

Article

# Symplectic Bregman Divergences

Frank Nielsen 

Sony Computer Science Laboratories Inc., Tokyo 141-0022, Japan; frank.nielsen.x@gmail.com

**Abstract:** We present a generalization of Bregman divergences in finite-dimensional symplectic vector spaces that we term symplectic Bregman divergences. Symplectic Bregman divergences are derived from a symplectic generalization of the Fenchel–Young inequality which relies on the notion of symplectic subdifferentials. The symplectic Fenchel–Young inequality is obtained using the symplectic Fenchel transform which is defined with respect to the symplectic form. Since symplectic forms can be built generically from pairings of dual systems, we obtain a generalization of Bregman divergences in dual systems obtained by equivalent symplectic Bregman divergences. In particular, when the symplectic form is derived from an inner product, we show that the corresponding symplectic Bregman divergences amount to ordinary Bregman divergences with respect to composite inner products. Some potential applications of symplectic divergences in geometric mechanics, information geometry, and learning dynamics in machine learning are touched upon.

**Keywords:** dual system; duality product; inner product; symplectic form; symplectic matrix group; symplectic subdifferential; symplectic Fenchel transform; Moreau proximation; geometric mechanics

## 1. Introduction

Symplectic geometry [1–3] was historically pioneered by Lagrange around 1808–1810 [4–6] where the motions and dynamics (evolution curves) of a finite set of  $m$  point mass particles in a time interval  $T$  are analyzed in the phase space by a 1D curve  $\mathcal{C} = \{c(t) = (q_1, p_1, \dots, q_m, p_m)(t) : t \in T \subset \mathbb{R}\} \subset \mathbb{R}^{2n}$ , where  $q_i(t) \in \mathbb{R}^n$ 's denote the point locations at time  $t$  and  $p_i(t) \in \mathbb{R}^n$ 's encode the momentum, i.e.,  $p_i(t) = m_i \dot{q}_i$  with  $\dot{q}_i = \frac{d}{dt} q_i(t)$ . See Figure 1. (Notice that Joseph-Louis Lagrange (1736–1813) was 72 years old in 1808, and is famous for his treatise on analytic mechanics [7,8] published first in french in 1788 when he was 52 years old).

The Hamiltonian coupled equations [9] governing the system motion are written in the phase space as follows:

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}, \quad (1)$$

where  $H(q, p, t)$  is the Hamiltonian describing the system. Lagrange originally started a new kind of calculus, “symplectic calculus”. Symplectic geometry can be thought as the first discovered non-Euclidean geometric structure since hyperbolic geometry is usually considered to be first studied by Lobachevsky and Bolyai around 1820–1930. We refer to the paper entitled “The symplectization of science” [10] for an outreach article on symplectic geometry.

The adverb “symplectic” stems from Greek: It means “braided together” to convey the interactions of point mass particle positions with their momenta. Its use in mathematics originated in the work of Hermann Weyl (see §6 on symplectic groups in [11]). Another synonym adverb of symplectic is “complex” which has been used to describe braided numbers  $z$  of  $\mathbb{C} = \{z = a + ib : (a, b) \in \mathbb{R}^2\}$ . Complex has its etymological root in Latin. In differential geometry, symplectic structures are closely related to (almost) complex structures on vector spaces and smooth manifolds [2].



**Citation:** Nielsen, F. Symplectic Bregman Divergences. *Entropy* **2024**, *26*, 1101. <https://doi.org/10.3390/e26121101>

Academic Editors: Stéphane Puechmorel and Florence Nicol

Received: 26 August 2024

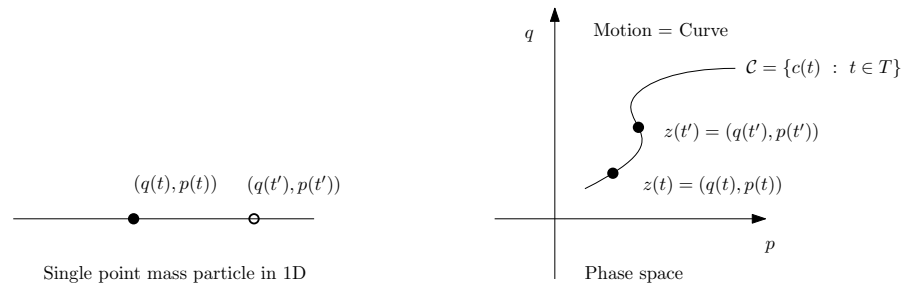
Revised: 5 December 2024

Accepted: 15 December 2024

Published: 16 December 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



**Figure 1.** The motion of a single point particle  $q(t)$  with mass  $m$  and momentum  $p(t) = m\dot{q}(t)$  on a 1D line can be modeled as a curve  $\mathcal{C} = \{c(t) = (q(t), p(t)) : t \in T \subset \mathbb{R}\}$  in the phase space  $\mathbb{R}^2$ .

In physics, symplectic geometry is not only at the core of classical mechanics (i.e., conservative reversible mechanics) and quantum mechanics [12], but has also recently been used to model and study dynamics of systems exhibiting dissipative terms [13,14] which are irreversible. As a pure geometry, symplectic geometry can be studied on its own by mathematicians, and gave birth to the field of symplectic topology [15]. Thus, symplectic geometry can be fruitfully applied to various areas beyond its original domain of geometric mechanics. For example, symplectic geometry has been considered in machine learning for accelerating numerical optimization methods based on symplectic integrators [16] and in physics-informed neural networks [17,18] (PINNs).

