Differential Geometry and its Applications 73 (2020) 101678

Contents lists available at ScienceDirect

Differential Geometry and its Applications

www.elsevier.com/locate/difgeo

Statistical mirror symmetry

Jun Zhang *, Gabriel Khan

University of Michigan, United States of America

A R T I C L E I N F O

Article history:
Received 18 October 2019
Received in revised form 6 August 2020
Accepted 14 August 2020
Available online xxxx
Communicated by L. Schwachhöfer

A B S T R A C T

In this paper, we investigate a duality between Hermitian and almost Kähler structures on the tangent manifold $T\mathcal{M}$ induced by pairs of conjugate connections on its base, affine Riemannian manifold $\mathcal{M}$. In the context of information geometry, the classical theory of statistical manifold (which we call $\mathcal{S}$-*geometry*) prescribes a parametrized family of probability distributions with a Fisher-Rao metric g and, using the Amari-Chensov tensor, a family of dualistic, torsion-free connections $\nabla^{(\alpha)}$, known as α-connections on $\mathcal{M}$. Here we prescribe an alternative geometric framework (which we call $\mathcal{P}$-*geometry* or partially flat geometry) by treating such parametrization as affine coordinates with respect to a flat connection ∇, and considering its g-conjugate connection ∇^* which is curvature-free but generally carries torsion. Under $\mathcal{P}$-geometry, the triplet (g, ∇, ∇^*) on $\mathcal{M}$ leads to a pair of complex and almost Kähler structures on $T\mathcal{M}$, in "mirror correspondence" to each other. Such complex-to-symplectic correspondence is reminiscent of mirror symmetry in string theory. We discuss the statistical meaning of mirror correspondence in terms of reference duality and representation duality in (various generalizations of) contrast/divergence functions characterizing proximity of probability distributions within a parametric statistical model.

1. Introduction

Information geometry provides a geometric characterization of parametric families of probability distributions or statistical models [2,5,7]. Based on statistical arguments, the second- and third-order invariants of a statistical model can be used to construct a Riemannian metric g and a pair of g-conjugate connections on the manifold $\mathcal{M}$ of a parametric family of probability distributions. Here, two connections ∇ and ∇^* are said to be conjugate if

$$Z(g(X,Y)) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$$

* Corresponding author.
E-mail addresses: junz@umich.edu (J. Zhang), gabekhan@umich.edu (G. Khan).

https://doi.org/10.1016/j.difgeo.2020.101678
0926-2245/

for any vectors fields X, Y and Z. The data $(\mathcal{M}, g, \nabla, \nabla^*)$ is said to be a *statistical manifold* (assuming both ∇, ∇^* are torsion-free). Conjugate connections play a central role in information geometry, and the duality reflects a fundamental duality in statistical inference and information theory.

In the classical setting, statistical manifolds are generated by divergence (contrast) functions. The resulting conjugate connections ∇ and ∇^* are necessarily torsion-free. The fact that both ∇ and ∇^* are torsion-free implies that the pair (∇, g), or equivalently the (∇^*, g) pair, is Codazzi-coupled. So classical information geometry $(\mathcal{M}, g, \nabla, \nabla^*)$ (which we call $\mathcal{S}$-*geometry* below, which stands for *standard* geometry or *statistical* manifold), involves torsion-free connections ∇ and ∇^* which are Codazzi-coupled to g. These torsion-free connections, along with the Fisher-Rao metric, are "canonical" objects once the parametric family is specified — in that they are unique second- and third-order invariants for a parametric statistical model.

In recent work [66] by the current authors, a different perspective about the manifold $\mathcal{M}$ of a parametric statistical model $p(\cdot|x)$ is taken. We treat x as a global coordinate, and associate a flat connection ∇ to it. This alternative construction of the manifold $\mathcal{M}$ of a parametric statistical model leads to a "partially flat" geometry where both conjugate connections are dually curvature-free yet not dually flat. A partially flat structure [32] is a relaxation of a dually flat structure, by allowing one of the connections (say, ∇^*) to carry torsion. In the partially flat structure, neither ∇ nor ∇^* is Codazzi-coupled to g; rather one of them is torsion-coupled to g instead (see below).

Requiring the conjugate connections ∇, ∇^* to be curvature-free but not both torsion-free means that g in a partially flat geometry is no longer necessarily a Hessian metric. We continue to take the Fisher-Rao metric g as the Riemannian metric on $\mathcal{M}$, and stipulate that ∇, the g-conjugate of the flat connection ∇^*, is allowed to carry torsion. We call this the $\mathcal{P}$-*geometry* or *partially flat* geometry of a parametric statistical model. It is a special case of a statistical manifold admitting torsion (SMAT, [40,47]) that can be generated by "pre-contrast functions" [31].

In our work [65], the conjugate connections in the $\mathcal{P}$-geometry was constructed as being the adapted connections to a pair of biorthogonal frames with respect to g. If we take one of the adapted connections, say ∇, to be flat (both curvature- and torsion-free), with $\partial/\partial x$ as its affine coordinate basis, then its biorthogonal frame is, in general, a non-coordinate frame, and the other adapted connection ∇^* (i.e., adapted to the biorthogonal frame) carries torsion T^{∇^*}. The pair of connections ∇, ∇^* are called "pseudo-Weitzenböck connections", and have been studied at length in [65]. Our subsequent work [66] further describes a canonical "pre-contrast function" to the $\mathcal{P}$-geometry, following a discovery by Henmi and Matsuzoe [31].

The key difference between the $\mathcal{S}$-geometry and the $\mathcal{P}$-geometry is the assumed coupling between (∇, g). $\mathcal{S}$-geometry is associated with *Codazzi coupling* of (∇, g), defined as

$$(\nabla_Z g)(X, Y) = (\nabla_X g)(Z, Y),$$

whereas $\mathcal{P}$-geometry is associated with *torsion coupling* of (∇, g), defined as[1]

$$(\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y) = g(T^{\nabla}(X, Z), Y).$$

In the $\mathcal{P}$-geometry, the torsion vanishes (i.e. $T^{\nabla} = 0$) if and only if $\mathcal{M}$ is a Hessian manifold. On the other hand, in the $\mathcal{S}$-geometry, the curvature of ∇ vanishes if and only if $\mathcal{M}$ is a Hessian manifold. So both $\mathcal{S}$-geometry and $\mathcal{P}$-geometry model parametric families of probability distributions, one assuming torsion-free but generally non-zero curvature, and the other assuming curvature-free but generally non-zero torsion.

[1] For applications to parametric probability models $p(\cdot|x)$, it is natural to designate ∇ (instead of ∇^*) as the flat connection and associate the parameter x to its affine coordinate chart, because neither depends on the metric g. However, when defining torsion coupling, we let ∇^* be torsion-free and let ∇ carry torsion. It is just a notational difference whether we designate ∇ or ∇^* as being flat in a partially flat geometry $(\mathcal{M}, g, \nabla, \nabla^*)$; the readers should not be alarmed that we make such designation quite freely based on surrounding context for ease of exposition.

It is worth noting that there are no topological obstruction in constructing $\mathcal{S}$-geometry, in that any Riemannian manifold can be equipped with $\mathcal{S}$-geometry. However, $\mathcal{P}$-geometry requires the manifold of probability distributions to be an affine manifold, which is the case for any parameterized family with the parameters serving as affine coordinates of the manifold.

In this paper, we will show that the $\mathcal{P}$-geometry of parametric probability distributions exhibits a kind of "mirror symmetry" in the sense of the phrase from string theory. Specifically, given a $\mathcal{P}$-structure $(\mathcal{M}, g, \nabla, \nabla^*)$ where now ∇ is taken as flat[2] and ∇^* is curvature-free but not necessarily torsion-free, we *canonically* prescribe *two* geometric structures on the tangent manifold $T\mathcal{M}$, a Hermitian structure $T\mathcal{M} = \mathbb{M}$ based on the data $(\mathcal{M}, g, \nabla)$ and an almost Kähler structure $T\mathcal{M} = \mathbb{W}$ based on the data $(\mathcal{M}, g, \nabla^*)$. The construction is based on the canonical split of the space $T(T\mathcal{M})$ into vertical and horizontal sub-distributions based on the connection ∇ or ∇^* on the base manifold $\mathcal{M}$, along with a Sasaki lift of g on $\mathcal{M}$ to a Riemannian metric G on $T\mathcal{M}$.

The correspondence between $T\mathcal{M}$ as a Hermitian manifold $\mathbb{M}$ and $T\mathcal{M}$ as an almost Kähler manifold $\mathbb{W}$ is based on the conjugate relationship of ∇ and ∇^* on $(\mathcal{M}, g)$. This correspondence is analogous to T-duality for semi-flat Calabi-Yau manifolds, and so we say that $\mathbb{M}$ and $\mathbb{W}$ are in "mirror correspondence." Note that through the musical isomorphism between the tangent bundle $T\mathcal{M}$ and the cotangent bundle $T^*\mathcal{M}$, such mirror symmetry can also be said to exist between $T\mathcal{M}$ as a complex manifold and $T^*\mathcal{M}$ as an almost Kähler manifold.

Because this $\mathbb{M} \longleftrightarrow \mathbb{W}$ correspondence arises out of the context of parametric statistical models, we called this "Statistical Mirror Symmetry" (SMS). However, just as "statistical structure" is a well-defined and now well-accepted geometric concept (see below) independent of statistical context which motivated its initial introduction, SMS as defined above is a geometric notion independent of its application to probability families and statistics. As a special case, when ∇^* (the g-conjugate of the flat ∇) is torsion-free (i.e. $\mathcal{M}$ is Hessian), we have a mirror pair of Kähler structures $\mathbb{M}$ and $\mathbb{W}$ on $T\mathcal{M}$. We call these "Kähler mirror pairs." We will show that, in a special case, this is precisely "mirror symmetry without corrections" [43], as defined for semi-flat Calabi-Yau manifolds. However, statistical mirror symmetry can be defined on Kähler manifolds which are not Ricci-flat, which makes it somewhat different to the Calabi-Yau case. An example of such statistical mirror built upon 2-D hyperbolic space (with constant negative curvature) is given, which reveals interesting correspondences between some familiar objects of complex geometry.

The paper is organized as follows. In Section 2, we first provide mathematical background on a slew of geometric structures built upon the notion of g-conjugate connections. In particular, we review

(1) Codazzi coupling and statistical structure,
(2) torsion coupling and partially flat structures,
(3) dually flat and Hessian structures,
(4) biorthogonal frames and adapted connections, and
(5) Hermitian, almost Kähler and Kähler structures.

In Section 3, we revisit the decomposition of $T(T\mathcal{M})$ into horizontal and vertical sub-distributions, and the construction of the Sasaki lift metric. We highlight that $T\mathcal{M}$ has the canonical construction as an almost Hermitian manifold based on (g, ∇) on the base manifold $\mathcal{M}$. Then we show, in Theorem 12, that based on the partially flat model of $(\mathcal{M}, g, \nabla, \nabla^*)$ (i.e., $\mathcal{P}$-geometry), the above structure gives rise to a Hermitian structure on the one hand (when the flat connection ∇ is used) and an almost Kähler structure on the other hand (when ∇^* is used). This leads to two distinct models $\mathbb{M}$ and $\mathbb{W}$ of the tangent manifold $T\mathcal{M}$, which we call "statistical mirror symmetry" (Definition 13). Its link to mirror symmetry for semi-flat Calabi-Yau manifolds is also discussed (Section 3.4). In Section 4, we discuss this work in the context of

[2] See footnote 1.

parametric statistical models, by explicating their $\mathcal{S}$-geometry and $\mathcal{P}$-geometry, and the contrast functions and pre-contrast functions generating each geometry. We discuss how in the Kähler case, statistical mirror symmetry is a geometric consequence of the so-called *reference-representation biduality*, introduced by the first author [63]. In Section 4.4, we provide a simple example of statistical mirror based on the well-studied univariate normal distribution as a 2-D hyperbolic manifold, which demonstrates an interesting connection to number theory. We also address (Section 4.5) how to find canonical coordinates ("balanced metrics") for an arbitrary parametric statistical model, such that they generalize the parameterizations by natural and expectation parameters in exponential family. We close the paper by summarizing properties of statistical mirror pairs ($\mathbb{M}$ and $\mathbb{W}$ geometries) and address some issues of global topological and geometric obstructions within our framework.

2. Mathematical background

2.1. g-Conjugate connection

We start by considering a Riemannian manifold $(\mathcal{M}, g)$ supplied with an affine connection ∇, which need not be the Levi-Civita connection. We recall that the g-conjugate connection ∇^* is defined as the (unique) connection that jointly preserves g with ∇:

$$Z(g(X,Y)) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y),$$

where X, Y, Z are all vector fields on $\mathcal{M}$.

The curvature and torsion of the conjugate connections ∇ and ∇^* are related as follows:

(a) their curvature tensors R^∇, R^{∇^*} satisfy

$$g(R^\nabla(Z,W)X,Y) + g(R^{\nabla^*}(Z,W)Y,X) \equiv 0; \tag{1}$$

(b) their torsion tensors T^∇, T^{∇^*} satisfy

$$g(T^{\nabla^*}(Z,X) - T^\nabla(Z,X), Y) \equiv (\nabla_Z g)(X,Y) - (\nabla_X g)(Z,Y). \tag{2}$$

Here, the curvature R^∇ and torsion T^∇ of an affine connection ∇ are defined by, respectively,

$$R^\nabla(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z; \tag{3}$$

$$T^\nabla(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{4}$$

Both (1) and (2) are well-known facts in information geometry. For an explicit derivation, see [65].

A connection is called *flat* if it is both curvature- and torsion-free. For any flat connection ∇, we can find local coordinate charts $x : V \to \mathbb{R}^n$ ($V \subset \mathcal{M}$) for which the associated Christoffel symbols Γ of ∇ vanish. We will call such coordinates the *affine coordinates*; they are uniquely specified apart from an affine transformation: $Ax + b$ for constants $A \in GL(n), b \in \mathbb{R}^n$.

A manifold $\mathcal{M}$ is said to be *affine* if it admits a flat affine connection ∇. As an equivalent definition, one can require that the manifold admit an atlas of coordinate charts such that the transition maps are all affine.

A manifold $(\mathcal{M}, g, \nabla, \nabla^*)$ is called *dually flat* when it carries two flat connections ∇ and ∇^* that form a conjugate pair with respect to the metric g. It is well known (see Section 2.3 for details) that in this case g must take the form of a Hessian metric constructed from either ∇ or ∇^*.

A consequence of Equation (1) is that if ∇ is curvature-free, then so is ∇^*. However, the torsion of ∇^* may not be zero even if ∇ has zero torsion. Therefore, an affine manifold is, in general, not dually flat with respect to an arbitrary metric g.

To investigate the torsion of the conjugate connections, let us recall and introduce the following definitions.

Definition 1. The pair (∇, g) is said to be

(a) Codazzi-coupled if $(\nabla_Z g)(X,Y) = (\nabla_X g)(Z,Y)$;
(b) torsion-coupled if $(\nabla_Z g)(X,Y) - (\nabla_X g)(Z,Y) = g(T^\nabla(X,Z),Y)$.

