Modern Physics Letters B
Vol. 26, No. 10 (2012) 1250063 (14 pages)

DOI: 10.1142/S0217984912500637

CONFORMAL GEOMETRY OF ESCORT PROBABILITY AND ITS APPLICATIONS

ATSUMI OHARA
Department of Electrical and Electronics Engineering,
University of Fukui, Bunkyo 3-9-1, Fukui 910-8507, Japan
ohara@fuee.u-fukui.ac.jp

HIROSHI MATSUZOE
Department of Computer Science and Engineering,
Graduate School of Engineering, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
matsuzoe@nitech.ac.jp

SHUN-ICHI AMARI
Riken Brain Science Institute, Wako-shi Hirosawa 2-1, Saitama 351-0198, Japan
amari@brain.riken.jp

Received 11 November 2011
Accepted 5 February 2012

Escort probability is a certain modification of ordinary probability and a conformally transformed structure can be introduced on the space of its distributions. In this contribution applications of escort probabilities and such a structure are focused on. We demonstrate that they naturally appear and play important roles for computationally efficient method to construct α-Voronoi partitions and analysis of related dynamical systems on the simplex.

Keywords: Voronoi partitions; dynamical systems; information geometry.

1. Introduction

In the research areas of multifractals and nonextensive statistical mechanics, escort probability[1–3] appears in many aspects and is widely recognized as an important concept. It has been known[4,5] that nonextensive entropies are closely connected with the α-geometry.[6,7] Further, we have geometrically studied the space of escort distributions and reported[8–10] that the well-established and abundant structure (called the *dually flat structure*) can be introduced by a conformal transformation of the α-geometry.

The purpose of this contribution is to show that escort probability and the associated conformal structure are also natural and useful to the other applications.

First, we discuss the Voronoi partition with respect to the α-*divergence* (or *Rényi divergence*). The Voronoi partitions on the space of probability distributions with the Kullback–Leibler,[11,12] or Bregman divergences[13] are useful tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on. See also the literature[14–16] for related problems. The largest advantage to take account of α-divergences is their *invariance under transformations by sufficient statistics*,[7,17] which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the α-geometry enables us to invoke the standard algorithm[18,19] using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates. As another application, we explore properties of dynamical systems defined by the escort transformation and the gradient with respect to the conformal metric. These flows are fundamental from geometrical viewpoints[20] and found to possess interesting properties.

The paper is organized as follows: Sec. 2 is a short review of properties of information geometric structure induced on the family of escort distributions obtained by the authors.[8] Section 3 describes the first application of escort probability and the conformal geometric structure to α-Voronoi partitions on the simplex. The properties including computational efficiency of a construction algorithm are discussed. Further, a formula for α-centroid is touched upon. In Sec. 4, we discuss properties of dynamical systems related with escort transformation and gradient flows in view of the conformal geometry.

In the sequel, we use two equivalent parameters q and α following to conventions of several research areas, but their relation is fixed as $q = (1+\alpha)/2$. Additionally, we assume that $q > 0$.

2. Preliminary Results

In this section, we review and summarize results in Ref. 8.

Let $\mathcal{S}^n$ denote the n-dimensional probability simplex, i.e.

$$\mathcal{S}^n := \left\{ \boldsymbol{p} = (p_i) \,\middle|\, p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right\}, \tag{1}$$

and $p_i, i = 1, \ldots, n+1$ denote probabilities of $n+1$ states. We introduce the α-geometric structure[6,7] on $\mathcal{S}^n$. Let $\{\partial_i\}, i = 1, \ldots, n$ be natural basis tangent vector fields on $\mathcal{S}^n$ defined by

$$\partial_i := \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}, \quad i = 1, \ldots, n, \tag{2}$$

where $p_{n+1} = 1 - \sum_{i=1}^{n} p_i$. Now we define a Riemannian metric g on $\mathcal{S}^n$ called the

Fisher metric:

$$g_{ij}(\boldsymbol{p}) := g(\partial_i, \partial_j) = \frac{1}{p_i}\delta_{ij} + \frac{1}{p_{n+1}}$$

$$= \sum_{k=1}^{n+1} p_k(\partial_i \log p_k)(\partial_j \log p_k), \quad i, j = 1, \ldots, n\,. \tag{3}$$