In this paper, we define symplectic Bregman divergences (Definition 5) which recover as special cases Bregman divergences [19] defined with respect to composite inner products. A Bregman divergence induced by a strictly convex and differentiable (potential) function  $F$  (called the Bregman generator) between  $x_1$  and  $x_2$  of  $X$  is defined in [19] (1967) by

$$B_F(x_1 : x_2) = F(x_1) - F(x_2) - \langle x_1 - x_2, \nabla F(x_2) \rangle, \tag{2}$$

where  $\langle \cdot, \cdot \rangle$  is an inner product on  $X$ . Let  $\Gamma_0(X)$  denote the set of functions which are lower semi-continuous convex with non-empty effective domains. The convex conjugate  $F^*(x^*)$  obtained by the Legendre–Fenchel transform  $F^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle - F(x)$  yields a dual Bregman divergence  $B_{F^*}$  when the function  $F \in \Gamma_0(X)$  is of Legendre type [20,21]:

$$B_{F^*}(x_1^* : x_2^*) = F^*(x_1^*) - F^*(x_2^*) - \langle x_1^* - x_2^*, \nabla F^*(x_2^*) \rangle,$$

such that  $B_F(x_1 : x_2) = B_{F^*}(x_2^* : x_1^*)$  with

$$F^*(x^*) = \langle x^*, (\nabla F)^{-1}(x^*) \rangle - F((\nabla F)^{-1}(x^*)).$$

This paper introduces and extends the work of Buliga and Saxcé [13,14] which is motivated by geometric irreversible mechanics. To contrast with [13,14], this expository paper is targeted to an audience familiar with Bregman divergences [19] in machine learning and information geometry [22] but does not assume any prior knowledge in geometric mechanics. Furthermore, we consider only finite-dimensional spaces in this study.

The paper is organized as follows: In Section 2, we define symplectic vector spaces and explain the representation of symplectic forms using dual pairings. We then define the symplectic Fenchel transform and the symplectic Fenchel–Young inequality in Section 3. The definitions of symplectic Fenchel–Young divergences (Definition 4) and symplectic Bregman divergences (Definition 5) are reported in Section 4. In particular, we show how to recover Bregman divergences with respect to composite inner products as special cases in Section 5 (Property 1). In general, symplectic Bregman divergences allow one to define Bregman divergences in dual systems equipped with pairing products. Finally, we recall the role of Bregman divergences in dually flat manifolds of information geometry in Section 6, and motivate the introduction of symplectic Bregman divergences in geometric mechanics (e.g., symplectic BEN principle of [13,14]) and learning dynamics in machine learning.

## 2. Dual Systems, Linear Symplectic Forms, and Symplectomorphisms

### 2.1. Symplectic Forms Derived from Dual Systems

We begin with two definitions:

**Definition 1** (Dual system). Let  $X$  and  $Y$  be finite  $m$ -dimensional vector spaces [23] equipped with a pairing product  $b(\cdot, \cdot)$ , i.e., a bilinear map:

$$b(\cdot, \cdot) : X \times Y \rightarrow \mathbb{R},$$

such that all continuous linear functionals on  $X$  and  $Y$  are expressed as  $x^\#(\cdot) = b(x, \cdot)$  and  $y^\#(\cdot) = b(\cdot, y)$ , respectively. The triplet  $(X, Y, b(\cdot, \cdot))$  forms a dual system.

(Notice that when the type of  $X$  is different from the type of  $Y$  then the bilinear map cannot be symmetric).

**Definition 2** (Symplectic vector space). A symplectic vector space  $(V, \omega)$  is a vector space equipped with a map [24]  $\omega : Z = V \times V \rightarrow \mathbb{R}$  which is

1. bilinear:  $\forall \alpha, \beta, \alpha', \beta' \in \mathbb{R}, \forall z_1, z_2 \in Z$ , we have

$$\omega(\alpha z_1 + \alpha' z'_1, \beta z_2 + \beta' z'_2) = \alpha \beta \omega(z_1, z_2) + \alpha \beta' \omega(z_1, z'_2) + \alpha' \beta \omega(z'_1, z_2) + \alpha' \beta' \omega(z'_1, z'_2),$$

2. skew-symmetric (or alternating):  $\omega(z_2, z_1) = -\omega(z_1, z_2)$ , and

3. non-degenerate: if for a  $z_0$ , we have  $\omega(z, z_0) = 0$  for all  $z \in Z$  then we have  $z_0 = 0$ .

Notice that skew-symmetry implies that  $\omega(z, z) = 0$  for all  $z \in Z$  since  $\omega(z, z) = -\omega(z, z)$  and hence  $2\omega(z, z) = 0$ . The map  $\omega$  is called a linear symplectic form [24,25].

We define the symplectic form  $\omega$  induced by the pairing product of a dual system as follows:

$$\omega(z_1, z_2) = b(x_1, y_2) - b(x_2, y_1), \tag{3}$$

where  $z_1 = (x_1, l_1)$  and  $z_2 = (x_2, l_2)$  belong to  $Z = X \oplus Y$ .

Let us report several examples of linear symplectic forms:

- Let  $X = V$  be a finite  $n$ -dimensional vector space with the dual space of linear functionals  $Y = V^*$  (space of covectors  $l$ ). The natural pairing  $((x, l)) = l(x) = \sum_i x^i l_i$  of a vector  $x \in V$  with a covector  $l \in V^*$  is an example of dual product. (We use the superscript index for indicating components of contravariant vectors and subscript index for specifying components of covariant vectors [9]). We define the symplectic form  $\omega$  induced by the natural pairing of vectors with covectors as follows:

$$\omega(z_1, z_2) = ((x_1, l_2)) - ((x_2, l_1)) = l_2(x_1) - l_1(x_2), \tag{4}$$

where  $z_1 = (x_1, l_1)$  and  $z_2 = (x_2, l_2)$  belong to  $Z = V \oplus V^*$ .