Note that Codazzi coupling is a well-known concept in affine differential geometry and information geometry. What we call here *torsion coupling* was first introduced by Kurose [40] in defining the so-called "statistical manifold admitting torsion" (SMAT). A manifold $(\mathcal{M}, g, \nabla)$ is called a SMAT if and only if (∇, g) is torsion-coupled. We mention in passing that the role of torsion in (para-)complexified structures has recently been extensively investigated [29,67].

2.2. Statistical versus partially flat structures

The two kinds of coupling of (∇, g) lead to two different geometries, which can both be used to describe the same parametric statistical model. We review them below.

2.2.1. Statistical structure

Due to Equation (2), the following are equivalent:

(i). (∇, g) is Codazzi-coupled;
(ii). (∇^*, g) is Codazzi-coupled;
(iii). $T^\nabla = T^{\nabla^*}$.

Lauritzen [42] introduced the definition of a statistical manifold as: $(\mathcal{M}, g, \nabla, \nabla^*)$ where the pair of g-conjugated connections ∇ and ∇^* are both torsion-free. Kurose [39] rephrased the definition as being equivalent to a Riemannian manifold $(\mathcal{M}, g, \nabla)$ where ∇ is torsion-free and Codazzi-coupled to g. Their equivalence is because a torsion-free ∇ is Codazzi-coupled to g if and only if $T^{\nabla^*} = 0$.

Note that Kurose's definition makes use of Codazzi coupling. Using the notion of torsion coupling of (∇, g), we can give a third definition of statistical manifold.

Remark 2. If ∇ and ∇^* are both torsion-coupled to g, then (g, ∇, ∇^*) is a statistical manifold.

We further remark that Codazzi coupling of (∇, g) is "invariant" under g-conjugate transformation of connection, in the sense that (∇^*, g) is Codazzi-coupled if and only if (∇, g) is, whereas torsion coupling is not — torsion coupling "selects" out the other (non-torsion-coupled) conjugate connection to be torsion-free.

2.2.2. Partially-flat structure

A manifold $(\mathcal{M}, g, \nabla, \nabla^*)$ is said to be *partially flat* if it is a statistical manifold admitting torsion (SMAT) with curvature-free connections. The following are two equivalent definitions of partial-flatness:

(i). $(\mathcal{M}, g, \nabla, \nabla^*)$ where ∇^* is flat (i.e., with zero curvature and torsion).
(ii). $(\mathcal{M}, g, \nabla, \nabla^*)$ where ∇ is torsion-coupled to g and has zero curvature.

Their equivalence is based on the curvature identity (1) and torsion identity (2). The term "partially flat" originates in the work of Henmi and Matsuzoe [32].

Note that while "flatness" of a connection does not depend on any metric g, the term "partial flatness" is used in the context of a connection-metric pair (∇, g). The notion of "partial flatness" is related to, but should not be confused with "semi-flatness" used in the context of string theory – see discussions in Section 3.5.

2.3. Dually flat (Hessian) structures

Recall that the Hessian of a function Φ (with respect to a connection ∇) is a bilinear form, denoted $\nabla^2\Phi$ that provides the second derivative of that function

$$(\nabla^2\Phi)(X,Y) := (\nabla_X d\Phi)(Y) = X(d\Phi(Y)) - d\Phi(\nabla_X Y).$$

In local coordinates using ∂_{x^j} as a shorthand for $\partial/\partial x^j$, the Hessian operator takes the form[3]

$$\nabla^2\Phi(\partial_{x^i}, \partial_{x^j}) = \frac{\partial^2\Phi}{\partial x^i \partial x^j} - \Gamma^k_{ij}\frac{\partial\Phi}{\partial x^k}.$$

Torsion-freeness of ∇ is equivalent to $\Gamma^k_{ij} = \Gamma^k_{ji}$, which implies that the Hessian is symmetric. When ∇ is further curvature-free (and hence flat), with x its affine coordinates (so that the Christoffel symbols satisfy $\Gamma^k_{ij} = 0$), the Hessian operator is the following:

$$\nabla^2\Phi(\partial_{x^i}, \partial_{x^j}) = \frac{\partial^2\Phi}{\partial x^i \partial x^j}.$$

For a manifold $\mathcal{M}$ with a flat connection ∇, a *Hessian metric* is a Riemannian metric for which there exists a local neighborhood $V \subset \mathcal{M}$ around each point so that the metric is of the form

$$g = \nabla^2\Phi \tag{5}$$

for a function $\Phi : V \to \mathbb{R}$. This is equivalent to insisting that in the ∇-affine coordinates, the metric is given by the second derivative of a convex function Φ. Such a Φ need not extend globally, and we will discuss global obstructions in Section 5.1. Metrics of the form (5) are sometimes also called *affine Kähler* [45], due to the formal correspondence with Kähler geometry. However, we will not use this terminology.

In this paper, we will mainly consider the case where there is a global coordinate chart and a single convex potential, and will denote the triple $(\mathcal{M}, \nabla, \nabla^2\Phi)$ as a *Hessian structure*, and (∇, Φ) as a *Hessian pair*.

It turns out ([65], Proposition 3) that given a torsion-free connection ∇ and a smooth function Φ on a manifold, then $(\mathcal{M}, \nabla, \nabla^2\Phi)$ is Codazzi-coupled iff $d\Phi(R^\nabla) = 0$. This implies the well known fact that any flat connection ∇ is always Codazzi-coupled to $\nabla^2\Phi$.

Denoting ∇^* to be the conjugate connection with respect to the Hessian metric $\nabla^2\Phi$, then ∇^* must also be flat (both curvature- and torsion-free), as a consequence of the above-referenced Proposition. So we denote a Hessian manifold as $(\mathcal{M}, g, \nabla, \nabla^*)$ where

$$g = \nabla^2\Phi = (\nabla^*)^2\Phi^*,$$

where Φ^* is the Legendre dual of Φ, defined as

[3] Here, and in the rest of the paper, Einstein summation notation is in effect.

$$\Phi^*(u) = \sup_{x \in V} \langle u, x \rangle - \Phi(x).$$

For any strictly convex function Ψ, its Legendre dual Ψ^* is also convex, and $(\Psi^*)^* = \Psi$.

2.4. The non-rigidity of Hessian structures

Given a Riemannian manifold $(\mathcal{M}, g)$, it may or may not admit any Hessian structure at all (see 5.1 for more details). However, even if one structure exists on a Riemannian manifold, it is generally not unique.

Given two distinct Hessian pairs (∇_1, Φ_1) and (∇_2, Φ_2) on a Riemannian manifold $(\mathcal{M}, g)$ that satisfy

$$\nabla_1^2 \Phi_1 = \nabla_2^2 \Phi_2, \tag{6}$$

it is not necessarily the case that $\Phi_2 = \Phi_1^*$. In particular, Kito [38] established that there are examples of metric g whose class of Hessian pairs (∇, Φ) satisfying $\nabla^2 \Phi = g$ (i.e., the equivalent class of (∇, Φ) giving rise to the same Hessian metric g) may have infinitely many solutions.

Stated differently, Equation (6) does not by itself guarantee that ∇_1, ∇_2 are conjugate connections with respect to the metric $\nabla_1^2 \Phi_1$ (equivalently $\nabla_2^2 \Phi_2$), nor that Φ_1 and Φ_2 are Legendre duals. As such, two arbitrary Hessian pairs (∇_1, Φ_1) and (∇_2, Φ_2) giving rise to the same Hessian metric $g = \nabla_1^2 \Phi_1 = \nabla_2^2 \Phi_2$ do not form a dually flat pair in general. However, it is the case that $\nabla_2 = \nabla_1^*$ whenever $\Phi_1^* = \Phi_2$.

2.5. $\mathcal{S}$- and $\mathcal{P}$-geometry both as relaxations of Hessian geometry

For a Riemannian manifold $(\mathcal{M}, g, \nabla)$ prescribed with a flat connection ∇ to become a Hessian manifold with *dually flat* connections, ∇ must be Codazzi-coupled to g. So for (∇, g) to deviate from a dually flat structure, we either (a) allow ∇ to carry curvature; or (b) no longer require ∇ to be Codazzi-coupled to g. In Section 4, we will show that:

(a) this case corresponds to the standard $\mathcal{S}$-*geometry* of a parametric statistical manifold with torsion-free connections;
(b) this case leads to a $\mathcal{P}$-*geometry* with a partially flat structure where ∇ is flat but ∇^* carries torsion — in fact (∇^*, g) will be torsion-coupled.

In both cases above, the Riemannian metric g will generally *not* be of the form $\nabla^2 \Phi$, for otherwise a dually flat (and hence g-Codazzi-coupled) structure emerges. For how Codazzi coupling in information geometry interacts with other geometric structures, see recent works [23,29,57].

2.6. Frames and adapted connections

As a generalization of biorthogonal coordinates in Hessian manifolds, [65] showed that on a Riemannian manifold, if we instead consider g-biorthogonal frames, this notion is compatible with g-conjugate connections.

Let us start by defining a local frame on a smooth manifold $\mathcal{M}$, with $n = \dim(\mathcal{M})$. A frame $\mathfrak{B} = \{\mathtt{b}_1, \cdots, \mathtt{b}_n\}$ is a collection of n local linearly independent vector fields $\{\mathtt{b}_i\}_{i=1}^n$ on $\mathcal{M}$. Given a coordinate chart $x = [x^1, \cdots, x^n]$, we can write the frame $\mathfrak{B} = \{\mathtt{b}_i\}_{i=1}^n$ using this coordinate system as $\mathtt{b}_i = B_i^j \partial_{x^j}$, where ∂_{x^j} is the shorthand for $\partial / \partial x^j$, and B_i^j is an $n \times n$ matrix (which is of full rank). Because B is full-rank, it is invertible. Below, B^{-1} denotes the matrix inverse of B_j^i:

$$(B^{-1})_l^i \, B_j^l = \delta_j^i = B_l^i \, (B^{-1})_j^l.$$

We say that the frame $\{\mathtt{b}_i\}_{i=1}^n$ is a coordinate frame when the B-matrix is taken to be the Jacobian matrix of a coordinate transform: $x \longrightarrow y$. In other words, a coordinate frame satisfies

$$(B^{-1})_j^\alpha = \frac{\partial y^\alpha}{\partial x^j} \longleftrightarrow B_\alpha^j = \frac{\partial x^j}{\partial y^\alpha}, \tag{7}$$

in which case we can simplify the frame as follows:

$$\mathtt{b}_i = \frac{\partial x^\alpha}{\partial y^i}\frac{\partial}{\partial x^\alpha} = \frac{\partial}{\partial y^i} = \partial_{y^i}.$$

In order for a frame to be a coordinate frame (locally), it is both necessary and sufficient for

$$\partial_{x^i}(B^{-1})_j^\alpha = \partial_{x^j}(B^{-1})_i^\alpha. \tag{8}$$

Definition 3 *(Adapted connection).* Given any frame $\mathfrak{B}$, the adapted connection $\nabla^{\mathfrak{B}}$ is defined by $\nabla^{\mathfrak{B}} = B\,\partial(B^{-1})$ or in component forms:

$$\Gamma_{k\alpha}^\beta = B_j^\beta(\partial_{x^\alpha}(B^{-1})_k^j) = -(B^{-1})_k^j(\partial_{x^\alpha}B_j^\beta). \tag{9}$$

We call $\nabla^{\mathfrak{B}}$ as given by Equation (9) the "pseudo-Weitzenböck" connection adapted to the frame $\mathfrak{B}$. It has the property that

(a) $\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j \equiv 0, \quad \forall i, j;$
(b) $R^{\nabla^{\mathfrak{B}}}(\mathtt{b}_i, \mathtt{b}_j)\mathtt{b}_k \equiv 0, \quad \forall i, j, k;$
(c) $T^{\nabla^{\mathfrak{B}}}(\mathtt{b}_i, \mathtt{b}_j) = -[\mathtt{b}_i, \mathtt{b}_j]$, so that $T^{\nabla^{\mathfrak{B}}} = 0$ iff $[\mathtt{b}_i, \mathtt{b}_j] = 0$, i.e., iff $\mathfrak{B}$ is a coordinate frame.

Definition 4 *(g-Biorthogonal frame).* Given any frame $\mathfrak{B} = \{\mathtt{b}_i\}_{i=1}^n$, the g-biorthogonal frame is defined as the (unique) frame $\mathfrak{B}^\star = \{\mathtt{b}_i^\star\}_{i=1}^n$ that is biorthogonal with respect to the given g:

$$g(\mathtt{b}_i, \mathtt{b}_j^\star) \equiv \delta_{ij}.$$

We have the following property.

Proposition 5 *([65], Theorem 10)). With respect to any Riemannian metric g, the g-conjugate of a connection adapted to a frame $\mathfrak{B}$ is precisely the connection adapted to the g-biorthogonal frame $\mathfrak{B}^\star$:*

$$\left(\nabla^{\mathfrak{B}}\right)^* = \nabla^{(\mathfrak{B}^\star)}.$$

Historically, an affine connection adapted to an orthonormal frame is called the *Weitzenböck connection*, and has been used in theoretical physics. For an application to information geometry, we here extend the construction of an adapted connection to an arbitrary frame, and hence the terminology "pseudo-Weitzenböck connections." Proposition 5 shows that the notion of biorthogonal frames is compatible with the notion of conjugate connections when the pair of connections are both adapted connections.

2.7. Almost complex structure, Hermitian and Kähler manifolds

Since statistical mirror symmetry involves a correspondence between Hermitian and almost Kähler manifolds, we first recall some background in complex geometry.

On a smooth manifold $\mathcal{N}$, an *almost complex structure* is an endomorphism J of the tangent bundle $T\mathcal{N}$ satisfying $J^2 = -id$. In this case, the pair $(\mathcal{N}, J)$ is said to be an almost complex manifold. It is well-known that any almost complex manifold must be even-dimensional and orientable (although these are not sufficient conditions for a manifold to admit such a structure).

Furthermore, a Riemannian manifold $(\mathcal{N}, G)$ with some prescribed almost complex structure J is said to be almost Hermitian if

$$G(J\xi^1, J\xi^2) = G(\xi^1, \xi^2) \tag{10}$$

for any vector fields ξ^1 and ξ^2 on $\mathcal{N}$. For clarity, we will denote Hermitian metrics by G so as to distinguish them from a generic Riemannian metric g. This will ease the notation when we discuss statistical mirror symmetry.

On an almost Hermitian manifold $\mathcal{N}$, the (skew-symmetric) 2-form

$$\Omega(\xi^1, \xi^2) = G(J\xi^1, \xi^2)$$

is called the *fundamental form* and satisfies $\Omega(\xi^1, J\xi^2) = \Omega(J\xi^1, \xi^2)$. The three structures (G, J, Ω) on $\mathcal{N}$ form a "compatible triple" such that given any two, the third one is uniquely specified.

A manifold is said to be *complex* if there exists an atlas of holomorphic coordinate charts. For a manifold with an arbitrarily given almost complex structure J, it can be very difficult to find holomorphic coordinates. However, there is a famous result of Newlander and Nirenberg [51], which shows that an almost complex manifold becomes a complex manifold if J is integrable, which is equivalent to the vanishing of the so-called Nijenhuis tensor N_J defined as

$$N_J(\xi^1, \xi^2) = [J\xi^1, \xi^2] - J[J\xi^1, \xi^2] - J[\xi^1, J\xi^2] - [\xi^1, \xi^2].$$

Definition 6. An almost Hermitian manifold $(\mathcal{N}, G, J)$ with $\Omega(\cdot, \cdot) = G(J(\cdot), \cdot)$ is called a(n)

(a) Hermitian manifold if J is integrable: $N_J = 0$.
(b) almost Kähler manifold if Ω is d-closed: $d\Omega = 0$.
(c) Kähler manifold if both $N_J = 0$ and $d\Omega = 0$.

