

Article

Generalized Legendre Transforms Have Roots in Information Geometry [†]

Frank Nielsen 

Sony Computer Science Laboratories, 3-14-13 Higashi Gotanda, Shinagawa Ku, Tokyo 141-0022, Japan;
frank.nielsen.x@gmail.com

[†] This paper is dedicated to Professor Shun-ichi Amari in honor of his 90th birthday.

Abstract

Artstein-Avidan and Milman [Annals of mathematics (2009), (169):661–674] characterized invertible reverse-ordering transforms in the space of lower, semi-continuous, extended, real-valued convex functions as affine deformations of the ordinary Legendre transform. In this work, we first prove that all those generalized Legendre transforms of functions correspond to the ordinary Legendre transform of dually corresponding affine-deformed functions. In short, generalized convex conjugates are ordinary convex conjugates of dually affine-deformed functions. Second, we explain how these generalized Legendre transforms can be derived from the dual Hessian structures of information geometry.

Keywords: reverse-ordering and convex duality; legendre transform; information geometry; dually flat space and Hessian manifold; Bregman and Fenchel–Young divergences; affine and curvilinear coordinate systems

1. Introduction

Let \mathbb{R} denote the field of real numbers and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ be the extended real line. A m -variate extended function $F : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ is proper when its efficient domain $\Theta = \text{dom}(F) := \{\theta : F(\theta) < +\infty\}$ is non-empty and lower semi-continuous (lsc.) when its epigraph $\text{epi}(F) := \{(\theta, y) : y \geq F(\theta), \theta \in \text{dom}(F)\}$ is closed with respect to the metric topology of \mathbb{R}^m . We denote by Γ_0 (shortcut for $\Gamma_0(\mathbb{R}^m)$) the space of proper, lower, semi-continuous, extended, real-valued convex functions.

Consider the Legendre–Fenchel transform [1] (LFT) $\mathcal{L}F$ of a function F :

$$(\mathcal{L}F)(\eta) := \sup_{\theta \in \mathbb{R}^m} \{\langle \theta, \eta \rangle - F(\theta)\}, \quad (1)$$

where $\langle v, v' \rangle := \sum_{i=1}^m v_i v'_i$ denotes the Euclidean inner product. The function $F^* = \mathcal{L}F$ is called the convex conjugate function. Some examples of convex conjugates are reported in Appendix B.

The Legendre transform originated in the work of Adrien-Marie Legendre [2] circa 1805. Legendre considered variational calculus problems and introduced the variable transform. The Legendre transform was later formalized by Hamilton [3] and Gibbs [4] as a way to switch between equivalent descriptions of a system by exchanging variables like position \leftrightarrow momentum or energy \leftrightarrow entropy. This variable duality principle induced by the Legendre transform turned out to be fundamental in many scientific fields, like thermodynamics [5] for relating thermodynamic potentials, quantum field theory [6] for connecting Lagrangian and Hamiltonian formulations (and underlying effective actions),



Academic Editors: Antonio M. Scarfone and Tatsuki Wada

Received: 5 December 2025

Revised: 28 December 2025

Accepted: 28 December 2025

Published: 30 December 2025

Copyright: © 2025 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

economics for relating expenditure with utility functions [7], and convex optimization [8], where it is the cornerstone of duality theory in convex analysis.

The Moreau–Fenchel–Rockafellar theorem [9] states that when a function $F \in \Gamma_0$, its biconjugate function $(F^*)^*$ coincides with the function F : $F^{**} = F$. That is, the Legendre transform is an involutive transform of Γ_0 . For general lsc. functions F (possibly non-convex), it is known that the biconjugate function $(F^*)^* := \mathcal{L}(\mathcal{L}F)$ lower bounds the function F by the largest possible lsc. convex function $(F^*)^* \leq F$. This lower bound convexification property has been proven useful in machine learning [10].

This study was motivated by the fundamental result of Artstein-Avidan and Milman [11,12], who proved the following theorem:

Theorem 1 ([12], Theorem 7). *Let \mathcal{T} be an invertible transform such that the following are true:*

- $F_1 \leq F_2 \Rightarrow \mathcal{T}F_2 \leq \mathcal{T}F_1$;
- $\mathcal{T}F_1 \leq \mathcal{T}F_2 \Rightarrow F_2 \leq F_1$.

Then, \mathcal{T} is a generalized Legendre–Fenchel transform (GLFT), written canonically as

$$(\mathcal{T}F)(\eta) = \lambda(\mathcal{L}F)(E\eta + f) + \langle \eta, g \rangle + h, \quad (2)$$

where $\lambda > 0$, $E \in \text{GL}(\mathbb{R}^m)$ (general linear group), $f, g \in \mathbb{R}^m$, and $h \in \mathbb{R}$.

We note in passing that Fenchel [13,14] interpreted the graph of the Legendre transform of an m -variate function $F \in \Gamma_0$ as the polarity with respect to the paraboloid surface of \mathbb{R}^{m+1}

$$\mathcal{P} = \left\{ \left(\theta, y = \frac{1}{2} \sum_{i=1}^m \theta_i^2 \right) : \theta \in \mathbb{R}^m \right\} \in \mathbb{R}^{m+1},$$

of the graph of F . This geometric polarity connection of the Legendre transform is all the more interesting since Böröczky and Schneider [15] characterized the duality of convex bodies in Euclidean spaces containing the origin in the interior into the same space [11].

In this work, we shall first prove that a generalized convex conjugate $\mathcal{T}F$ obtained by Equation (2) can always be expressed as the ordinary convex conjugate of a corresponding affine-deformed function. Namely, we shall show that

$$(\mathcal{T}F)(\eta) = \mathcal{L}(\lambda F(A\theta + b) + \langle \theta, c \rangle + d),$$

where $A \in \text{GL}(\mathbb{R}^m)$, $b, c \in \mathbb{R}^m$ and $d \in \mathbb{R}$ are defined according to E, f, g, h (details reported in Theorem 2). This equivalence result allows us to interpret the origin of the generalized Legendre transforms from the viewpoint of information geometry [16,17] and untangle the various degrees of freedom used when defining generalized Legendre transforms in Section 3.

