

Article

Group Invariance of Information Geometry on q -Gaussian Distributions Induced by Beta-Divergence

Atsumi Ohara ^{1,*} and Shinto Eguchi ²

¹ Department of Electrical and Electronics Engineering, University of Fukui, Fukui 910-8507, Japan

² The Institute of Statistical Mathematics and the Graduate University of Advanced Studies, Tachikawa Tokyo 190-8562, Japan; E-Mail: eguchi@ism.ac.jp

* Author to whom correspondence should be addressed; E-Mail: ohara@fuee.u-fukui.ac.jp; Tel.: +81-776-27-8570.

Received: 6 October 2013; in revised form: 26 October 2013 / Accepted: 29 October 2013 /

Published: 4 November 2013

Abstract: We demonstrate that the q -exponential family particularly admits natural geometrical structures among deformed exponential families. The property is the *invariance* of structures with respect to a general linear group, which transitively acts on the space of positive definite matrices. We prove this property via the correspondence between information geometry induced by a deformed potential on the space and the one induced by what we call β -divergence defined on the q -exponential family with $q = \beta + 1$. The results are fundamental in robust multivariate analysis using the q -Gaussian family.

Keywords: multivariate q -Gaussian family; β -divergence; information geometry; $GL(n; \mathbf{R})$ -invariance; V -potential

1. Introduction

Generalizations of the exponential family have recently had much attention paid to them in mathematical statistics or statistical physics [1–6]. One of their goals is to attack a wider class of statistical problems that ranges outside the ones solved via the well-established theory of the exponential family [7]. Among such generalizations, the q -exponential family, where the ordinary exponential function is replaced by the q -exponential [8], often naturally appears to play important roles in experimental and theoretical aspects.

For example, the family not only describes phenomena obeying the power-law well [6], but also, it is theoretically proven to include a velocity distribution of the classical gas with N particles [9,10], an

attracting invariant manifold of porous media flow [11], and so on. In statistics, it is reported to provide a reasonable statistical model in robust inference from data losing normality [1,12–14]. In addition, quite interesting and abundant mathematical structures have been developed [15–19] for the q -exponential function itself.

On certain families of elliptical distributions [20,21], we can introduce information geometric structures [22,23] starting from what is called the U -divergence [1] instead of the Kullback-Leibler divergence, in order to geometrically tackle various statistical problems that include the above robustness analysis. Zero-mean elliptical density functions $f_P(x) = u(-x^T P x/2 - c_U(\det P))$, with a fixed function, u , and the normalizing constant, $c_U(\det P)$, can be specified by positive definite matrices, P . Hence, we can discuss the geometric structures of such density functions explicitly, e.g., with no integrations, via the corresponding information geometry on the parameter space of positive definite matrices, called the V -geometry [24].

In the present paper, we focus on investigating the geometric structures of the q -Gaussian family induced from the β -divergence [1] via the V -geometry following the above idea. For this purpose, we establish the correspondence between two geometries and derive explicit formulas of important geometric quantities, such as the Riemannian metric and mutually dual affine connections. Consequently, we can prove that the information geometry on the q -Gaussian family enjoys fairly natural group invariance properties. These invariances or homogeneities are important in multivariate analysis (see, e.g., [21,25]). Further, they practically assure that the statistical inferences based on the geometrical theory are independent of linear transformations of multivariate random variables, such as scaling or numerical conditioning in computations. It should be additionally mentioned that our results might shed new light on the rich mathematical structures of the q -exponential or power functions.

The organization of the paper is as follows: Section 2 collects necessary results on the V -geometry of positive definite matrices and the U -divergence defined on elliptical distributions. In Section 3, we discuss the group invariances of fundamental structures of the V -geometry induced by the power potentials. We find that its pair of mutually dual connections and orthogonality are $GL(n, \mathbf{R})$ -invariant, which is a natural requirement for geometries on positive definite matrices. Section 4 is devoted to demonstrating that the dualistic geometry on the q -Gaussian family induced by the β -divergence coincides with the V -geometry with the power potentials. Finally, Section 5 gives concluding remarks.

2. Preliminaries: Geometries on Positive Definite Matrices and the U -Model

We recall the relation between information geometry on positive definite matrices induced by V -potentials and that on a multivariate statistical model called the U -model. Details described in this section and the ideas behind them can be found in [1,24].

2.1. V -Potential Function and the Induced Geometry on Positive Definite Matrices

Denote by $Sym(n, \mathbf{R})$ the vector space of $n \times n$ real symmetric matrices, and by $PD(n, \mathbf{R})$ the convex cone of $n \times n$ positive definite matrices in $Sym(n, \mathbf{R})$, respectively. For two matrices, X and Y , in $Sym(n, \mathbf{R})$, we denote $\text{tr}(XY)$ by $\langle X, Y \rangle$. For an arbitrary set of basis matrices $\{E_i\}_{i=1}^{n(n+1)/2}$ of $Sym(n, \mathbf{R})$, we can introduce a coordinate system, (x^i) , by $X = \sum_{i=1}^{n(n+1)/2} x^i E_i$. Note that $Sym(n, \mathbf{R})$

is isomorphic to a tangent vector space at each $P \in PD(n, \mathbf{R})$. Hence, we particularly identify E_i with a tangent vector, $(\partial/\partial x^i)_P$.

Definition 1. Let $V(s)$ be a smooth function of real numbers $s > 0$. The function defined by:

$$\varphi^{(V)}(P) = V(\det P) \quad (1)$$

is called a V -potential on $PD(n, \mathbf{R})$.

When $V(s) = -\log s$, the V -potential reduces to the standard one, called the characteristic function on $PD(n, \mathbf{R})$, which plays a fundamental role in the geometrical theory of $PD(n, \mathbf{R})$ [26–28].

Let $\nu_i(s), i = 1, 2, \dots$ be functions defined by

$$\nu_i(s) = \frac{d\nu_{i-1}(s)}{ds} s, \quad i = 1, 2, \dots, \quad \text{where } \nu_0(s) = V(s) \quad (2)$$

We assume that $V(s)$ satisfies the following two conditions:

$$\text{i) } \nu_1(s) < 0 \quad (s > 0), \quad \text{ii) } \beta^{(V)}(s) = \frac{\nu_2(s)}{\nu_1(s)} < \frac{1}{n} \quad (s > 0) \quad (3)$$

which are later shown to ensure the convexity of $\varphi^{(V)}(P)$ on $PD(n, \mathbf{R})$. Note that the first condition, $\nu_1(s) < 0$, for all $s > 0$ implies the function, $V(s)$, is strictly decreasing on $s > 0$.