An important and relevant fact is that, given a manifold $(\mathcal{M}, g, \nabla)$, there is a canonical process to construct, on its tangent manifold $T\mathcal{M} \equiv \mathcal{N}$, an almost Hermitian structure $(\mathcal{N}, G, J)$ by splitting the total space $T_{(x,v)}(T\mathcal{M})$ at any point $(x, v) \in T\mathcal{M}$ into a horizontal and a vertical subspace. This induces the Sasaki lift metric G (determined by both ∇ and g) and an almost complex operator J (determined by ∇) that swaps horizontal and vertical sub-distributions on $T\mathcal{M}$. This is the context under which we introduce statistical mirror symmetry, to be elaborated in the next section.

3. Canonical construction of statistical mirror on $T\mathcal{M}$

3.1. Canonical almost Hermitian structure on $T\mathcal{M}$ induced from $(\mathcal{M}, g, \nabla)$

Given the data (g, ∇) on the base manifold $\mathcal{M}$, there is a canonical construction[4] of an almost Hermitian structure on $T\mathcal{M}$. Below, we review the construction of $(T\mathcal{M}, G, J)$. We first review the notions of the horizontal lift, vertical lift, and the canonical almost complex structure involved.

[4] There appears to be a long history towards the understanding of this construction, starting from [19] showing that $T\mathcal{M}$ is Hermitian when a flat connection ∇ on $\mathcal{M}$ is used, to [56] showing that $T\mathcal{M}$ is almost Kähler when the Levi-Civita connection on $\mathcal{M}$ is used, and finally to [52] illuminating the role of torsion of the conjugate connection in the d-closedness of the fundamental form (see Proposition 9 below).

Suppose $x = [x^1, \cdots, x^n]$ is a local coordinate chart on $\mathcal{M}$, and we write $x \in \mathcal{M}$ for convenience. To specify an element $(x, v) = [x^1, \cdots, x^n, v^1, \cdots, v^n]$ of the tangent bundle $T\mathcal{M}$, denoted $(x, v) \in T\mathcal{M}$, we use $v = [v^1, \ldots, v^n]$ under the basis $[\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n}]$ of the tangent space $T_x\mathcal{M}$ to represent the fiber component $v = v^i \frac{\partial}{\partial x^i}$ of $T\mathcal{M}$ while the base component of $T\mathcal{M}$ is represented by x. The collection $\{x^i, v^j\}_{i,j=1,\cdots,n}$ form the set of coordinate functions for $T\mathcal{M}$, abbreviated as (x, v), with $[\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial v^1}, \cdots, \frac{\partial}{\partial v^n}] = \{\partial_{x^i}, \partial_{v^j}\}_{i,j=1,\cdots,n}$ as a coordinate frame for the tangent space $T_{(x,v)}(T\mathcal{M})$, and we will construct the Sasaki lift metric in terms of these coordinates.

Let ∇ be an affine connection on $\mathcal{M}$. Given a tangent vector X on $\mathcal{M}$ at point x, that is, $X \in T_x\mathcal{M}$, we can construct two tangent vectors on the tangent manifold $T\mathcal{M}$ at the point $(x, v) \in T\mathcal{M}$, $X^H \in T_{(x,v)}(T\mathcal{M})$ and $X^V \in T_{(x,v)}(T\mathcal{M})$. They are called the *horizontal* and *vertical* lifts of a vector on $\mathcal{M}$ to yield a vector $X^H \in T_{(x,v)}(T\mathcal{M}), X^V \in T_{(x,v)}(T\mathcal{M})$. Note we say a "vector on $\mathcal{M}$" to mean that it is a tangent vector of $\mathcal{M}$ and hence an element of $T_x\mathcal{M}$, and say a "vector on $T\mathcal{M}$" to mean it is a tangent vector of $T\mathcal{M}$ and hence an element of $T_{(x,v)}(T\mathcal{M})$.

To elaborate, a tangent vector X in $T_x\mathcal{M}$, or in coordinate form $X = X^i\partial_{x^i} \in T_x\mathcal{M}$, can be used to define the following two tangent vectors in $T_{(x,v)}(T\mathcal{M})$ at a point $(x, v) \in T\mathcal{M}$:

(a) the *vertical lift* vector $X^V \in \mathcal{V}_{(x,v)}(T\mathcal{M}) \subset T_{(x,v)}(T\mathcal{M})$, where the operator V is defined by

$$X^V = X^i \frac{\partial}{\partial v^i}, \quad \text{vertical lift of } X; \tag{11}$$

(b) the *horizontal lift* vector $X^H \in \mathcal{H}_{(x,v)}(T\mathcal{M}) \subset T_{(x,v)}(T\mathcal{M})$, where the operator H is defined by

$$X^H = X^i \frac{\partial}{\partial x^i} - X^i v^j \Gamma^k_{ij} \frac{\partial}{\partial v^k}, \quad \text{horizontal lift of } X. \tag{12}$$

Here Γ^i_{jk}'s are connection coefficients (Christoffel symbols) of ∇ with respect to x-coordinates, $\nabla_{\partial_{x^i}}\partial_{x^j} = \Gamma^k_{ij}\partial_{x^k}$, and is used in defining horizontal lift. Note also that the horizontal lift has both components (in the base $\frac{\partial}{\partial x^i}$ and fiber $\frac{\partial}{\partial v^i}$ direction), and it depends both on the affine connection Γ (defined on the base $\mathcal{M}$) and on the point (x, v) on $T\mathcal{M}$, whereas vertical lift results only in the fiber direction and is independent of the fiber-component v of the point. Intuitively, the vertical lift switches the vector from the tangent direction $\frac{\partial}{\partial x^i}$ to the "vertical" fiber direction $\frac{\partial}{\partial v^i}$:

$$X \equiv X^i \frac{\partial}{\partial x^i} \to X^i \frac{\partial}{\partial v^i} \equiv X^V.$$

These lifts define a splitting of the tangent space $T_{(x,v)}(T\mathcal{M})$ of $T\mathcal{M}$ at the specified point (x, v):

$$T_{(x,v)}(T\mathcal{M}) = \mathcal{H}_{(x,v)}(T\mathcal{M}) \oplus \mathcal{V}_{(x,v)}(T\mathcal{M}).$$

If we vary the base point and tangent vector, (x, v), the union of horizontal subspaces $\mathcal{H}_{(x,v)}(T\mathcal{M})$ and vertical subspaces $\mathcal{V}_{(x,v)}(T\mathcal{M})$ effectively leads to a splitting of the $T(T\mathcal{M})$:

$$T(T\mathcal{M}) = \mathcal{H}(T\mathcal{M}) \oplus \mathcal{V}(T\mathcal{M}),$$

where

$$\mathcal{H}(T\mathcal{M}) = \bigcup_{(x,v)\in T\mathcal{M}} \mathcal{H}_{(x,v)}(T\mathcal{M}), \qquad \mathcal{H}(T\mathcal{M}) = \bigcup_{(x,v)\in T\mathcal{M}} \mathcal{H}_{(x,v)}(T\mathcal{M}).$$

We also define a linear map J of $T_{(x,v)}(T\mathcal{M})$ by

$$J(X^H) = X^V, \qquad J(X^V) = -X^H,$$

such that $J^2 = -id$. Writing out explicitly in local coordinate basis $(\partial_{x^i}, \partial_{v^i})$, the J-operator has the action:

$$J\left(\frac{\partial}{\partial v^i}\right) = -\frac{\partial}{\partial x^i} + \Gamma^k_{ij} v^j \frac{\partial}{\partial v^k},$$
$$J\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial}{\partial v^i} - \Gamma^k_{ij}\, v^j \frac{\partial}{\partial x^k} + \Gamma^l_{ij}\, \Gamma^k_{lm}\, v^j\, v^m \frac{\partial}{\partial v^k}.$$

Adopting the base $\{\partial_{x^i}, \partial_{v^j}\}_{i,j=1,\cdots,n}$ of the tangent space of $T\mathcal{M}$, i.e., a basis frame of $T_{(x,v)}(T\mathcal{M})$, we write J in block-matrix form

$$\mathbf{J} = \begin{bmatrix} -\Gamma^k_{ij}\, v^j & \delta^k_i + \Gamma^l_{ij}\, \Gamma^k_{lm}\, v^j\, v^m \\ -\delta^k_i & \Gamma^k_{ij}\, v^j \end{bmatrix}. \tag{13}$$

J is called the *canonical almost complex structure* on $T\mathcal{M}$ (mapping $T_{(x,v)}(T\mathcal{M})$ to itself). Note that we can introduce $N^k_i = \Gamma^k_{ij} v^j$ and write the above as

$$\mathbf{J} = \begin{bmatrix} -N^k_i & \delta^k_i + N^l_i N^k_l \\ -\delta^k_i & N^k_i \end{bmatrix}. \tag{14}$$

A metric G on $T\mathcal{M}$ is said to be a Sasaki lift of g on $\mathcal{M}$ if $G = g \oplus g$. Working in local coordinates,

$$G\left(\left(\frac{\partial}{\partial x^i}\right)^V, \left(\frac{\partial}{\partial x^j}\right)^V\right) = G\left(\left(\frac{\partial}{\partial x^i}\right)^H, \left(\frac{\partial}{\partial x^j}\right)^H\right) = g_{ij},$$
$$G\left(\left(\frac{\partial}{\partial x^i}\right)^V, \left(\frac{\partial}{\partial x^j}\right)^H\right) = 0.$$

It can be easily verified that the Riemannian metric G and the almost complex structure J are compatible on $T\mathcal{M}$ — for any vector fields ξ^1, ξ^2 on $T\mathcal{M}$,

$$G(\xi^2, \xi^2) = G(J\xi^1, J\xi^2), \tag{15}$$

making G an almost Hermitian metric on $T\mathcal{M}$. We write out G in block-matrix form

$$\mathbf{G} = \begin{bmatrix} g_{ij} - N^k_i N^l_j g_{kl} & N^l_i g_{jl} \\ N^l_i g_{jl} & g_{ij} \end{bmatrix}. \tag{16}$$

We denote the fundamental form by $\Omega(\xi^1, \xi^2) = G(J\xi^1, \xi^2)$. Then

$$\Omega = (N^l_j g_{li} - N^l_i g_{lj}) dx^i \wedge dx^j + g_{ij} dx^i \wedge dv^j,$$

or, in block-matrix form

$$\mathbf{\Omega} = \begin{bmatrix} N^l_j g_{li} - N^l_i g_{lj} & g_{ij} \\ -g_{ji} & 0 \end{bmatrix}. \tag{17}$$

3.2. Almost Kähler versus Hermitian structure on $T\mathcal{M}$

Building upon the canonical almost Hermitian structure $(T\mathcal{M}, G, J)$, we have the following important results.

Proposition 7 *([19]). For $(T\mathcal{M}, G, J)$ induced from $(\mathcal{M}, g, \nabla)$, J is integrable if and only if ∇ is flat.*

Remark 8. Explicit calculation gives

$$\begin{aligned}
\left[X^H, Y^H\right] &= [X, Y]^H - (R^\nabla(X, Y)v)^V, \\
\left[X^H, Y^V\right] &= (\nabla_X Y)^V, \\
\left[X^V, Y^V\right] &= 0,
\end{aligned}$$

so that the Nijenhuis tensor N_J has the expression:

$$\begin{aligned}
N_J(X^H, Y^H) &= (T^\nabla(X, Y))^H + (R^\nabla(X, Y)v)^V, \\
N_J(X^H, Y^V) &= N_J(X^V, Y^H) = -(T^\nabla(X, Y))^V + (R^\nabla(X, Y)v)^H, \\
N_J(X^V, Y^V) &= -(T^\nabla(X, Y))^H - (R^\nabla(X, Y)v)^V.
\end{aligned}$$

For N_J to vanish on $T(T\mathcal{M})$, both R^∇ and T^∇ must vanish. Hence ∇ must be flat for J to be integrable.

Satoh [52] calculated $d\Omega$ (below $\mathfrak{S}$ denotes cyclic sum of X, Y, Z):

$$\begin{aligned}
d\Omega(X^H, Y^H, Z^H) &= g\left(\mathfrak{S} R^{\nabla^*}(X, Y)Z, v\right) \\
&= g\left(\mathfrak{S}\left(T^{\nabla^*}\left(T^{\nabla^*}(X, Y), Z\right) + (\nabla^*_X T)(Y, Z)\right), v\right) \\
d\Omega(X^H, Y^H, Z^V) &= g(T^{\nabla^*}(X, Y), Z); \\
d\Omega(X^H, Y^V, Z^V) &= d\Omega(X^V, Y^V, Z^V) = 0.
\end{aligned}$$

Proposition 9 *([52]). For $(T\mathcal{M}, G, J)$ induced from $(\mathcal{M}, g, \nabla)$, the following are equivalent.*

(i) $d\Omega = 0$;
(ii) $\Omega = \tilde{\Omega}$;
(iii) $T^{\nabla^*} = 0$;

where ∇^* is the g-conjugate connection, and $\tilde{\Omega}$ is the pullback by g of the canonical symplectic form Ω_0 on $T^*\mathcal{M}$. Note that the condition of $T^{\nabla^*} = 0$ is the condition of torsion coupling of (∇, g).

Remark 10. On $T^*\mathcal{M}$, there is a canonical symplectic form $\Omega_0 = dx^i \wedge dp_i$, induced from the tautological 1-form $\theta = p_i dx^i$ on $\mathcal{M}$, with $d\Omega_0 = 0$. The symplectic form $\tilde{\Omega}$ on $T\mathcal{M}$ is the pullback, by g, from this canonical symplectic form Ω_0 on $T^*\mathcal{M}$. That is,

$$\tilde{\Omega} = dx^i \wedge d(g_{il}v^l) = \left(\frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i}\right) v^k dx^i \wedge dx^j + g_{ij} dx^i \wedge dv^j.$$

In block-matrix form, it is

$$\tilde{\mathbf{\Omega}} = \begin{bmatrix} \left(\frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i}\right) v^k & g_{ij} \\ -g_{ji} & 0 \end{bmatrix}. \tag{18}$$

Comparing this with Equation (17), we find $\Omega = \tilde{\Omega}$ if and only if

$$\Gamma^l_{jk} g_{li} - \Gamma^l_{ik} g_{lj} = \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i}.$$

This is equivalent to

$$\Gamma^{*l}_{ij} = \Gamma^{*l}_{ji},$$

or

$$T^{*l}_{ij} = \Gamma^{*l}_{ij} - \Gamma^{*l}_{ji} = 0,$$

where T^* is the torsion of the conjugate connection ∇^*.

A consequence of Proposition 9 is the following.