2. Generalized Legendre Transforms as Ordinary Legendre Transforms

Consider a parameter

$$P = (\lambda, A, b, c, d) \in \mathbb{P} := \mathbb{R}_{>0} \times \text{GL}(\mathbb{R}^m) \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R},$$

and deform a function $F(\theta)$ by carrying out affine transformations of both the parameter argument and its output as follows:

$$F_P(\theta) := \lambda F(A\theta + b) + \langle \theta, c \rangle + d. \quad (3)$$

Those affine deformations preserve convexity:

Property 1 (Convexity-preserving affine deformations). *Let $F \in \Gamma_0$. Then, F_P belongs to Γ_0 for all $P \in \mathbb{P}$.*

Proof. Let us check the convexity of F_P :

$$\alpha F_P(\theta_1) + (1 - \alpha)F_P(\theta_2) = \lambda(\alpha F(\bar{\theta}_1) + (1 - \alpha)F(\bar{\theta}_2)) + \langle \alpha\theta_1 + (1 - \alpha)\theta_2, c \rangle + d,$$

where $\bar{\theta}_1 := A\theta_1 + b$ and $\bar{\theta}_2 := A\theta_2 + b$. Since F is convex, we have $\alpha F(\bar{\theta}_1) + (1 - \alpha)F(\bar{\theta}_2) \leq F(\alpha\bar{\theta}_1 + (1 - \alpha)\bar{\theta}_2)$. Let $\bar{\theta}'_\alpha := \alpha\bar{\theta}_1 + (1 - \alpha)\bar{\theta}_2 = A\bar{\theta}'_\alpha + b$ with $\theta'_\alpha := \alpha\theta_1 + (1 - \alpha)\theta_2$. We obtain

$$\alpha F_P(\theta_1) + (1 - \alpha)F_P(\theta_2) \leq \underbrace{\lambda F(A\bar{\theta}'_\alpha + b) + \langle \theta'_\alpha, c \rangle + d}_{=F_P(\alpha\theta_1 + (1 - \alpha)\theta_2)},$$

hence proving that F_P is convex. Since the lower semi-continuous property is ensured (i.e., the epigraph of F_P is closed), we conclude that $F_P \in \Gamma_0$. \square

Next, let us express the convex conjugate of a function F_P according to the ordinary convex conjugate $F^* = \mathcal{L}F$:

Proposition 1 (LFT of an affine-deformed function). *The Legendre transform of F_P with $P = (\lambda, A, b, c, d) \in \mathbb{P}$ is*

$$(F_P)^* = (F^*)_{P^\circ}$$

where

$$P^\circ := \left(\lambda, \frac{1}{\lambda}A^{-1}, -\frac{1}{\lambda}A^{-1}c, -A^{-1}b, \langle b, A^{-1}c \rangle - d \right) \in \mathbb{P}. \quad (4)$$

That is, we have $\mathcal{L}(F_P) = (\mathcal{L}F)_{P^\circ}$.

Proof. Let $\Gamma_1 \subset \Gamma_0$ be defined as the class of differentiable, strictly convex Legendre-type functions [18] (see Appendix A). In general, given a function $F(\theta) \in \Gamma_1$, we proceed as follows to calculate its convex conjugate. First, we find the inverse function $(\nabla F)^{-1}$ of its gradient ∇F to obtain the gradient of the convex conjugate $\nabla F^* = (\nabla F)^{-1}$. Then, we have

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

For parameter $P = (\lambda, A, b, c, d)$, consider the strictly convex and differentiable function $F_P(\theta) = \lambda F(A\theta + b) + \langle c, \theta \rangle + d$ for an invertible matrix $A \in \text{GL}(d, \mathbb{R})$, vectors $b, c \in \mathbb{R}^d$, and scalars $d \in \mathbb{R}$ and $\lambda \in \mathbb{R}_{>0}$. The gradient of F_P is

$$\eta = \nabla F_P(\theta) = \lambda A^\top \nabla F(A\theta + b) + c.$$

We denote by $G = F^*$ and $G_P = (F_P)^*$ the Legendre convex conjugates of F and F_P , respectively. By solving the equation $\nabla F_P(\theta) = \eta$, we find the reciprocal gradient $\theta(\eta) = \nabla G_P(\eta)$:

$$\nabla G_P(\eta) = A^{-1} \nabla G \left(\frac{1}{\lambda} A^{-\top} (\eta - c) \right) - b.$$

Therefore, the Legendre convex conjugate is obtained as follows:

$$\begin{aligned} G_P(\eta) &= \langle \eta, \nabla G_P(\eta) \rangle - F_P(\nabla G_P(\eta)), \\ &= \lambda' G(A'\eta + b') + \langle c', \eta \rangle + d', \end{aligned}$$

where

$$\begin{aligned}\lambda' &= \lambda, \\ A' &= \frac{1}{\lambda}A^{-1}, \\ b' &= -\frac{1}{\lambda}A^{-1}c, \\ c' &= -A^{-1}b, \\ d' &= \langle b, A^{-1}c \rangle - d.\end{aligned}$$

Hence, it follows that we have $P^\diamond = \left(\lambda, \frac{1}{\lambda}A^{-1}, -\frac{1}{\lambda}A^{-1}c, -A^{-1}b, \langle b, A^{-1}c \rangle - d\right) \in \mathbb{P}$. \square

Let us check that the (diamond) \diamond operator on affine deformation parameters is an involution:

Proposition 2 (\diamond involution). *The parameter transformation P^\diamond is an involution, where $(P^\diamond)^\diamond = P$.*

