Omega \left[ \frac{\partial \log p(\xi, x)}{\partial x^i} \frac{\partial \log p(\xi, x)}{\partial x^j} \right] p(\xi, x) \, d\xi \right)^{1/2} \left( d^i \, d^j + d^i \, d^j \right).
\]

Example 6.5. The Gaussian manifold \( M \) is a manifold of probability densities

\[
f(\xi, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(\xi - \mu)^2}{2\sigma^2} \right)
\]

on a real line. It is the two dimensional manifold with global coordinates \((\mu, \sigma)\) with \(\sigma > 0\) equipped with the metrics

\[
g = \frac{1}{\sigma^2} (d\mu \otimes d\mu + 2d\sigma \otimes d\sigma)
\]

and skewness tensor

\[
T = \frac{2}{\sigma^3} (d\mu \otimes d\mu \otimes d\sigma + d\mu \otimes d\sigma \otimes d\mu + d\sigma \otimes d\mu \otimes d\mu) + \frac{8}{\sigma^3} d\sigma \otimes d\sigma \otimes d\sigma.
\]

The Levi-Civita connection associated to \( g \) reads (in Christoffel symbols)

\[
\Gamma^\mu_{\mu\sigma} = \Gamma^\sigma_{\sigma\mu} = \Gamma^\sigma_{\mu\sigma} = 0, \quad \Gamma^\mu_{\mu\sigma} = \Gamma^\mu_{\sigma\mu} = \Gamma^\sigma_{\mu\sigma} = -\frac{1}{\sigma}, \quad \Gamma^\sigma_{\mu\mu} = \frac{1}{2\sigma}.
\]

In more compact way we can describe the above connection by specifying vector fields \( \mathcal{M} \) and \( \mathcal{S} \) on \( TM \) being horizontal lifts of \( \partial_\mu \) and \( \partial_\sigma \) respectively. In coordinates \((\mu, \sigma, \mu, \sigma)\) they read

\[
\mathcal{M} = \partial_\mu + \frac{\sigma}{\sigma} \partial_\mu - \frac{\mu}{2\sigma} \partial_\sigma, \quad \mathcal{S} = \partial_\sigma + \frac{\mu}{\sigma} \partial_\mu + \frac{\sigma}{\sigma} \partial_\mu.
\]

The \( \alpha \)-connection described by connection coefficients reads

\[
\Gamma^\mu_{\mu\sigma} = \Gamma^\sigma_{\sigma\mu} = \Gamma^\sigma_{\mu\sigma} = 0, \quad \Gamma^\mu_{\mu\sigma} = \Gamma^\mu_{\sigma\mu} = -\frac{\alpha + 1}{\sigma}, \quad \Gamma^\sigma_{\mu\sigma} = -\frac{\alpha - 1}{\sigma}, \quad \Gamma^\sigma_{\mu\mu} = \frac{1 - \alpha}{2\sigma},
\]

17
while horizontal vector fields are
\[ \mathcal{M} = \partial_{\mu} + \frac{(1+\alpha)\dot{\sigma}}{\sigma} \partial_{\mu} - \frac{(1-\alpha)\dot{\mu}}{2\sigma} \partial_{\sigma}, \quad \mathcal{\bar{S}} = \partial_{\sigma} + \frac{(1+\alpha)\dot{\mu}}{\sigma} \partial_{\mu} + \frac{(1+2\alpha)\dot{\sigma}}{\sigma} \partial_{\sigma}. \]

The commutator \([\mathcal{M}, \mathcal{\bar{S}}]\) reads
\[ [\mathcal{M}, \mathcal{\bar{S}}] = \frac{(1-\alpha^2)}{2\sigma^2} (2\dot{\sigma} \partial_{\mu} - \dot{\mu} \partial_{\sigma}). \]

From the above we can see that \(\alpha\)-connections for \(\alpha = \pm 1\) are flat. One can therefore find two global coordinate systems on \(M\) such that Christoffel symbols of 1-connection and for \((-1)\)-connection respectively are all zero. For the 1-connection the appropriate coordinate system is the following
\[ x^1 = \frac{\mu}{\sigma^2}, \quad x^2 = -\frac{1}{2\sigma^2}. \]

The above coordinate system is referred to in the literature as natural coordinates on Gaussian manifold. For the \((-1)\)-connection the appropriate coordinate system is given by the first two moments of the Gaussian distribution, namely
\[ m^1 = \mu, \quad m^2 = \mu^2 + \sigma^2. \]

The tangent lift \(g^c\) is a pseudo-Riemannian metrics on \(\mathcal{T}M\) described in coordinates by (11). It can be defined geometrically as follows. Metric tensor \(g\) gives rise to the isomorphism
\[ G : \mathcal{T}M \longrightarrow \mathcal{T}^*M, \quad G(v) = g(v, \cdot). \]