- Consider  $(X, \langle \cdot, \cdot \rangle)$  an inner product space of dimension  $n$ . The product space  $Z = X \oplus X$  of even dimension  $n$  can be equipped with the following map  $\omega : Z \times Z \rightarrow \mathbb{R}$  induced by the inner product:

$$\omega(z_1, z_2) = \langle x_1, y_2 \rangle - \langle x_2, y_1 \rangle, \tag{5}$$

where  $z_1 = (x_1, y_1) \in Z$  and  $z_2 = (x_2, y_2) \in Z$ .

For example, let  $X = \mathbb{R}$  and  $\langle x_1, x_2 \rangle = x_1 x_2$ . Then  $\omega(z_1, z_2) = x_1 y_2 - x_2 y_1$ . This symplectic form can be interpreted as the determinant of the matrix  $M = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$  which corresponds geometrically to the signed orientation of the parallelogram defined by the vectors  $z_1 = (x_1, y_1)$  and  $z_2 = (x_2, y_2)$ . See Figure 2. (This example indicates the

link with integration of 2D manifolds equipped with fields of symplectic forms smoothly varying called differential 2-forms [9]).

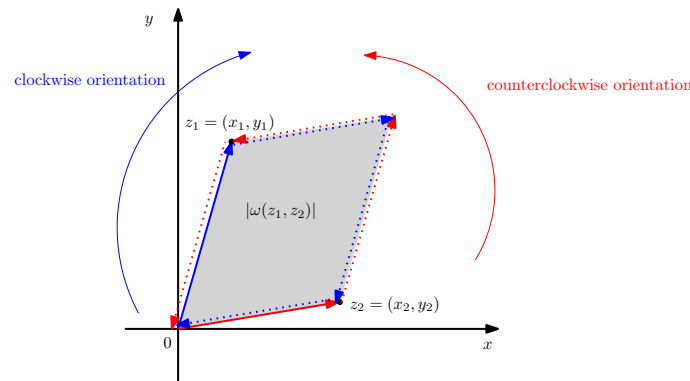
In a finite-dimensional vector space, we can express the inner product as  $\langle x, y \rangle = x^\top Q y$  for a symmetric positive-definite matrix  $Q \in \mathbb{R}^{n \times n}$ . Let  $Q = L^\top L$  be the Cholesky decomposition of  $Q$ . Then we have

$$\langle x, y \rangle = (L^\top x)^\top I (L^\top y) = \langle L^\top x, L^\top y \rangle_0,$$

where  $I$  is the  $n \times n$  identity matrix and  $\langle x, y \rangle_0 = x^\top y$  is the Euclidean inner product. Thus the form  $\omega_0$  induced by  $\langle \cdot, \cdot \rangle_0$  can be expressed using linear algebra as

$$\omega_0(z_1, z_2) = z_1^\top \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} z_2 = z_1^\top \Omega_0 z_2,$$

where  $\Omega_0 \in \mathbb{R}^{2n \times 2n}$  is a skew-symmetric matrix:  $\Omega_0^\top = -\Omega_0$ . More generally, we may consider skew-symmetric matrices of the form  $\Omega = \begin{bmatrix} 0 & L^\top \\ -L^\top & 0 \end{bmatrix}$  to define the symplectic form  $\omega_Q$  induced by the inner product  $\langle x, y \rangle_Q = x^\top Q y$ .



**Figure 2.** Interpreting a 2D symplectic form  $\omega(z_1, z_2)$  as the signed area of a parallelogram with first oriented edge  $z_1$  (grey). A pair of vectors defines two possible orientations of the parallelogram: The orientation compatible with  $z_1$  and the reverse orientation compatible with  $z_2$ .  $\omega$  is called the standard area form.

### 2.2. Linear Symplectomorphisms and the Groups of Symplectic Matrices

A symplectic form  $\omega$  can be expressed as a  $2n \times 2n$  matrix  $\Omega = [\omega_{ij}]$  such that  $\omega_{ij} = \omega(b_i, b_j)$  where  $b_1 = e_1, \dots, b_n = e_n, b_{n+1} = f_1, \dots, b_{2n} = f_n$  are the basis vectors, and  $\omega(z_1, z_2) = z_1^\top \Omega z_2$ .

The Darboux basis [2] of the canonical form  $\omega_0$  of  $\mathbb{R}^{2n}$  is such that  $\omega(e_i, f_j) = \delta_{ij}$  and  $\omega(e_i, e_j) = \omega(f_i, f_j) = 0$  where  $\delta$  denotes the Kronecker delta function.  $\Omega_0 \in \text{Sp}(2n)$  is the symplectic matrix  $\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$  corresponding to the canonical form  $\omega_0$  of  $\mathbb{R}^{2n}$ .

A transformation  $t : V \rightarrow V$  is called a linear symplectomorphism when  $\omega(t(z_1), t(z_2)) = \omega(z_1, z_2)$  (i.e.,  $t^* \omega = \omega$ ), i.e., when  $T^\top \Omega T = \Omega$  where  $T$  be the matrix representation of  $t$ . In particular  $t$  is a linear symplectomorphism with respect to  $\omega_0$  when  $T^\top \Omega_0 T = \Omega_0$ . Any symplectic vector space of  $(V, \omega)$  dimension  $2n$  is symplectomorphic to the canonical symplectic space  $(\mathbb{R}^{2n}, \omega_0)$ .

Linear symplectomorphisms can be represented by symplectic matrices of the symplectic group [11,26]  $\text{Sp}(2n)$ :

$$\begin{aligned} \text{Sp}(2n) &= \left\{ T : T^\top \Omega_0 T = \Omega_0 \right\} \subset \text{GL}(2n), \\ &= \left\{ T = \begin{bmatrix} A & B \\ C & D \end{bmatrix} : -C^\top A + A^\top C = 0, -C^\top B + A^\top D = I, -D^\top B + B^\top D = 0 \right\}, \end{aligned}$$

Transpose and inverse of symplectic matrices are symplectic matrices. The inverse of a symplectic matrix  $T$  is given by

$$\begin{aligned} T^{-1} &= -\Omega_0 T^\top \Omega_0, \\ &= \begin{bmatrix} D^\top & -B^\top \\ -C^\top & A^\top \end{bmatrix} \end{aligned}$$

Symplectic matrices of  $\text{Sp}(2n)$  have unit determinant ( $\text{Sp}(2n) \subset \text{SL}(2n) \subset \text{GL}(2n)$ ), and in the particular case of  $n = 1$ ,  $\text{Sp}(2)$  corresponds precisely to the set of matrices with unit determinant. Thus rotation matrices of  $\text{SO}(2)$  which have unit determinant form a subgroup of  $\text{Sp}(2)$ .