Proof. Let $P = (\lambda, A, b, c, d)$ and $P^\diamond = (\lambda', A', b', c', d')$ with

$$\begin{aligned}\lambda' &= \lambda, \\ A' &= \frac{1}{\lambda}A^{-1}, \\ b' &= -\frac{1}{\lambda}A^{-1}c, \\ c' &= -A^{-1}b, \\ d' &= \langle b, A^{-1}c \rangle - d.\end{aligned}$$

We check that $P^{\diamond\diamond} = (\lambda'', A'', b'', c'', d'') = (\lambda, A, b, c, d) = P$ component-wise as follows:

$$\begin{aligned}\lambda'' &= \lambda' &&= \lambda, \\ A'' &= \frac{1}{\lambda'}A'^{-1} &&= \frac{1}{\lambda} \left(\frac{1}{\lambda}A^{-1} \right)^{-1} &&= A, \\ b'' &= -\frac{1}{\lambda'}A'^{-1}c' &&= -\frac{1}{\lambda} \left(\frac{1}{\lambda}A^{-1} \right)^{-1} (-A^{-1}b) &&= b, \\ c'' &= -A'^{-1}b' &&= -\left(\frac{1}{\lambda}A^{-1} \right)^{-1} \left(-\frac{1}{\lambda}A^{-1}c \right) &&= c, \\ d'' &= \langle b', A'^{-1}c' \rangle - d' &&= \left\langle -\frac{1}{\lambda}A^{-1}c, \frac{1}{\lambda} \left(\frac{1}{\lambda}A^{-1} \right)^{-1} (-A^{-1}b) \right\rangle - \langle b, A^{-1}c \rangle + d &&= d.\end{aligned}$$

\square

The \diamond involution confirms that the Legendre–Fenchel transform is an involution:

$$((F_P)^*)^* = \mathcal{L}(\mathcal{L}F_P) = \mathcal{L}F_{P^\diamond} = F_{(P^\diamond)^\diamond} = F_P.$$

Let us now define the following notation to express the generalized Legendre convex conjugates:

Definition 1 (Generalized Legendre–Fenchel convex conjugates [12]). *Let $\mathcal{L}_{\lambda, E, f, g, h}$ denote a generalized Legendre–Fenchel transform*

$$\mathcal{L}_{\lambda, E, f, g, h}F := \mathcal{L}_P F := \lambda(\mathcal{L}F)(E\eta + f) + \langle \eta, g \rangle + h$$

for the parameter $P = (\lambda, E, f, g, h) \in \mathbb{P}$.

Our result is that we can interpret those generalized Legendre–Fenchel transforms of Equation (2) as the ordinary Legendre transform on corresponding affine-deformed convex functions:

Theorem 2. For any $F \in \Gamma_0$, we have $\mathcal{L}_P(F) := (F^*)_P = \mathcal{L}(F_{P^\diamond})$.

Proof. By definition, $\mathcal{L}_P F := (\mathcal{L}F)_P$. Since $P = (P^\diamond)^\diamond$ (Proposition 2), we have $(\mathcal{L}F)_P = (\mathcal{L}F)_{(P^\diamond)^\diamond}$, and by using Proposition 1, we obtain $(\mathcal{L}F)_{(P^\diamond)^\diamond} = \mathcal{L}(F_{P^\diamond})$. To summarize, we have

$$\mathcal{L}_P(F) = (\mathcal{L}F)_P = (\mathcal{L}F)_{(P^\diamond)^\diamond} = \mathcal{L}(F_{P^\diamond}).$$

□

That is, in plain words, the Artstein–Avidan–Milman generalized Legendre transforms [11,12] are ordinary Legendre transforms on affine-deformed functions evaluated on affine deformed arguments.

Next, we describe an information-geometric interpretation of Theorem 2 which casts light on the origin and meanings of the degrees of freedom used to define generalized Legendre transforms. Thus, understanding the ordinary Legendre transform from the information-geometric viewpoint allows one to understand how to recover the generalized Legendre transforms formalized axiomatically by Artstein–Avidan–Milman [11,12].

3. An Information-Geometric Interpretation of Generalized Legendre Transforms

We can interpret the fact that generalized Legendre convex conjugates are affine-deformed convex conjugates from the lens of information geometry [16]. Consider a smooth, strictly convex function $F : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ of the space Γ_2 of twice-differentiable lsc. Legendre-type convex functions (with $\Gamma_2 \subset \Gamma_1 \subset \Gamma_0$). This function $F(\theta)$ defines a dually flat m -dimensional space [16] (M, g, ∇, ∇^*) which is a global chart manifold $M = \{p\}$ of points p equipped with dual Hessian structures [19] (g, ∇) and (g, ∇^*) . The primal Hessian structure is induced by a potential function ψ on M with a torsion-free, flat, affine connection ∇ such that $\psi(p) = F(\theta(p))$, where $(M, \theta(\cdot))$ is the primal ∇ -affine coordinate system. The dual Hessian structure is induced by a dual potential function ϕ on M with a dual torsion-free, flat, affine connection ∇^* such that $\phi(p) = F^*(\eta(p))$, where $(M, \eta(\cdot))$ is the dual ∇^* -affine coordinate system. The Riemannian metric g can be expressed in the dual coordinate charts as $g(\theta) = \nabla^2 F(\theta)$ or $g(\eta) = \nabla^2 F^*(\eta)$, where F^* is the convex conjugate. These two potential functions ψ and ϕ living on the manifold M satisfy the Fenchel–Young inequality:

$$\psi(\theta(p)) + \phi(\eta(q)) \geq \sum_{i=1}^m \theta_i(p) \eta_i(q), \quad (5)$$

with equality if and only if $p = q$ on M . The metric tensor g can be expressed as $g = \nabla d\psi$ or equivalently as $g = \nabla^* d\phi$, where d denotes the exterior derivative and ψ and ϕ are 0 forms.

Now, the ∇ -affine coordinate system θ is defined up to an affine transformation. That is, if $\theta(\cdot)$ is a ∇ coordinate system, then the coordinate system is $\bar{\theta}(\cdot) = A\theta(\cdot) + b$. However, once parameters A and b are fixed, it fully determines the dual ∇^* coordinate system $\eta(\cdot)$.