The corresponding isomorphism \(G^c\) for \(g^c\) is a map \(G^c : \mathcal{T}\mathcal{T}M \rightarrow \mathcal{T}^*\mathcal{T}M\) defined as follows
\[ G^c = \alpha_M \circ \mathcal{T}G \circ \kappa_M, \]
where \(\kappa_M : \mathcal{T}\mathcal{T}M \rightarrow \mathcal{T}\mathcal{T}M\) is a canonical flip on \(\mathcal{T}\mathcal{T}M\) and \(\alpha_M : \mathcal{T}^*M \rightarrow \mathcal{T}^*\mathcal{T}M\) is its dual, called usually the Tulczyjew isomorphism. Translating \(G^c\) back to \(g^c\) we get in coordinates \((\mu, \sigma, \dot{\mu}, \dot{\sigma})\)
\[ g^c = \frac{1}{\sigma^4}[-2\dot{\sigma}d\mu \otimes d\mu - 4\dot{\sigma}d\sigma \otimes d\sigma + \sigma(d\dot{\mu} \otimes d\mu + d\mu \otimes d\dot{\mu}) + 2\sigma(d\dot{\sigma} \otimes d\sigma + d\sigma \otimes d\dot{\sigma})], \]
in matrix form
\[ \frac{1}{\sigma^2} \begin{bmatrix} -2\dot{\sigma} & 0 & \sigma & 0 \\ 0 & -4\dot{\sigma} & 0 & 2\sigma \\ \sigma & 0 & 0 & 0 \\ 0 & 2\sigma & 0 & 0 \end{bmatrix}. \]

It is easy to see that \(g^c\) has signature \((2, 2)\) as stated in section 4.

The components of the lifted skewness tensor \(T^c\) are either 0 or equal (by symmetric permutation of indices) to the following
\[ (T^c)_{\mu\mu\sigma} = \frac{-6\dot{\sigma}}{\sigma^4}, \quad (T^c)_{\sigma\sigma\sigma} = \frac{-24\dot{\sigma}}{\sigma^4}, \]
\[ (T^c)_{\mu\sigma\sigma} = (T^c)_{\mu\mu\sigma} = \frac{2}{\sigma^3}, \quad (T^c)_{\sigma\sigma\sigma} = \frac{8}{\sigma^3}. \]
Both Levi-Civita and ‘α’ connections can be lifted to $\mathbf{T}M$. Since the dimension of the manifold grows, so is the number of Christoffel symbols. It is convenient then to look at the horizontal distribution. The horizontal distribution of the lifted Levi-Civita connection is the distribution on $\mathbf{T}\mathbf{T}M$ obtained as the image by $\mathbf{T}_{\mathcal{K}M}$ of the subbundle of $\mathbf{T}(\mathbf{T}M)$ spanned by vertical and complete lifts of $\mathcal{M}$ and $\mathcal{S}$. Using coordinates $(\mu, \sigma, \hat{\mu}, \hat{\sigma})$ in $\mathbf{T}M$ and $(\mu, \sigma, \hat{\mu}, \hat{\sigma}, \delta\mu, \delta\sigma, \delta\hat{\mu}, \delta\hat{\sigma})$ in $\mathbf{T}\mathbf{T}M$ we can write horizontal lifts of $\partial_\mu$, $\partial_\sigma$, $\partial_{\hat{\mu}}$, $\partial_{\hat{\sigma}}$ with respect to the lifted Levi-Civita connection respectively as

\[
\begin{align*}
\mathcal{X}_\mu &= \partial_\mu - \frac{\delta\mu}{2\sigma} \partial_{\delta\sigma} + \frac{\delta\sigma}{\sigma} \partial_{\delta\mu} + \left(\frac{\delta\hat{\sigma}}{2\sigma^2} - \frac{\delta\hat{\mu}}{2\sigma}\right) \partial_{\delta\hat{\sigma}} + \left(-\frac{\delta\delta\sigma}{\sigma^2} + \frac{\delta\delta\hat{\sigma}}{2\sigma}\right) \partial_{\delta\hat{\mu}} \\
\mathcal{X}_\sigma &= \partial_\sigma + \frac{\delta\mu}{\sigma} \partial_{\delta\mu} + \frac{\delta\sigma}{\sigma} \partial_{\delta\sigma} + \left(-\frac{\delta\delta\hat{\sigma}}{\sigma^2} + \frac{\delta\delta\mu}{\sigma}\right) \partial_{\delta\hat{\mu}} + \left(-\frac{\delta\delta\hat{\sigma}}{\sigma^2} + \frac{\delta\delta\mu}{\sigma}\right) \partial_{\delta\hat{\sigma}} \\
\mathcal{X}_{\hat{\sigma}} &= \partial_{\hat{\sigma}} + \frac{\delta\mu}{\sigma} \partial_{\delta\mu} + \frac{\delta\sigma}{\sigma} \partial_{\delta\sigma} + \left(-\frac{\delta\delta\hat{\sigma}}{\sigma^2} + \frac{\delta\delta\mu}{\sigma}\right) \partial_{\delta\hat{\mu}} + \left(-\frac{\delta\delta\hat{\sigma}}{\sigma^2} + \frac{\delta\delta\mu}{\sigma}\right) \partial_{\delta\hat{\sigma}} \\
\mathcal{X}_{\hat{\mu}} &= \partial_{\hat{\mu}} - \frac{\delta\mu}{2\sigma} \partial_{\delta\sigma} + \frac{\delta\sigma}{\sigma} \partial_{\delta\mu} + \left(\frac{\delta\hat{\sigma}}{2\sigma^2} - \frac{\delta\hat{\mu}}{2\sigma}\right) \partial_{\delta\hat{\sigma}} + \left(-\frac{\delta\delta\sigma}{\sigma^2} + \frac{\delta\delta\hat{\sigma}}{2\sigma}\right) \partial_{\delta\hat{\mu}} \end{align*}
\]

