Ode to Legendre: Geometric and thermodynamic implications on curved statistical manifolds

Pablo A. Morales,† Jan Korbel,‡,§ and Fernando E. Rosas¶,†,

†Research Division, Araya Inc., Tokyo 107-6019, Japan
‡Section for Science of Complex Systems, CeMSIS, Medical University of Vienna, Spitalgasse, 23, 1090 Vienna, Austria
§Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria
¶Data Science Institute, Imperial College London, London SW7 2AZ, UK
¶Centre for Psychedelic Research, Department of Brain Science, Imperial College London, London SW7 2DD, UK

The recent link discovered between generalized Legendre transforms and curved (i.e. non-Euclidean) statistical manifolds suggests a fundamental reason behind the ubiquity of Rényi’s divergence and entropy in a wide range of physical phenomena. However, these early findings still provide little intuition on the nature of this relationship and its implications for physical systems. Here we shed new light on the Legendre transform by revealing the consequences of its deformation via symplectic geometry, complexification, and stochastic thermodynamics. These findings reveal a novel common framework that leads to a principled and unified understanding of physical systems that are not well-described by classic information-theoretic quantities.

I. INTRODUCTION

The Legendre transform plays a key — albeit perhaps not always transparent — role in many areas of mathematical physics. Specifically, it allows to identify dual coordinates and potentials that yield theories in terms of more convenient variables, being instrumental in the study of diverse problems in physics ranging from relativistic field theory to condensed matter physics. Applications of the transform have their roots on classical physics — in analytical mechanics serving as a link between its Lagrangian and Hamiltonian formulations, and in thermodynamics bridging intensive and extensive variables. These notions have led to more general frameworks which, in turn, gave rise to the development of symplectic topology [1].

Far from being a mere relic, the Legendre transform still plays an important role in contemporary physics. It is used in classical field theory where the index of pairs of components becomes continuous. Furthermore, in quantum field theory it relates the generator of connected Green functions to the quantum effective action — i.e. the generator of one-particle irreducible Green functions. The relevance of the Legendre transform has lead to generalizations in the context of perturbative quantum field theories [2, 3]. Hence, the transform continues to be at the core of important developments in current research.

The Legendre transform also plays a fundamental role in information geometry, where it mediates the relationship between primal and dual coordinates within the non-Riemannian geometry induced by dually-flat statistical manifolds [4]. These primal and dual coordinates establish orthogonality in these geometries, and correspond to alternative representations of physical systems based either on control parameters or expectation values of observables [5]. Interestingly, recent work has revealed how a generalized Legendre transform naturally arises in curved (i.e. non-Euclidean) statistical manifolds, which are directly linked with Rényi’s divergence and entropy [6, 7]. These findings suggest the existence of a fundamental reason that could explain why Rényi entropy and divergence naturally appears in a range of physical phenomena of interest. In effect, recent applications of Rényi to physics includes quantum systems [8, 9], strongly coupled or entangled systems [10–12], phase transitions [13–15], and multifractal thermodynamics [16, 17], among others. However, these early findings on the link between the generalized Legendre transforms and curved geometries still provide little intuition on the nature of this relationship, and its meaning and implications for physical systems in general.

The goal of this article is to shed new light on the generalized Legendre transform by developing its geometric and thermodynamic implications. For this purpose, we introduce the notion of link functions as a mean to characterize deformations in the Legendre transform, which in turn lead to generalizations of the Bregman divergence that are naturally associated with curved statistical manifolds. By building on these tools, our contribution focuses on two domains: geometrical aspects related to phase-space flow and manifold complexification, and thermodynamic aspects related to statistical mechanics. Take together, these results lead to a larger, unified picture that extend standard geometric and thermodynamic relationships associated with classic information-theoretic quantities such as Shannon’s entropy.

In the domain of geometry, our results establish a new perspective of the generalized Legendre transform by
studying its induced symplectic form and flow. In particular, we show how the symplectic structure induced by link functions related to the Rényi divergence leads to a modification of what is understood as a ‘canonical pair,’ which in turn leads to a deeper insight into the corresponding maximum entropy distributions. Furthermore, our results bring new insights to the relationship between the Kullback-Leiver divergence (related to the Shannon entropy), α-divergence (related to ‘Tsallis’ entropy), and the Rényi divergence via manifold complexification and Kalher manifolds. The complex geometry yields new conditions on the possible values of the manifold curvature, which are closely related to holomorphic polarization.

Additionally, our results show that the generalized Legendre transform naturally describes the thermodynamics of systems that obey non-linear constraints at the macroscopic level — i.e., at the level of probability distributions that reflect equilibrium conditions. The generalized Legendre structure on the macroscopic levels is found to play the same role relating thermodynamic quantities such as energy, temperature, and volume on such systems as the ordinary Legendre structure does on systems under linear constraints. Furthermore, our results show that systems under non-linear constraints also satisfy a number of standard thermodynamic results, as long as these are properly re-framed in terms of the Rényi entropy. In particular, these systems still minimize free energy, obey the second law of thermodynamics, and even satisfy a detailed fluctuation theorem at a stochastic level.

The rest of the paper is structured as follows. First, Section II provides a brief overview on the standard interpretation of the Legendre transform in mathematical physics. Then, Section III explores how the transform naturally arises in information geometry, and presents the intimate relationship that exists between a generalized Legendre transform and the curvature of statistical manifolds. Building on these foundations, Section IV investigates the consequences of generalized Legendre transforms on the symplectic structure and flows, and on the complexification of statistical manifolds. Then, Section V develops how generalized Legendre transforms lead to descriptions of the thermodynamics of systems under non-linear constraints. Finally, Section VI summarises our mains conclusions.

II. PRELIMINARIES

The Legendre transform is, at its core, an exploration of the properties of convex functions. Despite of its importance, the transform is unfortunately typically introduced as an obscure algebraic ‘trick,’ with nothing that can explain why it plays such an important role in many quite different areas of physics. For completion this section presents a basic standard interpretations of the Legendre transform in mathematical physics, which is then complemented by a perhaps deeper view based on information geometry in Section III.

The most straightforward interpretation of the Legendre transform comes from simple geometry [18]. In this view, the Legendre transform of a convex function $F(x)$ is another function $G(x)$ that corresponds to the (negative) height at which the tangent to $F(x)$ touches the x-axis — which is usually re-parametrized in terms of the slope of $F(x)$. This view is easy to grasp, but unfortunately makes the construction to seem arbitrary while failing to explain why this procedure is so fundamental.

A more principled view comes from an algebraic perspective as follows. If $F(x)$ with $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a strictly convex function (i.e. its Hessian is positive-definite), then the partial derivative $y_i(x) := \partial F/\partial x_i(x)$ is a monotonic function of $x_1, \ldots, x_n$ for $i = 1 \ldots n$. This means that there exists an isomorphism between $x$ and $y = (y_1, \ldots, y_n)$ — said differently, there exist functions $y_i(x)$ and $x_k(y)$ that transform one into the other. Using these, it would be natural to consider the possibility of “reparametrising” $F$ in terms of $y$ instead of $x$. However, instead of focusing on such reparametrization, an elegant move is to consider instead the function $G(y) = x \cdot y - F(x(y), y)$. Interestingly, the resulting pair $F(x)$ and $G(y)$ exhibit the following symmetry:

$$\frac{\partial G}{\partial y_k} = x_k, \quad \frac{\partial F}{\partial x_i} = y_i. \quad (1)$$

Useful properties of this transformation are that it preserves convexity (i.e. the transform of a convex function results into another convex function) and it is an ‘involution’ — so that the Legendre transform of the transform of a convex function is the function itself. The symmetry of these relationships is graphically represented in the right-hand-side of Figure 1.

Overall, one can think of the Legendre transform as acting on two inputs, $x$ and $F$, and providing two outputs: the dual variable $y$ and the convex conjugate $G$ [19]. Pairs of convex functions $\{F, G\}$ satisfying Eq. (1) are known as convex duals, with $(x, y)$ being known as dual variables. Additionally, convex functions and their duals satisfy the Fenchel inequality $F(x) + G(y) \geq x \cdot y$. The multiple useful properties of Legendre duals are leveraged in various areas of mathematics and engineering, particularly in convex optimization [20].

A more general definition of the Legendre transform of a convex function is given by

$$G(y) = \sup_x \{C(x, y) - F(x)\}. \quad (2)$$

This definition applies even when $F$ is not everywhere differentiable, and can be shown to recover the above procedure for the particular case where $C(x, y) = x \cdot y$. For other choices of $C$ this opens the door to so-called “deformed” Legendre transforms. Interestingly, dual functions according to these generalized Legendre transforms satisfy relationships analogous to Eq. (1), but where the role of the Euclidean gradient is replaced by a ‘Legendre derivative’ operator, which is denoted by $D_L$ and defined in Section III D. The goal of this paper is to explore the
relationship between deformed Legendre transforms and non-Euclidean geometry. Please note that the relationships in Figure 1 still hold for deformed Legendre transforms if the gradient is deformed as well, as explained in the following sections.

III. LEGENDRE TRANSFORM IN INFORMATION GEOMETRY

In this section we present the key role of the Legendre transform in statistical manifolds. For this purpose, Section III A first introduces necessary background about information geometry to the unfamiliar reader. Then, Section III B explains how the standard Legendre transform naturally describe the geometry of dually-flat spaces, which are associated with the Kullback-Leiver divergence and the Shannon entropy. Building on this, Section III C then presents how other divergences lead to more general geometries, and Section III D develops how generalized Legendre transforms are a natural way to build and describe curved manifolds. Please note that hereafter we use Einstein’s summation convention for convenience of the notation.

A. The dual structure of statistical manifolds

Our exposition is focused on statistical manifolds \( \mathcal{M} \), whose elements are probability distributions \( p_\xi(s) \) with \( s \in \mathbf{S} \) and \( \xi \in \mathbb{R}^d \). The geometry of such statistical manifolds is determined by two structures: a metric tensor \( g \), and a torsion-free affine connection pair \( (\nabla, \nabla^* ) \) that are dual with respect to \( g \). Intuitively, \( g \) establishes norms and angles between tangent vectors and, in turn, establishes curve length and the shortest curves.

On the other hand, the affine connection establishes covariant derivatives of vector fields establishing the notion of parallel transportation between neighbouring tangent spaces, which defines what is a straight curve.

Traditional Riemannian geometry is built on the assumption that the shortest and the straightest curves coincide — which is pivotal to the development of general relativity. This assumption leads to the study of metric-compatible Levi-Civita connections that satisfy \( \nabla = \nabla^* \), which can be directly derived from the metric. However, modern approaches motivated in information geometry [21] and gravitational theories [22, 23] consider more general scenarios, where connections may not be derivable from the metric. In such geometries, the parallel transport operator \( \Pi : T_p \mathcal{M} \rightarrow T_q \mathcal{M} \) and its dual \( \Pi^* \) [24] (induced by \( \nabla \) and \( \nabla^* \), respectively) might differ. The departure of \( \nabla \) (and \( \nabla^* \)) from self-duality can be shown to be proportional to Chentsov’s tensor, which allows for a single degree of freedom traditionally denoted by \( \alpha \in \mathbb{R} \) [21]. Put simply, \( \alpha \) captures the degree of asymmetry between short and straight curves, with \( \alpha = 0 \) corresponding to metric-compatible connections where \( \nabla = \nabla^* \).