The potential function ϕ is also reconstructed by solving differential modulo equations an affine term $\langle c, \theta \rangle + d$ (e.g., see the proofs relying on Poincaré lemma in [16,19–21]).

Fixing the degrees of freedom in the reconstruction of ψ will also fix the corresponding affine term in F^* which expresses ψ . The dual potential functions ψ and ϕ on the manifold M are related by the fiberwise Legendre transform [22] in geometric mechanics.

Thus, the information geometry of dually flat spaces allows one to explain the decoupling of the interactions of the dual set of parameters (A, b, c, d) with (E, f, g, h) . Lastly, the scalar parameter $\lambda > 0$ is the degree of freedom obtained from the fact that if (M, g, ∇, ∇^*) is a dually flat space, then so is $(M, \lambda g, \nabla, \nabla^*)$.

The Fenchel–Young inequality induced by the dual potential functions of Equation (5) defines a canonical divergence on a dually flat space that we term the dually flat Hessian divergence:

$$D_{g,\nabla}(p : q) = \psi(\theta(p)) + \phi(\eta(q)) - \sum_{i=1}^m \theta_i(p) \eta_i(q) \geq 0.$$

We have the following reference duality [23]: $D_{g,\nabla}(q : p) = D_{g,\nabla^*}(p : q)$.

The dually flat Hessian divergence can be expressed using the dual coordinate systems as a corresponding Fenchel–Young divergence defined by

$$Y_{F,F^*}(\theta : \eta') := F(\theta) + F^*(\eta') - \sum_{i=1}^m \theta_i \eta'_i.$$

A Fenchel–Young divergence can also be expressed equivalently as dual Bregman divergences B_F or B_{F^*} :

$$Y_{F,F^*}(\theta : \eta') = B_F(\theta : \theta') = B_{F^*}(\eta' : \eta),$$

where $\theta' = \nabla F^*(\eta')$ and $\eta = \nabla F(\theta)$ with

$$B_F(\theta : \theta') = F(\theta) - F(\theta') - \langle \theta - \theta', \nabla F(\theta') \rangle.$$

Thus, the dually flat Hessian divergence can be expressed as dual Fenchel–Young divergences (using the fact that $F^{**} = F$) in the mixed θ/η coordinate systems as

$$D_{g,\nabla}(p : q) = Y_{F,F^*}(\theta(p) : \eta(q)) = Y_{F^*,F}(\eta(q) : \theta(p)), \quad (6)$$

$$= \frac{1}{\lambda} Y_{F_P, F_{P^*}^*}(\tilde{\theta}(p) : \tilde{\eta}(q)) = \frac{1}{\lambda} Y_{F_{P^*}^*, F_P}(\tilde{\eta}(q) : \tilde{\theta}(p)), \forall P \in \mathbb{P}, \quad (7)$$

This follows from the fact that $\psi(p) = F_P(\tilde{\theta}(p)) = F(\theta(p))$ and $\phi(q) = F_{P^*}^*(\tilde{\eta}(q)) = F^*(\eta(q))$ for any $P \in \mathbb{P}$.

Thus, we can define an equivalence relation \sim between the functions of Γ_2 : $F \sim \tilde{F}$ if and only if $\tilde{F} = F_P$ for some $P \in \mathbb{P}$. Furthermore, a distance invariant under the Legendre transform [24] can be defined in the moduli space Γ_2 / \sim of dually flat spaces.

Lastly, we may consider some curvilinear dual coordinate systems instead of the mutually orthogonal, dually affine coordinate systems. Let us transform the affine θ coordinate system into some arbitrary curvilinear coordinate system $\tilde{\theta}$. The dual affine η coordinate system transforms correspondingly into the curvilinear coordinate system $\tilde{\eta}$. The underlying geometric structures have been called the (u, v) structures by Amari [16,25] or the (ρ, τ) structures by Zhang [23]. The dual Bregman divergences or, equivalently, the dual Fenchel–Young divergences of the underlying dually flat space need to be addressed as undeforming arguments [26,27] $\tilde{\theta}$ or $\tilde{\eta}$. Since the inverse transform $\tilde{\theta} \rightarrow \theta$ can be interpreted as a representation function, the Bregman divergences addressed as undeforming $\tilde{\theta}$ arguments were called representational Bregman divergences in [26]. For example, the β divergences are Bregman divergences [28–30] which can be transformed

by a nonlinear change of coordinates into α divergences [31]. Although the α divergences are Bregman divergences for any α on positive measures [26,27], α divergences are not Bregman divergences anymore when constrained on the probability simplex (except for the $\alpha = \pm 1$ case).

Remark 1. In information geometry [16], and more generally in Hessian geometry [19], a Legendre-type C^3 convex function $F(\theta)$ of Γ_3 induces a dually flat space or manifold. The convex function $F(\theta)$ (potential function) may stem from several modeling sources. We may consider the following settings to build dually flat spaces:

- The cumulant function of a natural exponential family [16,32]: Let $(\mathcal{X}, \Omega, \mu)$ be a measurable space with a sample space \mathcal{X} , σ algebra Ω , and a positive measure μ (e.g., counting or Lebesgue measure). An exponential family \mathcal{E} is a set of probability measures $\mathcal{E} = \{P_\theta(x) = \exp(\langle x, \theta \rangle - F(\theta)) d\mu(x) : \theta \in \Theta\}$, where

$$\Theta = \{\theta : F(\theta) = \log \int \exp(\langle x, \theta \rangle) d\mu(x) < \infty\},$$

is the natural parameter space. $Z(\theta) = \exp(F(\theta))$ is strictly log-convex [32], and hence $F(\theta) = \log Z(\theta)$ is strictly convex (and differentiable). In this case, the convex conjugate amounts to the negative Shannon or differential entropy [33] $F^*(\eta) = -S(p_\theta) = \int p_\theta(x) \log p_\theta(x) d\mu(x)$, where $S(p) = -\int p(x) \log p(x) d\mu(x)$ is the Shannon entropy (when μ is the counting measure) or the differential entropy (when μ is the Lebesgue measure).