From the above formulae we read the Christoffel symbols of the lifted Levi-Civita connection (symbols not listed below vanish):

\[
\begin{align*}
\Gamma^\mu_{\mu\sigma} &= \Gamma^\mu_{\sigma\mu} = \Gamma^\sigma_{\sigma\sigma} = -\frac{1}{\sigma}, \quad \Gamma^\sigma_{\mu\mu} = \frac{1}{2\sigma}, \\
\Gamma^\mu_{\mu\sigma} &= \Gamma^\mu_{\sigma\mu} = \frac{\hat{\sigma}}{\sigma^2}, \quad \Gamma^\mu_{\mu\sigma} = \Gamma^\mu_{\sigma\mu} = -\frac{1}{\sigma}, \quad \Gamma^\mu_{\mu\sigma} = \Gamma^\mu_{\sigma\mu} = -\frac{1}{\sigma}, \\
\Gamma^\sigma_{\mu\mu} &= \frac{1}{2\sigma}, \quad \Gamma^\sigma_{\sigma\sigma} = -\frac{1}{\sigma}, \\
\Gamma^\sigma_{\sigma\sigma} &= \frac{\hat{\sigma}}{\sigma^2}, \quad \Gamma^\sigma_{\mu\mu} = \frac{\hat{\sigma}}{\sigma^2}, \quad \Gamma^\sigma_{\mu\mu} = \frac{\hat{\sigma}}{\sigma^2}, \quad \Gamma^\sigma_{\sigma\sigma} = \Gamma^\sigma_{\sigma\sigma} = -\frac{1}{\sigma}.
\end{align*}
\]

If we are patient enough, we can do the same for $\alpha$-connection. We can also follow the rules of calculating Christoffel symbols for lifted connection that are the following: all Christoffel symbols with three dotted coordinates are zero; the symbols with one dotted up and one dotted down are equal to these with no dots, while the symbols with two dotted down and no dots up are zero; the symbols with one dotted coordinates are non-zero only when the dotted coordinate is up and then they are tangent lifts of non-dotted symbols; non-dotted symbols are the same as for the original connection on $M$. The table for Christoffel symbols for lifted $\alpha$-connection is then the following:

\[
\begin{align*}
\alpha^\mu_{\mu\sigma} &= \alpha^\mu_{\sigma\mu} = \alpha^\mu_{\mu\sigma} = \alpha^\mu_{\sigma\mu} = \alpha^\mu_{\mu\sigma} = \frac{\alpha}{\sigma} + 1 \\
\alpha^\sigma_{\mu\mu} &= \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = -\frac{2\alpha + 1}{\sigma} \\
\alpha^\sigma_{\mu\mu} &= \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = \alpha^\sigma_{\mu\mu} = -\frac{2\alpha + 1}{\sigma} \\
\alpha^\sigma_{\sigma\sigma} &= \alpha^\sigma_{\sigma\sigma} = \alpha^\sigma_{\sigma\sigma} = \alpha^\sigma_{\sigma\sigma} = \alpha^\sigma_{\sigma\sigma} = \frac{1 - \alpha}{2\sigma^2} \\
\alpha^\mu_{\mu\mu} &= \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \frac{(\alpha + 1)}{\sigma^2} \\
\alpha^\mu_{\mu\mu} &= \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \alpha^\mu_{\mu\mu} = \frac{(\alpha + 1)}{\sigma^2}.
\end{align*}
\]

**Example 6.6.** Let $x = (x^1, \ldots, x^n) \in \mathbb{R}^n = M$ and $\xi = (\xi^1, \ldots, \xi^n) \in \mathbb{R}^n$. The exponential family is a family of probability measures of the form

\[
p(\xi, x) = \exp (k(\xi) + (x | \xi) - \psi(x)) \, d^n\xi,
\]

19
where $k(\xi)$ is a normalizing factor such that $\int p(\xi, x) d^n \xi = 1$ and $\psi$ is some convex function on $\mathbb{R}^n$. The Fisher information matrix is calculated according to the usual formula

$$g_{ij}(x) = \int \left( \frac{\partial \log p(\xi, x)}{\partial x^i} \frac{\partial \log p(\xi, x)}{\partial x^j} \right) p(\xi, x) d^n \xi = \frac{\partial^2 \psi}{\partial x^j \partial x^i}(x).$$

We have used here some regularity assumptions, namely that we can differentiate under the integral sign. We have then

$$0 = \frac{\partial}{\partial x^i} \left( \int p(\xi, x) d^n \xi \right) = \int \frac{\partial \log p(\xi, x)}{\partial x^i} p(\xi, x) d^n \xi.$$