Proposition 11. *When $R^\nabla = 0$, the distributions $\mathcal{H}(T\mathcal{M})$ and $\mathcal{V}(T\mathcal{M})$ induce two foliations on $T\mathcal{M}$. The leaves of these foliations are Lagrangian submanifolds (with respect to $\tilde{\Omega}$) whenever T^{∇^*} is zero.*

Proof. From Remark 8, we can see that the vertical distributions are always integrable, without any assumptions on ∇. However, the horizontal distributions are integrable if and only if $R^\nabla = 0$.

Furthermore, $T\mathcal{M}$ is symplectic when T^{∇^*} is zero. In this case, for any vector fields $X^H, Y^H \in \mathcal{H}(T\mathcal{M})$ and $X^V, Y^V \in \mathcal{V}(T\mathcal{M})$, we have

$$\tilde{\Omega}(X^V, Y^V) = G(JX^V, Y^V) = G(-X^H, Y^V) = 0,$$

and

$$\tilde{\Omega}(X^H, Y^H) = G(JX^H, Y^H) = G(X^V, Y^H) = 0.$$

Combining with Proposition 9, this shows the desired result. □

The above Proposition is a slight generalization of a result [17], which states that $\mathcal{H}(T\mathcal{M})$ is a Lagrangian submanifold when ∇ on $\mathcal{M}$ is assumed to be both torsion-free and Codazzi-coupled to g — indeed, ∇ may carry torsion, and only needs to be torsion-coupled to g. Recall that when ∇ is torsion-free, torsion coupling of (∇, g) is equivalent to Codazzi coupling of (∇, g).

3.3. Statistical mirror symmetry

With the background covered, we can introduce “statistical mirror symmetry.” First, we combine Proposition 7 with Proposition 9 to obtain the following result.

Theorem 12 *(Statistical mirror pair $\mathbb{M}$ and $\mathbb{W}$). Suppose ∇ is a flat (curvature- and torsion-free) connection on a Riemannian manifold $(\mathcal{M}, g)$, and ∇^* is its g-conjugate connection. Then*

(a) $(T\mathcal{M}, G^\nabla, J^\nabla)$, denoted $\mathbb{M}$, is a Hermitian manifold induced from $(\mathcal{M}, g, \nabla)$;

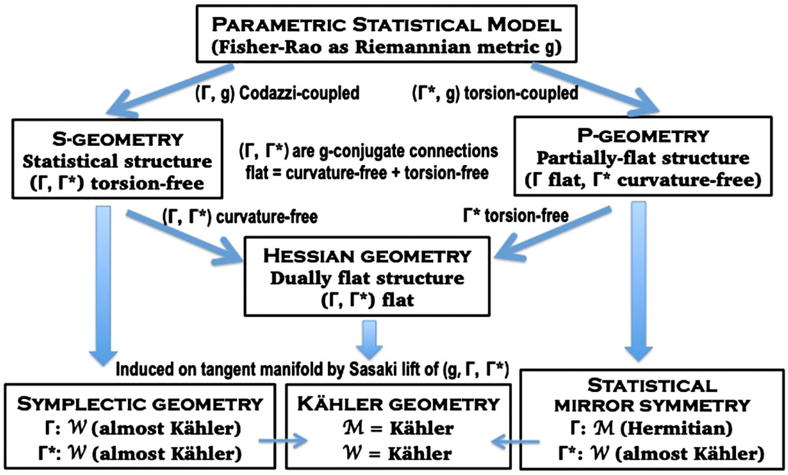

Fig. 1. *S*-geometry, *P*-geometry, and Statistical Mirror Symmetry.

(b) $(T\mathcal{M}, G^{\nabla^*}, J^{\nabla^*})$*, denoted* $\mathbb{W}$*, is an almost Kähler manifold induced from* $(\mathcal{M}, g, \nabla^*)$*;*
(c) $\mathbb{M}$ *and* $\mathbb{W}$ *are two Kähler manifolds, if and only if* $T^{\nabla^*} = 0$.

Proof. (a) The integrability of J^∇ is due to Proposition 7. Given $R^\nabla = 0$, we necessarily have $R^{\nabla^*} = 0$ according to Equation (1). However, it may be the case that $T^{\nabla^*} \neq 0$ even though $T^\nabla = 0$. By Proposition 9, $T^{\nabla^*} \neq 0$ is the obstruction to $d\Omega^\nabla = 0$, so $\mathbb{M}$ need not be Kähler.

(b) The fact that $\mathbb{W}$ is almost Kähler can be seen from the fact that $T^{(\nabla^*)^*} = T^\nabla = 0$, due to ∇ being assumed to be flat. Hence by Proposition 9 the fundamental form Ω^{∇^*} of the almost Hermitian manifold induced from $(\mathcal{M}, g, \nabla^*)$ is d-closed:

$$d\Omega^{\nabla^*} = 0.$$

And $\Omega^{\nabla^*} = \tilde{\Omega}$. So $\mathbb{W}$ is almost Kähler. However, since T^{∇^*} need not vanish so $N_{J^{\nabla^*}}$ may not be zero (by Proposition 7). As such, J^{∇^*} may not be integrable.

(c) In the case that $T^{\nabla^*} = 0$, then $\mathbb{M}$ and $\mathbb{W}$ are furthermore symplectic and Hermitian, respectively. As such, they are both Kähler manifolds (but are different, in general). □

Definition 13 *(Statistical Mirror Symmetry).* Given a partially flat manifold $(\mathcal{M}, g, \nabla, \nabla^*)$ with ∇ denoting the flat connection, let $\mathbb{M} = (T\mathcal{M}, G^\nabla, J^\nabla)$ be the Hermitian manifold induced from $(\mathcal{M}, g, \nabla)$ and $\mathbb{W} = (T\mathcal{M}, G^{\nabla^*}, J^{\nabla^*})$ be the almost Kähler manifold induced from $(\mathcal{M}, g, \nabla^*)$. We say that $\mathbb{M}$ and $\mathbb{W}$ are "statistical mirror pairs," and refer to this correspondence as "statistical mirror symmetry." Fig. 1 illustrates the construction of the statistical mirror.

For the special case when $T^{\nabla^*} = 0$, then both $\mathbb{M}$ and $\mathbb{W}$ are Kähler manifolds. They can be called *Kähler Sasaki manifolds*, and form a pair of "Kähler mirror manifolds." It should be noted that, in such a case, $\mathbb{M}$ and $\mathbb{W}$ are affine manifolds in addition to being Kähler manifolds. However they are not dually flat Hessian manifolds themselves, for such Kähler metric (which is a Sasaki lift metric) is generally not a Hessian metric on $T\mathcal{M}$ (see [1]). We will reserve the term "affine Kähler" for this situation, in contrast to how it is used elsewhere in the literature [45]. Neither of the two Kähler mirror manifolds $\mathbb{M}$ and $\mathbb{W}$ has to be a special Kähler manifold [24] in general. More details of Kähler mirror manifolds will be given in our sequel paper. An example of Kähler mirror will be given in Section 4.4.

3.4. Comparison with mirror symmetry of Calabi-Yau manifolds

Mirror symmetry [28] is a term used in string theory to refer to a duality between pairs of Calabi-Yau manifolds that model the intrinsic dimensions associated to space-time in fundamental physics [12]. Though there are many characterizations, the standard definition of Calabi-Yau manifolds is to treat them as compact Kähler manifolds with trivial canonical bundle. By a celebrated result of Yau [62], this is equivalent to the condition that the manifold admits a Kähler metric with zero Ricci curvature. We will use the term Calabi-Yau manifold to refer to a Kähler manifold with a Ricci-flat metric.

It was first observed by Greene and Plesser [28] that it is possible to construct "mirror" Calabi-Yau manifolds which are distinct, yet produce identical physics. This so-called *mirror symmetry* is an active field of research in mathematics and theoretical physics.

Mirror symmetry is still a mysterious phenomenon, but there are attempts to systematically understand it. Most famously, the SYZ mirror conjecture [55] (named after Strominger, Yau and Zaslow) states that a pair of Calabi-Yau manifolds can be constructed by dual fibrations by special Lagrangian tori in such a way that *symplectic geometry* as one of the mirror pair and *complex geometry* as the other are exchanged.

For statistical mirror symmetry, the simplest case of this duality is a "Kähler mirror" built upon a Hessian base manifold $\mathcal{M}$. However, we do not require the pair of Kähler mirror manifolds $(\mathbb{M}, \mathbb{W})$ to be Ricci-flat. The reason for dropping this requirement is that most statistical manifolds of interest do not induce (via Sasaki lift) Ricci-flat Kähler metric. The following Proposition gives a statistical interpretation for Ricci-flatness (that is mostly violated for parametric statistical families).

Proposition 14. *An exponential family*

$$p^{(e)}(\chi \mid x) = h(\chi) \exp\left(x^i t_i(\chi) - \Phi(x)\right)$$

has a Ricci-flat Kähler metric on $T\mathcal{M}$ if and only if its log-partition function Φ satisfies the Monge-Ampere equation

$$\det \nabla^2(\Phi) = e^{Ax+b} \tag{19}$$

for some constant $A \in GL(n), b \in \mathbb{R}^n$. Here χ denotes the random variable, $[t_1(\chi), \cdots, t_n(\chi)]$ are the sufficient statistics, and log-partition function Φ is

$$\Phi(x) = \log \int_\chi h(\chi) \exp\left(x^i t_i(\chi)\right).$$

Proof. For an exponential model (in canonical form), it is well-known that the Fisher-Rao metric is given by $g = \nabla^2\Phi$, and so is a Hessian metric. The Kähler form on $T\mathcal{M}$ for a base Hessian manifold $\mathcal{M}$ is $\Omega = g_{ij} dx^i \wedge dv^j = \frac{\partial^2\Phi}{\partial x^i \partial x^j} dx^i \wedge dv^j$ in the affine coordinates. The Ricci curvature of $T\mathcal{M}$ may be computed solely in terms of the Hessian potential Φ on $\mathcal{M}$. Explicitly, the Ricci form is

$$\begin{aligned}\rho &= \sqrt{-1}\partial\bar{\partial}\log(\det\Omega) \\ &= \sqrt{-1}\left(\frac{\partial}{\partial x^i} - \sqrt{-1}\frac{\partial}{\partial v^i}\right)\left(\frac{\partial}{\partial x^j} + \sqrt{-1}\frac{\partial}{\partial v^j}\right)\log(\det\nabla^2\Phi).\end{aligned}$$

Since $\log(\det\nabla^2\Phi)$ is independent of the v-coordinates,

$$\rho = \sqrt{-1}\,\frac{\partial^2 \log\det\nabla^2\Phi}{\partial x^i \partial x^j}.$$

Therefore, the Ricci-flat condition $\rho = 0$ amounts to requiring

$$\log\det\nabla^2\Phi = Ax + b,$$

which leads to (19). Therefore, Ricci-flatness for a Kähler Sasaki $T\mathcal{M}$ is equivalent to the requirement that the potential function Φ on $\mathbb{M}$ satisfies Equation (19). $\square$

For a general exponential family, the Riemannian volume density $\sqrt{\det\nabla^2\Phi}$ is known as the non-informative Jeffreys prior in Bayesian inference [34]. As such, Ricci-flatness corresponds to the Jeffreys' prior being an affine function (with respect to the natural parameters), a very stringent requirement.

There is an important special case of Equation (19), which is when Φ satisfies the following:

$$\det\nabla^2(\Phi) = C \tag{20}$$

for some constant $C > 0$. If Φ solves (20), its Legendre dual Φ^* satisfies

$$\det(\nabla^*)^2(\Phi^*) = C^{-1},$$

so in this case Ricci-flatness is preserved by mirror symmetry (i.e. if $\mathbb{M}$ is Ricci-flat, then $\mathbb{W}$ is as well). Furthermore, when $\mathcal{M}$ is a *compact* affine manifold, the maximum principle implies that the only solutions to (19) are actually of the form (20).

In general, the log-partition function Φ of an exponential family will not satisfy (19), much less so for (20), and hence the Ricci-flatness condition fails. Therefore, for statistical mirror symmetry, it is important to consider Kähler manifolds which are not Calabi-Yau. Even in the case where the log-partition function Φ does satisfy Equation (20), it is appropriate to think of $\mathbb{M}$ and $\mathbb{W}$ only as "pseudo" Calabi-Yau manifolds, because exponential families will not have a compact affine base.

It is possible to construct examples of parametric statistical models for which $\mathbb{M}$ and $\mathbb{W}$ are Kähler and Ricci-flat, but not otherwise flat. To do so, consider a smooth convex domain and solve the Monge-Ampere equation (20) with Dirichlet boundary conditions. By the work of Cheng and Yau [15], there is a unique potential solving this equation, which allows us to consider the domain as a statistical manifold whose associated Sasaki lift metric is Ricci-flat. Then, appealing to the embedding result of Le [58], we can find a parametric statistical model realizing this statistical manifold. As such, there exist parametric probability families which are not flat but which induce $\mathbb{M}$ with a Ricci-flat metric. However, we are not aware of a non-trivial example constructed via this approach.

3.5. Relation to semi-flat Calabi-Yau manifolds

Mirror symmetry on so-called "semi-flat" Calabi-Yau manifolds (i.e. Calabi-Yau manifolds which admit a smooth torus fibration over an affine base manifold) has been studied extensively, see [55,43,14]. In some sense, it is the geometrically simplest case of mirror symmetry, which motivates the SYZ conjecture (see [30] for a good survey on the conjecture).

Leung [43] studied the semi-flat case and constructed the mirror explicitly using the Legendre duality of the Hessian structure on the base manifold $\mathcal{M}$ and Fourier-Mukai transformation of the fiber space. Though Leung's construction quotients the fibers of $T\mathcal{M}$ by a discrete lattice so that the Kähler manifolds are compact, the essential arguments between symplectic-to-complex correspondence in the mirror Calabi-Yau manifolds can be made without compactifying the fiber space.

When $\mathbb{M}$ is Calabi-Yau (i.e. $\mathbb{M}$ is Ricci-flat) and has a compact affine base $\mathcal{M}$, the Kähler mirror manifolds we construct are equivalent to Leung's construction (apart from the lattice quotient operation to compactify the fibers). However, the conjugate connections Leung considered are defined differently:

the pair of conjugate connections on $\mathcal{M}$ is obtained through affine immersion in $\mathbb{R}^{n+1}$ using Blaskche normalization. Then, one of these conjugate connections (as induced from Blaskche immersion) is used to construct a complex connection (parallel to J) on one side of the mirror (i.e., $\mathbb{M}$), while the other a symplectic connection (parallel to Ω) on the other side of the mirror (i.e., $\mathbb{W}$). By doing so, a pair of mirror Calabi-Yau manifolds were obtained in which $\mathbb{M}$ corresponds to the so-called A-model and $\mathbb{W}$ corresponds to the so-called B-model.

Some main differences between our construction of statistical mirror symmetry and Leung's construction of semi-flat mirror symmetry are:

(a) Our connections on $\mathcal{M}$ are not Blaskche connections (which are necessarily parallel to the Riemannian volume on $\mathcal{M}$); Leung's Blaschke connections are in general not flat while one of our connections is;
(b) We do not assume any sort of Ricci-flatness, compactness or completeness for our mirror pairs;
(c) Our construction makes use of the Sasaki lift and canonical split of tangent bundle, which is not immediately obvious in Leung's case;
(d) Finally, we do not assume that $\mathbb{M}$ and $\mathbb{W}$ are Kähler.