- The negative Shannon entropy of a mixture family [16]: A mixture family \mathcal{M} is a set of probability measures

$$\mathcal{M} = \left\{ m_\theta(x) = \sum_{i=1}^{d-1} \theta_i p_i(x) + (1 - \sum_{i=1}^{d-1} \theta_i) p_0(x) \right\},$$

parameterized by a normalized positive weight vector $\theta \in \Theta = \Delta_d$ (the standard $(d-1)$ -dimensional simplex) such that the functions $1, p_0(x), \dots, p_{d-1}(x)$ are linearly independent. The function $F(\theta) = -S(m_\theta)$ is strictly convex and differentiable (see [34] for a proof). In this case, the convex conjugate $F^*(\eta) = S^\times(p_0, m_\theta)$ is the cross-entropy of $p_0(x)$ with the mixture $m_\theta(x)$, where $S^\times(p, q) = -\int p(x) \log q(x) d\mu(x)$.

- The characteristic function of a regular cone [19] (i.e., convex and pointed cone): Let \mathcal{K} be a regular cone and $\mathcal{K}^* = \cap_{\theta \in \mathcal{K}} \{\eta : \langle \theta, \eta \rangle \geq 0\}$ be its dual cone. We define $F(\theta) = \log \chi_{\mathcal{K}}(\theta)$, where $\chi_{\mathcal{K}}(\theta)$ is the characteristic function $\chi_{\mathcal{K}}(\theta) = \int_{\mathcal{K}^*} \exp(-\langle \theta, \eta \rangle) d\eta$. One can further build generalized Wishart exponential families on the cones [35]. (Note that Massieu [36] introduced the concept of characteristic functions and their convex conjugates in thermodynamics in the 19th century.) See also the barrier functions (logarithms of characteristic functions) on convex cones for interior point methods [37].
- The Guillemin potential of a (Delzant) polytope [38] which recovers the negative Shannon entropy when considering the standard simplex.

4. Conclusions

To summarize, we first proved that the generalized Legendre transforms obtained from the reverse-ordering involutive transform axiomatization of [11,12] can be interpreted as the ordinary Legendre transform of corresponding affine-deformed functions. Second, we explained how these generalized Legendre transforms are merely different expressions of the well-known geometric Legendre transforms on dually flat spaces (Hessian manifolds with global charts).

Figure 1 summarizes the various classes of convex functions used in this paper, with some of their key properties used to define information-geometric structures. In particular, the class of functions Γ_3 is thrice-differentiable Legendre-type lsc. convex functions (with $\Gamma_3 \subset \Gamma_2 \subset \Gamma_1 \subset \Gamma_0$), which allows one to construct α geometry [16] from the Amari–Chentsov cubic tensor [17].

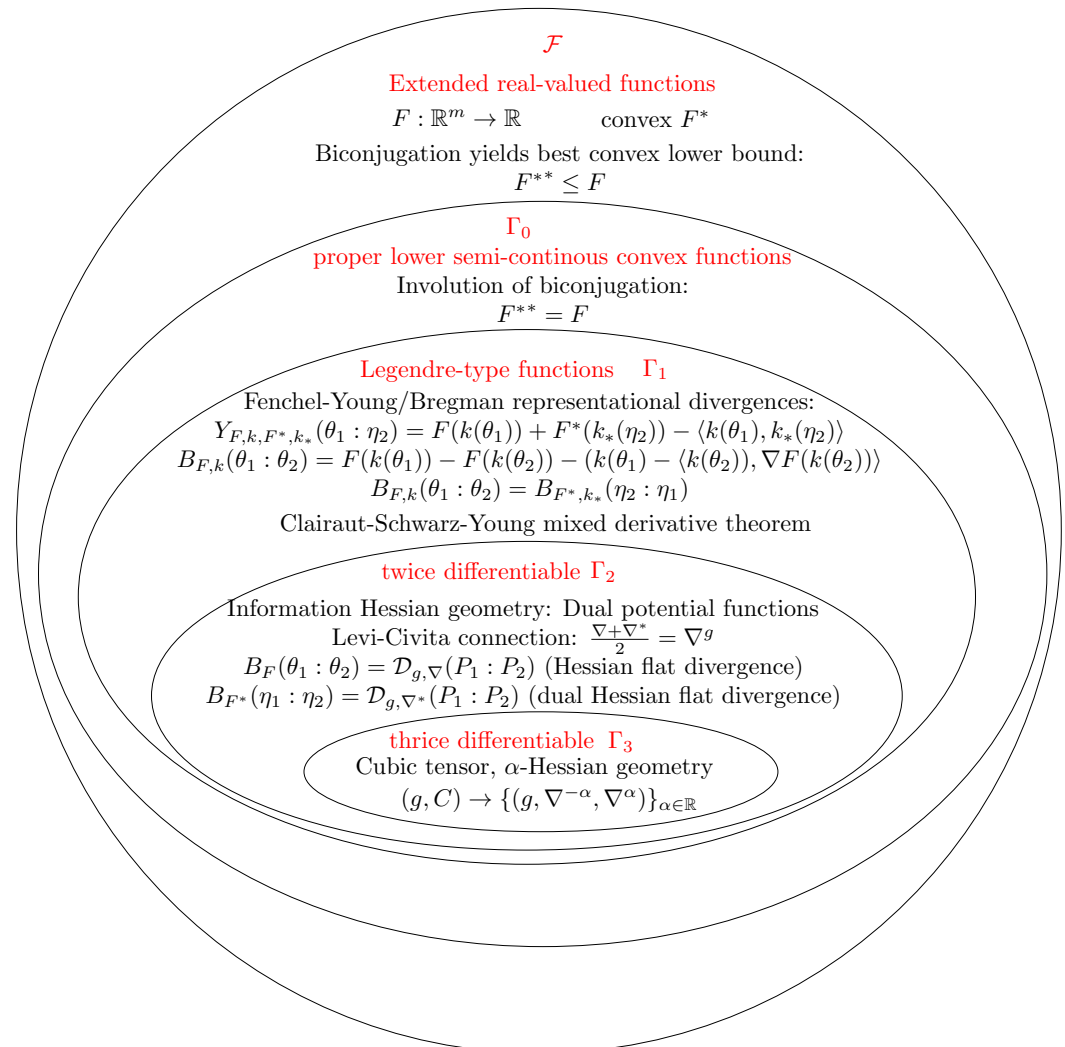


Figure 1. The ordinary Legendre transform for various classes of functions: relationships with Fenchel–Young and Bregman divergences, dually flat Hessian divergence, and α geometry in information geometry.