Differentiating once more we get that

$$0 = \frac{\partial}{\partial x^j} \left( \int \frac{\partial \log p(\xi, x)}{\partial x^i} \right) p(\xi, x) d^n \xi = \int \left( \frac{\partial^2 \log p(\xi, x)}{\partial x^i \partial x^j} \frac{\partial \log p(\xi, x)}{\partial x^i} \right) p(\xi, x) d^n \xi = -\frac{\partial^2 \psi}{\partial x^j \partial x^i}(x) + \int \left( \frac{\partial \log p(\xi, x)}{\partial x^i} \frac{\partial \log p(\xi, x)}{\partial x^j} \right) p(\xi, x) d^n \xi.$$

The Riemannian metric for the statistical model of exponential family in natural coordinates $(x^i)$ is given then is by the Hessian of the function $\psi$,

$$g(x) = \frac{\partial^2 \psi}{\partial x^j \partial x^i}(x) dx^i \otimes dx^j.$$

The complete lift of $g$ to $\mathcal{T}M$ reads then

$$g^c(x, \dot{x}) = \frac{\partial^3 \psi}{\partial x^k \partial x^i \partial x^j}(x) \dot{x}^k dx^i \otimes dx^j + \frac{\partial^2 \psi}{\partial x^j \partial x^i}(x) \left( d\dot{x}^i \otimes dx^j + dx^i \otimes d\dot{x}^j \right).$$

It is given by the Hessian of the complete lift $\psi^c$ of the function $\psi$,

$$\psi^c(x, \dot{x}) = \frac{\partial \psi}{\partial x^i}(x) \dot{x}^i.$$

The Levi-Civita connection for the exponential family is conveniently described by the Christoffel symbols $\Gamma_{ijk} = g_{kl} \Gamma^l_{ij}$. We have

$$\Gamma_{ijk} = \frac{1}{2} \frac{\partial^3 \psi}{\partial x^k \partial x^i \partial x^j}.$$

Let $\tilde{\Gamma}$ denote the Christoffel symbols of the Levi-Civita connection lifted to $\mathcal{T}M$. Using general rules we get

$$\tilde{\Gamma}_{jk}^i = \tilde{\Gamma}_{jk}^i = \tilde{\Gamma}_{jk}^i = \tilde{\Gamma}_{jk}^i = 0$$

$$\tilde{\Gamma}_{jk}^i = \tilde{\Gamma}_{jk}^i = \Gamma_{jk}^i, \quad \tilde{\Gamma}_{jk}^i = (\Gamma_{jk}^i)^{(1)}.$$

Taking into account the components of lifted metric we can calculate the Christoffel symbols of the lifted connection with lowered index. What we get is

$$\tilde{\Gamma}_{ijk} = (\Gamma_{ijk})^{(1)}, \quad \tilde{\Gamma}_{ijk} = \tilde{\Gamma}_{ijk} = \tilde{\Gamma}_{ijk} = \Gamma_{ijk},$$

20
while coefficients with two or three ‘lifted’ indices vanish. For the exponential family it means precisely that the coefficients of the lifted connection with lowered index are third order partial derivatives of the complete lift of $\psi$.

The skewness tensor $T$ is also given by the very simple formula

$$T_{ijk} = \frac{\partial^3 \psi}{\partial x^k \partial x^j \partial x^i}.$$  

It is easy to see from the above, that connection $\nabla$ in this case is flat and the natural coordinate system $(x^i)$ is an affine system for this connection. The tangent lift $T^c$ of $T$ is again given by the complete lift of the function $\psi$, namely

$$(T^c)_{ijk} = \frac{\partial^3 \psi^c}{\partial x^k \partial x^j \partial x^i} \quad \text{and} \quad (T^c)_{ijk} = \frac{\partial^3 \psi^c}{\partial x^i \partial x^j \partial x^k} = \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k}.$$  

Since $\psi^c$ is linear in $\dot{x}^i$ coordinates, the higher derivatives of $\psi^c$ with respect of these coordinates are 0, so the corresponding coefficients of $T^c$ vanish as well. Summarising, the geometry of exponential family in natural coordinates is defined by the function $\psi$: the metric is just the Hessian of $\psi$ while the Levi-Civita connection and skewness tensor are expressed by the third-order partial derivatives of $\psi$. The lifted geometry is associated to the complete lift of $\psi$. In adapted coordinates, the metric, the metric connection and skewness tensor are given by the same formulae but for the complete lift of $\psi$.

The above formulae for lifted metric, lifted connection and skewness were calculated ‘by hand’. We can however consider also higher lifts of the statistical model given by an exponential family. For any natural $r$, the lift to $T^r M$ of the metric tensor in coordinates reads

$$(g_{kl}(x) \, dx^k \otimes dx^l)^c = \sum_{\mu,\nu \geq 0} g^{(r-\mu-\nu)}_{kl}(x) \, dx^k_\mu \otimes dx^l_\nu,$$

which means that

$$g^{(r)}(\partial_{x^k_\mu}, \partial_{x^l_\nu}) = g^{(r-\mu-\nu)}_{kl}.$$  

We use the convention according to which $f^{(p)} = 0$ if $p < 0$. For $g$ given in privileged coordinates by the Hessian of a function $\psi$ we get