It is worth noting that semi-flatness of a Calabi-Yau manifold and partial-flatness for conjugate connections are deeply related. In non-Kähler string theory [41], the "semi-flatness" refers to the fact that the fiber directions of $\mathbb{M}$ are flat from the point of view of the complex geometry (though *not* in terms of the Hermitian metric) whereas "partially flat" means that the primal connection ∇ is flat. From this, we have the following observation.

Remark 15. The flat connection of a partially flat structure on $\mathcal{M}$ induces a semi-flat geometry on $\mathbb{M} = T\mathcal{M}$, in that $\mathbb{M}$ as a complex manifold is a tube domain over an affine manifold.

For more discussion on semi-flat geometry for non-Kähler Calabi-Yau manifolds, we refer the reader to [41] and [22].

Initially, it might seem that this construction can give rise to many semi-flat Calabi-Yau manifolds with compact affine base. However, that is not the case. As shown by Cheng and Yau [15], the only compact Calabi-Yau manifolds with Ricci-flat metrics which arise via pulling back solutions to Equation (20) over a compact affine base are complex tori (see [16] for more discussion).

4. Parametric statistical models

As the main motivation to study statistical mirror symmetry originates from the geometric characterization of statistical models, it is worth discussing the geometry of $\mathbb{M}$ and $\mathbb{W}$ in the particular case where $(\mathcal{M}, g, \nabla)$ is the manifold representing a parametric family of probability distributions. In this section, and throughout the rest of the paper, we will specialize to the case where the parameter values are in an open domain in Euclidean space, so that there are no obstructions to defining both $\mathcal{S}$-geometry and $\mathcal{P}$-geometry for an arbitrary parametric statistical model.

4.1. $\mathcal{S}$-geometry and $\mathcal{P}$-geometry

Depending on the affine connection chosen, a given parametric statistical model can have *both* $\mathcal{S}$-geometry *and* $\mathcal{P}$-geometry.

4.1.1. $\mathcal{S}$-geometry of $\mathcal{M}$

A parametric family of probability density functions or distributions, $p(\chi|x)$, all over a given sample space with random variable χ, is collectively called a *parametric statistical model*, where $x = [x^1, \cdots, x^n]$

is in a connected open subset $V \subset \mathbb{R}^n$. In classical information geometry, a manifold $\mathcal{M}$ is constructed by associating each $p(\cdot|x)$ to a point x under a local coordinate chart of $\mathcal{M}$:

$$\mathcal{M} = \left\{ p(\cdot|x) : p(\chi|x) \geq 0, \int_{\chi} p(\chi|x) = 1, x \in V \subset \mathbb{R}^n \right\}.$$

Statistical considerations lead to the following second- and third-order invariants on $\mathcal{M}$ (given in terms of x coordinates)

$$g_{ij}(x) = \int_{\chi} p(\chi|x) \frac{\partial \log p(\chi|x)}{\partial x^i} \frac{\partial \log p(\chi|x)}{\partial x^j} \quad \text{Fisher-Rao metric;}$$

$$C_{ijk} = \int_{\chi} p(\chi|x) \frac{\partial \log p(\chi|x)}{\partial x^i} \frac{\partial \log p(\chi|x)}{\partial x^j} \frac{\partial \log p(\chi|x)}{\partial x^k} \quad \text{Amari-Chensov tensor.}$$

From this data, a family of connections, called α-connections ($\alpha \in \mathbb{R}$), can be constructed as deformation to the Levi-Civita connection ∇^{LC} associated to the Fisher-Rao metric g (see [5]):

$$\nabla^{(\alpha)} = \nabla^{LC} - \frac{\alpha}{2} C,$$

or explicitly,

$$\Gamma^{(\alpha)}_{ij,k}(x) = \int_{\chi} \frac{\partial p(\chi|x)}{\partial x^k} \left(\frac{1-\alpha}{2} \frac{\partial \log p(\chi|x)}{\partial x^i} \frac{\partial \log p(\chi|x)}{\partial x^j} + \frac{\partial^2 \log p(\chi|x)}{\partial x^i \partial x^j} \right). \tag{21}$$

The α- and $(-\alpha)$-connection are conjugate to each other with respect to the Fisher-Rao metric g. Note that all α-connections are torsion-free; yet generally they have non-zero curvatures, with the curvature of $(\pm\alpha)$-connections equal but of opposite signs. When the curvatures of (± 1)-connections vanish, g takes the form of a Hessian metric. It is important to keep in mind that each member of the α-connection is Codazzi-coupled to the Fisher-Rao metric g. We call it $\mathcal{S}$*-geometry* of a parametric statistical model; it is the standard model in information geometry [2,5], with the so-called "statistical structure."

For application to parametric statistical models, the Riemannian metric is the Fisher-Rao metric and the pair of conjugate connections are the (± 1)-connections. These are "canonical" objects once the parametric statistical model $p(\cdot|x)$ is specified, and it has been shown that they are the unique second- and third-order invariants for the parametric statistical model (see [13,5,7]). Furthermore, this geometry can be generated using divergence (contrast) functions.

4.1.2. $T\mathcal{M}$ from $\mathcal{S}$-geometry

For $\mathcal{S}$-geometry, it is possible to construct a Sasaki lift metric on the tangent bundle $T\mathcal{M}$. In this case, both $(T\mathcal{M}, G^{\nabla}, J^{\nabla})$ and $(T\mathcal{M}, G^{\nabla^*}, J^{\nabla^*})$ are almost Kähler manifolds. In fact, we have an entire family of almost Kähler manifolds $\mathbb{W}^{(\alpha)} = (T\mathcal{M}, G^{\nabla^{(\alpha)}}, J^{\nabla^{(\alpha)}})$, $\alpha \in \mathbb{R}$, each with α-dependent metric G and α-dependent almost complex structure J. However, they all have the same symplectic form $\Omega(\cdot,\cdot) = G^{\nabla^{(\alpha)}}(J^{\nabla^{(\alpha)}}(\cdot),\cdot)$, which is the g-pullback of the canonical symplectic form Ω_0 on $T^*\mathcal{M}$. Furthermore, none of them are complex unless ∇ (and ∇^*) is curvature-free, in which case $\mathcal{M}$ becomes dually flat.

While there is a duality between $\mathbb{W}^{(\alpha)}$ and $\mathbb{W}^{(-\alpha)}$, it is a duality between two almost complex structures on symplectic manifolds, and lacks the essential complex-to-symplectic duality that appears in non-Kähler mirror symmetry. For that reason, we reserve the terminology "statistical mirror symmetry" to the context of the $\mathcal{P}$-geometry only.

4.1.3. $\mathcal{P}$-geometry of $\mathcal{M}$

In $\mathcal{P}$-geometry, we take $x = [x^1, \cdots, x^n]$, the parameter of the parametric statistical model $p(\cdot|x)$, to be affine coordinates for a flat connection which we denote as ∇, i.e., the Christoffel symbol Γ_{ij}^k vanishes: $\Gamma_{ij}^k(x) = 0$.

Writing out the equation of conjugate connection ∇^* under this coordinate chart

$$\frac{\partial g_{ij}}{\partial x^k} = g_{jl}\Gamma^l_{ki} + g_{il}\Gamma^{*l}_{kj} = g_{il}\Gamma^{*l}_{ki}\,.$$

Therefore, the conjugate connection ∇^*, expressed in terms of the x-coordinates, is

$$\Gamma^{*j}_{ki} = g^{jl}\frac{\partial g_{il}}{\partial x^k}\,, \tag{22}$$

with g^{ij} denoting the elements of the matrix inverse of g, the Fisher-Rao metric. It can be readily verified that such connection is always curvature-free, but carries torsion

$$T^{*j}_{ik} = \Gamma^{*j}_{ik} - \Gamma^{*j}_{ki} = g^{jl}\left(\frac{\partial g_{kl}}{\partial x^i} - \frac{\partial g_{il}}{\partial x^k}\right) \neq 0$$

unless $\frac{\partial g_{il}}{\partial x^k} = \frac{\partial g_{kl}}{\partial x^i}$, i.e., $\frac{\partial g_{il}}{\partial x^k}$ is totally symmetric, which is equivalent to g being Hessian. Note that the torsion of Γ^{*j}_{ki} is *not* captured by its geodesic curves themselves; it describes the "screw" component of the motion with axis of rotation precisely the tangent direction of the curve. The constructed biorthogonal frame $\{\mathbf{b}^i\}_{i=1}^n$ is (see [65])

$$\mathfrak{B} = \{\mathbf{b}^i\}_{i=1}^n = \{g^{ij}\partial_{x^j}\}_{i=1}^n\,.$$

This frame is nothing but "natural gradients" as referred to by the machine learning community after Amari [3].

4.1.4. $T\mathcal{M}$ from $\mathcal{P}$-geometry

The $\mathcal{P}$-geometry provides the natural setting to construct statistical mirror symmetry. In this context $\mathbb{M}$ naturally becomes a Hermitian manifold whereas $\mathbb{W}$ is almost Kähler.

More concretely, if we consider the x-coordinates as a global coordinate chart $V \to \mathcal{M}$ with $V \subset \mathbb{R}^n$, then as a complex manifold $\mathbb{M}$ is simply the set

$$\left\{x + \sqrt{-1}v \mid x \in V, v \in \mathbb{R}^n\right\} \subset \mathbb{C}^n.$$

Such a set is known as a tube domain with base V and the natural coordinates $z = x + \sqrt{-1}v$ are holomorphic. For a more complete discussion of tube domains, we refer the reader to [33], p. 41. Note that this complex structure on $\mathbb{M}$ is determined only by the choice of a flat connection on $\mathcal{M}$, and is independent of the Riemannian metric on $\mathcal{M}$.

On the symplectic side, $\mathbb{W}$ as a symplectic manifold has the pullback of the canonical symplectic structure on $T^*\mathcal{M}$. Both the horizontal sub-bundle $\mathcal{H}(T\mathcal{M})$ and vertical sub-bundle $\mathcal{V}(T\mathcal{M})$ of $T\mathcal{M}$ are Lagrangian sub-bundles. However, its associated almost complex structure, which swaps $\mathcal{H}(T\mathcal{M})$ and $\mathcal{V}(T\mathcal{M})$, is not integrable in general.

4.2. Contrast and pre-contrast functions

4.2.1. Contrast function and $\mathcal{S}$-geometry

A contrast function (a.k.a. divergence function) $D : \mathcal{M}\times\mathcal{M} \to \mathbb{R}_{\geq 0}$ on a manifold $\mathcal{M}$ under a local chart $V \subset \mathbb{R}^n$ is defined as a smooth function (three-times differentiable) which satisfies

(1) $D(x,y) \geq 0\ \forall x, y \in V$ with equality holding if and only if $x = y$;
(2) $\partial_{x^i} D(x,y)|_{x=y} = \partial_{y^j} D(x,y)\big|_{x=y} = 0, \forall i, j \in \{1, 2, \cdots, n\}$;
(3) $-\partial_{x^i}\partial_{y^j} D(x,y)\big|_{x=y}$ is positive semi-definite.

Note that (2) is included in the definition though it is a consequence of the smoothness assumption of D and (1). The mixed derivatives in (3) are not to be confused with Hessian of D with respect to either x or y.

On a manifold, contrast functions act as "pseudo-distance" functions that are non-negative but need not be symmetric nor satisfy the triangle inequality. The fact that a statistical manifold can be induced from a contrast function was first demonstrated by Eguchi [20,21].

Proposition 16 *([20,21]). A contrast function D induces a (semi-)Riemannian metric g and a pair of torsion-free conjugate connections Γ, Γ^* given as*

(a) $g_{ij}(x) = -\partial_{x^i}\partial_{y^j} D(x,y)\big|_{x=y}$;
(b) $\Gamma_{ij,k}(x) = -\partial_{x^i}\partial_{x^j}\partial_{y^k} D(x,y)\big|_{x=y}$;
(c) $\Gamma^*_{ij,k}(x) = -\partial_{x^k}\partial_{y^i}\partial_{y^j} D(x,y)\big|_{x=y}$.

Here, $\Gamma_{ij,k} = \Gamma^l_{ij} g_{lk} = g(\nabla_{\partial_{x^i}} \partial_{x^j}, \partial_{x^k})$. It is easily verified that $\Gamma_{ij,k}, \Gamma^*_{ij,k}$ as given above are torsion-free and satisfy the conjugacy condition with respect to the induced metric g_{ij}, which is semi-Riemannian (in the sense that g is positive semi-definite). Hence a contrast function induces a "statistical manifold", that is, the $\mathcal{S}$-geometry of the family of probability distributions.

4.2.2. Pre-contrast function and $\mathcal{P}$-geometry

Contrast functions induce a pair of conjugate connections without torsion. To consider connections with torsion, Henmi and Matsuzoe [31] proposed the idea of a pre-contrast function. They also constructed a canonical precontrast function ρ which induces a given partially flat structure [32].

Definition 17. Pre-contrast functions $\rho(X, x, y) : T\mathcal{M} \times \mathcal{M} \to \mathbb{R}$ are defined as [31]

(1) $\rho(f_1 X_1 + f_2 X_2, x, y) = f_1 \rho(X_1, x, y) + f_2 \rho(X_2, x, y)$, where f_1, f_2 are functions and X_1, X_2 are tangent vector fields evaluated at a point x in the manifold $\mathcal{M}$;
(2) $\rho(\partial_{x^i}, x, x) = 0$;
(3) $-\partial_{y^j} \rho(\partial_{x^i}, x, y)\big|_{x=y}$ is positive semi-definite.

A pre-contrast function $\rho(\partial_{x^i}, x, y)$ induces a dualistic geometry $(\mathcal{M}, g, \Gamma, \Gamma^*)$ as below:

(a) $\partial_{y^j} \rho(\partial_{x^i}, x, y)\big|_{x=y} = -g_{ij}(x)$;
(b) $\partial_{x^k}\partial_{y^j} \rho(\partial_{x^i}, x, y)\big|_{x=y} = -\Gamma_{ki,j}(x)$;
(c) $\partial_{y^k}\partial_{y^j} \rho(\partial_{x^i}, x, y)\big|_{x=y} = -\Gamma^*_{kj,i}(x)$.

Torsion-freeness of Γ^* is reflected in the fact that $\Gamma^*_{kj,i} = \Gamma^*_{jk,i}$. However, since $\Gamma_{ki,j} \neq \Gamma_{ik,j}$ in general, $\mathcal{M}$ as generated is a "statistical manifold admitting torsion" (SMAT), in which (∇, g) is torsion-coupled.

We can define the "dual" pre-contrast function $\rho^*(x, \partial_{y^i}, y) : \mathcal{M} \times T\mathcal{M} \to \mathbb{R}$ as one satisfying

(1) $\rho^*(x, f_1 Y_1 + f_2 Y_2, y) = f_1 \rho^*(x, Y_1, y) + f_2 \rho^*(x, Y_2, y)$, where f_1, f_2 are functions and Y_1, Y_2 are tangent vector fields evaluated at $y \in \mathcal{M}$;

(2) $\rho^*(y, \partial_{y^i}, y) = 0$;
(3) $-\,\partial_{x^j} \rho^*(x, \partial_{y^i}, y)\big|_{x=y}$ is positive semi-definite.