Using the formula $\text{grad } \det P = (\det P)P^{-1}$, we have the gradient mapping, $\text{grad}\varphi^{(V)}$:

$$\text{grad}\varphi^{(V)} : P \mapsto P^* = \nu_1(\det P)P^{-1} \quad (4)$$

The Hessian of $\varphi^{(V)}$ at $P \in PD(n, \mathbf{R})$, which we write as $g_P^{(V)}$, is given by:

$$g_P^{(V)}(X, Y) = -\nu_1(\det P) \text{tr}(P^{-1}XP^{-1}Y) + \nu_2(\det P) \text{tr}(P^{-1}X) \text{tr}(P^{-1}Y) \quad (5)$$

for arbitrary tangent vectors, X and Y , in $Sym(n, \mathbf{R})$.

Proposition 1. [24] The Hessian, $g^{(V)}$, is positive definite on $PD(n, \mathbf{R})$, if and only if the conditions in Equation (3) hold.

To establish the Legendre relation on $PD(n, \mathbf{R})$, we consider the conjugate function of $\varphi^{(V)}$ denoted by $\varphi^{(V)*}$. Define the Legendre transform:

$$\varphi^{(V)*}(P^*) = \sup_{P \in PD(n, \mathbf{R})} \{ \langle P^*, P \rangle - \varphi^{(V)}(P) \} \quad (6)$$

Since the extremal condition is:

$$P^* = \text{grad}\varphi^{(V)}(P) = \nu_1(\det P)P^{-1} \quad (7)$$

and $\text{grad}\varphi^{(V)}$ is invertible by the positive definiteness of $g^{(V)}$, we have the following expression for $\varphi^{(V)*}$ with respect to P :

$$\varphi^{(V)*}(P^*) = n\nu_1(\det P) - \varphi^{(V)}(P) \quad (8)$$

Hence, the canonical divergence [23] $D^{(V)}$ of $(PD(n, \mathbf{R}), \nabla, g^{(V)})$ is obtained as:

$$\begin{aligned} D^{(V)}(P, Q) &= \varphi^{(V)}(P) + \varphi^{(V)*}(Q^*) - \langle Q^*, P \rangle \\ &= V(\det P) - V(\det Q) + \langle Q^*, Q - P \rangle \end{aligned} \quad (9)$$

Regarding $g^{(V)}$ as a Riemannian metric, we can consider $PD(n, \mathbf{R})$ as a Riemannian manifold. Further, using the canonical flat affine connection on $Sym(n, \mathbf{R})$ denoted by ∇ , define the *dual affine connection* [23], ${}^*\nabla^{(V)}$, satisfying:

$$Xg^{(V)}(Y, Z) = g^{(V)}(\nabla_X Y, Z) + g^{(V)}(Y, {}^*\nabla^{(V)}_X Z) \quad (10)$$

for arbitrary tangent vector fields, X, Y and Z , on $PD(n, \mathbf{R})$; then, we can introduce a *dually flat structure* [23] or a *Hessian structure* [29] $(PD(n, \mathbf{R}), g^{(V)}, \nabla, {}^*\nabla^{(V)})$.

Their covariant derivatives at P are actually given by:

$$\begin{aligned} (\nabla_{\partial/\partial x^i} \partial/\partial x^j)_P &= 0 \\ ({}^*\nabla^{(V)}_{\partial/\partial x^i} \partial/\partial x^j)_P &= -E_i P^{-1} E_j - E_j P^{-1} E_i - \Phi(E_i, E_j, P) - \Phi^\perp(E_i, E_j, P) \end{aligned} \quad (11)$$

where:

$$\begin{aligned} \Phi(X, Y, P) &= \frac{\nu_2(s) \operatorname{tr}(P^{-1} X)}{\nu_1(s)} Y + \frac{\nu_2(s) \operatorname{tr}(P^{-1} Y)}{\nu_1(s)} X \\ \Phi^\perp(X, Y, P) &= \frac{\{\nu_3(s)\nu_1(s) - 2\nu_2^2(s)\} \operatorname{tr}(P^{-1} X) \operatorname{tr}(P^{-1} Y) + \nu_2(s)\nu_1(s) \operatorname{tr}(P^{-1} X P^{-1} Y)}{\nu_1(s)\{\nu_1(s) - n\nu_2(s)\}} P \end{aligned}$$

and $s = \det P$.

Since several properties of the pair of mutually dual connection, ∇ and ${}^*\nabla^{(V)}$, are stated in [24], we omit them here for the sake of simplicity. However, from a geometrical viewpoint, we should note that the following two important properties are related to the invariance of structures. In Section 3, we shall return to these points and discuss them in detail.

Proposition 2. [24]

1. The dually flat structure, $(PD(n, \mathbf{R}), g^{(V)}, \nabla, {}^*\nabla^{(V)})$, is $SL(n, \mathbf{R})$ -invariant for general $V(s)$ satisfying Equation (3), while it is $GL(n, \mathbf{R})$ -invariant for $V(s) = c_1 + c_2 \log s$ for real constants, c_1 and $c_2 < 0$.
2. When $V(s)$ is a power function (with a constant term) of the form:

$$V(s) = c_1 + c_2 s^\beta \quad (12)$$

for real constants, c_1, c_2 and β , satisfying Equation (3), both affine connections, ∇ and ${}^*\nabla^{(V)}$, are $GL(n, \mathbf{R})$ -invariant. Further, the orthogonality with respect to $g^{(V)}$ is also $GL(n, \mathbf{R})$ -invariant, while $g^{(V)}$ itself is not.

Remark 1. One of interesting implications is that the above second point means that for the power function in Equation (12), both ∇ - and ${}^*\nabla^{(V)}$ -projections [23], which can be also variationally characterized by the divergence in Equation (9) and its dual, are $GL(n, \mathbf{R})$ -invariant; hence, so is the Pythagorean theorem [23] as a one-dimensional special case for this implication. Conversely, $GL(n, \mathbf{R})$ -invariance of the Pythagorean theorem means those of both projections, because ∇ and ${}^*\nabla^{(V)}$ are torsion-free [23].

2.2. Relation between Information Geometries on the U -model and Positive Definite Matrices

We briefly introduce the U -divergence and U -model and show how the dualistic geometries induced from U -divergence and V -potential are related.

In the field of statistical inference, the well-established method is the maximum likelihood method, which is based on the Kullback-Leibler divergence. To improve the robustness performance of the method, maintaining its theoretical advantages, such as efficiency, the methods of minimizing general divergences have been proposed as alternatives to the maximum likelihood method [1,13,30–32].

Definition 2. Let $U(s)$ be a smooth convex function with the positive derivative $u(s) = U'(s) > 0$ on $\mathbf{R}$ or its (semi-infinite) interval and ξ be the inverse function of u there. If the following functional for two functions, $f(x)$ and $g(x)$, on $\mathbf{R}^n$:

$$D_U(f, g) = \int U(\xi(g)) - U(\xi(f)) - \{\xi(g) - \xi(f)\} f dx \quad (13)$$

exists, we call it the U -divergence.