$$g^{(r)}(\partial_{x^k_\mu}, \partial_{x^l_\nu}) = \left( \frac{\partial^2 \psi}{\partial x^k \partial x^l} \right)^{(r-\mu-\nu)} = \frac{\partial^2 \psi^c}{\partial x^k_\mu \partial x^l_\nu},$$

which means that the lifted metric is given by the Hessian of the lifted function. The same goes for the Levi-Civita connection and the skewness tensor – both are defined by $\psi^c$ by the formulæ

$$\Gamma^c_{(i,\mu)(j,\nu)(k,\lambda)} = \frac{1}{2} \frac{\partial^3 \psi^c}{\partial x^k_\mu \partial x^l_\nu \partial x^k_\lambda}, \quad T^c_{(i,\mu)(j,\nu)(k,\lambda)} = \frac{\partial^3 \psi^c}{\partial x^l_\mu \partial x^k_\nu \partial x^k_\lambda}.$$

**Remark 6.7.** In [37] the authors study lifts of Riemannian metrics and connections to the tangent bundle. Their lifted metric is a variant of the so called Sasaki metric. The Sasaki metric is a canonical almost Hermitian metric on the tangent bundle of a Riemannian manifold (with an affine connection). It was originally discovered by Sasaki [39] and expanded on by Dombrowski [16]. The advantage of this approach is that the Sasaki metric is again Riemannian. But the Sasaki metric depends on the chosen connection (originally the metric connection is taken) and the whole picture is not functorial. If $(M, g, \nabla)$ is a statistical manifold, then the obtained structure consisting of the Sasaki metric and the lifted connection is generally not a statistical manifold, so that the lifts in [37] do not produce statistical manifolds.
Our next observation is that we can lift also statistical manifolds admitting torsion (SMATs).

**Theorem 6.8.** If \((M, g, \nabla)\) is SMAT then \((\mathcal{T}^* M, g^c, \nabla^c)\) is a SMAT.

**Proof.** Let \(X, Y, Z\) be arbitrary vector fields on \(M\). Since If \((M, g, \nabla)\) is a SMAT, we have
\[
(\nabla_X g)(Y, Z) - (\nabla_Y g)(X, Z) = -g(\text{Tor}^\nabla(X, Y), Z).
\]

Hence, (cf. [13] and [16])
\[
(\nabla^c_{X(\lambda)} g^c)(Y^{(\mu)}, Z^{(\nu)}) - (\nabla^c_{Y(\mu)} g^c)(X^{(\lambda)}, Z^{(\nu)}) = (\nabla_X g)^{\lambda(\mu)}(Y^{(\mu)}, Z^{(\nu)}) - (\nabla_Y g)^{\mu(\nu)}(X^{(\lambda)}, Z^{(\nu)})
\]
\[
= -g^c((\text{Tor}^\nabla(X, Y))^{(\lambda+\mu+\nu-2r)}, Z^{(\nu)}) = -g^c((\text{Tor}^\nabla(X, Y))^{(\lambda+\mu+r)}, Z^{(\nu)})
\]
\[
= -g^c((\text{Tor}^\nabla)^c(X^{(\lambda)}, Y^{(\mu)}), Z^{(\nu)}) = -g^c((\text{Tor}^\nabla)^c(X^{(\lambda)}, Y^{(\mu)}), Z^{(\nu)}),
\]
so that the pair \((g^c, \nabla^c)\) is torsion coupled.

\(\square\)

# 7 Lifts of contrast functions

The main observation of this section is the following.

**Theorem 7.1.** If \(F : M \times M \to \mathbb{R}\) is a contrast function on \(M\), then
\[
F^c : \mathcal{T}^* (M \times M) = \mathcal{T}^* M \times \mathcal{T}^* M \to \mathbb{R}
\]
is a contrast function on \(\mathcal{T}^* M\). Moreover, the metric \(g^{F^c}\), the connection \(\nabla^{F^c}\), its dual \((\nabla^{F^c})^*\), and the \((0,3)\)-tensor \(T^{F^c}\) induced by \(F^c\) are the \(r\)-lifts of \(g^F\), \(\nabla^F\), \((\nabla^F)^*\), and \(T^F\).