Sesquilinear symplectic forms can also be defined on complex linear spaces [27].

### 3. Symplectic Fenchel Transform, Symplectic Subdifferentials, and Symplectic Fenchel–Young (in)Equality

Let  $F : Z = X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex lower semi-continuous (lsc) function called a potential function.

**Definition 3** (Symplectic Fenchel conjugate). *The symplectic Fenchel conjugate  $F^{*\omega}(z')$  is defined by*

$$F^{*\omega}(z') = \sup_{z \in Z} \{\omega(z', z) - F(z)\}.$$

Notice that since  $\omega$  is skew-symmetric, the order of the arguments in  $\omega$  is important: The symplectic Fenchel transform optimizes with respect to the second argument of  $\omega(\cdot, \cdot)$ .

The symplectic subdifferential of  $F$  at  $z$  is defined by

$$\partial^\omega F(z) = \{z_1 \in Z : \forall z_2 \in Z, \quad F(z + z_2) \geq F(z) + \omega(z_1, z_2)\}.$$

The differential operator  $\partial^\omega$  is a set-valued operator:  $\partial^\omega : \mathcal{F} \rightrightarrows Z$ , where  $\mathcal{F}$  is the set of potential functions. An element of the symplectic subdifferential of  $F$  at  $z$  is called a symplectic subgradient.

**Remark 1.** *Moreau generalized the Fenchel conjugate using a cost function [28]. In particular, the duality induced by logarithmic cost function was studied in [29], and led to a generalization of Bregman divergences called the logarithmic divergences which are canonical divergences of constant section curvature manifolds in information geometry.*

**Remark 2.** *In geometric mechanics [2], the symplectic gradient on a symplectic manifold  $(M, \omega)$  is the Hamiltonian vector field, i.e., the vector field  $X_H$  such that the Hamiltonian mechanics equation writes concisely as  $\omega(X_H, \cdot) = dH$ .*

**Theorem 1** (Symplectic Fenchel–Young inequality, Theorem 2.3 of [13,14]). *Let  $F(z)$  be a convex (i.e.,  $F(z) = F(x, y)$  is joint convex, i.e., convex with respect to  $z = (x, y)$ ) and lower semi-continuous function. Then the following inequality holds:*

$$\forall z, z' \in Z, \quad F(z) + F^{*\omega}(z') \geq \omega(z', z),$$

*with equality if and only if  $z' \in \partial^\omega(z)$ .*

Let us again notice that the argument order in  $\omega(\cdot, \cdot)$  is important.

Assume that the potential functions are smooth and that symplectic subdifferentials consist only of single-element sets (singletons). By abuse of language, we shall call in this paper the symplectic gradient of  $F$  the single element of the symplectic subdifferential  $\partial^\omega$ , and denote it by  $\nabla^\omega F$ :  $\partial^\omega F(z) = \{\nabla^\omega F(z)\}$ . (Our terminology and notation is thus not to be confused with the Hamiltonian vector field  $X_H$  of geometric mechanics).

#### 4. Symplectic Fenchel–Young Divergences and Symplectic Bregman Divergences

Divergences are smooth dissimilarity functions (see Section 4.2 of [30]). From the symplectic Fenchel–Young inequality of Theorem 1, we can define the symplectic Fenchel–Young divergence as follows:

**Definition 4** (Symplectic Fenchel–Young divergence). *Let  $F : Z = X \times Y \rightarrow \mathbb{R}$  be a smooth convex function. Then the symplectic Fenchel–Young divergence is the following non-negative measure of dissimilarity between  $z$  and  $z'$ :*

$$Y_F(z, z') = F(z) + F^{*\omega}(z') - \omega(z', z) \geq 0. \tag{6}$$

We have  $Y_F(z, z') = 0$  if and only if  $z' \in \partial^\omega F(z)$ , i.e.,  $z' = \nabla^\omega F(z)$  when  $F$  is smooth.

Let us now define the symplectic Bregman divergence  $B_F^\omega(z_1 : z_2)$  as  $Y_F(z_1, z'_2)$  where  $z'_2 = \nabla^\omega F(z_2)$ . Using the following identity derived from the symplectic Fenchel–Young equality:

$$F^{*\omega}(\nabla^\omega F(z)) = \omega(\nabla^\omega F(z), z) - F(z),$$

and the bilinearity of the symplectic form, we obtain:

$$\begin{aligned} B_F^\omega(z_1 : z_2) &= Y_F(z_1, z'_2), \\ &= F(z_1) + F^{*\omega}(z'_2) - \omega(z'_2, z_1), \\ &= F(z_1) + \omega(\nabla^\omega F(z_2), z_2) - F(z_2) - \omega(\nabla^\omega F(z_2), z_1), \\ &= F(z_1) - F(z_2) - \omega(\nabla^\omega F(z_2), z_1 - z_2). \end{aligned} \tag{7}$$

Since  $\omega$  is skew-symmetric, we can also rewrite Equation (7) equivalently as

$$B_F^\omega(z_1 : z_2) = F(z_1) - F(z_2) + \omega(z_1 - z_2, \nabla^\omega F(z_2)). \tag{8}$$

**Definition 5** (Symplectic Bregman divergence). *Let  $(Z = X \times Y, \omega)$  be a symplectic vector space. Then the symplectic Bregman divergence between  $z_1$  and  $z_2$  of  $Z$  induced by a smooth convex potential  $F(z)$  is*

$$B_F^\omega(z_1 : z_2) = F(z_1) - F(z_2) - \omega(\nabla^\omega F(z_2), z_1 - z_2),$$

where the symplectic subdifferential gradient is the singleton  $\partial^\omega F(z) = \{\nabla^\omega F(z)\}$ .