It follows that $D_U(f, g) \geq 0$ and $D_U(f, g) = 0$, if and only if $f = g$, because the integrand, $U(\xi_g) - \{U(\xi_f) + u(\xi_f)(\xi_g - \xi_f)\}$, where $\xi_f = \xi(f)$ and $\xi_g = \xi(g)$, is interpreted as the difference of the convex function, U , and its supporting function. If we set $U(s) = \frac{1}{\beta+1}(1 + \beta s)^{(\beta+1)/\beta}$ for $\beta \in \mathbf{R}$, then the corresponding U -divergence is the *beta-divergence* [1] defined by:

$$D_\beta(f, g) = \int \frac{g(x)^{\beta+1} - f(x)^{\beta+1}}{\beta + 1} - \frac{f(x)\{g(x)^\beta - f(x)^\beta\}}{\beta} dx \quad (14)$$

As β goes to zero, it reduces to the Kullback-Leibler divergence; on the other hand, as β goes to one, it reduces to the squared L^2 -distance. Thus, the efficiency increases as β goes to zero, while the robustness increases as β goes to one. In this sense, we could find an appropriate β between zero and one as a trade-off between efficiency and robustness. The beta-divergence is strongly connected to the Tsallis entropy [33].

When we consider the family of functions parametrized by elements in a manifold, $\mathcal{M}$, the divergences on the family induce the dualistic structure on $\mathcal{M}$ [1]. Concretely, we here confine our attention to the family of multivariate probability density functions specified by P in $\mathcal{M} = PD(n, \mathbf{R})$ and then study its structure on $PD(n, \mathbf{R})$. The family is natural in the sense that it is a dually flat statistical manifold with respect to the dualistic geometry induced by the U -divergence.

Definition 3. Let U and u be the functions given in Definition 2. The family of elliptical distributions with the following density functions:

$$\mathcal{M}_U = \left\{ f_U(x, P) = u \left(-\frac{1}{2} x^T P x - c_U(\det P) \right) \middle| P \in PD(n, \mathbf{R}) \right\} \quad (15)$$

is called the U -model associated with the U -divergence. Here, we set $f(x, P) = 0$ if the right-hand side is nonpositive or undefined, and $c_U(\det P)$ is a normalizing constant.

Note that if u satisfies the following self-similarity property:

$$\forall t = c_U(\det P), \exists a_t, \exists b_t, \quad u(s - t) = a_t u(b_t s) \quad (16)$$

where a_t and b_t are positive constants depending on $\det P$, the density function, f_U , in the U -model can be alternatively expressed in the usual form of an elliptical distribution [20,21], i.e.,

$$f_U(x, P) = a_t u(-x^T P' x / 2), \quad P' = b_t P \quad (17)$$

One such example is the β -model, to be discussed in Section 4 and the Appendix. Thus, the probability density function, $f_U(x, P)$, has the mean vector of zero and the variance matrix, cP^{-1} , where c is a positive constant obtained from the characteristic function of $f_U(x, P)$, and P is called the *precision matrix*.

Now, we consider the correspondence between the dualistic geometry induced by D_U on the U -model and that on $PD(n, \mathbf{R})$ induced by the V -potential function discussed.

Proposition 3. [24] Define the V -potential function, $\varphi^{(V)}$, via:

$$V(s) = s^{-\frac{1}{2}} \int U\left(-\frac{1}{2}x^T x - c_U(s)\right) dx + c_U(s), \quad s > 0 \quad (18)$$

Assume that V satisfies the conditions in Equation (3); then, the dually flat structure, $(g^{(V)}, \nabla, {}^*\nabla^{(V)})$, on $PD(n, \mathbf{R})$ coincides with that on U -model induced by the U -divergence, D_U .

2.3. Statistical Estimation on the U -Model

We discuss a statistical estimation for the precision matrix parameter, P , in the U -model, $\mathcal{M}_U$. The U -divergence, $D_U(f, g)$, is decomposed into the difference of U -cross entropy, $C_U(f, g)$, and U -entropy, $H_U(f)$ [1], that is, $D_U(f, g) = C_U(f, g) - H_U(f)$, where

$$C_U(f, g) = - \int \xi(g(x)) f(x) dx + \int U(\xi(g(x))) dx \quad (19)$$

and $H_U(f) = C_U(f, f)$. Consider a maximum U -entropy distribution on the moment equal space. Let $\mathcal{F}$ be the space of all probability density functions on $\mathbf{R}^n$ and

$$\mathcal{F}(P) = \{f \in \mathcal{F} : \mathbb{E}_f(X) = 0, \mathbb{E}_f(XX^T) = cP^{-1}\} \quad (20)$$

the zero-mean and equal variance space in $\mathcal{F}$, where $\mathbb{E}_f$ denotes the statistical expectation with respect to f , and c is the aforementioned constant. Then, we observe that:

$$f_U(\cdot, P) = \operatorname{argmin}_{f \in \mathcal{F}(P)} H_U(f) \quad (21)$$

In effect, for any f in $\mathcal{F}(P)$,

$$H(f_U(\cdot, P)) - H_U(f) = D_U(f_U(\cdot, P), f) \geq 0 \quad (22)$$

with equality only if $f = f_U(\cdot, P)$. Thus, the U -model, $\mathcal{M}_U$, is characterized by maximum U -entropy distributions.

Let $\{X_i\}_{1 \leq i \leq N}$ be random samples from a probability density function, $f_U(x, P)$. Then, the U -loss function is defined by

$$L_U(P) = -\frac{1}{N} \sum_{i=1}^N \xi(f_U(X_i, P)) + \int U(\xi(f_U(x, P))) dx \quad (23)$$

in which the U -estimator, $\hat{P}_U$, is defined by the minimizer of $L_U(P)$. The U -loss function is an empirical analogue of U -cross entropy, in the sense that $\mathbb{E}_f L_U(P) = C_U(f, f(\cdot, P))$. By definition, $\hat{P}_U$ is a solution of $\frac{\partial}{\partial P} L_U(P) = 0$ if the solution is unique. Hence, we conclude that if $N \geq n$, then $\hat{P}_U = cS^{-1}$, where S is the sample variance matrix defined by $\frac{1}{N} \sum_{i=1}^N X_i X_i^T$, because

$$\frac{\partial}{\partial P} L_U(P) = \frac{1}{2} \left(S - \int x x^T f_U(x, P) dx \right) \quad (24)$$

and $cP^{-1} = \int x x^T f_U(x, P) dx$. We remark that S is positive-definite with a probability of one if $N \geq n$. The derivation for $\hat{P}_U$ is confirmed by the following fact

$$L_U(P) - L_U(\hat{P}_U) = D_U(f_U(\cdot, \hat{P}_U), f_U(\cdot, P)) \geq 0 \quad (25)$$

with equality, only if $P = \hat{P}_U$.