**Proof.** Let \((x^i)\) be local coordinates on \(M\) in a neighbourhood of \(p \in M\), \(x^i(p) = 0\). Let \((y^i)\) be the same coordinates on another copy of \(M\), so that \((x^i, y^i)\) are local coordinates in a neighbourhood of \((p, p) \in M \times M\). Let \((\bar{x}, \bar{y})\) be the adapted coordinates \(\bar{x}^i = (x^i_0, \ldots, x^i)\) and \(\bar{y}^j = (y^j_0, \ldots, y^j)\) on \(\mathcal{T}^* M \times \mathcal{T}^* M\). According to Theorem 3.15, \(F\) is a contrast function in a neighbourhood of \((p, p)\) if and only if
\[
F(x, y) = \frac{1}{2} (x^i - y^i)(x^j - y^j) h_{ij}(x, y),
\]
where \(h_{ij} = h_{ji}\) and \([h_{ij}(0, 0)]\) is an invertible matrix, i.e. \([h_{ij}(x, y)]\) is invertible in a neighbourhood of \((0, 0)\). We have then (cf. [12] and Proposition 5.3)
\[
F^c(\bar{x}, \bar{y}) = \frac{1}{2} \sum_{\lambda+\mu+\nu=r} (x^i_{\lambda} - y^i_{\lambda})(x^j_{\mu} - y^j_{\mu}) (h_{ij})^{(\nu)}(\bar{x}, \bar{y}).
\]
As in the proof of Theorem 3.15, this implies that the first jets of \(F^c\) vanish on the diagonal \(\Delta_{\mathcal{T}^* M}\). To finish the proof, it suffices to show that the matrix \([h_{i(\lambda), j(\mu)}(\bar{0}, \bar{0})]\) is invertible, where \(h_{i(\lambda), j(\mu)} = (h_{ij})^{(\nu)}(\bar{0}, \bar{0})\). This is equivalent to the fact that the matrix \([h_{i(\lambda), j(\mu)}(\bar{x}, \bar{y})]\) is invertible for \((\bar{x}, \bar{y})\) in a neighbourhood of \((\bar{0}, \bar{0})\). This in turn is equivalent to the fact that the tensor field
\[
H(\bar{x}) = h_{i(\lambda), j(\mu)}(\bar{x}, \bar{x}) dx^i_{\lambda} \otimes dx^j_{\mu},
\]

22
is a pseudo-Riemannian metric in a neighbourhood of \( p \in M \). But

\[
H(\tilde{x}) = \left( h_{ij}^{(r-\lambda-\mu)} \, dx^i \otimes dx^j \right)(\tilde{x}, \tilde{x}) = \left( h_{ij} \, dx^i \otimes dx^j \right)^c(\tilde{x}).
\]

Since \( F \) is a contrast function, the tensor \( h(x) = h_{ij}(x, x) \, dx^i \otimes dx^j \) is the pseudo-Riemannian metric \((g^F)^c\). This proves that \( F^c \) is a contrast function and \( g^{F^c} = (g^F)^c \).

It is obvious that \((F^c)^* = (F^*)^c\). The connection \( \nabla^{F^c} \) and the skewness tensor \( T^{F^c} \) are defined as

\[
g^{F^c} \left( \nabla^{F^c}_{X^{(\lambda)}} Y^{(\mu)}, Z^{(\nu)} \right)(\tilde{x}) = -F^c[X^{(\lambda)} \, Y^{(\mu)}|Z^{(\nu)}](\tilde{x}, \tilde{x}),
\]

\[
T^{F^c}(X^{(\lambda)}, Y^{(\mu)}, Z^{(\nu)})(\tilde{x}) = (F^c - (F^c)^*)[X^{(\lambda)}|Y^{(\mu)} \, Z^{(\nu)}](\tilde{x}, \tilde{x}).
\]

It is easy to see that \( i(X^{(\lambda)}) = (X)^{\lambda} \) for \( i = 1, 2 \). We have (cf. Theorem 5.4)

\[
-F^c[X^{(\lambda)} \, Y^{(\mu)}|Z^{(\nu)}](\tilde{x}, \tilde{x}) = (1X)^{(\lambda)} \, (1Y)^{(\mu)} \, (2Z)^{(\nu)} \, (F^c)(\tilde{x}, \tilde{x})
\]

\[
= (1X \, 1Y \, 2Z \, (F))^{(\lambda+\mu+\nu-2r)}(\tilde{x}, \tilde{x}) = (g^F(\nabla_X \, Y, \, Z))^{(\lambda+\mu+\nu-2r)}(\tilde{x})
\]

\[
= (g^F)^c((\nabla^F_X \, Y)^{(\lambda+\mu-r)}, Z^{(\nu)})(\tilde{x}) = g^{F^c}((\nabla^{F^c})_{X^{(\lambda)}} Y^{(\mu)}, Z^{(\nu)})(\tilde{x}).
\]

This proves \( (\nabla^F)^c = \nabla^{F^c} \). Similarly one can prove \( (\nabla^F)^* = (\nabla^{F^c})^* \). For the skewness tensor we have

\[
(F^c - (F^c)^*)[X^{(\lambda)}|Y^{(\mu)} \, Z^{(\nu)}](\tilde{x}, \tilde{x}) = (1X)^{(\lambda)} \, (2Y)^{(\mu)} \, (2Z)^{(\nu)} \, (F^c - (F^c)^*)(\tilde{x}, \tilde{x})
\]

\[
= (1X \, 2Y \, 2Z \, (F - F^c))^{(\lambda+\mu+\nu-2r)}(\tilde{x}, \tilde{x}) = [T^F(X, Y, Z)]^{(\lambda+\mu+\nu-2r)}(\tilde{x})
\]

\[
= (T^F)^c(X^{(\lambda)}, Y^{(\mu)}, Z^{(\nu)}),
\]

that shows \( T^{F^c} = (T^F)^c \).

\[\square\]

**Example 7.2.** The tangent lift of the Kullback-Leibler divergence (2) reads

\[
D_{KL}^{(1)}(x, y, \dot{x}, \dot{y}) = \int_{\Omega} \left( \frac{\partial \, \log p(\xi, x)}{\partial x^k} \left[ \log \frac{p(\xi, x)}{p(\xi, y)} + 1 \right] \cdot \dot{x}^k - \frac{\partial \, \log p(\xi, y)}{\partial y^k} \cdot \dot{y}^k \right) p(\xi, x) \, d\xi.
\]

It is the matter of direct calculations to see that the pseudo-Riemannian metric \( g^{D_{KL}^{(1)}} \) is the tangent lift of the Fisher-Rao metric described in Example 6.4.