Funding: This research received no external funding

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 was employed by the company Sony Computer Science Laboratories. 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.

Appendix A. Legendre-Type Functions

Rockafellar [18] showed that there exist convex lsc. functions of Γ_0 with corresponding gradient domains that are not convex. For example, Rockafellar showed that the bivariate function $F(\theta) = F(\theta_1, \theta_2) = \frac{1}{4} \left(\frac{\theta_1^2}{\theta_2} + \theta_1^2 + \theta_2^2 \right)$ defined on the upper plane domain $\Theta =$

$\mathbb{R} \times \mathbb{R}_{>0}$ is strictly convex, but its gradient domain H is not convex. Thus, to rule out these functions, Rockafellar defined Legendre-type functions [18]:

Definition A1 (Legendre-type function [18]). A convex function $F : \Theta \subset \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ is of a Legendre type if the following are true:

- Θ is a non-empty open effective domain;
- F is strictly convex and differentiable on Θ ;
- F becomes infinitely steep close to boundary points of its effective domain:

$$\forall \theta \in \Theta, \forall \theta' \in \partial\Theta, \quad \lim_{\lambda \rightarrow 0} \frac{d}{d\lambda} F(\lambda\theta + (1-\lambda)\theta') = -\infty.$$

When F is of a Legendre type, so is its convex conjugate F^* , and moreover the gradients of convex conjugates are reciprocal to each other such that $\nabla F^* = (\nabla F)^{-1}$ and $\nabla F = (\nabla F^*)^{-1}$. This property of reciprocal gradients obtained from the LFT of Legendre-type functions is strong since, in general, the implicit function theorem only guarantees local inversion of multivariate functions and not the existence of global inverse functions.

We shall denote by $\Gamma_1(\Theta) \subset \Gamma_0(\Theta)$ the subset of proper lsc. strictly convex and differentiable functions of a Legendre type defined on the open effective domain Θ . We define the Legendre–Fenchel transform of $\mathcal{L}(\Theta, F) = (H, F^*)$ where H is the gradient domain, and we call F^* the convex conjugate of F . We have $\mathcal{L}\mathcal{L}(\Theta, F) = (\Theta, F)$ for Legendre-type functions $F \in \Gamma_1(\Theta)$. The concept of a Legendre-type function is related to the concept of steep exponential families in statistics.

In general, convex conjugate pairs enjoy the following reverse-ordering property:

Property A1 (Reverse-ordering). Let F_1 and F_2 in $\Gamma_0(\Theta)$. If $F_2 \leq F_1$, then $\mathcal{L}F_2 \geq \mathcal{L}F_1$. If $F_2 \geq F_1$, then $\mathcal{L}F_2 \leq \mathcal{L}F_1$.

Proof. Assume that $F_2 \leq F_1$, and let us prove that $F_2^* = \mathcal{L}F_2 \geq F_1^* = \mathcal{L}F_1$:

$$\begin{aligned} F_1^*(\eta) &= \sup_{\theta} \{ \langle \eta, \theta \rangle - F_1(\theta) \}, \\ &= \langle \eta, \nabla F^{-1}(\eta) \rangle - F_1(\nabla F^{-1}(\eta)), \\ &\leq \langle \eta, \nabla F^{-1}(\eta) \rangle - F_2(\nabla F^{-1}(\eta)), \\ &\leq \sup_{\theta} \{ \langle \eta, \theta \rangle - F_2(\theta) \} = F_2^*(\eta). \end{aligned}$$

The case $F_2 \geq F_1$ is similar to the case $F_1 \geq F_2$ by exchanging the role of F_1 and F_2 (i.e., $F_2 \leftrightarrow F_1$). \square

Appendix B. Some Examples of Convex Conjugates

Example A1. Let $F(\theta) = \langle a, \theta \rangle + b$ be an affine function. Then, we have

$$F^*(\eta) = \begin{cases} -b & \text{if } \eta = a \\ +\infty & \text{if } \eta \neq a \end{cases}$$

Let

$$1_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A \end{cases}$$

be the indicator function of a set A . Then, we can write $F^* = 1_{\{a\}} - b$.

Example A2. Let $F(\theta) = |\theta|$ for $\theta \in \mathbb{R}$. Then, we have the convex conjugate

$$F^*(\eta) = 1_{[-1,1]}(\eta) = \begin{cases} 0 & |\eta| \leq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Example A3. Consider $F(\theta) = \exp \theta$ for $\Theta = \mathbb{R}$. We have

$$F^*(\eta) = \begin{cases} \eta \log \eta - \eta, & \text{if } \eta > 0, \\ 0, & \text{if } \eta = 0, \\ +\infty & \text{if } \eta < 0. \end{cases}$$

The convex conjugate is the scalar Shannon entropy function extended to positive real values.