Surprisingly, such U -estimators are independent of the choice of U , which implies that a U -estimator for a U -model equals the maximum likelihood estimator for the Gaussian model. On the other hand, assume that $\{X_i\}_{1 \leq i \leq N}$ are random samples from a Gaussian density function, that is, $f_U(x, P)$ with $U = \exp$. Then, unless $U = \exp$, the U -estimator, $\hat{P}_U$, for the precision parameter, P , in the Gaussian model has no exact expression. This is rather a different aspect from the situation discussed above, and $\hat{P}_U$ solved by an iteration algorithm is shown to be robust against heavy outliers if the generator function, U , satisfies a tail condition.

For example, if we select as $U(s) = \frac{1}{\beta+1}(1 + \beta s)^{(\beta+1)/\beta}$ with a fixed $\beta > 0$, then the corresponding divergence is the β -divergence given in Equation (14), and the corresponding U -estimator is called the β -estimator. The estimator is associated with the iteration algorithm, say $\{P_t, t = 1, 2, \dots\}$, with an initial value, P_1 , in which the update, P_{t+1} , from the t -step, P_t , is given by:

$$P_{t+1}^{-1} = \frac{1}{\sum_{i=1}^N w(X_i, P_t) - d_\beta} \sum_{i=1}^N w(X_i, P_t) X_i X_i^T \quad (26)$$

where $w(x, P) = \exp(-\frac{\beta}{2} x^T P x)$ and $d_\beta = \beta/(\beta + 1)^{n+1}$. See [1] for a detailed discussion. We remark that the β -estimator, $\hat{P}_\beta$, satisfies the fixed-point condition by setting as $P_{t+1} = P_t = \hat{P}_\beta$ in Equation (26). Therefore, if the i -th observation, X_i , has an extremely large value, $X_i^T \hat{P}_\beta X_i$, then the i -th weight $w(X_i, \hat{P}_\beta)$ in the weighted variance form of Equation (26) becomes negligible, so that the β -estimator is automatically robust for these outliers. The degree of robustness for $\hat{P}_\beta$ depends on the value of β . In this way, it is also possible to introduce a dualistic structure on the product space of (U, U') for U -models and U' -estimators.

3. $GL(n, \mathbf{R})$ -Invariance Induced from the Power Function

The transformation group, τ_G , with $G \in GL(n, \mathbf{R})$, transitively acts on $PD(n, \mathbf{R})$, i.e., there exists $G \in GL(n, \mathbf{R})$ for all $P, P' \in PD(n, \mathbf{R})$, such that $\tau_G P = G P G^T = P'$. We denote by τ_{G*} the differential of τ_G . The invariance of geometry for these transformations is defined as follows:

Definition 4. We say a dually flat structure, $(PD(n, \mathbf{R}), g^{(V)}, \nabla, {}^*\nabla^{(V)})$, is $GL(n, \mathbf{R})$ -invariant if the Riemannian metric, $g^{(V)}$, and the pair of mutually dual connections, ∇ and ${}^*\nabla^{(V)}$, satisfy:

$$g_P^{(V)}(X, Y) = g_{P'}^{(V)}(X', Y') \quad (27)$$

$$\tau_{G*}(\nabla_X Y)_P = (\nabla_{X'} Y')_{P'}, \quad \tau_{G*}({}^*\nabla_X^{(V)} Y)_P = ({}^*\nabla_{X'}^{(V)} Y')_{P'} \quad (28)$$

for arbitrary $G \in GL(n, \mathbf{R})$, where $P' = \tau_G P$, $X' = \tau_{G*} X$ and $Y' = \tau_{G*} Y$, respectively.

We can similarly define $SL(n, \mathbf{R})$ -invariance while $SL(n, \mathbf{R})$ is not transitive on $PD(n, \mathbf{R})$. These invariances mean that the geometries are homogeneous to the corresponding transformations. They practically imply that the obtained geometrical results are not influenced by scaling (unit change), numerical conditioning, and so on. Note that $(PD(n, \mathbf{R}), g^{(V)}, \nabla, {}^*\nabla^{(V)})$ is $GL(n, \mathbf{R})$ -invariant, if and only if the corresponding canonical divergence is also, i.e.,

$$\forall G \in GL(n, \mathbf{R}), \quad D^{(V)}(P, Q) = D^{(V)}(\tau_G P, \tau_G Q) \quad (29)$$

because the dually flat structure can be alternatively derived from the canonical divergence [1,23].

Now, we fix the form of V as $V(s) = c_1 + c_2 s^\beta$ to actually confirm the invariance property described in the second statement of Proposition 2. First, the convexity conditions in Equation (3) reduce to:

$$(i) \beta c_2 < 0, \quad (ii) \beta < \frac{1}{n} \quad (30)$$

The dual variables, P^* , and the conjugate function, $\varphi^{(V)*}$, are expressed by P as:

$$P^* = c_2 \beta (\det P)^\beta P^{-1}, \quad \varphi^{(V)*}(P^*) = n c_2 \beta (\det P)^\beta - c_1 - c_2 (\det P)^\beta$$

The corresponding Riemannian metric in Equation (5) and divergence in Equation (9) are respectively given by:

$$g_P^{(V)}(X, Y) = c_2 \beta (\det P)^\beta \{ -\text{tr}(P^{-1} X P^{-1} Y) + \beta \text{tr}(P^{-1} X) \text{tr}(P^{-1} Y) \} \quad (31)$$

$$D^{(V)}(P, Q) = c_2 \{ (\det P)^\beta - (\det Q)^\beta + \beta (\det Q)^\beta \text{tr}(I - Q^{-1} P) \} \quad (32)$$

When we particularly set $c_1 = -c_2 = 1/\beta$, i.e., $V(s) = (1 - s^\beta)/\beta$, and move β to zero, we find that they converge to the standard Riemannian metric and divergence [27] for $V(s) = -\log s$. We immediately see that the above $g^{(V)}$ and $D^{(V)}$ are not $GL(n, \mathbf{R})$ - but $SL(n, \mathbf{R})$ -invariant. However, $g_P^{(V)}(X, Y) = 0$, if and only if $g_{P'}^{(V)}(X', Y') = 0$ for any $G \in GL(n, \mathbf{R})$. Thus, the orthogonality is $GL(n, \mathbf{R})$ -invariant.

Next, the covariant derivatives of ∇ vanish everywhere, since it is the canonical flat affine connection. For the covariant derivatives of ${}^*\nabla^{(V)}$, it is seen that the first and second terms in Equation (11) are $GL(n, \mathbf{R})$ -invariant. The third and fourth terms of the dual covariant derivatives respectively reduce to:

$$\Phi(X, Y, P) = \beta \text{tr}(P^{-1} X) Y + \beta \text{tr}(P^{-1} Y) X \quad (33)$$

$$\Phi^\perp(X, Y, P) = \frac{-\beta^2 \text{tr}(P^{-1} X) \text{tr}(P^{-1} Y) + \beta \text{tr}(P^{-1} X P^{-1} Y)}{1 - n\beta} P \quad (34)$$

which are independent of $\det P$. Thus, we can find that both ∇ and ${}^*\nabla^{(V)}$ are $GL(n, \mathbf{R})$ -invariant.