## 8 Conclusions and outlook

We have shown that the introduced lifting procedures provide new examples of statistical manifolds and that the lifted metric tensors and the lifted connections may be derivable from the lifted contrast function. Our construction is functorial and shows that to achieve this result we have to admit metric tensors which are not positive definite. Although the meaning of ‘null-varieties’ of probability distributions at the moment is unclear to us, we hope to come back to this aspect very soon.

As one possible application of our construction in the quantum setting we mention the evolution of quantum states for open systems. It is well known that when we consider a quantum system coupled to an environment, the evolution of the coupled system defines projected trajectories on the space of states of the subsystem, which in general are not trajectories of a vector field on the space of the quantum states of the subsystem. Indeed,
this happens only under particular assumptions of a weak coupling, and when it happens
the trajectories are solutions of a vector field which generates a semi-group of completely
positive maps (the infinitesimal generator being a GKLS-vector field).
When the family of trajectories is not associated with a vector field, one speaks of non-
Markovian evolution. We believe that by considering differential equations of higher order
it may be possible to describe families of projected trajectories as solutions of first order
vector fields on the higher tangent bundles we have introduced. A detailed discussion of
the construction of a vector field on $TM$ out of trajectories on $M$ is to be found in [31].
More likely, this approach will not be able to give all systems with memory, but we may
be able to capture a relevant part of them. These aspects will appear elsewhere.

Acknowledgments

G. Marmo is a member of the Gruppo Nazionale di Fisica Matematica (INDAM), Italy.
GM acknowledges financial support from the Spanish Ministry of Economy and Competitiveness,
through the Severo Ochoa Program for Centers of Excellence in RD (SEV-2015/0554). M. Kuś acknowledges support of the Polish national Science Center via the OPUS grant 2017/27/B/ST2/02959.

References

[1] S.-i. Amari. Differential Geometry of Curved Exponential Families Curvature and
Information Loss. Ann. Statist. 10, 357–385, 1982.
[2] S.-i. Amari. Differential geometric methods in statistics. Lect. Notes in Statist. 28,
Springer, Heidelberg, 1985.
[3] S.-i. Amari et al. Differential Geometry in Statistical Inference. IMS Lecture Notes–
Monograph Series 10, Institute of Mathematical Statistics, Hayward, California,
1987.
[4] S.-i. Amari and H. Nagaoka. Methods of information geometry. American Math.
Soc., Providence, R. I., 2007.
[5] S.-i. Amari. Differential-geometrical methods in statistics. Springer,
Berlin/Heidelberg, 2012.
[6] S.-i. Amari. Information geometry and its applications. Springer Japan,
Berlin/Heidelberg, 2016.
[7] O. E. Barndorff-Nielsen, D. R. Cox and N. Reid. The Role of Differential Geometry
in Statistical Theory. International Statistical Review 54, 83–96, 1986.
[8] O. E. Barndorff-Nielsen and P. E. Jupp. Statistics, yokes and symplectic geometry.
Ann. Fac. Sci. Toulouse Math., 6 (3), 389–427, 1997.
[9] P. Blæsild. Yokes and tensors derived from yokes. Ann. Inst. Statist. Math., 43,
95–113, 1991.
[10] N. C. Combe, Y. I. Manin, M. Marcolli. Geometry of information: Classical and
quantum aspects. Theoretical Computer Science, online.
[11] O. Calin and C. Udriște. Geometric modeling in probability and statistics. Springer, Cham, 2014.
[12] N. N. Chentsov. Statistical decision rules and optimal inferences. *Transl. Math. Monogr.* **53**, 117–120, 1982.
[13] F. M. Ciaglia, F. Di Cosmo, D. Felice, S. Mancini, G. Marmo, and J. M. Pérez-Pardo. Hamilton-Jacobi approach to potential functions in information geometry. *Journal of Mathematical Physics*, **58**, 063506, 2017.
[14] F. M. Ciaglia, F. Di Cosmo, M. Laudato, G. Marmo, F. M. Mele, F. Ventriglia, and P. Vitale. A pedagogical intrinsic approach to relative entropies as potential functions of quantum metrics: The q–z family. *Ann. Phys.*, **395**, 238 – 274, 2018.
[15] F. M. Ciaglia, G. Marmo, and J. M. Pérez-Pardo. Generalized potential functions in differential geometry and information geometry. *Int. J. Geom. Methods Mod. Phys.* **16** (supp01), 1940002, 2019.
[16] P. Dombrowski. On the geometry of the tangent bundle. *J. Reine Angew. Math.* **210**, 73–88, 1962.
[17] S. Eguchi. A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Math. J.* **15**, 341–391, 1985.
[18] S. Eguchi. Geometry of minimum contrast. *Hiroshima Math. J.* **22**, 631–647, 1992.
[19] P. Facchi, R. Kulkarni, V. I. Man’ko, G. Marmo, E. C. G. Sudarshan, and F. Ventriglia. Classical and quantum Fisher information in the geometrical formulation of quantum mechanics. *Phys. Lett. A* **374**(48), 4801–4803, 2010.
[20] K. Grabowska, J. Grabowski, M. Kuś, and G. Marmo. Lie groupoids in information geometry. *J. Phys. A* **52**, 505202, 2019.
[21] K. Grabowska, J. Grabowski, M. Kuś, and G. Marmo. Information geometry on groupoids: the case of singular metrics. *Open Syst. Inf. Dyn.* **27**, 2050015, 2020.
[22] J. Grabowski and P. Urbański. Tangent lifts of Poisson and related structures. *J. Phys. A* **28** (1995), no. 23, 6743–6777.
[23] J. Grabowski and P. Urbański. Tangent and cotangent lifts and graded Lie algebras associated with Lie algebroids. *Ann. Global Anal. Geom.* **15** (1997), no. 5, 447–486.
[24] J. Grabowski and P. Urbański. Algebroids–general differential calculi on vector bundles. *J. Geom. Phys.* **31** (1999), no. 2-3, 111–141.
[25] M. Henmi and H. Matsuzoe. Geometry of pre-contrast functions and non-conservative estimating functions. in: *AIP Conference Proceedings*, vol. **1340**, AIP, 32–41, 2011.
[26] M. Henmi, and H. Matsuzoe. Statistical manifolds admitting torsion and partially flat spaces. in: *Geometric Structures of Information*, Springer, Cham, 37–50, 2019.
[27] G. Khan and J. Zhang. A Hall of Statistical Mirrors. *arXiv:2109.13809*.
[28] T. Kurose. Statistical Manifolds Admitting Torsion (in Japanese). Geometry and Something. Fukuoka University, 2007.
[29] S. L. Lauritzen. Statistical manifolds. *Diff. Geom. Stat. Inference* **10**, 163–216, 1987.