An important property of the geometry of a statistical manifold \( (\mathcal{M}, g, \nabla, \nabla^* ) \) is its curvature, which can be of two types: the (Riemann-Christoffel) metric curvature, or the curvature associated to the connection. Both quantities capture the distortion induced by parallel transport over closed curves — the former with respect to the Levi-Civita connection, and the latter with respect to \( \nabla \) and \( \nabla^* \). In the sequel we use the term “curvature” to refer exclusively to the latter type. Statistical manifolds with zero curvature (equivalently, with connections that satisfy \( \Gamma_{ijk}(\xi) = \Gamma^*_{ijk}(\xi) = 0 \)) are said to be Euclidean, or more simply, flat.
B. Dually flat geometry, Bregman divergences, and the Legendre transform

The geometry of Riemannian manifolds can be formulated in terms of a single set of local coordinates. However, the nature of statistical manifolds suggest to consider them being non-Riemannian — i.e. having two dissimilar affine connections $\nabla$ and $\nabla^*$, which in turn are associated to dual coordinates $\xi$ and $\eta$ [21]. Specifically, while in Riemannian geometry orthogonality can be assessed between dimensions of a single set of coordinates, in statistical manifolds it is more fruitful to consider orthogonality between elements of the primal $\xi$ and dual coordinates $\eta$ [5, 7]. A standard example of this is statistical manifolds where $\xi$ correspond to the natural parameters of an exponential family distribution and $\eta$ correspond to the corresponding expectation values. In the sequel, we follow Schouten’s notation in which upper indices are reserved for dual coordinates, i.e.

$$\partial_i = \frac{\partial}{\partial \xi^i} \quad \text{and} \quad \partial^i = \frac{\partial}{\partial \eta^i}. \quad (3)$$

Under this notation, $\partial_i$ gives rise to a basis for the tangent space $T_p \mathcal{M}$ while $\partial^i$ is related to a natural dual basis of the cotangent space $T^*_p \mathcal{M}$.

A Riemannian metric is always “locally flat” — i.e., it can be brought down to its signature (a Kronecker delta) at a given point $p \in \mathcal{M}$ by choosing an appropriate coordinate chart. It is not guaranteed, however, that such a chart would preserve the delta at a neighborhood of $p$; finding a chart that satisfies this property globally is the hallmark of a flat geometry. Analogously, affine geometries are also locally flat when considering its dual entry, therefore satisfying $g(\partial_i, \partial^j) = \delta^j_i$ for an appropriate pair of primal and dual coordinate charts $\{\xi^i, \eta_i\}$ at some point $p$. In a similar fashion, this property in general only holds locally; dually flat geometries are characterized by the fact that one can find a pair of coordinates that satisfy this condition of orthogonality on the whole manifold [25].

For an orthogonal pair $\{\xi, \eta\}$ of a given dually flat manifold, the gradients of the mappings $\xi \mapsto \eta$ and $\eta \mapsto \xi$ are both symmetric. To confirm this, let’s first note that

$$g_{ij} = g(\partial_i \eta_k \partial^k, \partial_j) = \partial_i \eta_k g(\partial^k, \partial_j) = \partial_i \eta_k \delta^k_j = \partial_i \eta_j,$$

where the first equality follows from the chain rule of derivatives $\partial_k = \partial_i \eta_k \partial^k$. Then, using the fact that Riemannian metrics are always symmetric, one can see that $\partial_i \eta_j = g_{ij} = g_{ji} = \partial_j \eta_i$. A similar derivation shows that $g^{ij} = \partial^i \xi^j$, and hence $\partial^i \xi^j = \partial^j \xi^i$ [26].

There is an intimate relationship between an orthogonal pair of coordinates in a dually flat manifold and the Legendre transform. To see this, we first note that the symmetry of the Jacobian of $\xi \mapsto \eta$ implies the existence of a closed 1-form $d\omega = 0$, and this — via Poincare Lemma — implies in turn the existence of scalar potential $\psi \in C^\infty$ that satisfies

$$\eta_i = \partial_i \psi \quad \text{and} \quad g_{ij} = \partial_i \partial_j \psi. \quad (4)$$

Note that the second condition, combined that the fact that $g_{ij}$ is positive-semidefinite, implies that $\psi$ is convex. By a similar line of reasoning, the symmetry of $g^{ij}$ induces a dual convex potential $\varphi$ that satisfies

$$\xi^i = \partial^i \varphi \quad \text{and} \quad g^{ij} = \partial^i \partial^j \varphi. \quad (5)$$

Furthermore, a direct calculation shows that the dual potentials $\psi(\xi^1, ..., \xi^n)$ and $\varphi(\eta_1, ..., \eta_n)$ always satisfies $d\{\psi + \varphi - \xi^i \eta_i\} = 0$. This implies that, modulo an unimportant constant, the following relationship holds over any dually flat manifold [27]:

$$\psi + \varphi - \xi^i \eta_i = 0. \quad (6)$$

Let us now consider the behaviour of Eq. (6) on dually flat spaces when the coordinates and potentials are evaluated at different points of the manifold. For this, let us denote as $\xi(p)$ and $\eta(q)$ the coordinates and dual coordinates of $p, q \in \mathcal{M}$, respectively, and define the so-called Bregman divergence $D$ as

$$D(p|q) := \varphi(\eta(p)) + \psi(\xi(q)) - \xi^i(\eta_i(p)). \quad (7)$$

Then, the differential of the mapping $q \mapsto D(p_0|q)$ is

$$d\{D(p_0, q)\} = \left(\partial_i \psi(\xi(q)) - \eta_i(p_0)\right) d\xi^i(q) = \left(\eta_i(q) - \eta_i(p_0)\right) d\xi^i(q). \quad (8)$$

From this, and considering that $D$ by definition is a difference between a linear and two convex functions, one can verify that this mapping attains its unique minimum when $q = p_0$. Interestingly, at this minimal value one recovers Eq. (6) which implies that $D = 0$. This shows that Bregman divergences are non-negative.

These results suggest an alternative definition for $\varphi$ and $\psi$, conceiving them as a maximum of the following maps:

$$\varphi(\eta(p)) = \max_{q \in \mathcal{M}} \{\xi^i(q) \eta_i(p) - \psi(\xi(q))\}, \quad (9)$$

$$\psi(\xi(p)) = \max_{q \in \mathcal{M}} \{\eta_i(q) \xi^i(p) - \varphi(\eta(q))\}. \quad (10)$$

This reveal that the orthogonal coordinate pair is always dual in the Legendre sense — or equivalently, that dual flatness implies that the potential are convex duals. This property generalizes the well-known Legendre duality between the natural and expectation parameters of an exponential family [28], showing that the same holds to any coordinate pair as long as they satisfy local flatness.

C. Divergences as a general tool to establish geometries

This subsection explains how divergences, such as the one introduced in Eq. (7), can be used as a convenient tool to establish a geometry on a statistical manifold [29],
Divergences are a general class of functions that assess the dissimilarity of their arguments. More specifically, a divergence is a smooth, distance-like function that satisfy $\mathcal{D}[x; x'] \geq 0$ and vanish only when $x = x'$. Divergences are more general — hence weaker — notions than distances, as they don’t need to be symmetric in their arguments and may not respect the triangle inequality. Of the various types of divergences explored in the literature (see [32] and references within), two are particularly important: $f$-divergences (which are monotonic with respect to coarse-grainings of the domain of events $S$ [33]), and Bregman divergences (studied in previous section).

Let us show how divergences can be used to establish metrics and connections over manifolds. For this, let us use the shorthand notation $\mathcal{D}[\xi; \xi'] := \mathcal{D}(p||q)$ when expressing $\mathcal{D}$ in terms of coordinates $\xi = \xi(p)$ and $\xi' = \xi(q)$. Then, the Riemannian metric of the manifold is recovered from the second-order expansion of the divergence as follows:

$$g_{ij}(\xi) = \langle \partial_i, \partial_j \rangle = -\partial_i \partial_j \mathcal{D}[\xi; \xi']|_{\xi = \xi'} ,$$

which is positive-definite due to the non-negativity of $\mathcal{D}$. This construction leads to the Fisher’s metric, which is the unique metric that emerges from a broad class of divergences [29, Th. 5], with this being this closely related Chentsov’s theorem [34–37]. Similarly, connections emerge at the third-order expansion of the divergence as follows:

$$\Gamma_{ijk}(\xi) = \langle \nabla_{\partial_i} \partial_j, \partial_k \rangle = -\partial_i \partial_j \partial_k \mathcal{D}[\xi; \xi']|_{\xi = \xi'} ,$$

(12a)

$$\Gamma^\ast_{ijk}(\xi) = \langle \nabla^\ast_{\partial_i} \partial_j, \partial_k \rangle = -\partial_i \partial_j \partial_k \mathcal{D}[\xi; \xi']|_{\xi = \xi'} .$$

(12b)

In summary, Fisher’s metric is insensible the choice of divergence but the resulting connections are, and therefore the effects of a particular $\mathcal{D}$ manifest only at third-order.

Bregman divergences always give rise to flat geometries (as for them $\partial_i \partial_j \partial_k \mathcal{D}[\xi; \xi'] = \partial_k \partial_i \partial_j \mathcal{D}[\xi; \xi'] = 0$), and therefore other type of divergences are needed in order to establish curved non-Riemannian geometries. As mentioned in Section III A, the deviation of a given connection $\nabla$ from its corresponding metric-compatible (i.e. Levi-Civita) counterpart can be measured by $\alpha T$, where $T$ corresponds to the invariant Amari-Chentsov tensor [38, 39] and $\alpha \in \mathbb{R}$ is a free parameter. The invariance of $T$ implies that the value of $\alpha$ entirely determines the connection, and the corresponding geometry can be obtained from a divergence of the form

$$\mathcal{D}_\alpha(p||q) = \frac{4}{1-\alpha^2} \int_S \left(1 - p^{\frac{1-\alpha}{2}}(s)q^{\frac{1+\alpha}{2}}(s)\right) \, d\mu(s) ,$$

(13)

which is known as $\alpha$-divergence. As important particular cases, if $\alpha = 0$ then $\mathcal{D}_\alpha$ becomes the square of Hellinger’s distance, and if $\alpha = \pm 1$ then it gives the well-known Kullback-Leibler divergence. Furthermore, it can be shown that the Kullback-Leiver divergence is a Bregman divergence, which in turn implies that for those cases the resulting geometry is flat. This illustrates the fact that being Riemmanian (i.e. $\alpha = 0$) and Euclidean ($\alpha = \pm 1$) are independent features of a geometry.

We finish this subsection by noting that multiple divergences can give rise to the same geometry. A 1-1 relationship between divergence and geometries is obtained when considering conformal-projective equivalent classes of divergences, which are related both via conformal and projective transformations. For a more detailed explanation, we refer the interested reader to Ref. [7, Sec. 2-D].

**D. Generalized Legendre transforms as a natural way to describe curved manifolds**

Sections III B and III C clarified the 1-1 relationship that exists between dually flat manifolds, Bregman divergences, and the Legendre transform. Here we explain how these relationships are modified in curved manifolds.

In curved geometries it is impossible to construct dual potentials that satisfy Eq. (6) on the whole manifold. This impossibility is a symptom of the fact that the divergences that gives rise to this geometry — e.g. the $\alpha$-divergence given in Eq. (13) — is not a Bregman divergence, but only an $f$-divergence [33].

To better understand the nature of the $\alpha$-divergence, let us consider in detail its relationship with Bregman divergences. Bregman divergences, as given in Eq. (7), can also be expressed as

$$\mathcal{D}_\Phi[\xi; \xi'] = \Phi(\xi') - \Phi(\xi) - D\Phi(\xi) \cdot (\xi' - \xi).$$

(14)

Hence, $\mathcal{D}_\Phi[\xi; \xi']$ measures how convex the function $\Phi$ is at $\xi$ in the direction of $\xi' - \xi$ [40], and exploits the fact that a first-order approximation of a convex function always underestimates its value — i.e., that $\Phi(\xi') \geq \Phi(\xi) + D\Phi(\xi) \cdot (\xi' - \xi)$, where $D$ is the (Euclidean) gradient. Interestingly, such first-order approximation can also be built on an intermediate point between $\xi$ and $\xi'$, which leads to

$$\frac{1 - \alpha}{2} \Phi(\xi) + \frac{1 + \alpha}{2} \Phi(\xi') \geq \Phi(\xi_\alpha) ,$$

(15)

where $\xi_\alpha = \frac{1-\alpha}{2} \xi + \frac{1+\alpha}{2} \xi'$, with $\alpha \in (-1, 1)$ being a one-dimensional parameter that regulates how close $\xi_\alpha$ is to $\xi$ and $\xi'$. This inequality leads to a family of divergences indexed by $\alpha$, given by

$$\mathcal{D}^{(\alpha)}\Phi[\xi; \xi'] := \frac{4}{1-\alpha^2} \left[\frac{1 - \alpha}{2} \Phi(\xi) + \frac{1 + \alpha}{2} \Phi(\xi') - \Phi(\xi_\alpha)\right] ,$$

(16)

where the factor $4/(1 - \alpha^2)$ is introduced so that the limit $\lim_{\alpha \to 1} \mathcal{D}^{(\alpha)}\Phi = \mathcal{D}_\Phi$ gives a Bregman divergence. In particular, if $\Phi(\xi) = \sum_i c_i e_i$ then $\mathcal{D}^{(\alpha)}\Phi$ becomes the $\alpha$-divergence. Importantly, divergences of the form of
Eq. (16) with \( \alpha \neq \pm 1 \) are not Bregman divergences (as they cannot be expressed in terms of convex conjugates as in Eq. (7)), and hence they don’t lead to flat geometries (see Section III C).