Finally, consider two smooth curves, $\gamma = \{C_\gamma(t) | -\epsilon < t < \epsilon\}$ and $\gamma^* = \{C_{\gamma^*}(t) | -\epsilon < t < \epsilon\}$, in $PD(n, \mathbf{R})$, satisfying:

$$C_\gamma(0) = C_{\gamma^*}(0) = Q, \quad \frac{dC_\gamma}{dt}(0) = X, \quad \frac{dC_{\gamma^*}}{dt}(0) = Y \quad (35)$$

Since ∇ is the canonical flat connection on $Sym(n, \mathbf{R})$, γ is ∇ -geodesic iff it is represented by $C_\gamma(t) = Q + tX$. On the other hand, γ^* is ${}^*\nabla^{(V)}$ -geodesic iff it is represented as a straight line in the dual variable, $C_{\gamma^*}^*(t)$ [23]. Let us obtain its explicit form. Since it follows that:

$$\begin{aligned} \frac{d}{dt}C_{\gamma^*}^* &= c_2\beta \left\{ \left(\frac{d}{dt}(\det C_{\gamma^*})^\beta \right) C_{\gamma^*}^{-1} + (\det C_{\gamma^*}) \left(\frac{d}{dt}C_{\gamma^*}^{-1} \right) \right\} \\ &= c_2\beta^2(\det C_{\gamma^*})^{\beta-1} \left\langle \text{grad } \det C_{\gamma^*}, \frac{dC_{\gamma^*}}{dt} \right\rangle + c_2\beta(\det C_{\gamma^*})^\beta \left(-C_{\gamma^*}^{-1} \frac{dC_{\gamma^*}}{dt} C_{\gamma^*}^{-1} \right) \end{aligned} \quad (36)$$

we have:

$$\tilde{Y} = \frac{dC_{\gamma^*}^*}{dt}(0) = c_2\beta(\det Q)^\beta \{ \beta \text{tr}(Q^{-1}Y)Q^{-1} - Q^{-1}YQ^{-1} \} \quad (37)$$

by substituting Equation (35). Thus, γ^* is ${}^*\nabla^{(V)}$ -geodesic iff it is represented in the dual variables by:

$$C_{\gamma^*}^*(t) = Q^* + t\tilde{Y} \quad (38)$$

Assume that X and Y are mutually orthogonal at Q , and two points, P and R , in $PD(n, \mathbf{R})$ are, respectively, located on the ∇ -geodesic γ and ${}^*\nabla^{(V)}$ -geodesic γ^* satisfying Equation (35), i.e.,

$$P = Q + t_1X, \quad R^* = Q^* + t_2\tilde{Y}, \quad g_Q^{(V)}(X, Y) = 0 \quad (39)$$

for some real numbers, t_1 and t_2 . Then, we have:

$$\begin{aligned} D^{(V)}(P, Q) + D^{(V)}(Q, R) - D^{(V)}(P, R) &= \langle R^* - Q^*, P - Q \rangle \\ &= t_1t_2\langle \tilde{Y}, X \rangle \\ &= t_1t_2g_Q^{(V)}(X, Y) = 0 \end{aligned} \quad (40)$$

which results in the Pythagorean theorem. If we, respectively, transform Q, X, Y in Equation (35) to $Q' = \tau_G Q, X' = \tau_{G*} X, Y' = \tau_{G*} Y$, we see that Equations (39) and (40) hold by replacing t_2 with $t_2(\det G)^{2\beta}$. Thus, even if such three points, P, Q, R , are, respectively, transformed by τ_G to P', Q', R' with arbitrary, but common, $G \in GL(n, \mathbf{R})$, the Pythagorean theorem still holds for P', Q', R' .

The $GL(n, \mathbf{R})$ -invariances of both ∇ - and ${}^*\nabla^{(V)}$ -projections are similarly confirmed.

Thus, we have confirmed the second statement of Proposition 2, i.e., the above $GL(n, \mathbf{R})$ -invariance holds if $V(s) = c_1 + c_2s^\beta$. In fact, we can show that the converse of Proposition 2 is also true.

Theorem 1. Assume that the function, V , meets Equation (3). Mutually dual connections, ∇ and ${}^*\nabla^{(V)}$, and the orthogonality with respect to $g^{(V)}$ are $GL(n, \mathbf{R})$ -invariant, if and only if $V(s) = c_1 + c_2s^\beta$ or $V(s) = c_1 + c_2 \log s$.

Proof. We only show the if part. The covariant derivatives of ∇ are independent of V and clearly invariant. For those of $^*\nabla^{(V)}$ in Equation (11), the first and second terms can be readily understood to be invariant. The third term, Φ , and the fourth term, $\Phi^\perp$, are invariant, only if the ratios, $\nu_2(s)/\nu_1(s)$ and $\nu_3(s)/\nu_1(s)$, also are, because the coefficients in these terms are respectively expressed by:

$$\frac{\nu_2}{\nu_1} \operatorname{tr}(P^{-1}X), \quad \frac{\nu_2}{\nu_1} \operatorname{tr}(P^{-1}Y) \quad (41)$$

$$\frac{\nu_3/\nu_1 - 2(\nu_2/\nu_1)^2}{1 - n\nu_2/\nu_1} \operatorname{tr}(P^{-1}X) \operatorname{tr}(P^{-1}Y) + \frac{\nu_2/\nu_1}{1 - n\nu_2/\nu_1} \operatorname{tr}(P^{-1}XP^{-1}Y) \quad (42)$$

where $\operatorname{tr}(P^{-1}X)$, $\operatorname{tr}(P^{-1}Y)$ and $\operatorname{tr}(P^{-1}XP^{-1}Y)$ are invariant.

From the definition of $\nu_i(s)$'s, the invariance of $\nu_2(s)/\nu_1(s)$ is satisfied by the solutions, $\nu_1(s)$, of the ordinary differential equation:

$$\frac{1}{\nu_1} \frac{d\nu_1}{ds} = \frac{\beta}{s} \quad (43)$$

for a real constant, β . By solving the ODE and integrating again, we have $V(s) = c_1 + c_2 s^\beta$ or $V(s) = c_1 + c_2 \log s$ for real constants, c_i , $i = 1, 2$. These forms also meet the invariance of $\nu_3(s)/\nu_1(s)$. The invariance of the orthogonality for such functions, V , have been already confirmed. $\square$

Note that the above theorem does not reject the possibility of the $GL(n, \mathbf{R})$ -invariances for the potentials of the other forms, except the V -potentials.

4. Geometry on the q -Gaussian Family Induced by the β -Divergence and the V -Geometry

This section demonstrates the main result.