$\rho^*(x, \partial_{y^i}, y)$ will induce on $\mathcal{M}$ the following (g, Γ, Γ^*):

(a) $\partial_{x^j} \rho^*(x, \partial_{y^i}, y)\big|_{x=y} = -g_{ij}(x)$;
(b) $\partial_{x^k} \partial_{x^j} \rho^*(x, \partial_{y^i}, y)\big|_{x=y} = -\Gamma_{kj,i}(x)$,
(c) $\partial_{y^k} \partial_{x^j} \rho^*(x, \partial_{y^i}, y)\big|_{x=y} = -\Gamma^*_{ki,j}(x)$.

In such a case, ∇ is torsion-free and (∇^*, g) is torsion-coupled, so the induced $(\mathcal{M}, g, \nabla, \nabla^*)$ is a SMAT.

A special case of SMAT is partially flat geometry, in which the conjugate connections have zero curvatures. A closed-form expression for the pre-contrast function generating the partially flat geometry was given in [66],

$$\rho(\partial_{x^i}, x, x') = (x^j - x'^j)\, g_{ij}(x).$$

The above pre-contrast function induces a partially flat manifold, that is, the $\mathcal{P}$-geometry of the family of probability distributions.

4.2.3. Super-contrast function

Pre-contrast functions can generate conjugate connections in which one of the connections is torsion-free and the other may have torsion. In order to induce dualistic geometry $(\mathcal{M}, g, \nabla, \nabla^*)$ in which both ∇ and ∇^* may have torsion, we resort to a new kind of function, which we call “super-contrast function” defined below.

Definition 18. A super-contrast function $S : T\mathcal{M} \times T\mathcal{M} \to \mathbb{R}$ is a smooth function such that

(1) For $X \in T_x\mathcal{M}$, $Y \in T_yM$, $S(X, x, Y, y)$ is bilinear in X and Y, and
(2) $-\,S(\partial_{x^i}, x, \partial_{y^j}, y)\big|_{x=y}$ is positive semi-definite.

Super-contrast functions induce the dualistic geometry $(\mathcal{M}, g, \nabla, \nabla^*)$ as follows:

(a) $S(\partial_{x^i}, x, \partial_{y^j}, y)\big|_{x=y} = -g_{ij}(x)$;
(b) $\partial_{x^k} S(\partial_{x^i}, x, \partial_{y^j}, y)\big|_{x=y} = -\Gamma_{ki,j}(x)$;
(c) $\partial_{y^k} S(\partial_{x^i}, x, \partial_{y^j}, y)\big|_{x=y} = -\Gamma^*_{kj,i}(x)$.

Note that (a) says that super-contrast functions S on a manifold $\mathcal{M}$ induce a (positive semi-definite) Riemannian metric on $\mathcal{M}$. S takes in one vector at point x and another vector at point y to form a two point function. S can also be thought of as a one-form upon fixing a vector either at x or at y.

In some sense, super-contrast functions induce a more flexible geometric framework than pre-contrast functions, which is in turn more flexible than contrast functions. To make this observation precise, we make the following remark.

Remark 19.

(1) Every contrast function $D(\cdot, \cdot) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$ induces *two* associated pre-contrast functions, $\rho(\cdot, \cdot, \cdot) : T\mathcal{M} \times \mathcal{M} \to \mathbb{R}$ and $\rho^*(\cdot, \cdot, \cdot) : \mathcal{M} \times T\mathcal{M} \to \mathbb{R}$

$$\begin{tikzcd}
& T\mathcal{M}\times T\mathcal{M} \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"] & \\
\mathcal{M}\times T\mathcal{M} \arrow[dr, "\pi_2"] & & T\mathcal{M}\times\mathcal{M} \arrow[dl, "\pi_1"] \\
& \mathcal{M}\times\mathcal{M} &
\end{tikzcd}$$

Fig. 2. Pre- and super-contrast functions can be defined along pullbacks.

$$\rho(\partial_{x^i}, x, y) = \partial_{x^i} D(x, y)$$
$$\rho^*(x, \partial_{y^i}, y) = \partial_{y^i} D(x, y)$$

and an associated super-contrast function $S(\cdot,\cdot,\cdot,\cdot) : T\mathcal{M} \times T\mathcal{M} \to \mathbb{R}$

$$S(\partial_{x^i}, x, \partial_{y^j}, y) = \partial_{x^i}\partial_{y^j} D(x, y). \tag{23}$$

(2) Every pre-contrast function $\rho(\cdot,\cdot,\cdot) : T\mathcal{M}\times\mathcal{M} \to \mathbb{R}$ or dual pre-contrast function $\rho^*(\cdot,\cdot,\cdot) : \mathcal{M}\times T\mathcal{M} \to \mathbb{R}$ induces an associated super-contrast function S

$$S(\partial_{x^i}, x, \partial_{y^j}, y) = \partial_{y^j}\rho(\partial_{x^i}, x, y) = \partial_{x^i}\rho^*(x, \partial_{y^j}, y). \tag{24}$$

(3) The inducing function and the induced function(s) in Part (1) and (2) of this Remark generate the same dualistic geometry $(\mathcal{M}, g, \nabla, \nabla^*)$.

Whereas reference duality is encoded by the presence of a dual pair of pre-contrast functions, super-contrast functions have a natural reference duality, namely

$$S(\partial_{x^i}, x, \partial_{y^i}, y) \mapsto S^*(\partial_{x^i}, x, \partial_{y^i}, y) = S(\partial_{y^i}, y, \partial_{x^i}, x).$$

Here we nominally define S^* as switching the role of the two copies of $T\mathcal{M}$ in $T\mathcal{M} \times T\mathcal{M}$ ("reference duality"). This duality switches the roles of ∇ and ∇^*, but leaves g unchanged.

It is easily seen that these inducing relationships among pre/super/contrast functions can be "pulled back" along the projections within the following diagram (Fig. 2).

4.3. Reference-representation biduality and statistical mirror symmetry

Given a Hessian pair (∇, Φ) on $\mathcal{M}$, namely, a flat connection ∇ on and a convex function Φ in its affine coordinate, we can construct

(1) a Hessian manifold $(\mathcal{M}, \nabla, \nabla^2\Phi)$ leading to a Kähler manifold $\mathbb{M} = (T\mathcal{M}, G^{\nabla}, J^{\nabla})$;
(2) a canonical divergence function, D_Φ, the so-called Bregman divergence, defined as

$$D_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle \partial\Phi(y), x - y\rangle.$$

It is well known in information geometry that the Bregman divergence is a surrogate to a metric in an affine space — in fact $D_\Phi(x, y)$ is the affine support function [53]. It is non-negative, non-symmetric, and obeys a generalized Pythagorean relation instead of the triangle inequality. With respect to the Bregman divergence, reference duality pertains to a switch of reference point in the non-symmetric divergence

$$D_\Phi(x, y) \mapsto D_\Phi(y, x),$$

$$\begin{tikzcd}
(\mathcal{M} \times \mathcal{M}, D_\Phi) \arrow[r, mapsto, squiggly] \arrow[d, dashed, leftrightarrow, "\mathrm{RRB}"'] & (T\mathcal{M}, G^{\nabla}, J^{\nabla}) \arrow[d, dashed, leftrightarrow, "\mathrm{SMS}"] \\
(\mathcal{M} \times \mathcal{M}, D_{\Phi^*}) \arrow[r, mapsto, squiggly] & \left(T\mathcal{M}, G^{\nabla^*}, J^{\nabla^*}\right)
\end{tikzcd}$$

Fig. 3. Commutative diagram for SMS and RRB, where SMS = Statistical Mirror Symmetry, RRB = Reference-Representation Biduality.

and representation duality pertains to a change to the convex conjugate representation

$$D_\Phi(x, y) \mapsto D_{\Phi^*}(x^*, y^*),$$

where $x^* = \partial\Phi(x), y^* = \partial\Phi(y)$, or $x = \partial\Phi^*(x^*), y = \partial\Phi^*(y^*)$. Combining these two transformations of the divergence function yields the following identity known as "reference-representation biduality" [64]:

$$D_\Phi(x, y) = D_{\Phi^*}(y^*, x^*). \tag{25}$$

Proposition 20. *Given a Kähler manifold $\mathbb{M}$ on the tangent bundle $T\mathcal{M}$ of a Hessian manifold $(\mathcal{M}, \nabla, \nabla^2\Phi)$, a statistical mirror is obtained by applying reference-representation biduality to the associated Bregman divergence. See Fig. 3.*

In other words, statistical mirror symmetry for dually flat statistical manifolds is a geometric consequence of the reference-representation biduality. Here, D_Φ and D_{Φ^*} denote the Bregman divergences and the horizontal arrows correspond to using a Kähler metric on $T\mathcal{M}$ to find an associated Hessian structure on $\mathcal{M}$ and furthermore find the canonically defined Bregman divergence D_Φ on this Hessian structure.

We will now show that this diagram (Fig. 3) is commutative, which makes precise the observation that statistical mirror symmetry is a consequence of reference-representation biduality.

Proof. With reference to the diagram (Fig. 3).

If we are given a Hessian manifold $(\mathcal{M}, \Phi, \nabla)$ with Bregman divergence D_Φ, we can construct *both* a Kähler metric $(T\mathcal{M}, g^{\nabla}, J^{\nabla}) := \mathbb{M}$ on its tangent bundle *and* a canonical divergence on $(\mathcal{M} \times \mathcal{M}, D_\Phi)$. Applying reference-representation biduality (down dashed arrow) to the left side of the diagram, we express the same canonical divergence D_Φ based on Φ using its Legendre dual Φ^*, due to the identity (25). We then construct the dual connection ∇^* on $\mathcal{M}$ with respect to $g = \nabla^2\Phi$. Using $(\mathcal{M}, \Phi^*, \nabla^*)$, we can construct a mirror Kähler metric $(T\mathcal{M}, g^{\nabla^*}, J^{\nabla^*})$ (which is denoted by the lower right squiggle arrow). This space is exactly what we call $\mathbb{W}$, which shows that the diagram commutes. As such, we can construct Kähler statistical mirror symmetry using reference-representation biduality. □

It is important to emphasize that correspondences in the above proof are not maps between points in the corresponding spaces, but rather correspondences between the spaces themselves. As such, these arrows should be considered as "maps to."

When a statistical manifold is not dually flat, we cannot use reference-representation biduality to induce statistical mirror symmetry. This is due to two issues.

(1) There is not yet a complete theory of canonical divergences for the $\mathcal{S}$-geometry (see but [6]). In the case of $\mathcal{S}$-geometry, which uses a dual pair of torsion-free connections on the base manifold $\mathcal{M}$, the almost complex structures on $T\mathcal{M}$ are almost Kähler, but are not necessarily J-integrable.
(2) In the partially flat case (the $\mathcal{P}$-geometry), we cannot induce the dualistic geometry using contrast functions which necessarily lead to torsion-free connections. Instead, it seems necessary to consider pre-

$$\begin{tikzcd}
(T\mathcal{M} \times T\mathcal{M}, S) \arrow[r, mapsto, squiggly] \arrow[d, dashed, leftrightarrow, bend right, "\text{Reference Duality}"'] & \mathbb{M} \arrow[d, dashed, leftrightarrow, bend left, "\text{SMS}"] \\
(T\mathcal{M} \times T\mathcal{M}, S^*) \arrow[r, mapsto, squiggly] & \mathbb{W}
\end{tikzcd}$$

Fig. 4. Commutative diagram for the partially flat case.

contrast functions $\rho : T\mathcal{M} \times \mathcal{M} \to \mathbb{R}$. However, since the domain is not symmetric, we cannot switch variables to induce reference-representation biduality.

Proposition 21. *In the partially flat case, statistical mirror symmetry is a geometric consequence of the reference duality for the canonical super-contrast function.*

Proof. As before, we construct the following diagram (Fig. 4), using the canonical super-contrast function and its dual. As the reference duality for super-contrast functions switches ∇ and ∇^*, this diagram is automatically commutative, and so we can obtain the statistical mirror pair of $\mathbb{M}$ by moving counter-clockwise around the diagram. □

It is worth remarking that in the dually flat case, we can differentiate the Bregman divergence twice to obtain the super-contrast function and obtain statistical mirror symmetry from the reference duality of the super-contrast function. However, this status of representation duality is yet to be understood.

4.4. An example: univariate normal distribution

In order to get an intuitive understanding of Statistical Mirror Symmetry (SMS), we give a concrete example of parametric probability family. We consider the statistical manifold $\mathcal{M}$ of univariate normal distributions

$$\mathbf{N}(\chi|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\chi-\mu)^2}{2\sigma^2}\right).$$

Here, the random variable χ defined on the real line and the parameters are $m = \sqrt{2}\,\mu$ and σ (the factor of $\sqrt{2}$ is for later convenience). The manifold $\mathcal{M}$ is the upper half-plane

$$\mathbb{H} = \{(m,\sigma) \in \mathbb{R}^2 | \sigma \geq 0\},$$

and the Fisher-Rao metric is $g = \frac{dm^2+d\sigma^2}{\sigma^2}$. This is a metric of constant negative curvature. In other words, $\mathcal{M}$ is the hyperbolic geometry (the Poincaré half-plane model), which was first observed by Amari.

Below, we consider both the statistical structure ($\mathcal{S}$-geometry) and partially flat structure ($\mathcal{P}$-geometry) of this manifold $\mathcal{M}$, as well as the statistical mirror pairs $\mathbb{M}$ and $\mathbb{W}$ on its $T\mathcal{M}$.

4.4.1. $\mathcal{P}$-geometry

One possible $\mathcal{P}$-geometry treats the univariate normal distributions $\mathbf{N}(\chi|\mu,\sigma)$ as a location-scale family, using the mean μ as the location parameter and variance σ as the dispersion parameter. These parameters of the univariate normal model are intrinsically meaningful in statistics, so it is natural to consider this parameterization $(m = \sqrt{2}\mu, \sigma)$ as affine coordinates for their associated flat connection. In these coordinates, the Fisher-Rao metric is

$$g = \frac{1}{\sigma^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The Fisher-Rao metric, when expressed in the (m, σ)-coordinates, is *not* a Hessian metric.

Starting from this coordinate frame,

$$\left\{ \frac{\partial}{\partial m}, \frac{\partial}{\partial \sigma} \right\},$$

its biorthogonal frame is

$$\left\{ \frac{\sigma^2}{2} \frac{\partial}{\partial m}, \frac{\sigma^2}{2} \frac{\partial}{\partial \sigma} \right\}.$$

By computing the Lie bracket of the biorthogonal frame, we find that

$$\left[\frac{\sigma^2}{2} \frac{\partial}{\partial m}, \frac{\sigma^2}{2} \frac{\partial}{\partial \sigma} \right] = -\frac{\sigma^3}{2} \frac{\partial}{\partial m}.$$

This is the torsion of the conjugate connection adapted to the g-biorthogonal frame. The biorthogonal frame is not a coordinate frame due to this non-zero torsion. Some properties of the connection adapted to this biorthogonal frame were discussed in [65], and we refer the reader to that paper for more details.