Fortunately, recent results suggest a ways to express non-Bregman divergences in terms of generalized Legendre transforms [6]. The generalized Legendre transform is based on a link function — a smooth function \( C : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R} \) that links generalized potentials \( \varphi \) and \( \psi \) via the following relationship:

\[
\psi(\xi) + \varphi(\eta) - C(\xi, \eta) = 0. \tag{17}
\]

Put differently, for a given link function \( C \), a pair of generalized potentials are functions \( \varphi, \psi \) that are related via a generalized Fenchel-Legendre \( C \)-transform as follows:

\[
\varphi(\xi(p)) = \inf_{q \in \mathcal{M}} \{ \psi(\eta(q)) - C(\xi(p), \eta(q)) \}, \tag{18}
\]

\[
\psi(\eta(q)) = \inf_{p \in \mathcal{M}} \{ \varphi(\xi(p)) - C(\xi(p), \eta(q)) \}. \tag{19}
\]

Note that these equations use a different sign than Eq. (2), which leads to the consideration of concave instead of convex functions. Arguments for adopting this choice are discussed in Ref. [6].

Following the rationale that lead to Eq. (7), for a given function \( C \) and a \( C \)-conjugate potentials \( \varphi, \psi \), one can follow a similar rationale can define a generalized Bregman divergence

\[
\Delta(p|q) = C(\xi(p), \eta(q)) - \varphi(\xi(p)) - \psi(\eta(q)). \tag{20}
\]

Interestingly, Eqs. (18) and (18) imply that \( \Delta(p|q) \geq 0 \), with equality if an only if \( p = q \). As all divergences, the metric induced by generalized Bregman divergences is the Fisher metric; interestingly, Eqs. (12a) and (12b) imply that the connections are given by

\[
\Gamma_{ijk}(\xi) = -\partial_{i}\partial_{j}C(\xi; \eta(\xi'))|_{\xi=\xi'}, \tag{21a}
\]

\[
\Gamma_{ijk}(\xi) = -\partial_{k}\partial_{i}\partial_{j}C(\xi; \eta(\xi'))|_{\xi=\xi'}. \tag{21b}
\]

Note that if \( C(\xi, \eta) = \xi \cdot \eta \) then \( \Gamma_{ijk}(\xi) = 0 \), and hence curved geometries only arise from non-trivial link functions — i.e. from generalized Legendre transforms.

For the dual geometries that arise from the \( \alpha \)-divergence one can identifies the corresponding link function following a two-steps procedure. First, one applies a monotonous transformation that turns the \( \alpha \)-divergence into the Rényi divergence [41]

\[
\Delta_{\gamma}(p|q) = \frac{1}{\gamma} \log \int_{S} p^{\gamma+1}(s)q^{-\gamma}(s)d\mu(s), \tag{22}
\]

leveraging the fact that both divergences generate the same geometry being part of the same conformal-projective equivalent class [7, Sec. 2-D]. Note that when \( \gamma \to 0 \) then \( C \) tends to \( \xi \cdot \eta \), and the Rényi divergence tends to the Kullback-Leiver divergence [42]. As a second step, one uses the fact that the Rényi divergence can be expressed in terms of generalized convex conjugates as [6, Th. 13] can be recovered as a generalized Bregman divergence when using the link function

\[
C(\xi, \eta) = \frac{1}{\gamma} \log(1 + \gamma \xi \cdot \eta), \tag{23}
\]

and a corresponding generalized potential given by

\[
\varphi_{\gamma}(\xi) = \log \int_{S} (1 + \gamma \xi \cdot h(s))^{-\frac{1}{\gamma}} d\mu(s). \tag{24}
\]

Interestingly, it has been shown that this non-trivial logarithmic link function — or, equivalently, the Rényi divergence — gives rise to dual geometries of constant curvature [7]. Therefore, they are the natural first step in the exploration of more general curved statistical manifolds.

To finish this section, let us introduce the notion of Legendre derivative. For given generalized potentials \( \varphi \) and \( \psi \), the corresponding Legendre derivative acting on a smooth function \( \varphi \) is given by

\[
D_{L}\varphi(\xi) = \eta \quad \text{and} \quad D_{L}\psi(\eta) = \xi. \tag{25}
\]

The functional form for \( D_{L} \) is determined by the corresponding link function. For example, for the case of \( C(\xi, \eta) = \xi \cdot \eta \) then Eqs. (4) and (5) show that \( D_{L} \) is given by the Euclidean gradient. On the other hand, for a logarithmic link function as in Eq. (23) one can find that the corresponding (non-Euclidean) Legendre derivative acting on a smooth function \( \varphi \) is given by

\[
D_{L}(\gamma)\varphi = \frac{1}{1 - \gamma \xi \cdot D\varphi}, \tag{26}
\]

with \( D \) denoting the Euclidean gradient.

IV. SYMPLECTIC AND KÄHLER STRUCTURES IN INFORMATION GEOMETRY

This section studies the realization of symplectic structures in statistical manifolds. This naturally lead towards considering the complexification of statistical manifolds, which in turn bring new insights into the Legendre transform. Complex manifolds are a ‘bigger’ bundle that possesses a richer structure benefited by greater symmetry. These complex structures are quintessential to physics; from quantization of the spin and coherent states [43] to entanglement [44] and string theory [45], Kähler oscillators [46, 47]. Put simply, the reasoning pursued here is that by recasting the manifold as a complex structure with higher symmetry one can obtain a clearer and more systematic understanding of the geometry that arises from the deformed Legendre transform.

For this purpose, we first establish a parallel between statistical manifolds and phase spaces. In doing this, it is important to note that while in statistical manifolds the dual coordinates \( \xi \) and \( \eta \) usually refer to the same point, in phase spaces they typically refer to canonical
pairs (e.g. position and momentum) and hence correspond to different dimensions. This naturally leads us to consider product manifolds of double dimensionality of the original one.

A. Establishing dynamics on phase-space

In analytical mechanics the Legendre transforms enables us to derive the Hamiltonian formalism from the Lagrangian — a smooth function of $n$ generalized coordinates $q^i$, velocity $\dot{q}$, and time $t$. By doing this one trades $n$ second-order equations of motion for $2n$ first-order differential equations of the form

$$\frac{\partial H}{\partial p_k} = \dot{q}^k, \quad \frac{\partial H}{\partial q^k} = -\dot{p}_k, \quad (27)$$

Notice that the transformation $(q,p) \mapsto (-q,p)$ preserves the form of the above equations. This symmetry due to a rich mathematical structure that provides the foundations of classical mechanics, which we introduce in the rest of this subsection.

We start by reviewing the standard method to establish dynamics over a manifold based on the Hamiltonian formulation of classical mechanics — as described for instance in Refs. [1, 48, 49]. For this, let us consider a phase-space $\mathcal{M}$ that describes a system of interest. More specifically, each point in $\mathcal{M}$ has the form $z = (q^1, \ldots, q^n, p_1, \ldots, p_n)$, with $(q^1, \ldots, q^n) \in \mathbb{R}^n$ corresponding to a configuration manifold $Q$, and $(p_1, \ldots, p_n) \in \mathbb{R}^n$ corresponding to its (generalized) conjugate momenta. Dynamics over the phase-space $\mathcal{M}$ are be established by a Hamiltonian $H : \mathcal{M} \to \mathbb{R}$ via the following equations of motion:

$$\dot{z} = X_H, \quad (28)$$

where the Hamiltonian vector field is given by

$$X_H = JD^{(0)}H(z), \quad \text{with} \quad J := \begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix} \quad (29)$$

and $D$ denotes the standard gradient (see Eq. (26)). In this way, dynamics are established flowing the integral curves of $X_H$. At any point $z \in \mathcal{M}$ there is a trajectory governed by the dynamics induced by the Hamiltonian, which is unique due to the linearity of the equations involved.

Above, the role Eq. (29) — which turns the Hamiltonian into a vector field — can be re-framed in a more principled manner via symplectic geometry, as follows [50]. A symplectic form $\omega$ is a 2-form on $\mathcal{M}$ that is closed ($d\omega = 0$) and nondegenerate ($\forall v \neq 0 \exists u : \omega(v,u) \neq 0$). On a symplectic manifold (i.e. a manifold equipped with a symplectic form) the flow of the Hamiltonian $H$ can be defined as the vector field $X_H$ that satisfies the following relationship:

$$-dH = \iota_{X_H} \omega, \quad (30)$$

where $\iota_{X_H} \omega = \omega(X_H, \cdot)$ is the 1-form that results from the interior contraction of $\omega$. Above, $dH$ is the differential of $H$ and the sign corresponds to a convention in the definition of the symplectic form. The fact that $\omega$ is non-degenerate guarantees that one can always find a unique $X_H$ that satisfies Eq. (30). Additionally, the closure of the symplectic form locally implies — by the Poincare lemma — the existence of a tautological 1-form $\theta$ (also known as the canonical 1-form or symplectic potential), which satisfies the condition $\omega = d\theta$. This coordinate-invariant expression for $\omega$ emphasizes its topological nature.

Symplectic manifolds belongs to equivalent classes established defined via symplectomorphims (i.e. difeomorphisms that preserve the symplectic form), which are equivalent to canonical transformation in the context of analytical mechanics context. Correspondence between diffeomorphisms that preserve the symplectic form $\mathcal{L}_{X_H} \omega = 0$. Furthermore, the geometry of the phase-space account for important properties of the underlying system. Indeed, while for an unconstrained system may be described by a phase-space of the form $\mathcal{M} = \mathbb{R}^{2n}$, more complicated systems usually require more convoluted geometries. As a simple example, a pendulum is described a phase-space of the form of a cylinder — which has a flat internal geometry but a non-trivial topology. Next subsections explore the role of non-zero curvature on the phase-space.

B. Symplectic structure under the deformed Legendre transform

Section III C shows that, from an information geometric perspective, divergences can be used to determine the metric and connections of a manifold. In this subsection we will show how divergences also generate a symplectic 2-form, from which much of the insights from Hamiltonian mechanics are inherited. This will, in turn, allow us to study probability distributions in phase space and discuss the flow induced by divergences. Our results will show that the symplectic 2-form induced by the divergence on the phase space and the product manifold (and hence the induced Hamiltonian dynamics) are different when the geometry is curved — or equivalently, when the Legendre transform has been deformed.