Further, define a torsion-free affine connection $\nabla^{(\alpha)}$ called the *α-connection*, which is represented in its coefficients with a real parameter α by

$$\Gamma_{ij}^{(\alpha)k}(\boldsymbol{p}) = \frac{1+\alpha}{2}\left(-\frac{1}{p_k}\delta_{ij}^k + p_k g_{ij}\right), \quad i, j, k = 1, \ldots, n\,, \tag{4}$$

where δ_{ij}^k is equal to one if $i = j = k$ and zero otherwise. Then we have the α-covariant derivative $\nabla^{(\alpha)}$, which gives

$$\nabla_{\partial_i}^{(\alpha)}\partial_j = \sum_{k=1}^{n}\Gamma_{ij}^{(\alpha)k}\partial_k\,,$$

when it is applied to the vector fields ∂_i and ∂_j. We can define a distance-like function on $\mathcal{S}^n \times \mathcal{S}^n$ for $\alpha \neq \pm 1$ by

$$D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = \frac{4}{1-\alpha^2}\left\{1 - \sum_{i=1}^{n+1}(p_i)^{(1-\alpha)/2}(r_i)^{(1+\alpha)/2}\right\},$$

which we call the *α-divergence*. The Fisher metric g and the α-connection $\nabla^{(\alpha)}$ can be derived from the α-divergence.[7,21]

Since $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ geometrically play dualistic roles[6,7] with respect to g, we consider the triple $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$, which is called the α-geometric structure on $\mathcal{S}^n$. The properties of the Tsallis entropy are studied through the α-geometry.[4,5]

While the α-geometric structure for $\alpha \neq \pm 1$ is not flat, we reported[8] that it can be flattened via a certain conformal transformation[22–25] to a nonstandard dually flat structure[6,7] denoted by (h, ∇, ∇^*). The theoretical advantage or interesting aspect of such a conformally flattening is that we can obtain the Legendre structure on $\mathcal{S}^n$ preserving several properties of the α-geometric structure. We summarize the result in the following proposition by preparing some notation: the *escort probability*[1] P_i and a function Z_q are respectively defined for $q \in \mathbf{R}$ by

$$P_i(\boldsymbol{p}) := \frac{(p_i)^q}{\sum_{j=1}^{n+1}(p_j)^q}, \quad i = 1, \ldots, n+1, \quad Z_q(\boldsymbol{p}) := \sum_{i=1}^{n+1}\frac{(p_i)^q}{q}\,. \tag{5}$$

For $0 < q$ with $q \neq 1$, we define two functions by

$$\ln_q(s) := \frac{s^{1-q}-1}{1-q}, \quad s \geq 0, \quad \exp_q(t) := [1 + (1-q)t]_+^{1/(1-q)}, \quad t \in \mathbf{R}\,,$$

where $[t]_+ := \max\{0, t\}$, and the so-called Tsallis entropy[26] by

$$S_q(\boldsymbol{p}) := \frac{\sum_{i=1}^{n+1}(p_i)^q - 1}{1-q}\,.$$

Note that $s = \exp_q(\ln_q(s))$ holds and they respectively recover the usual logarithmic, exponential function and the Boltzmann–Gibbs–Shannon entropy $-\sum_{i=1}^{n+1} p_i \ln p_i$ when $q \to 1$. For $q > 0$, $\ln_q(s)$ is concave on $s > 0$.

Proposition 1. *The dually flat structure (h, ∇, ∇^*) on $\mathcal{S}^n$ is induced via a conformal transformation from the α-structure $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ on $\mathcal{S}^n$. The induced potential functions ψ, ψ^*, and dually flat affine coordinate systems $(\theta^1, \dots, \theta^n)$ and $(\eta_1, \dots, \eta_n)$ are represented as follows*:

$$\begin{aligned} \theta^i(\boldsymbol{p}) &= \ln_q(p_i) - \ln_q(p_{n+1}), \quad i = 1, \dots, n\,, \\ \eta_i(\boldsymbol{p}) &= P_i(\boldsymbol{p}), \quad i = 1, \dots, n\,, \\ \psi(\boldsymbol{\theta}(\boldsymbol{p})) &= -\ln_q(p_{n+1})\,, \\ \psi^*(\boldsymbol{\eta}(\boldsymbol{p})) &= \frac{1}{\kappa}\left(\lambda(\boldsymbol{p}) - q\right)\,, \end{aligned}$$

where $\kappa = (1-\alpha^2)/4 = q(1-q)$ is the scalar curvature of the α-structure, $\theta^{n+1} \equiv 0$, $\eta_{n+1} := P_{n+1}(\boldsymbol{p}) = 1 - \sum_{i=1}^n P_i(\boldsymbol{p})$ and $\lambda = 1/Z_q$ is a conformal factor, i.e. $h = \lambda g$.

Further, the coordinate systems $(\theta^1, \dots, \theta^n)$ and $(\eta_1, \dots, \eta_n)$ are ∇- and ∇^*-affine, respectively.

For the proofs of Proposition 1 and necessary lemmas, see Ref. 27. The result is extended to the q-exponential family with continuous random variables.[9,10]

Note that by defining what we call the *conformal divergence* ρ,

$$\begin{aligned} \rho(\boldsymbol{p}, \boldsymbol{r}) := \lambda(\boldsymbol{r}) D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) &= \sum_{i=1}^{n+1} -P_i(\boldsymbol{r})\left(\ln_q(p_i) - \ln_q(r_i)\right) \\ &= \psi(\boldsymbol{\theta}(\boldsymbol{p})) + \psi^*(\boldsymbol{\eta}(\boldsymbol{r})) - \sum_{i=1}^{n} \theta^i(\boldsymbol{p})\eta_i(\boldsymbol{r}), \end{aligned} \tag{6}$$

we can confirm the Legendre structure, i.e. relations $\rho(\boldsymbol{p}, \boldsymbol{p}) = 0,\ \forall\, \boldsymbol{p} \in \mathcal{S}^n$ and

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \psi^*}{\partial \eta_i}, \quad i = 1, \dots, n\,. \tag{7}$$

The dual potential ψ^* can be alternatively represented[8] in $\boldsymbol{p}$ by

$$\psi^* = \ln_q\left(\frac{1}{\exp_q(S_q(\boldsymbol{p}))}\right),$$

which is known as the negative of the *normalized Tsallis entropy*.[28–30] Thus, when $q \to 1$, we have the standard dually flat structure on $\mathcal{S}^n$ as follows:

$$\psi \to -\ln p_{n+1}, \quad \psi^* \to \sum_{i=1}^{n+1} p_i \ln p_i \quad \theta^i \to \ln(p_i/p_{n+1}), \quad \eta_i \to p_i, \quad i = 1, \dots, n\,.$$

Finally, it should be remarked that the both structures (h, ∇, ∇^*) and $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ are related in terms of not only the conformality of the metrics $h = \lambda g$ but also the *projective equivalence*[31] between the connections ∇^* and $\nabla^{(-\alpha)}$,[a] which implies that a curve on $\mathcal{S}^n$ is ∇^*-geodesic if and only if it is $\nabla^{(-\alpha)}$-geodesic.[b] More generally, a submanifold in $\mathcal{S}^n$ is ∇^*-autoparallel if and only if it is $\nabla^{(-\alpha)}$-autoparallel. For (h, ∇, ∇^*), in particular, a submanifold is ∇- (resp. ∇^*-) autoparallel when the affine coordinates θ^i (resp. η_i) are affinely parametrized by $\beta^j, j = 1, \dots, m \leq n$ as $\theta^i = \sum_{j=1}^m a^i_j \beta^j + c^i$, for $i = 1, \dots, n+1$ (similarly for η_i). For example, the q-exponential family

$$p_i = \exp_q\{\theta^i - \tilde{\psi}(\boldsymbol{\beta})\}, \quad i = 1, \dots, n+1\,, \tag{8}$$

where $\tilde{\psi}$ is a normalizing term defined by $\tilde{\psi} = \theta^{n+1} + \psi$, is ∇-autoparallel in a proper domain of $\boldsymbol{\beta}$. These properties are crucially used in the following sections.