**Remark 3.** *The ordinary Bregman divergences (BDs) have been generalized to non-smooth strictly convex potential functions using a subdifferential map in [31–33] to choose among several potential subgradients at a given location. Similarly, we can extend symplectic Bregman divergences to non-smooth strictly convex potential functions using a symplectic subdifferential map.*

#### 5. Particular Cases Recover Composite Bregman Divergences

When  $Y = X$  and  $(X, \langle \cdot, \cdot \rangle)$  is an inner-product space, we may consider the composite inner-product on  $Z = X \times X$ :

$$\langle\langle z_1, z_2 \rangle\rangle = \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle,$$

with  $z_1 = (x_1, y_1)$  and  $z_2 = (x_2, y_2)$ .

Let  $I : Z \rightarrow Z$  be the linear function  $I(z) = z$  and denote by  $J : Z \rightarrow Z$  the linear function defined by

$$J(z) = J(x, y) = (-y, x).$$

Notice that this definition of  $J$  makes sense because  $X = Y$  and thus  $(-y, x) \in Z$ . We check that we have  $J^2(x, y) = J(-y, x) = (-x, -y) = -(x, y)$ , i.e.,  $J^2 = -I$ . Furthermore, we have  $g(z_1, z_2) = \omega(z_1, Jz_2) = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle$  that is a positive definite inner product. That is, the automorphism  $J$  is a complex structure  $\omega$ -compatible ( $J$  is a symplectomorphism).

We can express the symplectic form  $\omega(z_1, z_2) = \langle x_1, y_2 \rangle - \langle x_2, y_1 \rangle$  induced by the inner product using the composite inner product as follows:

$$\begin{aligned} \omega(z_1, z_2) &= \langle x_1, y_2 \rangle - \langle x_2, y_1 \rangle = \langle \langle J(z_1), z_2 \rangle \rangle, \\ \omega(-Jz_1, z_2) &= \langle \langle z_1, z_2 \rangle \rangle. \end{aligned}$$

Similarly, the symplectic subdifferential of  $F$  can be expressed using the ordinary subdifferential (and vice versa) as follows:

$$\begin{aligned} z' \in \partial^\omega F(z) &\Leftrightarrow J(z') \in \partial F(z), \\ z' \in \partial F(x) &\Leftrightarrow -J(z') \in \partial^\omega F(z). \end{aligned}$$

When subdifferentials are singletons, we thus have

$$\begin{aligned} J(\nabla^\omega F(z)) &= \nabla F(z), \\ \nabla^\omega F(z) &= -J(\nabla F(z)). \end{aligned}$$

Last, the symplectic Fenchel conjugate of  $F$  is related by the ordinary Fenchel conjugate  $F^*$  of  $F$  as follows:

$$F^{*\omega}(z) = F^*(J(z)).$$

Thus in that case the symplectic Bregman divergence amounts to an ordinary Bregman divergence:

$$\begin{aligned} B_F^\omega(z_1 : z_2) &= F(z_1) - F(z_2) - \omega(\nabla^\omega F(z_2), z_1 - z_2), \\ &= F(z_1) - F(z_2) + \omega(-\nabla^\omega F(z_2), z_1 - z_2), \\ &= F(z_1) - F(z_2) + \omega(J(\nabla F(z_2)), z_1 - z_2), \\ &= F(z_1) - F(z_2) - \langle \langle z_1 - z_2, \nabla F(z_2) \rangle \rangle, \\ &= B_F(z_1 : z_2). \end{aligned}$$

**Property 1.** When the symplectic form  $\omega$  is induced by an inner product  $\langle \cdot, \cdot \rangle$  of  $X$ , the symplectic Bregman divergence  $B_F^\omega(z_1 : z_2)$  between  $z_1 = (x_1, y_1)$  and  $z_2 = (x_2, y_2)$  of  $Z = X \times X$  amounts to an ordinary Bregman divergence with respect to the composite inner-product  $\langle \langle z_1, z_2 \rangle \rangle = \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle$ :

$$B_F^\omega(z_1 : z_2) = B_F(z_1 : z_2) = F(z_1) - F(z_2) - \langle \langle z_1 - z_2, \nabla F(z_2) \rangle \rangle.$$

Furthermore, if the potential function  $F(z)$  is separable, i.e.,  $F(z) = F_1(x) + F_2(y)$  for Bregman generators  $F_1$  and  $F_2$ , then we have  $B_F^\omega(z_1 : z_2) = B_{F_1}(x_1 : x_2) + B_{F_2}(y_1, y_2)$  where the Bregman divergences  $B_{F_1}$  and  $B_{F_2}$  are defined with respect to the inner product of  $X$ .

Notice that the symplectic Fenchel–Young inequality can be rewritten using the ordinary Fenchel–Young inequality and the linear function  $J$  as:

$$\begin{aligned} F(z) + F^{*\omega}(z') &\geq \omega(z', z), \\ F(z) + F^*(J(z')) &\geq \langle \langle J(z'), z \rangle \rangle. \end{aligned}$$

### 6. Summary, Discussion, and Perspectives

Since its inception in operations research, Bregman divergences [19] have proven instrumental in many scientific fields including information theory, statistics, and machine learning, just to cite a few. Let  $(X, \langle \cdot, \cdot \rangle)$  be a Hilbert space, and  $F : X \rightarrow \mathbb{R}$  a strictly convex and smooth real-valued function. Then the Bregman divergence induced by  $F$  is defined in [19] (1967) by

$$B_F(x_1 : x_2) = F(x_1) - F(x_2) - \langle x_1 - x_2, \nabla F(x_2) \rangle.$$

In this work, we consider finite-dimensional vector spaces equipped with an inner product.