To start, let us introduce some terminology. We will contrast structures on the cotangent bundle of statistical manifolds with structures in the product manifold $\mathcal{M} \times \mathcal{M}$, which is made of pairs $(p,q)$. The product manifold is often parametrized using dual coordinates as $(\xi, \eta) := (\xi(p), \eta(q))$ [51]. Also, we use the projection operators over the left and right elements, $\pi_l(p,q) = p$ and $\pi_r(p,q) = q$, to define the following sub-manifolds: $\mathcal{M}_q := \pi_l^{-1}(p,q) = \mathcal{M} \times \{p\} \simeq \mathcal{M}$, and similarly $\mathcal{M}_p := \pi_r^{-1}(p,q) = \{q\} \times \mathcal{M} \simeq \mathcal{M}$. The diagonal of the product manifold will be denoted as $\Delta_p \subset \mathcal{M} \times \mathcal{M}$, being made by pairs of the form $(p,p)$. 
Divergences are smooth functions mapping $\mathcal{M} \times \mathcal{M}$ into $\mathbb{R}$, and we are interested in the geometrical structure it induces. To probe this, let us consider the canonical symplectic form $\omega_p$ on $T^*\mathcal{M}_p$, which by taking a local chart $(U, \xi^k, \nu_k)$ can be written as

$$\omega_p := d\xi^i \wedge d\nu_i$$  \hspace{1cm} (31)

with $\nu_k$ being the conjugate coordinate to $\xi^k$. Note that, thanks to Darboux’s theorem [1], such canonical pairs are guaranteed to always exist locally. Let us then recast the map presented in Eq. (8) as the symplectomorphism $L_D : \mathcal{M} \times \mathcal{M} \rightarrow T^*\mathcal{M}$ given by

$$L_D : (\xi, \eta) \mapsto (\xi, \nu) = (\xi, \partial_i D(\xi, \eta) d\xi^i) .$$  \hspace{1cm} (32)

As shown in [52], this map induces — via the pull-back $L_D^* \omega_p = \omega_D$ — the following symplectic form on $\mathcal{M} \times \mathcal{M}$:

$$L_D^* \omega_p = L_D^*[d\xi^i \wedge d\nu_i] = d\xi^i \wedge (\partial_i D(\xi, \eta) d\xi + \partial_k^l D(\xi, \eta) d\eta_k) = d\xi^i \wedge \partial_k^l D(\xi, \eta) d\xi^i \wedge d\eta_k,$$  \hspace{1cm} (33a)

where the vanishing of the first expression $\partial_k^l D(\xi, \eta)$ is a result of the commutativity of the second derivatives of the divergence. Note that $\partial_k^l D(\xi, \eta)$ reduces to the Fisher metric when evaluated on $\Delta_p$ (i.e. when $\xi$ and $\eta$ are evaluated at the same element $p$). Importantly, the same symplectic form on $\mathcal{M} \times \mathcal{M}$ is obtained by pulling back the canonical symplectic form $\omega_q := d\nu_k \wedge d\lambda^k$ on $T^*\mathcal{M}_q$ (where $(\eta, \lambda)$ form a canonical pair) in an analogous fashion, using this time the symplectomorphisms given by

$$R_D : (\xi, \eta) \mapsto (\xi, \lambda) = (\xi, \partial_k D(\xi, \eta) d\xi^i) .$$  \hspace{1cm} (34)

Now that the symplectic form given by Eq. (33d) has been identified as the natural one on $\mathcal{M} \times \mathcal{M}$, our next step is to investigate how is it influenced by the manifold’s curvature. For this, we note that if the divergence $D$ is a generalized Bregman divergence, then this symplectic form depends solely on the link function. Concretely, a direct calculation shows that for this case

$$\omega_D = \partial_k^l C(\xi, \eta)d\xi^i \wedge d\eta_k.$$  \hspace{1cm} (35)

This clarifies how, although identical on the cotangent bundle $T^*\mathcal{M}$, the symplectic structure induced by different divergences may differ on $\mathcal{M} \times \mathcal{M}$. In particular, while the dually-flat geometry established by Bregman divergences leads to a symplectic form given by $\omega_D = d\xi^i \wedge d\eta_i$, for $\gamma$-curved geometry the Rényi divergence induces the following symplectic form:

$$\omega_D = \frac{1}{1 + \gamma \xi^i \eta_i} \left( \delta^i_k - \frac{\gamma e^k \eta_i}{1 + \gamma \xi^i \eta_i} \right) d\eta_k \wedge d\xi^i .$$  \hspace{1cm} (36)

The coefficients of this symplectic form coincide with the metric tensor at Ref. [6, Proposition 4], now on the product manifold $\mathcal{M} \times \mathcal{M}$. 

The symplectic form shown in Eq. (36) is closed, as can be confirmed by a direct calculation leading to $\omega_D = 0$. This, in turn, implies the local existence of a tautological 1-form via Poincaré lemma, as explained in the previous section. Similar to the derivation that lead to Eq. (36), we define the canonical 1-form $\theta_p = \nu_k d\xi^i$ on $T^*\mathcal{M}_p$ and take its pull-back onto $\mathcal{M} \times \mathcal{M}$, yielding

$$\theta = \frac{1}{2} \frac{\eta_i d\xi^i - \xi^i d\eta_i}{1 + \gamma \xi^i \eta_i} .$$  \hspace{1cm} (37)

which can be directly identified as the 1-form emerging from connections that describe the projective-flat geometry induced by Rényi’s divergence.

As a last step, we use the symplectic form $\omega_D$ to evaluate the action of the smooth function $D_\gamma$ on the product manifold $\mathcal{M} \times \mathcal{M}$. This function is of particular interest, as it generate integral curves of constant $D$, and hence the induced flow is closed within the diagonal $\Delta_p \simeq \mathcal{M}$. For this purpose, let us denote as $X_\gamma = X_0 \partial_0^\gamma + X_\eta \partial_\eta^\gamma$ the vector field generated by the observable $D_\gamma$, and the corresponding symplectic form. We are interested in the vector fields that preserve the symplectic form $\omega_D$, that is $X_\gamma$ satisfies $L_{X_\gamma} \omega_D = 0$ (where $L_{X_\gamma} \omega_D$ denotes the Lie derivative of $\omega_D$ in the direction of $X_\gamma$), then one can find that

$$L_{X_\gamma} \omega_D = \iota_{X_\gamma} d\omega_D + d(\iota_{X_\gamma} \omega_D) = d(\iota_{X_\gamma} \omega_D),$$  \hspace{1cm} (38)

where the last equality is a consequence of the fact that $\omega_D$ is closed. Therefore, the Lie derivative (38) vanishes only if $X_\gamma$ is Hamiltonian (30), i.e. if $X_H$ satisfies $\iota_{X_H} \omega_D + dD_\gamma = 0$. One can then determine the Rényi vector field via explicit evaluation of the interior product as follows:

$$-dD_\gamma = (\iota_{X_H} g_k^l d\eta_l) \wedge d\xi^k - g_k^l d\eta_l \wedge (\iota_{X_H} d\xi^k) = g_k^l (X_{\gamma_0} d\xi^k - X_{\gamma_0} d\eta_l),$$  \hspace{1cm} (39a)

which results in a Hamiltonian flow generated by Rényi’s divergences of the form

$$X_{\gamma k} = -g_a^k \partial_{\gamma a} D_\gamma, \quad X_{\gamma} = g_a^k \partial_{\gamma a} D_\gamma .$$  \hspace{1cm} (40)

Then, the corresponding Rényi vector field can be found to be equal to

$$X_\gamma = g_a^k \partial_{\gamma a} D_\gamma \partial_0^\gamma - g_k^l \partial_0^\gamma D_\gamma \partial_{\gamma_0} + [\eta_\gamma - D_L^{(\gamma)} \psi(q)]_k \partial_0^\gamma$$  \hspace{1cm} (41a)

$$- [\xi_\gamma - D_L^{(\gamma)} \varphi(q)]^k \partial_{\gamma_0} .$$  \hspace{1cm} (41b)

Importantly, this Rényi flow is closed within the diagonal $\Delta_p$. Moreover, the above result shows, on the diagonal, flows follow the geodesic respect to the primal and dual connections — which naturally satisfy Eq. (25).
In this way, we gain a new understanding where the deformed exponential family obtained from Rényi’s divergence [6, Section 4] describe the sets of points flowing along the integral curves at \( X_\gamma \) and external points diverging from it.

### C. Complexification of statistical manifolds

This section discusses some fundamental aspects of complex geometry, followed by the complexification of statistical manifolds. Then, the next section focuses on the complex structure induced by the Rényi divergence.

A complex manifold can be depicted as a topological space that locally looks like \( \mathbb{C}^n \). One way to try building a complex manifold would be to first consider a \( 2n \)-dimensional real manifold, and then arrange a set of coordinates \( \{ x^k_x \} \) into complex combinations such as \( x^k_x \) for all integers \( k \). Unfortunately, such arrangement is arbitrary and — more importantly — is coordinate dependent. In effect, additional structure of the manifold is required in order to ‘complexify’ it properly.

One way to build a complex manifold is via a tensor field \( J_{ab} \) of real components satisfying \( J^2 = -1 \), which provides a linear endomorphism \( J : T_p M \to T_p M \). First, note that diagonalizing such a tensor cannot be done in a vector space of real values, hence the coefficients of vectors in \( T_p M \) must be allowed to be complex-valued, i.e. \( T_p^C M = T_p M \otimes \mathbb{C} \). By arranging \( 2n \)-local coordinates into complex coordinates \( x^k + iy^k, \) e.g. via \( x = x^k_x \), \( y = y^k_x \), then one can express \( J \) in complex coordinates as

\[
J = i dz^a \otimes \frac{\partial}{\partial z^a} - idz^a \otimes \frac{\partial}{\partial z^a} .
\]

Hereon, \( a \) and \( \check{a} \) are indices in \( \{ 1, \ldots, n \} \), with the bar being used to distinguish between holomorphic and anti-holomorphic components. The manifold \( M \) together with the tensor \( J \) are known as an “almost complex structure.” With the aid of \( J \), such complexified \( T_p^M \) can now be decomposed into a holomorphic and anti-holomorphic parts via projection operators given by \( P_{\pm} \). These projection operators can be used to decomposed any \( k \)-form into \( (p, q) \)-forms with \( p + q = k \).

While every complex manifold is also a real manifold, the converse does not always hold. A necessary and sufficient condition on \( J \) to allow a real manifold to be a complex one is given by \( N_{ab}^c = 0 \) [52], where \( N_{ab}^c \) stands for the Nijenhuis tensor given by [53]

\[
N_{ab}^c := 2 \left( J^d_a \nabla_{[d} J^c_b] - J^d_a \nabla_{[d} J^c_b] \right) ,
\]

with squared brackets denoting the antisymmetrization of indices.

Up to this point, a complex manifold \( (\mathcal{M}, J) \) has not been equipped with a metric — in fact, a \( J \)-compatible metric may not exist (e.g. in Hopf manifolds). When such metric does exist, this imposes the following compatibility conditions:

\[
g_{\mu\nu} J_\mu^\rho J_\nu^\sigma = g_{\rho\sigma} \quad \text{and} \quad \nabla_\mu J_\mu^\nu = 0 . \tag{44}
\]

The first condition above implies that the pure holomorphic and anti-holomorphic components of the metric vanish, hence \( ds^2 = g_{\mu\nu} dx^\mu \otimes dx^\nu = g_{\rho\sigma} dz^a \otimes d\bar{z}^b \) is hermitian. The second condition enforces the vanishing of Nijenhuis tensor [43] not only to guarantee complexification, but also implies that the Kähler \( 2 \)-form \( k \) given by

\[
k = \frac{1}{2} g_{\mu\nu} J_\mu^\rho d\bar{z}^b \wedge dx^\nu = i g_{ab} dz^a \wedge d\bar{z}^b \tag{45}
\]

is closed, serving as the manifold’s symplectic form \( \omega \). In components, Eq. (45) means that \( \partial_d g_{bc} = \partial_b g_{dc} \) and \( \partial_b g_{ac} = \partial_c g_{ab} \). Analogously as in (5), these expressions can be locally integrated out revealing the metric

\[
g_{ab} = \partial_a \partial_b \mathcal{K}(z, \bar{z}) , \tag{46}
\]

with \( \mathcal{K} \) a real-valued smooth function also known as a Kähler potential. Please note that this potential \( \mathcal{K} \) is not unique, as it is only determined only up to the addition of an holomorphic and an anti-holomorphic function:

\[
\mathcal{K}(z, \bar{z}) \to \mathcal{K}(z, \bar{z}) + U(z) + \bar{U}(\bar{z}) \tag{47}
\]

Also, \( \mathcal{K} \) may not be globally defined [54]. In this way, a Riemannian metric as well as the symplectic form determined by \( \mathcal{K} \), as \( \omega = k = \frac{1}{2} \partial \bar{\partial} \mathcal{K} \), where \( \partial, \bar{\partial} \) denote Dolbeault operators \( \partial = dz \wedge \partial a \) and \( \bar{\partial} = d\bar{z} \wedge \partial z \) respectively.

The similarities between the expressions found here for the metric in terms of a scalar potential Section III D and Kähler’s is no coincidence as \( \mathcal{K} \) itself must convex. This relation is in fact, the catalyst of a set of more intimate relations between the space of Kähler metrics and convexity [58].