4.4.2. Statistical mirror under $\mathcal{P}$-geometry

On $\mathbb{M}$, we can write (z_1, z_2) as the (holomorphic) complex coordinates, where

$$z_1 = m + \sqrt{-1}\,\zeta_1,$$
$$z_2 = \sigma + \sqrt{-1}\,\zeta_2.$$

As such, the complex manifold $\mathbb{M}$ is biholomorphic to a half-space in $\mathbb{C}^2$. As a Riemannian manifold, its metric is diagonal (and complete), but it is not Kähler due to the fact that the Fisher-Rao metric g on the base $\mathcal{M} = \mathbb{H}$ is not Hessian. The associated Kähler form in $(m, \sigma, \zeta_1, \zeta_2)$ coordinates is given by

$$\Omega_{\mathbb{M}} = \frac{1}{\sigma^2} \left(dm \wedge d\zeta_1 + d\sigma \wedge d\zeta_2 \right).$$

Computing the exterior derivative, we find that

$$d\Omega_{\mathbb{M}} = \frac{2}{\sigma^3} d\zeta_1 \wedge dm \wedge d\sigma.$$

This shows directly that the Kähler form is not closed, $d\Omega_{\mathbb{M}} \neq 0$.

On the other hand, $\mathbb{W}$ is an almost Kähler manifold. It does not have holomorphic coordinates, but in the $(m, \sigma, \zeta_1, \zeta_2)$-coordinates, its symplectic form is

$$\Omega_{\mathbb{W}} = \frac{1}{\sigma^2} \left(dm \wedge d\zeta_1 + d\sigma \wedge d\zeta_2 \right) - \frac{2}{\sigma^3} \zeta_1 \, dm \wedge d\sigma,$$

which is closed — one can directly verify $d\Omega_{\mathbb{W}} = 0$.

It is also possible to explicitly calculate the almost-complex structure. In the (m, σ)-coordinates, the Christoffel symbols are ${\Gamma^*}^j_{ki} = \frac{-2}{\sigma} \delta^j_i \delta_{kj}$. Using the general form for the almost complex structure, we find that in the $(m, \sigma, \zeta_1, \zeta_2)$-coordinates, the almost complex structure acts on the basis $\left(\frac{\partial}{\partial m}, \frac{\partial}{\partial \sigma}, \frac{\partial}{\partial \zeta_1}, \frac{\partial}{\partial \zeta_2} \right)$ as follows:

$$\mathbf{J}_{\mathbb{W}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ \frac{2}{\sigma}\zeta_1 & \frac{2}{\sigma}\zeta_2 & \frac{4}{\sigma^2}\zeta_2\zeta_1 & \frac{4}{\sigma^2}\zeta_2^2+1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -\frac{2}{\sigma}\zeta_1 & -\frac{2}{\sigma}\zeta_2 \end{bmatrix}.$$

It is a tedious verification that $N_{J_{\mathbb{W}}} \neq 0$.

4.4.3. $\mathcal{S}$-geometry

Since the univariate normal distribution $\mathbf{N}(\chi|\mu,\sigma)$ is an exponential model,

$$\mathbf{N}(\chi|\mu,\sigma) = \exp\left(\frac{\mu}{\sigma^2}\chi + \left(-\frac{1}{2\sigma^2}\right)\chi^2 - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)\right),$$

the $\mathcal{S}$-geometry will consist of a dually flat Hessian geometry on the base manifold.

The natural coordinates $x=(x^1,x^2)$ and expectation coordinates $u=(u_1,u_2)$ are the following:

$$x^1 = \frac{\mu}{\sigma^2}, \qquad x^2 = -\frac{1}{2\sigma^2};$$
$$u_1 = \mu, \qquad u_2 = \mu^2+\sigma^2.$$

When treating those coordinates as affine coordinates for the dually flat connections, the Fisher-Rao metric g becomes a Hessian metric with potential Φ given by

$$\Phi(x) = -\frac{x^1\cdot x^1}{4x^2} - \frac{1}{2}\log(-x^2).$$

The conjugate potential Φ^* (expressed in u-coordinates) is

$$\Phi^*(u) = \frac{1}{2}\log(u_2 - u_1\cdot u_1).$$

The Fisher-Rao metric, given in the x-coordinates is

$$g = \frac{1}{2x^2}\begin{bmatrix} -1 & \frac{x^1}{x^2} \\ \frac{x^1}{x^2} & \frac{-x^1\cdot x^1+x^2}{x^2\cdot x^2} \end{bmatrix},$$

and in the u-coordinates, it is

$$g = \frac{1}{(u_1\cdot u_1 - u_2)^2}\begin{bmatrix} u_1\cdot u_1+u_2 & -u_1 \\ -u_1 & \frac{1}{2} \end{bmatrix}.$$

4.4.4. Kähler mirror pair under $\mathcal{S}$-geometry

On $T\mathcal{M}$, there are two Kähler manifolds $\mathbb{M}$ and $\mathbb{W}$ linked by statistical mirror symmetry, which are interesting spaces in their own right.

The Kähler manifold associated with the natural parameters (i.e., x-coordinates) on the base $\mathcal{M}$ is a half-space in $\mathbb{C}^2$ (i.e. a tube domain over a half plane). The associated Kähler metric is known as the *Kähler-Berndt metric*, and its geometry was studied in depth by Molitor [49]. He further observed that $\mathbb{M}$ is known as the *Siegel-Jacobi space*, and is a complete space of constant negative scalar curvature and non-positive Ricci curvature.

This metric was previously studied by the current authors in the setting of optimal transport [37]. In that work, it was shown that this metric has no definite sign for holomorphic sectional curvature, bisectional curvature or orthogonal bisectional curvature.[5] However, it does have a subtle non-negativity property known as *non-negative orthogonal anti-bisectional curvature*, which is relevant for optimal transport.

The mirror Kähler manifold $\mathbb{W}$ is also of independent interest. As a complex manifold, it is a tube domain over the convex set $S = \{u_2 - u_1 \cdot u_1 > 0\}$. In fact, this is a classic domain in the study of several complex variables, known as the *Siegel half-space*. Although not immediately obvious, it is biholomorphic to the unit ball in $\mathbb{C}^2$ [61]. As a Kähler manifold, $\mathbb{W}$ is a complete complex space form with constant negative holomorphic sectional curvature. As a result, it is Hermitian symmetric, Kähler-Einstein, and has negative quarter-pinched sectional curvature.

It is worth noting that in this example $\mathbb{M}$ and $\mathbb{W}$ are distinct as complex manifolds — for $n > 1$ the half space contains a complex line so cannot map surjectively to the unit ball (or indeed any bounded domain). As is the case with generic statistical mirrors, $\mathbb{M}$ and $\mathbb{W}$ are diffeomorphic (as smooth manifolds, they are both $T\mathcal{M}$). However, in general they are not biholomorphic, even when $\mathbb{W}$ is complex.

Finally, we note that both the Siegel-Jacobi space (i.e. $\mathbb{M}$) and the Siegel half-space (i.e. $\mathbb{W}$) have been studied in the context of number theory (see, e.g. [9] [60]). One of the main motivations for studying mirror symmetry is that it has been used to solve difficult problems in algebraic geometry by translating them to questions at the other side of the mirror (see, e.g. [18]). Similarly, this example suggests possible applications of statistical mirror symmetry for analytic number theory. We will discuss this further in future work.

4.5. Finding canonical coordinates: balanced metrics

The last example showed that the family of univariate normal distributions $\mathbf{N}(\chi|\mu, \sigma)$ has the generic statistical mirror with a complex $\mathbb{M}$ and a symplectic $\mathbb{W}$ under the $\mathcal{P}$-geometry where the (μ, σ) parameter is treated as affine coordinates of a flat connection; this original parameterization reflects the location-scale nature of $\mathbf{N}(\chi|\mu, \sigma)$. On the other hands, under the $\mathcal{S}$-geometry with (x^1, x^2) and (u_1, u_2) as biorthogonal coordinates, the resulting $\mathbb{M}$ and $\mathbb{W}$ are both Kähler manifolds, with the Kähler mirror reflecting a Siegel-Jacobi space on the one side ($\mathbb{M}$) and a Siegel half-space on the other ($\mathbb{W}$). Such parameterization is due to the fact that the family $\mathbf{N}(\chi|x^1, x^2)$ of univariate normal distributions is also an exponential family. The $\mathcal{S}$-geometry treats $\mathcal{M}_{\mathbf{N}}$ as dually flat under the x and u parameterizations. In other words, though using an identical Fisher-Rao metric for $\mathcal{M}_{\mathbf{N}}$, $\mathcal{S}$-geometry and $\mathcal{P}$-geometry pick different flat connections: just as (μ, σ) are not affine coordinates under $\mathcal{S}$-geometry, neither x nor u parameterizations are affine coordinates under $\mathcal{P}$-geometry. In our theory of Statistical Mirror Symmetry, the Riemannian manifold $(\mathcal{M}_{\mathbf{N}}, g)$ for the same parametric statistical model $\mathbf{N}$ (in this case univariate normal distributions) may be prescribed with different flat affine connections or, equivalently, may treat different parameterizations as affine coordinates.

For any exponential family, it is well-known to information geometers that there is a canonical dually flat structure induced by the natural and expectation parameters. For more general statistical models, there is no guarantee that the Fisher-Rao metric can be written in Hessian form, and it is of interest to try to generalize this construction to a more general (affine) Riemannian manifold.

QUESTION. Given an affine Riemannian manifold $(\mathcal{M}, g)$, can we find a pair of dual connections which are in some sense *canonical*?

Let us start with the partially flat construction ($\mathcal{P}$-geometry) based on one coordinate chart (treated as affine coordinates of some flat connection). However, any affine manifold admits flat connections in spades. For instance, when $\mathcal{M}$ is a simply connected domain, each flat connection corresponds to a choice of global coordinates (modulo affine transformations). As such, there is no canonical choice of $\mathcal{P}$-geometry without further assumptions.

[5] We have made available a Mathematica notebook to calculate these quantities [36].

Based on the example of exponential families, one natural approach would be to try to obtain dually flat connections on $(\mathcal{M}, g)$ (i.e. try to find a Hessian structure for a given metric). In other words, we would try to find a Kähler metric on $T\mathcal{M}$. However, there are two problems with this approach.

(1) For $n > 2$, a generic Riemannian metric does not admit any dually flat structure at all. In fact, for $n \geq 4$, there are pointwise curvature obstructions to a Riemannian metric being Hessian [4].
(2) Even for metrics which admit a dually flat structure, this often is highly non-unique. For instance, Kito [38] showed that on the upper half plane, the set of Hessian pairs (∇, Φ) with hyperbolic metric admits at least one functional degree of freedom.

In this paper, we will not address the second issue (which we will discuss in future work), and focus only on the first issue.

As mentioned before, there are many possible $\mathcal{P}$-geometries that can be associated with any affine Riemannian manifold. From the previous discussion, we know it may not be possible to find a $\mathcal{P}$-geometry such that $\mathbb{M}$ is Kähler. One approach to reconcile this is to impose a condition on the complex manifold $\mathbb{M}$ which is less restrictive than Kähler, but more restrictive than J-integrability (i.e., $\mathbb{M}$ being Hermitian).

It turns out that there is a natural geometric condition that can be imposed, i.e., to require $\mathbb{M}$ to be *balanced*. A Hermitian manifold (with complex dimension n) is said to be balanced if it satisfies

$$d\Omega^{n-1} = 0.$$

For comparison, recall that a Kähler metric satisfies $d\Omega = 0$, so $\mathbb{M}$ being balanced is a weaker condition than $\mathbb{M}$ being Kähler. Balanced metrics need not be Kähler, though for complex surfaces the two notions are equivalent. Similar to the way that existence of a Kähler metric imposes topological restrictions on a manifold, so does the existence of a balanced metric [48]. While the balanced condition is just one possible relaxation of the Kähler condition; it is special in that considering $d\Omega^k = 0$ for any other k $(1 < k < n-1)$ will force $d\Omega = 0$ (i.e., the metric to be Kähler). Many alternative conditions have been studied, such as the Gauduchon condition $(\partial\bar{\partial}\Omega^{n-1} = 0)$ [27], the k-Gauduchon condition $(\partial\bar{\partial}\Omega^k \wedge \Omega^{n-1-k} = 0)$ [25] [35] and the Astheno-Kähler condition $(\partial\bar{\partial}\Omega^{n-2} = 0)$ [46]. However, balanced metrics are natural to consider in this context, for two reasons.

(1) The statistical manifolds of interest in this paper are modeled on open domains in Euclidean space. In this situation, flat connections correspond to a choice of coordinates (modulo affine transformations), which require specifying n coordinate functions.

    On the other hand, $d\Omega^{n-1}$ is a (real) $2n-1$ form (which has $2n$ components). However, for a partially flat structure in its affine coordinates, n of these components vanish automatically (Ω is constant in terms of the fiber directions v). As such, if we require $\mathbb{M}$ to be balanced, the flat connection must satisfy a further system of n equations, which is exactly the number of unknowns. As a result, the system of equations to induce a balanced metric is determined. As such, the balanced condition is natural from a partial differential equation point of view. We will provide a more detailed discussion with examples in a future paper.

    By comparison, for $n > 2$, the corresponding system of equations required for $\mathbb{M}$ to be Kähler is over-determined. In other words, the tangent bundle of a generic Riemannian manifold does not admit a Kähler metric, even locally. Similarly, the other special complex metric conditions mentioned above are either under-determined or over-determined.
(2) If we consider statistical mirror symmetry as being analogous to the mirror symmetry of Calabi-Yau's, balanced metrics are natural objects to investigate. In non-Kähler string theory, the Strominger system was introduced by Strominger as a non-Kähler Calabi-Yau manifold to model superstrings admitting

torsion [54,26]. This system provides an intricate coupling of various structures that is desirable from a physics perspective. It was shown by Li and Yau [44] that one of the equations in the Strominger system forces the complex metric to be conformally-balanced (i.e. it is conformal to a balanced metric). From an information geometry standpoint, conformal deformations of Hessian metrics have been studied in the context of "deformed exponential families" upon fixing a particular gauge [50]. This further motivates the investigation of statistical mirror symmetry with balanced Hermitian metrics.

5. Conclusions and discussions

In this paper, we propose and differentiate between two geometries for the manifold of parametric statistical model. $\mathcal{S}$-geometry is the now-standard statistical structure of parametric probability manifold. Its tangent manifold $T\mathcal{M}$ can take on dualistic $(\mathbb{W}, \mathbb{W}^*)$ structures, each corresponding to using one of the pairs of torsion-free connections ∇, ∇^* on the base $\mathcal{M}$. Though the two induced J's and induced G's are different, the induced Ω is the same regardless which connection on $\mathcal{M}$ is used — it is the g-pullback of the canonical symplectic form on $T^*\mathcal{M}$.

Extending to alpha-connections, each α-connection of $\mathcal{S}$-geometry induces a distinct almost Kähler structure on $T\mathcal{M}$. Because all α-connections are torsion-free, all of them share the one and same symplectic form Ω, which is the g-pullback canonical symplectic form. Hence, different α-connections induce α-dependent $T(T\mathcal{M})$-splitting, α-dependent $J^{(\alpha)}$, and α-dependent $G^{(\alpha)}$. The presence of curvature of the connections is the obstruction to integrability of $J^{(\alpha)}$ in the $\mathbb{W}$-structures of $\mathcal{S}$-geometry.