[30] Hồng Văn Lê. Statistical manifolds are statistical models. *J. Geom.* **84**, 83–93, 2006.

[31] G. Marmo, E. J. Saletan, A. Simoni, B. Vitale. Dynamical Systems: Differential Geometric Approach to Symmetry and Reduction. J. Wiley, Chichester, 1985.

[32] H. Matsuzoe. Geometry of statistical manifolds and its generalization. in: *Topics in contemporary differential geometry, complex analysis and mathematical physics*, 244–251, World Sci. Publ., Hackensack, NJ, 2007.

[33] H. Matsuzoe. Statistical manifolds and affine differential geometry. *Probabilistic approach to geometry*, 303–321, Adv. Stud. Pure Math., **57**, Math. Soc. Japan, Tokyo, 2010.

[34] T. Matumoto. Any statistical manifold has a contrast function: On the $C^3$-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J.* **23**, 327–332, 1993.

[35] A. Morimoto, Liftings of tensor fields and connections to tangent bundles of higher order, *Nagoya Math. J.* **40** (1970), 99–120.

[36] M. K. Murray and J. W. Rice. Differential geometry and statistics. *Monographs on Statistics and Applied Probability* **48**, Chapman & Hall, London, 1993.

[37] E. Peyghan, D. Seifipour, and A. M. Blaga. Geometry of lift metrics and lift connections on the tangent bundle. *arXiv:2112.07202*.

[38] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91, 1945.

[39] S. Sasaki. On the differential geometry of tangent bundles of Riemannian manifolds. *Tohoku Math J.* **10**, 338–354, 1958.

[40] J. A. Wheeler. Information, Physics, Quantum: the search for links. Proc. 3rd Int. Symp. Foundations of Quantum Mechanics, Tokyo, pp. 354–368, 1989.

[41] K. Yano and S. Ishihara. Horizontal lifts of tensor fields and connections to tangent bundles. *J. Math. Mech.* **16** (1967), 1015–1029.

[42] K. Yano and S. Ishihara. Tangent and Cotangent Bundles. Marcel Dekker, Inc., 1973.

[43] J. Zhang and G. Khan. Statistical mirror symmetry. *Diff. Geom. Appl.* **73**, 101678, 2020.

Katarzyna Grabowska
*Faculty of Physics, University of Warsaw,*
ul. Pasteur 5, 02-093 Warszawa, Poland
konieczn@fuw.edu.pl

Janusz Grabowski
*Institute of Mathematics, Polish Academy of Sciences*
ul. Śniadeckich 8, 00-656 Warszawa, Poland
jagrab@impan.pl

Marek Kuś
Center for Theoretical Physics, Polish Academy of Sciences,
Aleja Lotników 32/46, 02-668 Warszawa, Poland
marek.kus@cft.edu.pl

Giuseppe Marmo
Dipartimento di Fisica “Ettore Pancini”, Università “Federico II” di Napoli
and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli,
Complesso Universitario di Monte Sant’Angelo,
Via Cintia, I-80126 Napoli, Italy
marmo@na.infn.it