In information geometry [22,34,35], a smooth dissimilarity  $\mathcal{D}(p, q)$  between two points  $p$  and  $q$  on an  $n$ -dimensional smooth manifold  $M$  induces a statistical structure on the manifold [36], i.e., a triplet  $(g, \nabla, \nabla^*)$  where the Riemannian metric  $g$  and the torsion-free affine connections  $\nabla$  and  $\nabla^*$  are induced by the divergence  $\mathcal{D}$ . The duality in information geometry is expressed by the fact that the mid-connection  $\frac{\nabla + \nabla^*}{2}$  corresponds to the Levi-Civita connection induced by  $g$ . To build the divergence-based information geometry [37], the divergence  $\mathcal{D}(p, q)$  is interpreted as a scalar function on the product manifold  $M \times M$  of dimension  $2n$ . Thus, the divergence  $\mathcal{D}$  is called a contrast function [36] or yoke [38]. Conversely, a statistical structure  $(g, \nabla, \nabla^*)$  on an  $n$ -dimensional manifold  $M$  induces a contrast function [39]. When the statistical manifold  $(M, g, \nabla, \nabla^*)$  is dually flat with  $\theta(\cdot)$  the global  $\nabla$ -affine coordinate system and  $\eta(\cdot)$  the global  $\nabla^*$ -affine coordinate system [40], there exists two dual global potential functions  $\phi$  and  $\phi^*$  on the manifold  $M$  such that  $\phi(p) = F(\theta(p))$  and  $\phi^*(\eta(p)) = F^*(\eta(p))$  where  $F^*(\eta)$  is the Legendre–Fenchel convex conjugate of  $F(\theta)$ . The canonical dually flat divergence on  $M$  is then defined by

$$\mathcal{D}(p, q) = F(\theta(p)) + F^*(\eta(q)) - \sum_{i=1}^n \theta_i(p)\eta_i(q),$$

and amounts to a Fenchel–Young divergence or equivalently a Bregman divergence:

$$\mathcal{D}(p, q) = Y_F(\theta(p) : \theta(q)) = B_F(\theta(p) : \theta(q)),$$

where the Fenchel–Young divergence is defined by

$$Y_F(\theta : \eta') = F(\theta) + F^*(\eta') - \sum_{i=1}^n \theta^i \eta'_i.$$

The Riemannian metric  $g$  of a dually flat space can be expressed as  $g = \nabla d\phi = \nabla^* d\phi^*$  or in the  $\theta$ -coordinates by  $g_{ij}(\theta) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} F(\theta)$  and in the  $\eta$ -coordinates by  $g_{ij}(\eta) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} F^*(\eta)$ . That is,  $g$  is a Hessian metric [40],  $(g, \nabla)$  a Hessian structure and  $(g, \nabla^*)$  a dual Hessian structure. In differential geometry,  $(M, g, \nabla)$  is called a Hessian manifold which admits a dual Hessian structure  $(g, \nabla^*)$ . In particular, a Hessian manifold is of Koszul type [40] when there exists a closed 1-form  $\alpha$  such that  $g = \nabla \alpha$ .

**Remark 4.** Notice that the potential functions  $F$  and  $F^*$  are not defined uniquely although the potential functions  $\phi$  and  $\phi^*$  on the manifold are. Indeed, consider the generator  $\bar{F}(\theta) = F(A\theta + b) + \langle c, \theta \rangle + d$  for invertible matrix  $A \in \text{GL}(d, \mathbb{R})$ , vectors  $b, c \in \mathbb{R}^d$  and scalars  $d \in \mathbb{R}$ . The gradient of the generator  $\bar{F}$  is  $\eta = \nabla \bar{F}(\theta) = A^\top \nabla F(A\theta + b) + c$ . Solving the equation  $\nabla \bar{F}(\theta) = \eta$  yields the reciprocal gradient  $\theta(\eta) = \nabla \bar{G}(\eta) = A^{-1} \nabla G(A^{-\top}(\eta - c)) - b$  from which the Legendre convex conjugate is obtained as  $\bar{G}(\eta) = \langle \eta, \nabla \bar{G}(\eta) \rangle - F(\nabla \bar{G}(\eta))$ . We have  $B_{\bar{F}}(\theta_1 : \theta_1) = B_{\bar{F}}(\bar{\theta}_1 : \bar{\theta}_2)$  where  $\bar{\theta} = A^{-1}(\theta - b)$ .

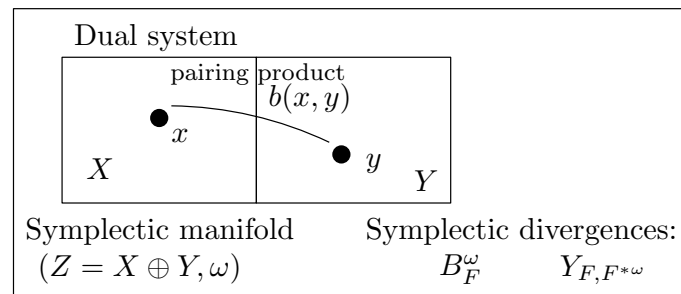
It has been shown that a divergence  $\mathcal{D}$  also allows one to define a symplectic structure  $\omega$  on a statistical manifold [38,41]. The symplectic vector space  $(\mathbb{R}^{2n}, \omega_0)$  viewed as a symplectic manifold has symplectic form  $\omega_0 = \sum_{i=1}^n dx^i \wedge dy_i = -d(\sum_{i=1}^n y_i dx^i)$ . There are no local invariants but only global invariants on symplectic manifolds (symplectic topology). That is, a symplectic structure is flat.