Let β be a real parameter satisfying $\beta \neq 0$ and $\beta \neq -1$ and define a function, U , by:

$$U(s) = \begin{cases} \frac{1}{\beta+1} (1+\beta s)^{(\beta+1)/\beta}, & s \in I_\beta = \{s \in \mathbf{R} | 1+\beta s > 0\} \\ 0, & \text{otherwise} \end{cases} \quad (44)$$

Using its derivative, we define a function, u , by:

$$u(s) = \begin{cases} \frac{dU(s)}{ds} = (1+\beta s)^{1/\beta} & s \in I_\beta = \{s \in \mathbf{R} | 1+\beta s > 0\} \\ 0, & \text{otherwise} \end{cases} \quad (45)$$

and the inverse, ξ of u , on I_β by:

$$\xi(t) = \frac{t^\beta - 1}{\beta}, \quad t > 0 \quad (46)$$

Note that U is convex and u is positive where $s > -1/\beta$ if $\beta > 0$ and $s < -1/\beta$ if $\beta < 0$, respectively. Further, u and ξ respectively approach the usual exponential and logarithmic functions when β goes to zero. Hence, by introducing a parameter, $q = 1 + \beta$, they are called the q -exponential and q -logarithmic functions in the literature of nonextensive statistical physics [4,6].

Let us fix the parameter, β , arbitrarily and consider an elliptical density function, $f(x, P)$, specified by $P \in PD(n, \mathbf{R})$ using the q -exponential function, u :

$$f(x, P) = u \left(-\frac{1}{2} x^T P x - c_\beta (\det P) \right) \quad (47)$$

where $c_\beta(\det P)$ is the normalizing constant. The density function, f , is said to be the (zero-mean) q -Gaussian [4,6], and we call the family of all such f 's denoted by $\mathcal{M}_\beta$ the q -Gaussian family or β -model.

By starting from the β -divergence, we can define a dualistic structure on $\mathcal{M}_\beta$, which invokes the corresponding V -geometry on the parameter space, $PD(n, \mathbf{R})$. The V -potential for the V -geometry is obtained as follows:

Theorem 2. *The information geometry on $\mathcal{M}_\beta$ induced from the β -divergence is characterized by a dually flat structure, $(g^{(V)}, \nabla, {}^*\nabla^{(V)})$, on $PD(n, \mathbf{R})$ induced by the V -potentials:*

$$V(s) = \begin{cases} \frac{1}{\beta} + c^+ s^{1/(2n_\beta)}, & \beta > 0 \\ \frac{1}{\beta} + c^- s^{1/(2n_\beta)}, & -\frac{2}{n+2} < \beta < 0 \end{cases} \quad (48)$$

satisfying Equation (30), where $s = \det P$, $n_\beta = n/2 + 1/\beta$ and $c^\pm$ are constants depending on β and n .

The proof can be found in the Appendix.

The above theorem implies that geometric structure on $\mathcal{M}_\beta$ induced from the β -divergence admits the natural invariance properties discussed in Section 3.

5. Conclusions

We prove that information geometry on the q -Gaussian family induced by the β -divergence is equivalently characterized by the V -geometry on the space of positive definite matrices induced by the power potential. Studying the corresponding V -geometry, we show that some of the dually flat structures of the q -Gaussian family admit the $GL(n, \mathbf{R})$ -invariances. This fact implies the importance of the family in multivariate statistical analysis, as well as gives a geometrical viewpoint to mathematical properties of the q -exponential functions. Following the way given in Section 2.1, we can introduce the other dually flat structures via any convex potentials in addition to the V -potential, by defining Riemannian metrics as their Hessians and dual flat connections satisfying Equation (10) for ∇ . The relations between such dually flat structures and the ones on the other deformed exponential families are left for future work.

Robustness in statistical estimations involving the q -Gaussian family (β -model) and the β -divergence, which is another important aspect, is roughly explained at the end of Section 2.3.

Recently, the theory of optimal transportation has made great developments in which a geometrical insight is founded with a close relation to the problem of Poincaré's conjecture [34]. We can find some ideas and arguments similar to those established in this paper, while no direct link exists between their objectives. In effect, a divergence is defined in [35] and shown to play a significant role on checking a condition for the existence of the optimal transportation, where geometry is deeply explored to consider a family of probability density functions on a space in place of investigating properties of the space directly. It is expected that a coupling of the theory and information geometry will give fruitful results in the near future.

Acknowledgments

We thank the anonymous referees for their useful comments and suggestions, in particular, for pointing out a relation with the optimal transportation theory. Atsumi Ohara was partially supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid (C) 23540134. Shinto Eguchi was supported by Japan Science and Technology Agency (JST), Core Research for Evolutionary Science and Technology (CREST).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Eguchi, S. Information geometry and statistical pattern recognition. *Sugaku Expos.* **2006**, *19*, 197–216.
2. Grünwald, P.D.; Dawid, A.P. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Ann. Stat.* **2004**, *32*, 1367–1433.
3. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Deformed logarithms and entropies. *Physica A* **2004**, *340*, 41–49.
4. Naudts, J. *Generalized thermostatics*; Springer: London, UK, 2010.
5. Ollila, E.; Tyler, D.; Koivunen, V.; Poor, V. Complex elliptically symmetric distributions : Survey, new results and applications. *IEEE Trans. Signal Process.* **2012**, *60*, 5597–5623.
6. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
7. Barndorff-Nielsen, O.E. *Information and Exponential Families in Statistical Theory*; Wiley: Chichester, UK, 1978.
8. Tsallis, C. What are the numbers that experiments provide? *Quimica Nova* **1994**, *17*, 468–471.
9. Naudts, J. The q -exponential family in statistical physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405–413.
10. Naudts, J.; Baeten, M. Non-extensivity of the configuration density distribution in the classical microcanonical emsenble. *Entropy* **2009**, *11*, 285–294.
11. Ohara, A.; Wada, T. Information geometry of q -Gaussian densities and behaviors of solutions to related diffusion equations. *J. Phys. A: Math. Theor.* **2010**, *43*, 035002.
12. Eguchi, S.; Copas, J. A class of logistic-type discriminant functions. *Biometrika* **2002**, *89*, 1–22.
13. Fujisawa, H.; Eguchi, S. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* **2008**, *99*, 2053–2081.
14. Higuchi, I.; Eguchi, S. Robust principal component analysis with adaptive selection for Tuning Parameters. *J. Mach. Learn. Res.* **2004**, *5*, 453–471.
15. Borges, E.P. A possible deformed algebra and calculus inspired in nonextensive thermostatics. *Physica A*, **2004**, *340*, 95–101.
16. Nivanen, L.; Le Mehaute, A; Wang, Q.A. Generalized algebra within a nonextensive statistics. *Rep. Math. Phys.*, **2003**, *52*, 437–444.