In statistical manifolds the fundamental object is its divergence \( D \), and so, the constraints on the metric are ultimately enforced onto \( D \). Therefore, the conditions for complexification of a manifold translate into two conditions over the corresponding divergence [59]:

1. \( \partial_{i,j} D = \partial_{j,i} D \) on \( \mathcal{M} \times \mathcal{M} \).
2. \( \partial_{i,j} D + \partial_{j,i} D = \kappa \partial_{i,j} D \) for some \( \kappa \in \mathbb{R} \).

here the primed indices denote differentiation with respect to \( y \in \mathcal{M}_q \) as opposed to regular derivatives with respect to \( x \in \mathcal{M}_p \). Note that, although condition 1 is trivially satisfied when evaluated at the diagonal (as shown in Eq. (11)), it is not automatic for it to hold on the whole \( \mathcal{M} \times \mathcal{M} \). Both conditions arise from the construction of of an invariant arc element \( ds^2 \) from the symmetric and antisymmetric parts as:

\[
ds^2 = g_D - i\omega_D . \tag{48}
\]
where \( g_\mathcal{D} \) and \( \omega_\mathcal{D} \) denote the metric and symplectic form (1st and 2nd line at (48) respectively) induced by the divergence on \( \mathcal{M} \times \mathcal{M} \). The resulting closed two-form \( \omega_\mathcal{D} = 0 \) is equivalent to that derived in (37). Similar for its symmetric counterpart \( g_\mathcal{D} \) whose components are (46).

The second condition motivated by the fact that we are interested into expressing the above quantity as \( \partial \bar{\partial} D \), the conditions discussed above, and hence the geometries should be satisfied on \( \mathcal{M} \) respectively).



This means that the geometry that arises from the Rényi divergence, \( \gamma \), belongs to the family of divergences that can be expressed as in Eq. (16) using \( \Phi(x) \) as given by

\[
\Phi(x) = \log \sum_{s \in S} e^{x(s)}, \quad \text{with} \quad x(s) := \log p(s). \tag{51}
\]

This means that the geometry that arises from the Rényi divergence is susceptible of being complexified. Furthermore, when evaluated on arguments that correspond to probability distributions (i.e. \( x^a = \log p^a \) and \( y^a = \log q^a \)), we note that the first two terms in Eq. (16) vanish, and therefore Rényi’s divergence itself serves as the Kähler potential.

We now show that the two conditions for complexification, discussed in the previous subsection, are satisfied by product manifolds \( \mathcal{M} \times \mathcal{M} \) endowed by a geometry induced by Rényi’s divergence. For this, we adopt the complex coordinate \( w^a = x^a + iy^a \in \mathcal{C} \) with \( x^a = \log p^a \) and \( y^a = \log q^a \) for \( p, q \in \mathcal{M} \). Using these coordinates, one finds that

\[
- \frac{1}{\kappa} \mathcal{D}_\gamma(x, y) = \log \sum_{a=1}^{n} \exp(\bar{\Gamma} w^a + \gamma \bar{w}^a) \tag{52a}
\]

where we are using the shorthand notations \( \Gamma = \frac{1}{2}(\alpha_- + i \alpha_+) \) and \( z^a = \exp(x^a) \). In this manner, \( \Phi(\alpha_+ x + \alpha_- y) \) (or, equivalently, Rényi, \( \mathcal{D}_\gamma(x, y) \)) can be identified as the Kähler potential for the product manifold.

The resemblance between the induced symplectic form in Eq. (36) and the connections (37) at the previous section to the well-known Fubini-Study metric and its connection are suggestive of the complex-projective spaces, \( \mathbb{C}P^n \) \cite{60}. Unfortunately complexification of the local charts do not preserve the functional form of (36) and (37). However, under special circumstances such as \( \gamma = 1 \) restriction to the diagonal \( \Delta_p \) does lead to \( \mathbb{C}P^n \) upon complexification. Disregarding the pure holomorphic and anti-holomorphic functions of the divergence, the link function of the deformed Legendre transform can be directly read as the Kähler potential as follows:

\[
\mathcal{K}(z, \bar{z}) = C(z, \bar{z}) = \log(1 + z_a \bar{z}^a), \tag{53}
\]

hence generating the Fubini-Study metric given by

\[
g_{ab} = \frac{1}{1 + z_a \bar{z}^a} \left( \delta_{ab} - \frac{z^a \bar{z}^b}{(1 + z_a \bar{z}^a)} \right). \tag{54}
\]

The case of complex dimension \( n = \dim_{\mathbb{C}} \mathcal{M} = 1 \) (two real dimensions), that is \( \mathcal{M} = \mathbb{C}P^1 \subset \mathbb{C}^2 \) is of particular interest to physical systems. Indeed, from a group-theoretic perspective, this manifold corresponding to the coset group \( SU(2)/U(1) \) (isomorphic to the Riemann sphere \( S^2 \simeq \mathbb{C}P^1 \)) is crucially important to the coherent state formulation \cite{43,61} and the geometric quantization of the spin \cite{48}. In addition, \( \mathbb{C}P^1 \) describes pure quantum states whose direct product enables a nice geometric formulation of entangled systems \cite{44}, to name a few.

The connection on this manifold corresponds to the canonical 1-form, which is now determined by its Kähler potential

\[
\mathcal{A} = i \left( \partial - \bar{\partial} \right) \mathcal{K} = i \left( \frac{z_a d \bar{z}^a - \bar{z}_a dz^a}{1 + z_a \bar{z}^a} \right) \tag{55}
\]

via Dolbeault operators \cite{62}. This gauge-field is consistent with the expression obtained for the connection 1-form found in Eq. (36).

Let us now show how quantization of the 2-sphere restricts the allowed values for \( \gamma \). As the Poincare lemma tells us every closed form is locally exact, the existence of closed forms failing to be exact is as an aspect of the nontrivial topology of the manifold. This featured is captured by cohomology classes \( \mathcal{H}^k(\mathcal{M}, \mathbb{R}) \) whose members are closed yet globally not exact k-forms. In this sense, the Kähler form belongs to \( \mathcal{H}^2(\mathcal{M}, \mathbb{R}) \). The single-valuedness of points on the manifold would require the \( \omega_\mathcal{D} \) to belong to a cohomology class \( \mathcal{H}^2(\mathcal{M}, \mathbb{R}) \). Therefore it’s symplectic two-form must be an integer multiple of \( \omega_\mathcal{D} \). Hence the covariant derivative is \( \nabla_a = \partial_a - ik \partial_a \mathcal{A}_a \) with \( k \in \mathbb{Z} \) \cite{63}, the same for its anti-holomorphic counterpart.
The holomorphic polarization (see Appendix A) imposes the condition $\nabla_{\bar{z}} \psi = 0$ for $\psi$ a wave function, with $\psi$ a function whose squared module gives a the probability density — closely resembling wave functions in quantum mechanics. This results in
\[
\left( \partial_z + \frac{k}{2} \frac{z^a}{1 + z_a \bar{z}^a} \right) \psi = 0 .
\] (56)
This implicit equation is solved by physical solutions $\psi_{\text{phys}}$ of the form
\[
\psi_{\text{phys}} = \exp \left( - \frac{k}{2} \log(1 + z_a \bar{z}^a) \right) f(z) ,
\] (57)
with $f(z)$ an holomorphic function. The resulting probability density $|\psi_{\text{phys}}|^2$, reads
\[
P(z) = \frac{|f(z)|^2}{(1 + z_a \bar{z}^a)^k} .
\] (58)
The holomorphic function $f(z)$ can be expanded in the basis $\{1, z, z^2, ..., z^k\}$ as higher powers would imply $P(z)$ to diverge, hence a Hilbert space of finite dimension as $\psi_{\text{phys}}$ is defined over the 2-sphere.

Just as holomorphic polarization results into exponential family for $\gamma = 0$ (Appendix A), one recovers the Rényi max-entropy distributions as a polarization of the manifold. Identifying $\gamma = \frac{1}{k}$, we quickly realize (keeping the sign of $\gamma$) that $k \in \mathbb{Z}_+$ restricts $\gamma \in (0, 1]$ coinciding with $\alpha \in (-1, 1)$ of positive curvature discussed in Ref. [7]. Although ruled out by the polarization, it is interesting to notice that considering $\gamma \notin (0, 1]$ would result into the manifold having hyperbolic topology and becoming non-compact, hence not being susceptible to complexification. These results establish $\gamma \in (0, 1]$ as values of special interest, which do not include $\gamma = 0$ — that corresponds to conventional dually-flat geometry, being related to the Shannon entropy.

V. THERMODYNAMICS OF EXPONENTIAL KOLMOGOROV-NAGUMO AVERAGES

In thermodynamics and statistical mechanics, the Legendre transform plays a crucial role relating intensive and extensive variables — where often one is easier to measure and control than its conjugate counterpart. Special attention was paid to universality of Legendre structure for the case of so-called generalized entropies [64]. In this section we investigate how generalized thermodynamic relationships can be recovered using non-trivial link functions — i.e. related to non-Shannon divergences.

Let us consider the study of systems subject to constraints on non-linear averages. A natural extension of standard linear (arithmetic) averages is found in the Kolmogorov-Nagumo average given by [65, 66]
\[
\langle X \rangle_\gamma = g^{-1} \left( \sum_i p_i g(x_i) \right) ,
\] (59)
where $g$ is an increasing function. The linear average is recovered by setting $g(x) = x$, which can be shown to be the only one that satisfies two properties:

1. **Homogeneity**: $\langle aX \rangle = a \langle X \rangle$.

2. **Translation homogeneity**: $\langle X + c \rangle = \langle X \rangle + c$.

However, one can also consider averages satisfying only one of both conditions. The first requirement leads to the well-known class of Hölder means, with $g(x) = x^p$ [67]. In contrast, the second property leads to the class of so-called exponential means, where the $g(x) = \exp_p(x) = (e^{\gamma x} - 1)/\gamma$ , with the inverse function being given by $\ln_\gamma(x) = \frac{1}{\gamma} \ln(1 + \gamma x)$.

This second property above is of particular interest for thermodynamics, as it makes the thermodynamic relations independent of the particular value of the ground energy state (i.e., one can always subtract the ground energy value). Also, note that the arithmetic mean is recovered for $\gamma = 0$. Hence, we will focus on developing the thermodynamic implications of this second class of averages. Thermodynamics of Kolmogorov-Nagumo averages has been discussed in [68, 69]. However, the connection between exponential Kolmogorov-Nagumo means and generalized Legendre transforms has not been explored yet.

The exponential means lead to exponential Kolmogorov-Nagumo averages, which we will denote as
\[
\langle X \rangle_\gamma = \ln_\gamma \left( \sum_i p_i \exp_\gamma(x_i) \right) = \frac{1}{\gamma} \ln \left( \sum_i p_i e^{\gamma x_i} \right) .
\] (60)

Interestingly, the Rényi entropy can be naturally formulated by the exponential Kolmogorov-Nagumo average of Hartley information [70] (also known as the Shannon pointwise entropy [71]) $\ln 1/p_k$, as confirmed by a direct calculation:
\[
R_\gamma(P) = \frac{1}{1 - \gamma} \ln \left( \sum_i p_i e^{\gamma (\ln 1/p_i)} \right)
\]
\[
= \frac{1}{\gamma(1 - \gamma)} \ln \sum_i p_i^{1 - \gamma}
\] (61)
which is the Rényi entropy of order $1 - \gamma$ [72]. Note that while the $1/(1 - \gamma)$ factor in Eq. (61) is often not considered, there are two reasons for including it. First, the limit $\gamma \rightarrow 1$ leads to well-known Burg entropy given by $R_1 = -\sum_i \ln p_i$ [73]. Additionally, adding this factor simplifies the calculations presented in the next sections.