In this expository paper, we have defined symplectic Fenchel–Young divergences and equivalent symplectic Bregman divergences by following the study of geometric mechanics reported in [13,14]. The symplectic Bregman divergence between two points  $z_1$  and  $z_2$  on a symplectic vector space  $(Z, \omega)$  induced by a convex potential function  $F : Z \rightarrow \mathbb{R}$  is defined by

$$B_F^\omega(z_1 : z_2) = F(z_1) - F(z_2) + \omega(z_1 - z_2, \nabla^\omega F(z_2)),$$



where  $\nabla^\omega F$  has been called the symplectic gradient in this paper, and assumed to be the unique symplectic subdifferential at any  $z \in Z$ , i.e.,  $\partial^\omega F(z) = \{\nabla^\omega F(z)\}$ . Symplectic Bregman divergences are used to define Bregman divergences on dual systems (Figure 3). In the particular case of dual system  $(X, X, \langle \cdot, \cdot \rangle)$ , we recover ordinary Bregman divergences with composite inner products.



**Figure 3.** Bregman divergences generalized to dual systems  $(X, Y, b(\cdot, \cdot))$ : A symplectic form  $\omega$  on the space  $Z = X \oplus Y$  is induced by the pairing product. The Bregman divergence on the dual system is then defined as the symplectic Bregman divergence on the symplectic vector space  $(Z, \omega)$ .

In finite  $2n$ -dimensional symplectic vector spaces, linear symplectic forms  $\omega$  can be represented by symplectic matrices of the matrix group  $\text{Sp}(2n)$ . Buliga and de Saxcé [13,14] considered geometric mechanics with dissipative terms, and stated the following “symplectic Brezis–Ekeland–Nayroles principle” (SBEN principle for short):

**Definition 6** (SBEN principle [13,14]). *The natural evolution path  $z(t) = z_{\text{rev}}(t) + z_{\text{irr}}(t) \in Z$  for  $t \in [0, T]$  in a geometric mechanic system with convex dissipation potential  $\phi(z)$  minimizes among all admissible paths  $\int_0^T Y_F^\omega(\dot{z}(t), \dot{z}_{\text{irr}}(t)) dt$  and satisfies  $Y_F^\omega(\dot{z}(t), \dot{z}_{\text{irr}}(t)) = 0$  for all  $t \in [0, T]$ , where  $Y_F^\omega$  denotes the symplectic Fenchel–Young divergence induced by  $\phi$ , and  $z_{\text{rev}}(t)$  and  $z_{\text{irr}}(t)$  are the reversible and irreversible parts of the particle  $z(t)$ , respectively.*

The decomposition of  $z = z_{\text{rev}} + z_{\text{irr}}$  into two parts can be interpreted as Moreau’s proximation [42,43] associated to the potential function  $\phi$ : Indeed, let  $F(z)$  be a convex function of  $Z = \mathbb{R}^d$ . Then for all  $z \in \mathbb{R}^d$ , we can uniquely decompose  $z$  as  $z = z + z^*$  such that  $F(z) + F^*(z^*) = \langle z, z^* \rangle$  (Fenchel–Young equality) where  $z^* = \nabla F(z)$  (see Proposition in Section 4 of [42]). The part  $z$  is called the proximation with respect to  $F$ , and the part  $z^*$  is the proximation with respect to the convex conjugate  $F^*$ .

We may consider the non-separable potential functions  $F(z) = F(x, y) = x f(y/x)$  which are obtained from the perspective transform [44,45] of arbitrary convex functions  $f(u)$  to define symplectic Bregman divergences. The perspective functions  $F(x, y)$  are jointly convex if and only if their corresponding generators  $f$  are convex. Such perspective transforms play a fundamental role in information theory [46] and information geometry [22].

In machine learning, symplectic geometry has been used for designing accelerated optimization methods [16,47] (Bregman–Lagrangian framework) and physics-informed neural networks [17,18] (PINNs).

This paper aims to spur interest in either designing or defining symplectic divergences from first principles, and to demonstrate their roles when studying thermodynamics [48] or the learning dynamics of ML and AI systems.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** Author Frank Nielsen is employed by the company Sony Computer Science Laboratories Inc. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The author declares no conflicts of interest.