17. Suyari, H. Mathematical structures derived from the q -multinomial coefficient in Tsallis statistics. *Physica A* **2006**, *368*, 63–82.
18. Suyari, H.; Wada, T. Multiplicative duality, q -triplet and μ, ν, q -relation derived from the one-to-one correspondence between the (μ, ν) -multinomial coefficient and Tsallis entropy S_q . *Physica A* **2008**, *387*, 71–83.
19. Tsallis, C.; Gell-Mann, G.; Sato, Y. Asymptotically scale-invariant occupancy of phase space makes the entropy S_q . *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15377–15382.
20. Fang, K.T.; Kotz, S.; Ng, K.W. *Symmetric multivariate and related distributions*; Chapman and Hall: London, UK, 1990.
21. Muirhead, R.J. *Aspects of Multivariate Statistical Theory*; Wiley: New York, NY, USA, 1982.
22. Amari, S. *Differential-Geometrical Methods in Statistics*; Lecture notes in statistics, Volume 28; Springer: New York, NY, USA, 1985.
23. Amari, S.; Nagaoka, H. *Methods of Information Geometry (Translations of Mathematical Monographs)*; Oxford University Press: Oxford, UK, 2000.
24. Ohara, A.; Eguchi, S. *Geometry on Positive Definite Matrices Induced from V-Potential Function*; Nielsen, F., Barbaresco, F., Eds.; Geometric Science of Information; Lecture Notes in Computer Science 8085; Springer: Berlin/Heidelberg, Germany, 2013; pp. 621–629.
25. Eaton, M.L. *Group Invariance Applications in Statistics*; Institute of Mathematical Statistics: Hayward, CA, USA, 1989.
26. Faraut, J.; Korányi, A. *Analysis on Symmetric Cones*; Oxford University Press: New York, NY, USA, 1994.
27. Ohara, A.; Suda, N.; Amari, S. Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra Appl.* **1996**, *247*, 31–53.
28. Vinberg, E.B. The theory of convex homogeneous cones. *Trans. Moscow Math. Soc.* **1963**, *12*, 340–430.
29. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, Singapore, 2007.
30. Cichocki, A.; Cruces, S.; Amari, S. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy* **2011**, *13*, 134–170.
31. Eguchi, S.; Komori, O.; Kato, S.; Projective power entropy and maximum Tsallis entropy Distributions. *Entropy* **2011**, *13*, 1746–1764.
32. Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195.
33. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
34. Villani, C. *Optimal Transport: Old and New*; Springer: Heidelberg, Germany, 2009.
35. Lott, J.; Villani, C. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **2009**, *3*, 903–991.

Appendix: Proof of Theorem 2

Proof. We prove the theorem by solving the corresponding V -potential according to Equation (18).

Derivation for the case $\beta > 0$

In the case $\beta > 0$, we see that the density function, $f(x, P)$, has a bounded support, and $f(0, P) > 0$ implies $c_\beta(\det P) < 1/\beta$ by considering the positivity of u .

We first calculate $c_\beta(\det P)$. Set $y = P^{1/2}x$; then the normalizing condition is:

$$\int_{\mathbf{R}^n} f(x, P) dx = (\det P)^{-\frac{1}{2}} \int_{\mathbf{R}^n} u\left(-\frac{1}{2}y^T y - c_\beta(\det P)\right) dy = 1 \quad (49)$$

The equality is modified as:

$$(\det P)^{-\frac{1}{2}} \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^\rho u\left(-\frac{r}{2} - c_\beta(\det P)\right) r^{\frac{n}{2}-1} dr = 1 \quad (50)$$

where $\rho = \min\{r | u(-r/2 - c_\beta) = 0\} = -2c_\beta + 2/\beta > 0$ and Γ is the Euler's gamma function.

For a positive number, α , let us define the integral appearing in the above as:

$$I_{\alpha, \beta}(c_\beta) = \int_0^\rho u\left(-\frac{r}{2} - c_\beta\right) r^{\alpha-1} dr = \int_0^\rho \left(1 - \beta c_\beta - \frac{\beta}{2}r\right)^{1/\beta} r^{\alpha-1} dr \quad (51)$$

then, by a change of variables $r = \rho z$, we have:

$$\begin{aligned} I_{\alpha, \beta}(c_\beta) &= (\beta/2)^{1/\beta} \int_0^\rho (\rho - r)^{1/\beta} r^{\alpha-1} dr \\ &= (\beta/2)^{1/\beta} \int_0^1 \rho^{1/\beta} (1 - z)^{1/\beta} \rho^{\alpha-1} z^{\alpha-1} \rho dz \\ &= (\beta/2)^{1/\beta} \rho^{\alpha+1/\beta} B\left(\frac{1}{\beta} + 1, \alpha\right) \\ &= 2^\alpha \beta^{1/\beta} B\left(\frac{1}{\beta} + 1, \alpha\right) \left(\frac{1}{\beta} - c_\beta\right)^{\alpha+1/\beta} \end{aligned} \quad (52)$$

where B is Euler's beta function.

From Equations (50) and (52), the normalizing constant for $\alpha = n/2$ is given by:

$$c_\beta(s) = \frac{1}{\beta} - (k^+ s^{1/2})^{1/n_\beta} \quad (53)$$

where $s = \det P$, $n_\beta = n/2 + 1/\beta > 0$ and:

$$k^+ = \frac{(\frac{1}{\beta})^{1/\beta}}{(2\pi)^{\frac{n}{2}}} \frac{\Gamma(\frac{n}{2})}{B(\frac{1}{\beta} + 1, \frac{n}{2})} = \frac{(\frac{1}{\beta})^{1/\beta}}{(2\pi)^{\frac{n}{2}}} \frac{\Gamma(n_\beta + 1)}{\Gamma(\frac{1}{\beta} + 1)} > 0 \quad (54)$$

Now, to derive $V(s)$ for $\beta > 0$, substitute $U(s)$ into Equation (18). Then, we have:

$$V(s) = s^{-1/2} \frac{\pi^{\frac{n}{2}}}{(\beta + 1)\Gamma(\frac{n}{2})} \int_0^\rho \left(1 - \beta c_\beta(s) - \frac{\beta}{2}r\right)^{(\beta+1)/\beta} r^{\frac{n}{2}-1} dr + c_\beta(s) \quad (55)$$

Since the integral in the above equation can be computed similarly to Equation (52), it holds that:

$$V(s) = s^{-1/2} \frac{(2\pi)^{\frac{n}{2}} \beta^{1+\frac{1}{\beta}}}{(\beta+1)\Gamma(\frac{n}{2})} B\left(\frac{1}{\beta} + 2, \frac{n}{2}\right) \left(\frac{1}{\beta} - c_{\beta}(s)\right)^{n_{\beta}+1} + c_{\beta}(s) \quad (56)$$