A. Thermodynamic equilibrium under the Rényi divergence

Let us consider now a system whose internal energy at state $i$ is given by $\epsilon_i$. Then, the Kolmogorov-Nagumo $\gamma$-average of energy is given as $U = \langle \epsilon \rangle_\gamma$. As discussed
above, the γ-average is not invariant to rescaling energies by a factor $\epsilon_i \mapsto \beta \epsilon_i$. To this end, let us define the rescaled internal energy

$$U_\gamma^\beta := \frac{1}{\beta}(\beta \epsilon)\gamma = \frac{1}{\beta} \ln \gamma \left( \sum_i p_i e^{\gamma (\beta \epsilon_i)} \right).$$  \hspace{1cm} (62)

The reason is that due to generalized mean inhomogeneity, $U$ does not have properly defined units (since units of $\exp(\epsilon_i)$ are not $J$.) By considering units of $[\beta] = J^{-1}$, $\beta \epsilon_i$ is dimensionless and $U_\gamma^\beta$ is the properly defined mean.

Let us now focus on the distribution that corresponds to $U_\gamma^\beta$ according to the maximum entropy principle. The distribution $\pi_i$ that maximizes the Rényi entropy while satisfying a given γ-average level of energy can be found by using the method of Lagrange multipliers with the Lagrange function $\mathcal{L} = R_\gamma - \alpha_0 \sum_i p_i - \alpha_1 \frac{1}{\beta}(\beta \epsilon)\gamma$. A direct calculation shows that $\pi_i$ is the solution of the following equation:

$$\frac{1}{\gamma} \sum_k \frac{\pi_k^{-\gamma}}{\pi_k} - \alpha_0 - \alpha_1 \frac{\gamma}{\beta \gamma} \sum_k \pi_k e^{\beta \epsilon_k} = 0. \hspace{1cm} (63)$$

By multiplying the equation by $\pi_i$ and summing over $i$, one obtains that $\alpha_0 = 1 - \frac{\alpha_1}{\beta}$, which leads to

$$\frac{\pi_k^{-\gamma}}{\sum_k \pi_k^{-\gamma}} - 1 = \frac{\alpha_1}{\beta} \left( \frac{e^{\gamma \beta \epsilon_k}}{\sum_k \pi_k e^{\gamma \beta \epsilon_k}} - 1 \right). \hspace{1cm} (64)$$

Above, the Lagrange parameter $\alpha_1$ is chosen such that one recovers standard thermodynamic relationships. To this end, we identify that $\alpha_1 = \beta$ (which is the standard relation between the Lagrange multiplier and inverse temperature), which gives us

$$\pi_i = \left( \frac{ \sum_k \pi_k e^{\gamma \beta \epsilon_k} }{ \sum_k \pi_k^{-1} } \right)^{1/\gamma} \frac{1}{\gamma} \exp(-\beta \epsilon_i). \hspace{1cm} (65)$$

From the fact that $\sum_k \pi_k = 1$, one finds that

$$\ln \sum_k e^{-\beta \epsilon_k} = \frac{1}{\gamma} \ln \sum_k \pi_k^{-1} - \frac{1}{\gamma} \ln \sum_k \pi_k e^{\gamma \beta \epsilon_k} = (1 - \gamma) \ln \sum_k e^{-\beta \epsilon_k}, \hspace{1cm} (66)$$

where $\Psi_\gamma = R_\gamma - \beta U_\gamma^\beta$ is the free entropy (also called Massieu function).

Note that while the obtained equilibrium distribution in Eq. (65) is Boltzmann, the thermodynamic quantities are nonetheless different from the case of ordinary thermodynamics based on Shannon entropy and linear averages. In fact, Eq. (66) implies that the relation between the free entropy and the logarithm of the partition function differs by the term $-\gamma R_\gamma$, which vanishes only for $\gamma = 0$. Interestingly, it is still possible to derive the free energy by plugging in the equilibrium distribution into internal energy and Rényi entropy as

$$F_\gamma(\pi) = U_\gamma^\beta(\pi) - \frac{1}{\beta} R_\gamma(\pi) = \frac{1}{\beta (\gamma - 1)} \ln \sum_k e^{(\gamma - 1)\beta \epsilon_k}. \hspace{1cm} (67)$$

Above, note that one recovers the ordinary relation between the free energy and the partition function for $\gamma = 0$.

Let us now derive thermodynamic relationships, which will be intrinsically linked with a generalized Legendre structure. The partial derivative of $R_\gamma$ evaluated at the equilibrium distribution is given by

$$\frac{\partial R_\gamma}{\partial p_i}(\pi) = \frac{1}{\gamma} \ln \left( \frac{\sum_i p_i e^{\gamma \beta \epsilon_i}}{\gamma \sum_k \pi_k e^{\gamma \beta \epsilon_k}} \right) = \frac{1}{\gamma} \left( \frac{\sum_i p_i e^{\gamma \beta \epsilon_i}}{\gamma \sum_k \pi_k e^{\gamma \beta \epsilon_k}} \right) = 1. \hspace{1cm} (68)$$

Thus, a direct calculation shows that the link function of the Rényi divergence (see Eq. (23)) between the thermodynamic gradient $\nabla R_\gamma$ and $p - \pi$ is closely related to the energy difference

$$C(\nabla R_\gamma, p - \pi) = \frac{1}{\gamma} \ln (1 + \gamma \nabla R_\gamma \cdot (p - \pi)) = \frac{1}{\gamma} \ln \left( \frac{\sum_i p_i e^{\gamma \beta \epsilon_i}}{\sum_k \pi_k e^{\gamma \beta \epsilon_k}} \right) = \frac{1}{\gamma} \delta_\gamma(\pi \mid \pi). \hspace{1cm} (69)$$

Therefore, the Bregman divergence obtained from the generalized Legendre transform of Rényi entropy can be expressed as the free energy difference between the initial distribution and equilibrium distribution:

$$F_\gamma(p) - F_\gamma(\pi) = \Delta U_\gamma^\beta - \frac{1}{\beta} \Delta R_\gamma = \frac{1}{\gamma} \ln \left( \frac{\sum_i p_i e^{\gamma \beta \epsilon_i}}{\gamma \sum_k \pi_k e^{\gamma \beta \epsilon_k}} \right) = \frac{1}{\gamma} \ln \left( \frac{\sum_i p_i e^{\gamma \beta \epsilon_i}}{\gamma \sum_k \pi_k e^{\gamma \beta \epsilon_k}} \right) = \frac{1}{\gamma} \delta_\gamma(\pi \mid \pi). \hspace{1cm} (70)$$

The fact that $\delta_\gamma(\pi \mid \pi) \geq 0$ implies that $\Delta F_\gamma$ is always positive, and therefore the free energy is minimized by the equilibrium distribution. The implications of this for thermodynamic scenarios are developed in the next section.

Let us finally note that while the free energy is obtained by a regular Legendre transform on the macroscopic level (i.e., it is still defined as $F = U - TS$), due to the generalized mean, it is obtained as a generalized Legendre transform on the mesoscopic (i.e., probability) level, compare Eqs. (67) and (70).

### B. Non-equilibrium stochastic thermodynamics under the Rényi divergence

Let us now focus on the case of non-equilibrium stochastic thermodynamics for the case of Rényi entropy with exponential Kolmogorov-Nagumo average. Stochastic thermodynamics [74] recently became an important topic of non-equilibrium statistical physics. There have been several attempts to apply stochastic thermodynamics for the case of generalized entropies, as [75]. Nonetheless, there is no study that would be focused on the stochastic thermodynamics of Kolmogorov-Nagumo averages yet.
| Quantity                          | Shannon framework | Rényi framework |
|---------------------------------|-------------------|-----------------|
| Link function                   | $C(\xi, \eta) = \xi^k \eta_k$ | $C_\gamma(\xi, \eta) = \frac{1}{\gamma} \log(1 + \gamma \xi^k \eta_k)$ |
| Divergence                      | $\mathcal{D}_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$ | $\mathcal{D}_\gamma(p||q) = \frac{1}{\gamma} \int p^{\gamma+1}(x) e^{-\gamma(x)} dx$ |
| Entropy                          | $S(p) = -\sum_k p_k \log p_k$ | $R_\gamma(p) = \frac{1}{\gamma} \log \sum_k p_k^{\gamma}$ |
| Kolmogorov-Nagumo function      | $g(x) = x - g^{-1}(x)$ | $g(x) = \frac{x}{\gamma+1}, \quad g^{-1}(x) = \frac{1}{\gamma} \log(1 + \gamma x)$ |
| Internal energy                 | $U = \sum_k p_k e_k$ | $U_\gamma = \frac{1}{\gamma} \log \left( \sum_k p_k e^{\gamma e_k} \right)$ |
| Heat rate                       | $Q = \sum_k \dot{p}_k e_k$ | $Q_\gamma^2 = \frac{1}{\gamma^2} \sum_k p_k e^{\gamma^2 e_k}$ |
| Work rate                       | $W = \sum_k p_k \dot{e}_k$ | $\dot{W}_\gamma^\beta = \sum_k p_k e^{\gamma e_k}$ |
| Equilibrium free energy         | $F(\pi) = -\frac{1}{\beta} \log \sum_k e^{-\beta e_k}$ | $F_\gamma(\pi) = \frac{1}{\gamma} \log \sum_k e^{(\gamma-1) e_k}$ |
| Non-equilibrium free energy     | $F(p) = F(\pi) + \frac{1}{2} \mathcal{D}_{KL}(p||\pi)$ | $F_\gamma(p) = F_\gamma(\pi) + \frac{1}{2} \mathcal{D}_\gamma(p||\pi)$ |

TABLE I. Comparison of the relevant quantities for Shannon’s and Rényi’s framework. Quantities from the Shannon framework can be recovered by taking $\gamma \to 0$.

We start by re-stating the first law of thermodynamics for the total energy measured as a Kolmogorov-Nagumo average, which reads as follows:

$$\frac{dU^\beta}{dt} = \dot{Q}_\gamma^\beta + \dot{W}_\gamma^\beta,$$  \hspace{1cm} (71)

where the work and heat flow into the system of interest are given by

$$\dot{W}_\gamma^\beta = \sum_i \dot{p}_i(t) e^{\gamma \beta \epsilon_i} \quad \text{and} \quad \dot{Q}_\gamma^\beta = \frac{1}{\beta \gamma} \sum_i \dot{p}_i(t) e^{\gamma \beta \epsilon_i}.$$  \hspace{1cm} (72)

Let us show validity of second law for the case of Rényi entropy with Kolmogorov-Nagumo exponential mean for the case of arbitrary non-equilibrium process driven by the linear Markov dynamics. To this end, we consider (linear) master equation given by

$$\dot{p}_i(t) = \sum_j \left( w_{ij} p_j(t) - w_{ji} p_i(t) \right),$$  \hspace{1cm} (73)

where $w_{ij}$ is the transition rate. Additionally, let us assume that the stationary distribution satisfies detailed balance, i.e.

$$\frac{w_{ij}}{w_{ji}} = \frac{\pi_i}{\pi_j} = e^{\beta (\epsilon_j - \epsilon_i)}.$$  \hspace{1cm} (74)

In appendix B, we will show that the entropy production rate, i.e.,

$$\dot{\Sigma}_\gamma = \dot{R}_\gamma - \beta \dot{Q}_\gamma^\beta \geq 0,$$  \hspace{1cm} (75)

which is nothing else than the second law of thermodynamics. Please note that Eq. (75) reveals a family of inequalities — indexed by $\gamma$ — that hold for any stochastic system whose dynamics are governed by a master equation as in Eq. (73) and satisfy detail balance as in Eq. (74). However, the constraints of the system will determine which values of $\gamma$ are physically useful: linear constraints yield to a Shannon-type second law (with $\gamma = 0$), while non-linear Kolmogorov-Nagumo constraints lead to Rényi-type second law (with $\gamma \neq 0$).

Additionally, it is possible to combine the first and second law of thermodynamics (Eqs. (71) and (75)) to find an expression for the total work of the system, denoted by $W^{\text{tot}}$. By associating the changes in free energy $\Delta F_\gamma$ with reversible work, denoted by $W^{\text{rev}}$, a direct calculation shows that

$$W^{\text{tot}} = W^{\text{rev}} + W^{\text{irr}} = \frac{1}{\beta} \mathcal{D}_\gamma(p||\pi) + \frac{1}{\beta} \Delta \Sigma_\gamma,$$  \hspace{1cm} (76)

where $W^{\text{irr}}$ corresponds to irreversible work. Overall, this result shows that — similarly as for the case of standard thermodynamics [76, 77] — the reversible work due to changes in free energy corresponds to an informational difference between non-equilibrium and equilibrium distributions, while the irreversible work is related to the entropy production.