Proposition 1 with (7) implies that

$$P_i = \frac{\partial \psi}{\partial \theta^i}, \quad i = 1, \dots, n \tag{9}$$

for $p_i = \exp_q(\theta^i - \psi)$, $i = 1, \dots, n$ and $p_{n+1} = \exp_q(-\psi)$. This relation can be regarded as a special case of a known one[3,32] for the q-exponential family (8), using the *escort expectation*,[2]

$$\langle\langle a_j \rangle\rangle_q := \sum_{i=1}^{n+1} P_i a^i_j = \frac{1}{qZ_q} \sum_{i=1}^{n+1} (p_i)^q \frac{\partial}{\partial \beta^j} (\ln_q(p_i) + \tilde{\psi} - c^i) = \frac{\partial \tilde{\psi}}{\partial \beta^j}\,,$$

because (9) is derived when $a^i_j = \delta^i_j, j = 1, \dots, n$ and $a^i_{n+1} = c^i = 0$.

3. Applications to Construction of Alpha-Voronoi Partitions and Alpha-Centroids

For given m points $\boldsymbol{p}_1, \dots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ we define α-*Voronoi regions* on $\mathcal{S}^n$ using the α-divergence as follows:

$$\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}) < D^{(\alpha)}(\boldsymbol{p}_l, \boldsymbol{p})\}, \quad k = 1, \dots, m\,.$$

An α-*Voronoi partition* (*diagram*) on $\mathcal{S}^n$ is a collection of the α-Voronoi regions and their boundaries. Note that $D^{(\alpha)}$ approaches the Kullback–Leibler (KL) divergence if $\alpha \to -1$, and $D^{(0)}$ is called the Hellinger distance. If we use the *Rényi divergence*[33] *of order* $\alpha \neq 1$ defined by

$$D_\alpha(\boldsymbol{p}, \boldsymbol{r}) := \frac{1}{\alpha - 1} \ln \sum_{i=1}^{n+1} (p_i)^\alpha (r_i)^{1-\alpha}$$

[a]Note that ∇^* is projectively equivalent with $\nabla^{(\alpha)}$ in Ref. 8 because there we adopted a different correspondence of parameters: $q = (1 - \alpha)/2$.

[b]Precisely speaking, the term "geodesic" should be replaced by "pre-geodesic".

instead of the α-divergence, $\mathrm{Vor}^{(1-2\alpha)}(\boldsymbol{p}_k)$ gives the corresponding Voronoi region because of their one-to-one functional relationship.

The standard algorithm using projection of a polyhedron[18,19] commonly works well to construct Voronoi partitions for the Euclidean distance,[19] the KL divergence.[12] The algorithm is generally applicable if a divergence function is of *Bregman type*,[13] which is represented by the remainder of the first order Taylor expansion of a convex potential function in a suitable coordinate system. Geometrically speaking, this implies that i) the divergence is of the form (6) in a dually flat structure and ii) its affine coordinate system is chosen to realize the corresponding Voronoi partitions. In this coordinate system with one extra complementary coordinate the polyhedron is expressed as the upper envelop of m hyperplanes tangent to the potential function.

A problem for the case of the α-Voronoi partition is that the α-divergence on $\mathcal{S}^n$ *cannot* be represented as a remainder of any convex potentials. The following theorem, however, claims that the problem is resolved by Proposition 1, i.e. conformally transforming the α-geometry to the dually flat structure (h, ∇, ∇^*) and using the conformal divergence ρ and escort probabilities as a coordinate system.

Here, we denote the space of escort distributions by $\mathcal{E}^n$ and represent the point on $\mathcal{E}^n$ by $\boldsymbol{P} = (P_1, \dots, P_n)$ because $P_{n+1} = 1 - \sum_{i=1}^n P_i$.

Theorem 1.

(i) *The bisector of* $\boldsymbol{p}_k$ *and* $\boldsymbol{p}_l$ *defined by* $\{\boldsymbol{p} | D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}) = D^{(\alpha)}(\boldsymbol{p}_l, \boldsymbol{p})\}$ *is a simultaneously* $\nabla^{(-\alpha)}$- *and* ∇^*-*autoparallel hypersurface on* $\mathcal{S}^n$.