Substituting $c_{\beta}(s)$ in Equation (53), we see that:

$$\begin{aligned} V(s) &= \kappa_1 s^{-1/2} (\kappa_2 s^{1/(2n_{\beta})})^{n_{\beta}+1} + \frac{1}{\beta} - \kappa_2 s^{1/(2n_{\beta})} \\ &= \frac{1}{\beta} + (\kappa_1 \kappa_2^{n_{\beta}} - 1) \kappa_2 s^{1/(2n_{\beta})} \end{aligned} \quad (57)$$

where κ_i , $i = 1, 2$ are constants respectively represented by:

$$\kappa_1 = \frac{(2\pi)^{\frac{n}{2}} \beta^{1+\frac{1}{\beta}}}{(\beta+1)} \frac{B\left(\frac{1}{\beta} + 2, \frac{n}{2}\right)}{\Gamma(\frac{n}{2})} = \frac{(2\pi)^{\frac{n}{2}} \beta^{1+\frac{1}{\beta}}}{(\beta+1)} \frac{\Gamma(\frac{1}{\beta} + 2)}{\Gamma(n_{\beta} + 2)}, \quad \kappa_2 = (k^+)^{1/n_{\beta}} \quad (58)$$

Using the well-known formula $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$, we see that the coefficient meets $\beta(\kappa_1 \kappa_2^{n_{\beta}} - 1) \kappa_2 < 0$. Further, it holds that the exponent meets $1/(2n_{\beta}) < 1/n$, since β is positive. These two conditions assure the positive definiteness of the Riemannian metric, $g^{(V)}$, given by Equation (30).

Derivation for the Case $\beta < 0$

In the case $\beta < 0$, the support of $f(x, P)$ is $\mathbf{R}^n$, and $f(0, P) > 0$ implies that $c_{\beta}(\det P) > 1/\beta$.

Again, we first calculate $c_{\beta}(\det P)$. The normalizing condition in Equation (49) is modified as:

$$(\det P)^{-\frac{1}{2}} \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{\infty} \frac{r^{\frac{n}{2}-1}}{(1 - \beta c_{\beta}(\det P) - \frac{\beta}{2}r)^{-1/\beta}} dr = 1 \quad (59)$$

where we assume that the integral converges. Let us define $\rho = -2c_{\beta} + 2/\beta < 0$ and consider changes of the variables $r = (-\rho)z$ and $z = (1 - \tilde{z})/\tilde{z}$; then, for a positive number, α , the integral in the above is:

$$\begin{aligned} \int_0^{\infty} \frac{r^{\alpha-1}}{(1 - \beta c_{\beta} - \frac{\beta}{2}r)^{-1/\beta}} dr &= (-\beta/2)^{1/\beta} \int_0^{\infty} \frac{r^{\alpha-1}}{((- \rho) + r)^{-1/\beta}} dr \\ &= (-\beta/2)^{1/\beta} (-\rho)^{\alpha+1/\beta} \int_0^{\infty} \frac{z^{\alpha-1}}{(1+z)^{-1/\beta}} dz \\ &= (-\beta/2)^{1/\beta} (-\rho)^{\alpha+1/\beta} \int_0^1 \tilde{z}^{-1/\beta-\alpha-1} (1-\tilde{z})^{\alpha-1} d\tilde{z} \\ &= 2^{\alpha} (-\beta)^{1/\beta} B\left(-\frac{1}{\beta} - \alpha, \alpha\right) \left(c_{\beta} - \frac{1}{\beta}\right)^{\alpha+1/\beta} \end{aligned} \quad (60)$$

Thus, by $\alpha = n/2$ the convergence condition for β is:

$$\beta > -2/n \quad (61)$$

From Equations (59) and (60), the normalizing constant is given by:

$$c_{\beta}(s) = \frac{1}{\beta} + (k^- s^{1/2})^{1/n_{\beta}} \quad (62)$$

where $s = \det P$, $n_\beta = n/2 + 1/\beta < 0$ and:

$$k^- = \frac{(-\frac{1}{\beta})^{1/\beta}}{(2\pi)^{\frac{n}{2}}} \frac{\Gamma(\frac{n}{2})}{B(-\frac{1}{\beta} - \frac{n}{2}, \frac{n}{2})} = \frac{(-\frac{1}{\beta})^{1/\beta}}{(2\pi)^{\frac{n}{2}}} \frac{\Gamma(-\frac{1}{\beta})}{\Gamma(-n_\beta)} > 0 \quad (63)$$

Now, to derive $V(s)$ for $\beta < 0$, substitute $U(s)$ into Equation (18):

$$V(s) = s^{-1/2} \frac{\pi^{\frac{n}{2}}}{(\beta + 1)\Gamma(\frac{n}{2})} \int_0^\infty \left(1 - \beta c_\beta(s) - \frac{\beta}{2}r\right)^{1+1/\beta} r^{\frac{n}{2}-1} dr + c_\beta(s) \quad (64)$$

Here, we assume that the exponent at least meets $1 + 1/\beta < 0$ for the above integral to converge. Then, the integral can be computed similarly to Equation (60) as:

$$V(s) = s^{-1/2} \frac{(2\pi)^{\frac{n}{2}} (-\beta)^{1+\frac{1}{\beta}}}{(\beta + 1)\Gamma(\frac{n}{2})} B\left(-n_\beta - 1, \frac{n}{2}\right) \left(c_\beta(s) - \frac{1}{\beta}\right)^{n_\beta+1} + c_\beta(s) \quad (65)$$

Thus, the exact convergence condition is $n_\beta + 1 < 0$ or, equivalently, $\beta > -2/(n+2)$, which is stronger than Equation (61) and $1 + 1/\beta < 0$. Substituting the expression of $c_\beta(s)$ in Equation (62), we see:

$$\begin{aligned} V(s) &= \kappa_3 s^{-1/2} (\kappa_4 s^{1/(2n_\beta)})^{n_\beta+1} + \frac{1}{\beta} + \kappa_4 s^{1/(2n_\beta)} \\ &= \frac{1}{\beta} + (\kappa_3 \kappa_4^{n_\beta} + 1) \kappa_4 s^{1/(2n_\beta)} \end{aligned} \quad (66)$$

where κ_i , $i = 3, 4$ are constants respectively represented by:

$$\kappa_3 = \frac{(2\pi)^{\frac{n}{2}} (-\beta)^{1+\frac{1}{\beta}}}{\beta + 1} \frac{B(-n_\beta - 1, \frac{n}{2})}{\Gamma(\frac{n}{2})} = \frac{(2\pi)^{\frac{n}{2}} (-\beta)^{1+\frac{1}{\beta}}}{(\beta + 1)} \frac{\Gamma(-n_\beta - 1)}{\Gamma(-\frac{1}{\beta} - 1)}, \quad \kappa_4 = (k^-)^{1/n_\beta}. \quad (67)$$

It is readily confirmed that $\beta (\kappa_3 \kappa_4^{n_\beta} + 1) \kappa_4 < 0$ holds. Further, we see $1/(2n_\beta) < 1/n$, since n_β is negative. These two conditions again assure the positive definiteness of $g^{(V)}$ given by Equation (30). $\square$