Let us finally mention that the trajectory thermodynamics remains in the same form as in the ordinary stochastic thermodynamics. For this, we define the trajectory entropy is defined in a standard way as $s_t = -\log p(x(t), t)$, which is justified by the fact that its Kolmogorov-Nagumo exponential average gives back the Rényi entropy (see Section V A). The trajectory entropy production is then defined as $\Delta s_t = \Delta s_t - \beta \Delta q_t$, where $\Delta q_t$ is the stochastic heat. Detailed fluctuation theorem then holds in the common form [78], i.e.,

$$\frac{P(\Delta s_t)}{P(-\Delta s_t)} = e^{\Delta s_t},$$  \hspace{1cm} (77)

where $P$ is the probability distribution under a reverse protocol. Therefore, on the trajectory level the relations remain exactly the same as for the case of ordinary Shannon-Boltzmann-Gibbs framework.

The relation to the integrated fluctuation theorem is then given by the connection between trajectory quantities and ensemble quantities, since the ensemble average is given by

$$\langle Y \rangle_\gamma = \ln \gamma \int P(x(t)) \exp_\gamma(Y(x(t))),$$  \hspace{1cm} (78)
where \( \mathcal{P} x(t) \) is the the path integral measure. The integrated fluctuation theorem can then be formulated as

\[
\langle e^{\Delta s_i} \rangle_\gamma = 1. \tag{79}
\]

A direct application of Jensen’s inequality yields \( \langle \Delta s_i \rangle \geq 0 \), thus recovering again the second law of thermodynamics.

VI. CONCLUSIONS

The Legendre transform, a fundamental piece of classic and contemporary physics, has a direct but non-trivial correspondence with the dually-flat geometry of statistical manifolds induced by Shannon’s entropy and the Kullback-Leibler divergence. This paper explores deformations of the Legendre transform and its consequences on symplectic geometry, complexification, and stochastic thermodynamics. Taken together, these results pave the way towards a novel, rigorous, and encompassing understanding of physical systems that are not well-described by classic information-theoretic quantities.

A. Towards a better understanding of non-cannonical pairs and Rényi’s curvature

The role of the Legendre transform on analytical mechanics differs from that in information geometry: in the latter dual coordinates refer to different descriptions of the same point, whereas in the former they refer to an isomorphism between the tangent and cotangent bundles. In flat geometry the symplectic form of the cotangent bundle is equivalent to a canonical area form at the product manifold. In contrast, our results show that this equivalence is broken if the manifold is curved. Interestingly, this implies that a deformation of the regular Legendre transform results into the failure of the natural coordinates to form a canonical pair. Furthermore, an analysis of the deformed symplectic form and flow that arises in curved manifolds reveals a new understanding of the family of maximum Rényi entropy distributions, which are found to form sets of points flowing along the integral curves of the flow.

The departure of the symplectic form of the product manifold from the cotangent bundle provides a promising lead to study coupled physical systems, with non-canonical coordinates — like the pair induced by the Rényi geometry — being subjects of special interest. For instance, there have been studies on the consequences of deformations in the symplectic form in field theory [79], and in \( \mathbb{C}P^n \) Kähler oscillators where deformations to the symplectic structure via magnetic field are explored [46, 47]. Also related phenomena has been studied in Fermi liquids under an external magnetic field, where the the magnetic field couples to Berry’s curvature deforming the symplectic form. Such deformations have been shown to have strong consequences for observables, as the invariant phase volume is modified via a topological invariant [80, 81]. An interesting avenue for future research is to investigate if there are divergences that can recapitulate these deformations, providing a mathematical scaffolding for the study of such systems.

In this work we have established a broad range of nonzero \( \gamma \) values relevant not only from a mathematical perspective; Both symplectic topology and Kähler manifolds are sensitive to the topology rather than local changes in geometry. But also, to the physical systems to which it now connects to. In particular, our results show that \( \gamma = 1 \) correspond to a special case that is associated with the \( \mathbb{C}P^1 \) manifolds relevant across various fields such as coherent states [43], worldline formalism [82], Kähler oscillators [46, 47] and entanglement [44] to name a few. Via geometric quantization methods, our results show that holomorphic polarization leads to \( \gamma \in (0, 1] \). This reveals a further array of values of interest outside of the conventional \( \gamma = 0 \) that characterizes the conventional dually-flat, Shannon systems.

B. Unveiling the impact of constraints and dynamics on thermodynamic relationships

From the point of view of thermodynamics, the Rényi entropy allows a natural description of systems where the internal energy corresponds to a Kolmogorov-Nagumo average. In practice, this means that the relevant constraints that characterise the system are non-linear. We found that the generalized Legendre transform of Rényi entropy leads to the free entropy, which is in turn directly related to the entropy production of the system. Furthermore, our results show that most thermodynamic relations, including the second law of thermodynamics and fluctuation theorems, can be naturally extended for these systems when their relationships are adequately recast in terms of the deformed Legendre transform.

It is relevant to mention that generalized entropies — as the Rényi entropy — do not only emerge in the case of non-arithmetic means as constraints. As discussed in Ref. [75], generalized entropies also naturally emerge in systems that obey a non-linear master equation with ordinary arithmetic average (e.g. as the constraint for the internal energy). Perhaps surprising, it has been argued that in equilibrium thermodynamics different entropies and constraints can lead to the same equilibrium distribution [83]. Similarly, for the case of non-equilibrium processes, our results imply that a combination of particular dynamics (exact form of the Fokker-Planck/master equation), detailed balance (connecting stationary distribution with the equilibrium distribution), and energetic constraint can also lead to the same entropic functional. The exact combination of thermodynamic and dynamic relations depends on the particular choice of physical system. The investigation of how to better characterise equivalent classes of dynamics and constraints that lead
to similar phenomena is an important topic that deserves further investigation.

C. Departing thoughts

The results presented here establish a first step in uncovering the consequences of the relationship between generalized Legendre transforms and curved statistical manifolds have for physical systems. We hope that this investigation may foster future work on these important implications, which may reveal other hidden threads connecting seemingly dissimilar approaches — such as the one revealed here relating non-Shannon entropies and non-cannonical coordinates. Such investigations may lead towards a principled and unified understanding of physical systems that are not well-described by traditional approaches, providing solid foundations to support and guide some of today’s effective but ad-hoc procedures of analysis.

ACKNOWLEDGMENTS

P.A.M. was supported by GoodAI for this project. J.K. acknowledges support by the Austrian Science Fund (FWF) project No. P 34994. F.E.R. is supported by the Ad Astra Chandaria foundation.

Appendix A: Complex Polarizations

This appendix illustrates the method of holomorphic polarization, which establishes a intimate relation between link functions and natural families. For a given Kähler manifold one can choose a polarization. A holomorphic polarization has the consequence that physical states are represented as holomorphic functions, and thereby generalizing Bargmann-Segal’s (Fock) spaces that are relevant to coherent states. The complex polarization is a condition determined by

$$\nabla_a \psi = \left( \partial_a + \frac{1}{2} \partial_a \mathcal{K}(z, \bar{z}) \right) \psi = 0,$$  
(A1)

where the connection is determined by the Kähler potential over the manifold. This polarization implies that the commutator $[\nabla_a, \nabla_b] = 0$ and, hence the system described at (A1), is integrable. It’s general solution is given by

$$\psi_{\text{phys}} = \exp\left[ - \frac{1}{2} \mathcal{K}(z, \bar{z}) \right] \phi(z).$$  
(A2)

In the context of statistical manifolds, $\mathcal{K}(z, \bar{z})$ corresponds to a link function $C(z, \bar{z})$. Therefore, Eq. (A2) corresponds to a natural family of distributions — e.g., the flat geometry $C^0$ is described by $C(z, \bar{z}) = z^a \bar{z}^a$ which leads to the exponential family, whereas a link function of the form of Eq. (53) yields Rényi’s natural family. The resulting physical Hilbert space is

$$\mathcal{H}_{\text{phys}} = \left\{ \phi(z) \left| \int_{\mathcal{M}} |\phi|^2 e^{-C(z, \bar{z})} \omega^n < \infty \right. \right\},$$  
(A3)

where $\omega^n$ denotes the manifold’s volume form. In other words, one considers square-integrable global sections that are covariantly constant along $\nabla_a$.

Appendix B: Proof of the second law of thermodynamics for systems described by Kolmogorov-Nagumo averages

For simplicity, let us consider a scenario of pure relaxation, where no work is exerted on the system — thus, $\epsilon_i$ is time-independent. The time derivative of entropy and heat rate over temperature can be found to be

$$\dot{R}_\gamma = \frac{1}{\gamma} \sum_{ij} \left( w_{ij} p_j - w_{ji} p_i \right),$$  
$$\frac{\beta Q^\gamma_i}{\gamma} = \frac{1}{\gamma} \sum_{ijk} \left( w_{ij} p_j - w_{ji} p_i \right) \frac{e^{\gamma \beta_{i}}}{\sum_k p_k e^{\gamma \beta_k}}$$  

where $i = 1, 2, \ldots, n$.

Above, the explicit dependence of $p_i(t)$ on time is omitted for simplicity. The entropy production rate can be then expressed as

$$\dot{\Sigma}_\gamma = \frac{1}{\gamma} \sum_{ij} \left( w_{ij} p_j - w_{ji} p_i \right) \left[ \frac{p_i^{-\gamma}}{\sum_k p_k^{-\gamma}} - \frac{e^{\gamma \beta_{i}}}{\sum_k p_k e^{\gamma \beta_k}} \right]$$
$$\times \left[ \frac{p_i^{-\gamma}}{\sum_k p_k^{-\gamma}} - \frac{e^{\gamma \beta_{i}}}{\sum_k p_k e^{\gamma \beta_k}} \right]$$  
(B3)

Furthermore, by using Eq. (68) one can find that

$$\sum_k \pi_k e^{\gamma \beta_k} \pi_i^{-\gamma} = \frac{\pi_i^{-\gamma}}{\sum_k \pi_k e^{\gamma \beta_k}}.$$  
(B4)

Let us now define $P_i = \frac{\pi_i^{-\gamma}}{\sum_j \pi_j \pi_i^{-\gamma}}$ and $\Pi_i = \frac{\pi_i^{-\gamma}}{\sum_j \pi_j \pi_i^{-\gamma}}$ as shorthand notations. We can then find that

$$\dot{\Sigma}_\gamma = \frac{1}{\gamma} \sum_{ij} w_{ij} p_j \left( \frac{\Pi_i P_j}{P_i} \right)^{-1/\gamma} - 1 \left( P_j - \Pi_j \right).$$  
(B5)
By using inequality $\gamma(x^{-1/\gamma} - 1) \geq \log(1/x)$ for $\gamma > 0$ we obtain

$$\sum_{x} \geq \frac{1}{\gamma} \sum_{ij} w_{ij} p_i \log \left( \frac{P_i}{P_j} \right) (P_i - P_j) =: \mathcal{J}$$

which can be expressed as

$$\mathcal{J} = \frac{1}{\gamma} \sum_{ij} w_{ij} p_i \log \left( \frac{w_{ij} p_i}{w_{ij} p_j} \right) (P_i - P_j)$$

By then using $\log(1/x) \geq 1 - x$, one can obtain the following inequality:

$$\mathcal{J} \geq \sum_{ij} w_{ij} p_i (P_i - \Pi_i) - \sum_{ij} w_{ij} p_j (P_j - \Pi_j)$$

Above, the first term is zero since $\sum_j w_{ij} = 0$, and the second term is equal to $-\Sigma_{\gamma}$. Thus, one finally finds that $\Sigma_{\gamma} \geq -\Sigma_{\gamma}$, which in turn leads to $\Sigma_{\gamma} \geq 0$.

[1] Dusa McDuff and Dietmar Salamon, *Introduction to symplectic topology*, Vol. 27 (Oxford University Press, 2017).

[2] David M Jackson, Achim Kempf, and Alejandro H Morales, “A robust generalization of the legendre transform for qft,” Journal of Physics A: Mathematical and Theoretical 50, 225201 (2017).