Starting with a flat connection on a parametric probability manifold, we can induce a $\mathcal{P}$-geometry using the partially flat construction, with one of the connections flat and the other connection torsion-coupled to g. $T\mathcal{M}$ can take on the form of $\mathbb{M}$ or $\mathbb{W}$ for those two connections on $\mathcal{M}$ — the flat connection ∇ turning $T\mathcal{M}$ into a Hermitian manifold $\mathbb{M}$, while torsion coupling of ∇^* with g turning $T\mathcal{M}$ into an almost Kähler manifold $\mathbb{W}$. The presence of torsion of the connections is the obstruction to d-closedness of Ω^∇ in the $\mathbb{M}$-structure and to J^{∇^*}-integrability in the $\mathbb{W}$-structure.

The term "statistical mirror symmetry" we coined in this paper is inspired by the work of Leung [43]. Our construction captures a correspondence between $\mathbb{M} \longleftrightarrow \mathbb{W}$, namely correspondence between a complex manifold and an almost Kähler manifold which are constructed from the pair of conjugate connections on the base manifold $\mathcal{M}$. It turns out that, when $\mathcal{M}$ is Hessian, the statistical mirror symmetry between two Kähler metrics on $T\mathcal{M}$ is a geometric consequence of Reference-Representation Biduality (RRB) [64] of canonical divergence on $\mathcal{M}$ built upon the Legendre duality.

There are some important differences between statistical mirror symmetry and the mirror symmetry of Calabi-Yau manifolds. First, many statistical manifolds of interest are neither compact nor complete. It turns out that in the Kähler case, $\mathbb{M}$ and $\mathbb{W}$ are complete as metric spaces if and only if $\mathcal{M}$ is (for a proof of this fact, see [49]). Second, there are no Ricci-flatness conditions imposed on this statistical mirror pair of manifolds. As a result, neither $\mathbb{M}$ not $\mathbb{W}$ should be considered as Calabi-Yau manifolds. This also means that general cohomological tools are more difficult to use and the geometric considerations thus far are local in nature.

5.1. Global obstructions

All the calculations in this paper are local in nature, and we have not discussed global existence issues at all. With respect to $\mathcal{S}$-geometry, since a Riemannian manifold is a special kind of statistical structure (in which the torsion-free Levi-Civita connection is self-conjugate), there are no geometric obstructions for a Riemannian manifold to be considered as a statistical manifold (and hence admits $\mathcal{S}$-geometry). On the other hand, with respect to $\mathcal{P}$-geometry, for a manifold to *globally* admit a partially flat structure requires the manifold to be *affine*. A manifold is affine if it admits an atlas of coordinate charts whose transition maps are affine functions. It is non-trivial to find compact examples of affine manifolds other than tori and there are topological obstructions. For instance, the fundamental group of a compact affine manifold is infinite [8]. To study the global geometry of locally flat spaces, a cohomology known as KV cohomology has been investigated [10].

A manifold is parallelizable when its tangent bundle is trivial; only then the manifold can admit a global pseudo-Weitzenböck connection. As such, there are topological obstruction to the existence of an affine atlas on the primal connection side and a global frame on the conjugate connection side. In practice, parametric families of probability distributions are defined over open (often convex) domains in Euclidean space upon having chosen of parameters, and serve as a global affine chart for a flat connection. Furthermore, by choosing a locally strongly convex function on such a domain, we can find a Hessian metric. There are examples of statistical manifolds with non-trivial topology. For example, the von Mises-Fisher distribution with fixed concentration rate has parameter space $\mathbb{S}^{n-1}$. For $n > 2$, this is not an affine manifold so it admits no partially flat geometry. For $n \neq 2, 4, 8$, it is not parallelizable so it admits no global frame. It is not possible to construct $\mathcal{P}$-geometry in this case.

The existence of a dually flat structure (i.e. a Hessian metric) is even more restrictive. Clearly, it is necessary for such a manifold to be affine, but this is not sufficient. Yagi [59] found a rigidity theorem for Hessian metrics on compact affine manifolds under a certain cohomological assumption. Using this, he constructed an affine manifold which admits no Hessian metric. The local existence question of Hessian metrics was studied by [4] and by [11] independently. Their work shows that for $n \geq 3$, generically a Riemannian metric admits no dually flat structure and that for $n \geq 4$, there are pointwise curvature obstructions. However, a positive result exists for $n = 2$, in that any real analytic Riemannian metric locally admits such a structure.

Acknowledgement

The project is supported by DARPA/ARO Grant W911NF-16-1-0383 ("Information Geometry: Geometrization of Science of Information", PI: Zhang). The writing of this manuscript is also supported by AFOSR Grant FA9550-19-1-0213 ("Brain-Inspired Networks for Multifunctional Intelligent Systems and Aerial Vehicles", MURI subcontract from UCLA).

References

[1] Dmitri V. Alekseevsky, Vicente Cortés, Geometric construction of the r-map: from affine special real to special Kähler manifolds, Commun. Math. Phys. 291 (2) (2009) 579–590.
[2] Shun-ichi Amari, Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics, Springer, New York, 1985.
[3] Shun-Ichi Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (2) (1998) 251–276.
[4] Shun-ichi Amari, John Armstrong, Curvature of Hessian manifolds, Differ. Geom. Appl. 33 (2014) 1–12.
[5] Shun-ichi Amari, Hiroshi Nagaoka, Methods of Information Geometry, vol. 191, American Mathematical Soc., 2007.
[6] Nihat Ay, Shun-ichi Amari, A novel approach to canonical divergences within information geometry, Entropy 17 (12) (2015) 8111–8129.
[7] Nihat Ay, Jürgen Jost, Hông Vân Lê, Lorenz Schwachhöfer, Information geometry and sufficient statistics, Probab. Theory Relat. Fields 162 (1–2) (2015) 327–364.
[8] Nihat Ay, Wilderich Tuschmann, Dually flat manifolds and global information geometry, Open Syst. Inf. Dyn. 9 (2) (2002) 195–200.
[9] Rolf Berndt, Sur l'arithmétique du corps des fonctions elliptiques de niveau N, in: Seminar on Number Theory, Paris 1982–83, vol. 51, 1984, pp. 21–32.
[10] Michel Nguiffo Boyom, Paul Mirabeau Byande, KV cohomology in information geometry, in: Matrix Information Geometry, Springer, 2013, pp. 69–92.
[11] Robert Bryant, http://mathoverflow.net/questions/122308/, 2013.
[12] Philip Candelas, Gary T. Horowitz, Andrew Strominger, Edward Witten, Vacuum configurations for superstrings, in: Current Physics–Sources and Comments, vol. 1, Elsevier, 1988, pp. 49–77.
[13] N.N. Censov, Statistical decision rules and optimal inferences, Transl. Math. Monogr. 53 (1982) 117–120.
[14] Kwokwai Chan, Naichung Conan Leung, et al., On SYZ mirror transformations, in: New Developments in Algebraic Geometry, Integrable Systems and Mirror Symmetry, RIMS, Kyoto, 2008, Mathematical Society of Japan, 2010, pp. 1–30.

[15] Shiu-Yuen Cheng, Shing-Tung Yau, On the regularity of the Monge-Ampère equation det $\frac{\partial^2 u}{\partial x^i \partial x^j} = f(x, u)$, Commun. Pure Appl. Math. 30 (1) (1977) 41–68.

[16] Tristan C. Collins, Adam Jacob, Shing-Tung Yau, Poisson metrics on flat vector bundles over non-compact curves, arXiv preprint, arXiv:1403.7825, 2014.

[17] Philippe Delanoë, Remarques sur les variétés localement hessiennes, Osaka J. Math. 26 (1) (1989) 65–69.

[18] Robbert Dijkgraaf, Mirror symmetry and elliptic curves, in: The Moduli Space of Curves, Springer, 1995, pp. 149–163.

[19] Peter Dombrowski, On the geometry of the tangent bündle, J. Reine Angew. Math. 210 (1962) 73–88.

[20] Shinto Eguchi, Second order efficiency of minimum contrast estimators in a curved exponential family, Ann. Stat. (1983) 793–803.

[21] Shinto Eguchi, et al., Geometry of minimum contrast, Hiroshima Math. J. 22 (3) (1992) 631–647.

[22] Teng Fei, Sebastien Picard, Anomaly flow and T-duality, arXiv preprint, arXiv:1903.08768, 2019.

[23] Teng Fei, Jun Zhang, Interaction of Codazzi couplings with (para-) Kähler geometry, Results Math. 72 (4) (2017) 2037–2056.

[24] Daniel S. Freed, Special Kähler manifolds, Commun. Math. Phys. 203 (1) (1999) 31–52.

[25] Jixiang Fu, Zhizhang Wang, Damin Wu, Semilinear equations, the γ_k function, and generalized Gauduchon metrics, arXiv preprint, arXiv:1010.2013, 2010.

[26] Mario Garcia-Fernandez, Lectures on the Strominger system, arXiv preprint, arXiv:1609.02615, 2016.

[27] Paul Gauduchon, La 1-forme de torsion d'une variété hermitienne compacte, Math. Ann. 267 (4) (1984) 495–518.

[28] Brian R. Greene, M. Ronen Plesser, Duality in Calabi-Yau moduli space, Nucl. Phys. B 338 (1) (1990) 15–37.

[29] Sergey Grigorian, Jun Zhang, (Para-)holomorphic and conjugate connections on (para-)Hermitian and (para-)Kähler manifolds, Results Math. 74 (4) (2019) 150.

[30] Mark Gross, The Strominger-Yau-Zaslow Conjecture: From Torus Fibrations to Degenerations, 2009.

[31] Masayuki Henmi, Hiroshi Matsuzoe, Geometry of pre-contrast functions and non-conservative estimating functions, in: AIP Conference Proceedings, vol. 1340, AIP, 2011, pp. 32–41.

[32] Masayuki Henmi, Hiroshi Matsuzoe, Statistical manifolds admitting torsion and partially flat spaces, in: Geometric Structures of Information, Springer, Cham, 2019, pp. 37–50.

[33] Lars Hörmander, An Introduction to Complex Analysis in Several Variables, North-Holland Mathematical Library, North-Holland, 1990.

[34] Harold Jeffreys, An invariant form for the prior probability in estimation problems, Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci. 186 (1007) (1946) 453–461.

[35] Gabriel Khan, On the Hermitian geometry of k-Gauduchon orthogonal complex structures, arXiv preprint, arXiv:1811.01037, 2018.

[36] Gabriel Khan, The Curvature of the Sasaki Metric, Mathematica Notebook, 2019, https://www.wolframcloud.com/obj/gabekhan/Published/Curvature_of_Sasaki_metric.nb.

[37] Gabriel Khan, Jun Zhang, The Kähler geometry of certain optimal transport problems, Pure Appl. Anal. 2 (2) (2020) 397–426.

[38] Hideyuki Kito, On Hessian structures on the Euclidean space and the hyperbolic space, Osaka J. Math. 36 (1) (1999) 51–62.

[39] Takashi Kurose, Dual connections and affine geometry, Math. Z. 203 (1) (1990) 115–121.

[40] Takashi Kurose, Statistical Manifolds Admitting Torsion. Geometry and Something, Fukuoka University, 2007.

[41] Siu-Cheong Lau, Li-Sheng Tseng, Shing-Tung Yau, Non-Kähler SYZ mirror symmetry, Commun. Math. Phys. 340 (1) (2015) 145–170.

[42] Stefan L. Lauritzen, Statistical manifolds, Differ. Geom. Stat. Inference 10 (1987) 163–216.

[43] Naichung Conan Leung, Mirror symmetry without corrections, Commun. Anal. Geom. 13 (2) (2005) 287–331.

[44] Jun Li, Shing-Tung Yau, The existence of supersymmetric string theory with torsion, J. Differ. Geom. 70 (1) (2005) 143–181.

[45] John C. Loftin, Affine spheres and Kähler-Einstein metrics, Math. Res. Lett. 9 (4) (2002) 425–432.

[46] Koji Matsuo, Takao Takahashi, On compact astheno-Kähler manifolds, in: Colloquium Mathematicum, vol. 2, 2001, pp. 213–221.

[47] Hiroshi Matsuzoe, Geometry of statistical manifolds and its generalization, in: Topics in Contemporary Differential Geometry, Complex Analysis and Mathematical Physics, World Scientific, 2007, pp. 244–251.

[48] Marie-Louise Michelsohn, On the existence of special metrics in complex geometry, Acta Math. 149 (1) (1982) 261–295.

[49] Mathieu Molitor, Gaussian distributions, Jacobi group, and Siegel-Jacobi space, J. Math. Phys. 55 (12) (2014) 122102.

[50] Jan Naudts, Jun Zhang, Rho–tau embedding and gauge freedom in information geometry, Inf. Geom. 1 (1) (2018) 79–115.

[51] August Newlander, Louis Nirenberg, Complex analytic coordinates in almost complex manifolds, Ann. Math. (1957) 391–404.

[52] Hiroyasu Satoh, Almost Hermitian structures on tangent bundles, in: Proceedings of the Eleventh International Workshop on Differential Geometry, Kyungpook Nat. Univ., Taegu 11 (2007) 105–118, arXiv preprint, arXiv:1908.10824.

[53] Udo Simon, Angela Schwenk-Schellschmidt, Helmut Viesel, Introduction to the Affine Differential Geometry of Hypersurfaces, Science University of Tokyo, 1991.

[54] Andrew Strominger, Superstrings with torsion, Nucl. Phys. B 274 (2) (1986) 253–284.

[55] Andrew Strominger, Shing-Tung Yau, Eric Zaslow, Mirror symmetry is T-duality, Nucl. Phys. B 479 (1–2) (1996) 243–259.

[56] Shun-ichi Tachibana, Masafumi Okumura, On the almost-complex structure of tangent bundles of Riemannian spaces, Tohoku Math. J. (2) 14 (2) (1962) 156–161.

[57] James Tao, Jun Zhang, Transformations and coupling relations for affine connections, Differ. Geom. Appl. 49 (2016) 111–130.

[58] Hông Vân Lê, Statistical manifolds are statistical models, J. Geom. 84 (1–2) (2006) 83–93.

[59] Katsumi Yagi, On Hessian structures on an affine manifold, in: Manifolds and Lie Groups, Springer, 1981, pp. 449–459.
[60] Jae-Hyun Yang, Geometry and arithmetic on the Siegel–Jacobi space, in: Geometry and Analysis on Manifolds, Springer, 2015, pp. 275–325.
[61] Paul C. Yang, Automorphism of tube domains, Am. J. Math. 104 (5) (1982) 1005–1024.
[62] Shing-Tung Yau, On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation, I, Commun. Pure Appl. Math. 31 (3) (1978) 339–411.
[63] Jun Zhang, Divergence function, duality, and convex analysis, Neural Comput. 16 (1) (2004) 159–195.
[64] Jun Zhang, Dual scaling of comparison and reference stimuli in multi-dimensional psychological space, J. Math. Psychol. 48 (6) (2004) 409–424.
[65] Jun Zhang, Gabriel Khan, From Hessian to Weitzenböck: manifolds with torsion-carrying connections, Inf. Geom. 2 (1) (Jun 2019) 77–98.
[66] Jun Zhang, Gabriel Khan, New geometry of parametric statistical models, in: International Conference on Geometric Science of Information, GSI'2019, Springer, 2019, pp. 288–296.
[67] Jun Zhang, Gabriel Khan, Connections with torsion in (para-)complexified structures, in: Geometric Structure of Information, Springer, 2020 (to appear).