## References

1. McDuff, D. Symplectic structures—A new approach to geometry. *Not. AMS* **1998**, *45*, 952–960.
2. Da Silva, A.C. *Lectures on Symplectic Geometry*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 3575.
3. Libermann, P.; Marle, C.M. *Symplectic Geometry and Analytical Mechanics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 35.
4. de Lagrange, J.L. Mémoire sur la théorie des variations des éléments des planètes, et en particulier des variations des grands axes de leurs orbites. *Paris* **1808**, *VI*, 713–768.
5. Lagrange, J.L. Second mémoire sur la théorie générale de la variation des constantes arbitraires dans tous les problèmes de la mécanique. *Mémoires Prem. Cl. l'Institut Fr.* **1810**, *19*, 809–816.
6. Marle, C.M. The inception of symplectic geometry: The works of Lagrange and Poisson during the years 1808–1810. *Lett. Math. Phys.* **2009**, *90*, 3–21. [[CrossRef](#)]
7. Lagrange, J.L. *Mécanique Analytique*; First Published by La Veuve Desaint, Paris in French in 1788 by Joseph-Louis De La Grange with title “Méchanique analitique”; Mallet-Bachelier: Paris, France, 1811.
8. Lagrange, J.L. *Analytical Mechanics*; First Published in French in 1811; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 191.
9. Godinho, L.; Natário, J. An introduction to Riemannian geometry. In *With Applications*; Springer: Berlin/Heidelberg, Germany, 2012.
10. Gotay, M.J.; Isenberg, G. The symplectization of science. *Gaz. Mathématiciens* **1992**, *54*, 59–79. (In French)
11. Weyl, H. *The Classical Groups: Their Invariants and Representations*; Number 1; Princeton University Press: Princeton, NJ, USA, 1946.
12. Souriau, J.M. *Structure of Dynamical Systems: A Symplectic View of Physics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1997; Volume 149.
13. Buliga, M.; de Saxcé, G. A symplectic Brezis–Ekeland–Nayroles principle. *Math. Mech. Solids* **2017**, *22*, 1288–1302. [[CrossRef](#)]
14. de Saxcé, G. A variational principle of minimum for Navier–Stokes equation and Bingham fluids based on the symplectic formalism. In *Information Geometry*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–22.
15. Audin, M. Vladimir Igorevich Arnold and the invention of symplectic topology. In *Contact and Symplectic Topology*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–25.
16. Jordan, M.I. Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018, Rio de Janeiro, Brazil, 1–9 August 2018*; World Scientific: Singapore, 2018; pp. 523–549.
17. Chen, Y.; Matsubara, T.; Yaguchi, T. Neural symplectic form: Learning Hamiltonian equations on general coordinate systems. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 16659–16670.
18. Matsubara, T.; Miyatake, Y.; Yaguchi, T. Symplectic adjoint method for exact gradient of neural ODE with minimal memory. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 20772–20784.
19. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [[CrossRef](#)]
20. Rockafellar, R.T. Conjugates and Legendre transforms of convex functions. *Can. J. Math.* **1967**, *19*, 200–205. [[CrossRef](#)]
21. Bauschke, H.H.; Borwein, J.M.; Combettes, P.L. Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. *Commun. Contemp. Math.* **2001**, *3*, 615–647. [[CrossRef](#)]
22. Amari, S.i. *Information Geometry and Its Applications*; Applied Mathematical Sciences; Springer: Tokyo, Japan, 2016.
23. Horváth, J. *Topological Vector Spaces and Distributions*; Courier Corporation: New York, NY, USA, 2013.
24. McInerney, A. First Steps in Differential Geometry: Riemannian, Contact, Symplectic. In *Undergraduate Texts in Mathematics*; Springer: New York, NY, USA, 2013.
25. Bourguignon, J.P. *Variational Calculus*; Springer: Berlin/Heidelberg, Germany, 2022.
26. Siegel, C.L. *Symplectic Geometry*; Elsevier: Amsterdam, The Netherlands, 1964.
27. Everitt, W.; Markus, L. Complex symplectic geometry with applications to ordinary differential operators. *Trans. Am. Math. Soc.* **1999**, *351*, 4905–4945. [[CrossRef](#)]
28. Moreau, J.J. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Mathématiques Pures Appliquées* **1970**. Available online: <https://hal.science/hal-02162006/> (accessed on 25 August 2024).
29. Wong, T.K.L. Logarithmic divergences from optimal transport and Rényi geometry. *Inf. Geom.* **2018**, *1*, 39–78. [[CrossRef](#)]
30. Leok, M.; Zhang, J. Connecting information geometry and geometric mechanics. *Entropy* **2017**, *19*, 518. [[CrossRef](#)]
31. Kiwiel, K.C. Free-steering relaxation methods for problems with strictly convex costs and linear constraints. *Math. Oper. Res.* **1997**, *22*, 326–349. [[CrossRef](#)]

32. Gordon, G.J. Approximate Solutions to Markov Decision Processes. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 1999.
33. Iyer, R.; Bilmes, J.A. Submodular-Bregman and the Lovász-Bregman divergences with applications. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2933–2941.
34. Nielsen, F. An elementary introduction to information geometry. *Entropy* **2020**, *22*, 1100. [[CrossRef](#)]
35. Nielsen, F. The many faces of information geometry. *Not. Am. Math. Soc* **2022**, *69*, 36–45. [[CrossRef](#)]
36. Eguchi, S. A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Math. J.* **1985**, *15*, 341–391. [[CrossRef](#)]
37. Amari, S.i.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci. Tech. Sci.* **2010**, *58*, 183–195. [[CrossRef](#)]
38. Barndorff-Nielsen, O.E.; Jupp, P.E. Statistics, yokes and symplectic geometry. In *Annales de la Faculté des Sciences de Toulouse: Mathématiques*; Université Paul Sabatier: Toulouse, France, 1997; Volume 6, pp. 389–427.
39. Matumoto, T. Any statistical manifold has a contrast function: On the  $C^3$ -functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J* **1993**, *23*, 327–332. [[CrossRef](#)]
40. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
41. Zhang, J. Divergence functions and geometric structures they induce on a manifold. In *Geometric Theory of Information*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–30.
42. Moreau, J.J. Proximité et dualité dans un espace hilbertien. *Bull. Société Mathématique Fr.* **1965**, *93*, 273–299. [[CrossRef](#)]
43. Rockafellar, R. Integrals which are convex functionals. *Pac. J. Math.* **1968**, *24*, 525–539. [[CrossRef](#)]
44. Dacorogna, B.; Maréchal, P. The role of perspective functions in convexity, polyconvexity, rank-one convexity and separate convexity. *J. Convex Anal.* **2008**, *15*, 271–284.
45. Combettes, P.L. Perspective functions: Properties, constructions, and examples. *Set-Valued Var. Anal.* **2018**, *26*, 247–264. [[CrossRef](#)]
46. Csiszár, I.; Shields, P.C. Information theory and statistics: A tutorial. *Found. Trends<sup>®</sup> Commun. Inf. Theory* **2004**, *1*, 417–528. [[CrossRef](#)]
47. Shi, B.; Du, S.S.; Su, W.; Jordan, M.I. Acceleration via symplectic discretization of high-resolution differential equations. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 5744–5752.
48. Barbaresco, F. Symplectic theory of heat and information geometry. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2022; Volume 46, pp. 107–143.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.