[3] Olga Krupková and Dana Smetanová, “Legendre transformation for regularizable lagrangians in field theory,” Letters in Mathematical Physics 87, 901–910 (2010).

[4] Shun-ichi Amari, *Information geometry on hierarchy of probability distributions* (Cambridge university press, 2004).

[5] S-I Amari, “Information geometry on hierarchy of probability distributions,” IEEE transactions on information theory 47, 1701–1711 (2001).

[6] Ting-Kam Leonard Wong, “Logarithmic divergences from optimal transport and Rényi geometry,” Information Geometry 1, 39–78 (2018).

[7] Pablo A. Morales and Fernando E. Rosas, “Generalization of the maximum entropy principle for curved statistical manifolds,” Phys. Rev. Research 3, 033216 (2021).

[8] Jean-Marie Stéphan, Stephen Inglis, Paul Fendley, and Roger G. Melko, “Geometric mutual information at classical critical points,” Physical review letters 112, 127204 (2014).

[9] Jean-Marie Stéphan, “Shannon and Rényi mutual information in quantum critical spin chains,” Phys. Rev. B 90, 045424 (2014).

[10] Xi Dong, “The Gravity Dual of Renyi Entropy,” Nature Comm. 7, 12472 (2016), arXiv:1601.06788 [hep-th].

[11] Taylor Barrella, Xi Dong, Sean A. Hartnoll, and Victoria L. Martin, “Holographic entanglement beyond classical gravity,” JHEP 09, 109 (2013), arXiv:1306.4682 [hep-th].

[12] Petr Jizba and Jan Korbel, “Maximum entropy principle in statistical inference: Case for non-Shannonian entropies,” Physical review letters 122, 120601 (2019).

[13] Jean-Marie Stéphan, Stephen Inglis, Paul Fendley, and Roger G. Melko, “Geometric Mutual Information at Classical Critical Points,” Phys. Rev. Lett. 112, 127204 (2014).

[14] Jason Iaconis, Stephen Inglis, Ann B. Kallin, and Roger G. Melko, “Detecting classical phase transitions with renyi mutual information,” Phys. Rev. B 87, 195134 (2013).

[15] Michael P. Zaletel, Jens H. Bardarson, and Joel E. Moore, “Logarithmic Terms in Entanglement Entropies of 2D Quantum Critical Points and Shannon Entropies of Spin Chains,” Phys. Rev. Lett. 107, 020402 (2011).

[16] Petr Jizba and Toshihiko Arimitsu, “The world according to rényi: thermodynamics of multifractal systems,” Annals of Physics 312, 17–59 (2004).

[17] P. Jizba and T. Arimitsu, “Observability of rényi’s entropy,” Physical Review E 69, 026128 (2004).

[18] Royce KP Zia, Edward F Redish, and Susan R McKay, “Making sense of the legendre transform,” American Journal of Physics 77, 614–622 (2010).

[19] Similarly, the Fourier transform of a time series $F(t)$ can be thought of as giving two outputs too: the spectrum $s(t)$ (analogous to the conjugate function) and the frequency domain $f(t)$ (analogous to the dual variable).

[20] Stephen Boyd, Stephen P. Boyd, and Lieven Vandenberghe, *Convex optimization* (Cambridge university press, 2004).

[21] Shun-ichi Amari, *Convex optimization* (Cambridge university press, 2004).

[22] Shun-ichi Amari, Shiro Ikeda, and Hidetoshi Shimokawa, “Information geometry and its applications, Vol. 194 (Springer, 2016).

[23] Eq. (6) holds on any manifold but only locally. In contrast, dually-flat spaces are a special case in which dual potentials $\varphi, \psi$ that satisfy Eqs. (4) and (5) can be defined over the whole manifold.
“Information geometry of α-projection in mean-field approximation,” Recent Developments of Mean Field Approximation, M. Opper, D. Saad, Eds., MIT Press, Cambridge (2000).

[9] Shun-ichi Amari and Andrzej Cichocki, “Information geometry of divergence functions,” Bulletin of the polish academy of sciences. Technical sciences 58, 183–195 (2010).

[10] Takeo Matumoto et al., “Any statistical manifold has a contrast function—On the C3-functions taking the minimum at the diagonal of the product manifold,” Hiroshima Math. J 23, 327–332 (1993).

[11] Nihat Ay and Shun-ichi Amari, “A novel approach to canonical divergences within information geometry,” Entropy 17, 8111–8129 (2015).

[12] F. Liese and I. Vajda, “On Divergences and Informations in Statistics and Information Theory,” IEEE Transactions on Information Theory 52, 4394–4412 (2006).

[13] Shun-ichi Amari, “α-Divergence is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes,” IEEE Trans. Inf. Theor. 55, 4925–4931 (2009).

[14] NN Chentsov, “Statistical decision rules and optimal inference. Transl. Math.,” Monographs, American Mathematical Society, Providence, RI (1982).

[15] Nihat Ay, Jürgen Jost, Hồng Vân Lê, and Lorenz Schwachhöfer, “Information geometry and sufficient statistics,” Probability Theory and Related Fields 162, 327–364 (2015).

[16] Hồng Vân Lê, “The uniqueness of the Fisher metric as information metric,” Annals of the Institute of Statistical Mathematics 69, 879–896 (2017).

[17] James G Dowty, “Chentsov’s theorem for exponential families,” Information Geometry 1, 117–135 (2018).

[18] Nikolai Nikolaevich Cencov, Statistical decision rules and optimal inference, 53 (American Mathematical Soc., 2000).

[19] Shun-ichi Amari, “Differential geometry of curved exponential families-curvatures and information loss,” The Annals of Statistics 35, 375–385 (1982).

[20] This also explains the asymmetry that exists in the arguments of a Bregman divergence.

[21] A. Rényi, Selected Papers of Alfréd Rényi, Selected Papers of Alfréd Rényi No. 2 (Akadémiai Kiadó, 1976).

[22] Note that we follow Ref. [84] in adopting a shifted indexing.

[23] EA Kochetov, “Su (2) coherent-state path integral,” Journal of Mathematical Physics 36, 4667–4679 (1995).

[24] Dorje C Brody and Lane P Hughston, “Geometric quantum mechanics,” Journal of geometry and physics 38, 19–53 (2001).

[25] Krzysztof Gawędzki, “Non-compact wzw conformal field theories,” in New Symmetry Principles in Quantum Field Theory (Springer, 1992) pp. 247–274.

[26] Stefano Bellucci and Armen Nersessian, “(super) oscillator on cp5 and a constant magnetic field,” Phys. Rev. D 67, 065013 (2003).

[27] Stefano Bellucci and Armen Nersessian, “Erratum: (super) oscillator on cp5 and a constant magnetic field [phys. rev. d 67, 065013 (2003)],” Phys. Rev. D 71, 089901 (2005).

[28] Nicholas Michael John Woodhouse, Geometric quantization (Oxford university press, 1997).

[29] Sean Bates and Alan Weinstein, Lectures on the Geometry of Quantization, Vol. 8 (American Mathematical Soc., 1997).

[30] Vladimir Igorevich Arnol’d, Mathematical methods of classical mechanics, Vol. 60 (Springer Science & Business Media, 2013).

[31] As a consequence, in this section ξ and η refer to different points in the manifold unless it is explicitly specified to be otherwise.

[32] Jun Zhang and Fubo Li, “Symplectic and Kähler structures on statistical manifolds induced from divergence functions,” in International Conference on Geometric Science of Information (Springer, 2013) pp. 596–603.

[33] Please note that the connections appearing from the covariant derivatives cancels out, which is why it is often found written in terms of partial derivatives in spite of being a tensor.

[34] If it were ω form would be exact and so its manifold’s volume form implying that vanishing of its integral, violating the non-degeneracy condition for the metric.

[35] Philip Candelas, “Lectures on complex manifolds,” in Superstrings and grand unification (1988).

[36] Vincent Bouchard, “Lectures on complex geometry, calabi-yau manifolds and toric geometry,” arXiv preprint hep-th/0702063 (2007).

[37] Mikio Nakahara, Geometry, topology and physics (CRC press, 2018).

[38] Bo Berndtsson, “Convexity on the space of Kähler metrics,” Annales de la Faculté des sciences de Toulouse: Mathématiques Ser. 6, 22, 713–746 (2013).

[39] Jun Zhang, “Divergence functions and geometric structures they induce on a manifold,” in Geometric Theory of Information (Springer, 2014) pp. 1–30.

[40] For an overview on CPn spaces, please refer to Refs. [55–57].

[41] Wei-Min Zhang, Da Hsuan Feng, and Robert Gilmore, “Coherent states: Theory and some applications,” Rev. Mod. Phys. 62, 867–927 (1990).

[42] Here the index takes only one entry a = 1, with trivial generalization to CPn.

[43] Not to be confused with the manifolds complex dimension n.

[44] A Plastino and AR Plastino, “On the universality of thermodynamics’ legendre transform structure,” Physics Letters A 226, 257–263 (1997).

[45] Kolmogorov A.N., “Sur la notion de la moyenne,” Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. 12, 388–391 (1930).

[46] Nagumo M., “Über eine klasse der mittelwerte,” Japanese Journal of Mathematics 7, 71 – 79 (1930).

[47] Godfrey Harold Hardy, John Edensor Littlewood, George Pólya, György Pólya, et al., Inequalities (Cambridge university press, 1952).

[48] Marek Czachor and Jan Naudts, “Thermostatistics based on kolmogorov–nagumo averages: unifying framework for extensive and nonextensive generalizations,” Physics Letters A 298, 369–374 (2002).

[49] A.M. Scarfone, H. Matsuzoe, and T. Wada, “Consistency of the structure of legendre transform in thermodynamics with the kolmogorov–nagumo average,” Physics Letters A 380, 3022–3025 (2016).

[50] Ralph VL Hartley, “Transmission of information 1,” Bell System technical journal 7, 535–563 (1928).

[51] Robin AA Ince, “Measuring multivariate redundant information with pointwise common change in surprisal,” Entropy 19, 318 (2017).
This is not surprising, since Rényi entropy appears in the Campbell coding theorem \cite{Campbell1965} as the minimal price that one must pay to encode a message, where the price is the exponential function of the message length.

John Parker Burg, “The relationship between maximum entropy spectra and maximum likelihood spectra,” Geophysics 37, 375–376 (1972).

Udo Seifert, “Stochastic thermodynamics, fluctuation theorems and molecular machines,” Reports on progress in physics 75, 126001 (2012).

Jan Korbel and David H Wolpert, “Stochastic thermodynamics and fluctuation theorems for non-linear systems,” New Journal of Physics 23, 033049 (2021).

Massimiliano Esposito and Christian Van den Broeck, “Second law and landauer principle far from equilibrium,” EPL (Europhysics Letters) 95, 40004 (2011).

Juan MR Parrondo, Jordan M Horowitz, and Takahiro Sagawa, “Thermodynamics of information,” Nature physics 11, 131–139 (2015).

Massimiliano Esposito and Christian Van den Broeck, “Three detailed fluctuation theorems,” Physical review letters 104, 090601 (2010).

F. O. Kazinski, “Stochastic deformation of a thermodynamic symplectic structure,” Phys. Rev. E 79, 011105 (2009).

C. Duval, Z. Horvath, P. A. Horvathy, L. Martina, and P. Stichel, “Berry phase correction to electron density in solids and ’exotic’ dynamics,” Mod. Phys. Lett. B 20, 373–378 (2006), arXiv:cond-mat/0506051.

Dam Thanh Son and Naoki Yamamoto, “Berry curvature, triangle anomalies, and the chiral magnetic effect in fermi liquids,” Phys. Rev. Lett. 109, 181602 (2012).

Patrick Copinger and Pablo Morales, “Schwinger pair production in SL(2,C) topologically nontrivial fields via non-abelian worldline instantons,” Phys. Rev. D 103, 036004 (2021).

Jan Korbel, “Calibration invariance of the maxent distribution in the maximum entropy principle,” Entropy 23 (2021), 10.3390/e23010096.

Francisco Valverde-Albacete and Carmen Peláez-Moreno, “The Case for Shifting the Renyi Entropy,” Entropy 21, 46 (2019).

L.L. Campbell, “A coding theorem and rényi’s entropy,” Information and Control 8, 423 – 429 (1965).