Example A4. For $p \in [1, +\infty)$, let $F_p(\theta) = \frac{1}{p} \|\theta\|^p$ for $\theta \in \Theta = \mathbb{R}^m$. Then, we have $F_p^*(\eta) := F_q(\eta)$, where $\frac{1}{p} + \frac{1}{q} = 1$. The powers p and q are called a conjugate pair of exponents. Notice that the Fenchel–Young inequality reduces to the Young’s inequality in the scalar case:

$$F_p(\theta') + F_q(\eta) \geq \langle \theta', \eta \rangle.$$

Furthermore, by integrating Young’s inequality on both sides for $\eta = g(x)$ and $\theta = f(x)$ for $x \in \mathcal{X}$ and $f \in L_p$ and $g \in L_q$ (Lebesgue spaces), we recover Hölder’s inequality:

$$\frac{1}{p} \|f\|_p + \frac{1}{q} \|g\|_q \geq \|fg\|_1.$$

In fact, we have the self-duality $F = F^*$ only for $F(\theta) = F_2(\theta) = \frac{1}{2} \|\theta\|^2$.

When a function $F \in \Gamma_0(\Theta)$ is not of the Legendre type, the Fenchel–Young inequality may be tight for several pairs (θ', η) instead of a single pair $(\theta = \nabla F^*(\eta), \eta = \nabla F(\theta))$. A subgradient η of $F(\theta)$ at θ_0 satisfies

$$F(\theta) - F(\theta_0) \geq \langle \eta, \theta - \theta_0 \rangle.$$

The subdifferential $\partial_{\theta_0} F$ of F at θ_0 is the set of subgradients. The subdifferential operator $\partial : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is multivalued:

$$\partial F(\theta) = \begin{cases} \{\eta : F(\theta) - F(\theta_0) \geq \langle \eta, \theta - \theta_0 \rangle\}, & \text{if } \theta \in \Theta = \text{dom}(F) \\ \emptyset & \text{if } \theta \notin \Theta. \end{cases}$$

The following example illustrates the LFT of a non-Legendre type function and the LFT of a Legendre-type function obtained by restricting the function to a specified domain.

Example A5. Let $F(\theta) := \exp(|\theta|) - |\theta| - 1$ be defined on $\Theta = \mathbb{R}$. Then, $F^*(\eta) = (1 + |\eta|) \log(1 + |\eta|) - |\eta|$, defined on $H = \mathbb{R}$. These two functions F and F^* are even and convex, and they form a conjugate pair (Figure A1). The function $F(\theta)$ is not differentiable at $\theta = 0$, and thus we have

$$\partial F(\theta) = \begin{cases} \nabla F(\theta) = \exp(\theta) - 1, & \theta > 0, \\ \partial F(0) = [-1, 1], & \theta = 0, \\ \nabla F(\theta) = -\exp(-\theta) + 1, & \theta < 0. \end{cases}$$

The function $F^*(\eta)$ is differentiable everywhere, and we have

$$\partial F^*(\eta) = \begin{cases} \nabla F^*(\eta) = \log(1 + \eta), & \eta \geq 0, \\ \nabla F^*(\eta) = -\log(1 - \eta), & \eta \leq 0. \end{cases}$$

For $\theta \neq 0$, the gradients $\nabla F(\theta)$ and $\nabla F^*(\eta)$ are reciprocal to each other. The function $(\mathbb{R}, F(\theta))$ is not of a Legendre type since it is not differentiable at $\theta = 0$, but the functions $(\mathbb{R}_{>0}, F(\theta))$ and $(\mathbb{R}_{>0}, F^*(\eta))$ form a Legendre-type pair when restricted to the positive domain $\mathbb{R}_{>0}: F, F^* \in \Gamma_1(\mathbb{R}_{>0})$.

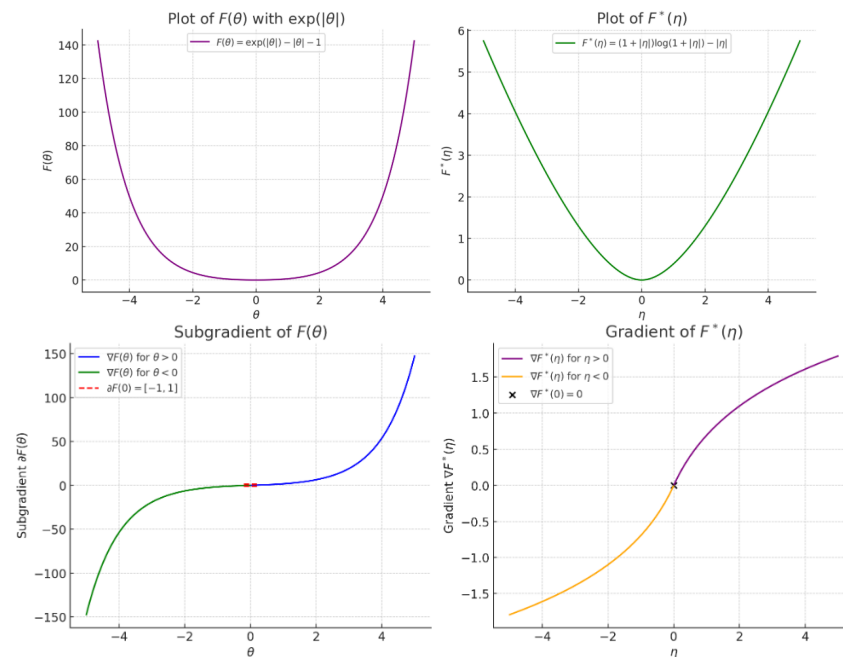


Figure A1. A pair $(F(\theta), F^*(\eta))$ of conjugate functions (**top**) with their subgradients plotted (**bottom**). The function $F(\theta)$ is not differentiable at $\theta = 0$ and thus admits a subgradient $\partial F(0)$ at $\theta = 0$. The function $F^*(\eta)$ is differentiable everywhere, and when $\theta \neq 0$, we have $\nabla F^* = (\nabla F)^{-1}$, visualized by rotating the xy axis 90 degrees).

References

- Bauschke, H.H.; Lucet, Y. What is... a Fenchel conjugate? *Not. AMS* **2012**, *59*, 44–46. [[CrossRef](#)]
- Legendre, A.M. *Nouvelles méthodes pour la détermination des orbites des comètes*; Firmin Didot: Paris, France, 1805.
- Hamilton, W.R. On a General Method in Dynamics. In *Philosophical Transactions of the Royal Society of London*; Cambridge University: Cambridge, UK, 1834; pp. 247–308.
- Gibbs, J.W. *Elementary Principles of Statistical Mechanics*; Yale University Press: New Haven, CT, USA, 1902.
- Callen, H.B. *Thermodynamics and an Introduction to Thermostatistics*; John Wiley & Sons: Hoboken, NJ, USA, 1985.
- Weinberg, S. *The Quantum Theory of Fields*; Cambridge University Press: Cambridge, UK, 1995; Volume 2.
- Mas-Colell, A.; Whinston, M.D.; Green, J.R. *Microeconomic Theory*; Oxford University Press: New York, NY, USA, 1995; Volume 1.
- Rockafellar, R.T. Convex Analysis. *Princet. Math. Ser.* **1970**, *28*, 470.
- Correa, R.; Hantoute, A.; López, M.A. *Fundamentals of Convex Analysis and Optimization*; Springer: Berlin/Heidelberg, Germany, 2023.
- Möllenhoff, T.; Khan, M.E. SAM as an Optimal Relaxation of Bayes. In Proceedings of the The Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023; ICLR: Appleton, WI, USA, 2023.
- Artstein-Avidan, S.; Milman, V. A characterization of the concept of duality. *Electron. Res. Announc.* **2007**, *14*, 42–59.
- Artstein-Avidan, S.; Milman, V. The concept of duality in convex analysis, and the characterization of the Legendre transform. *Ann. Math.* **2009**, *169*, 661–674. [[CrossRef](#)]
- Fenchel, W. On Conjugate Convex Functions. *Can. J. Math.* **1949**, *1*, 73–77. [[CrossRef](#)]
- Fenchel, W. On conjugate convex functions. In *Traces and Emergence of Nonlinear Programming*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 125–129.

15. Böröczky, K.J.; Schneider, R. A characterization of the duality mapping for convex bodies. *Geom. Funct. Anal.* **2008**, *18*, 657–667. [[CrossRef](#)]
16. Amari, S.i. *Information Geometry and its Applications*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
17. Ay, N.; Jost, J.; Vân Lê, H.; Schwachhöfer, L. *Information Geometry*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 64.
18. Rockafellar, R.T. Conjugates and Legendre transforms of convex functions. *Can. J. Math.* **1967**, *19*, 200–205. [[CrossRef](#)]
19. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
20. Amari, S.i.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000; Volume 191.
21. Morales, P.A.; Korbel, J.; Rosas, F.E. Geometric structures induced by deformations of the Legendre transform. *Entropy* **2023**, *25*, 678. [[CrossRef](#)] [[PubMed](#)]
22. Leok, M.; Zhang, J. Connecting information geometry and geometric mechanics. *Entropy* **2017**, *19*, 518. [[CrossRef](#)]
23. Naudts, J.; Zhang, J. Legendre duality: From thermodynamics to information geometry. *Inf. Geom.* **2024**, *7*, 623–649. [[CrossRef](#)]
24. Attouch, H.; Wets, R.J.B. Isometries for the Legendre-Fenchel transform. *Trans. Am. Math. Soc.* **1986**, *296*, 33–60. [[CrossRef](#)]
25. Nock, R.; Nielsen, F.; Amari, S.I. On conformal divergences and their population minimizers. *IEEE Trans. Inf. Theory* **2015**, *62*, 527–538. [[CrossRef](#)]
26. Nielsen, F.; Nock, R. The dual Voronoi diagrams with respect to representational Bregman divergences. In Proceedings of the Sixth International Symposium on Voronoi Diagrams (ISVD), Copenhagen, Denmark, 23–26 June 2009; IEEE: New York, NY, USA, 2009; pp. 71–78.
27. Amari, S.i. α -Divergence is unique, belonging to both f -divergence and Bregman divergence classes. *IEEE Trans. Inf. Theory* **2009**, *55*, 4925–4931. [[CrossRef](#)]
28. Csiszar, I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.* **1991**, *19*, 2032–2066. [[CrossRef](#)]
29. Murata, N.; Takenouchi, T.; Kanamori, T.; Eguchi, S. Information geometry of U -Boost and Bregman divergence. *Neural Comput.* **2004**, *16*, 1437–1481. [[CrossRef](#)] [[PubMed](#)]
30. Hennequin, R.; David, B.; Badeau, R. Beta-divergence as a subclass of Bregman divergence. *IEEE Signal Process. Lett.* **2010**, *18*, 83–86. [[CrossRef](#)]
31. Dikmen, O.; Yang, Z.; Oja, E. Learning the information divergence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1442–1454. [[CrossRef](#)] [[PubMed](#)]
32. Barndorff-Nielsen, O. *Information and Exponential Families: In Statistical Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
33. Nielsen, F.; Nock, R. Entropies and cross-entropies of exponential families. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; IEEE: New York, NY, USA, 2010; pp. 3621–3624.
34. Nielsen, F.; Hadjeres, G. Monte Carlo information-geometric structures. In *Geometric Structures of Information*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 69–103.
35. Graczyk, P.; Ishi, H.; Mamane, S. Wishart exponential families on cones related to tridiagonal matrices. *Ann. Inst. Stat. Math.* **2019**, *71*, 439–471. [[CrossRef](#)]
36. Massieu, F. *Thermodynamique: Mémoire sur les Fonctions Caractéristiques des Divers Fluides et sur la Théorie des Vapeurs*; Académie des Sciences de L’Institut National de France: Paris, France, 1876; Volume 22.
37. Güler, O. Barrier functions in interior point methods. *Math. Oper. Res.* **1996**, *21*, 860–885. [[CrossRef](#)]
38. Fujita, H. The generalized Pythagorean theorem on the compactifications of certain dually flat spaces via toric geometry. *Inf. Geom.* **2024**, *7*, 33–58. [[CrossRef](#)]

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.