(ii) *Let* $\mathcal{H}_k, k = 1, \dots, m$ *be the hyperplane in* $\mathcal{E}^n \times \mathbf{R}$ *which is respectively tangent at* $(\boldsymbol{P}_k, \psi^*(\boldsymbol{P}_k))$ *to the hypersurface* $\{(\boldsymbol{P}, y) | y = \psi^*(\boldsymbol{P})\}$, *where* $\boldsymbol{P}_k = \boldsymbol{P}(\boldsymbol{p}_k)$. *The* α-*Voronoi diagram can be constructed on* $\mathcal{E}^n$ *as the projection of the upper envelope of* $\mathcal{H}_k$*'s along the* y-*axis.*

Proof. (i) Consider the $\nabla^{(\alpha)}$-geodesic $\gamma^{(\alpha)}$ connecting $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$, and let $\bar{\boldsymbol{p}}$ be the midpoint on $\gamma^{(\alpha)}$ satisfying $D^{(\alpha)}(\boldsymbol{p}_k, \bar{\boldsymbol{p}}) = D^{(\alpha)}(\boldsymbol{p}_l, \bar{\boldsymbol{p}})$. Denote by $\mathcal{B}$ the $\nabla^{(-\alpha)}$-autoparallel hypersurface that is orthogonal to $\gamma^{(\alpha)}$ and passes $\bar{\boldsymbol{p}}$. Then, for all $\boldsymbol{r} \in \mathcal{B}$, the modified Pythagorean theorem[4,23] implies the following equality:

$$\begin{aligned} D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{r}) &= D^{(\alpha)}(\boldsymbol{p}_k, \bar{\boldsymbol{p}}) + D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{r}) - \kappa D^{(\alpha)}(\boldsymbol{p}_k, \bar{\boldsymbol{p}}) D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{r}) \\ &= D^{(\alpha)}(\boldsymbol{p}_l, \bar{\boldsymbol{p}}) + D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{r}) - \kappa D^{(\alpha)}(\boldsymbol{p}_l, \bar{\boldsymbol{p}}) D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{r}) = D^{(\alpha)}(\boldsymbol{p}_l, \boldsymbol{r}) . \end{aligned}$$

Hence, $\mathcal{B}$ is a bisector of $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$. The projective equivalence ensures that $\mathcal{B}$ is also ∇^*-autoparallel.

(ii) Recall the conformal relation (6) between $D^{(\alpha)}$ and ρ, then we see that $\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k) = \mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k)$ holds on $\mathcal{S}^n$, where

$$\mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | \rho(\boldsymbol{p}_k, \boldsymbol{p}) < \rho(\boldsymbol{p}_l, \boldsymbol{p})\} .$$

Proposition 1 and the Legendre relations (6) and (7) imply that $\rho(\boldsymbol{p}_k, \boldsymbol{p})$ is represented with the coordinates (P_i) by

$$\rho(\boldsymbol{p}_k, \boldsymbol{p}) = \psi^*(\boldsymbol{P}) - \left(\psi^*(\boldsymbol{P}_k) + \sum_{i=1}^{n} \frac{\partial \psi^*}{\partial P_i}(\boldsymbol{P}_k)(P_i(\boldsymbol{p}) - P_i(\boldsymbol{p}_k))\right),$$

where $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{p})$. Note that a point $(\boldsymbol{P}, y_k(\boldsymbol{P}))$ in $\mathcal{H}_k$ is expressed by

$$y_k(\boldsymbol{P}) := \psi^*(\boldsymbol{P}_k) + \sum_{i=1}^{n} \frac{\partial \psi^*}{\partial P_i}(\boldsymbol{P}_k)(P_i(\boldsymbol{p}) - P_i(\boldsymbol{p}_k)).$$

Hence, we have $\rho(\boldsymbol{p}_k, \boldsymbol{p}) = \psi^*(\boldsymbol{P}) - y_k(\boldsymbol{P})$. We see, for example, that the bisector on $\mathcal{E}^n$ for $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$ is represented as a projection of $\mathcal{H}_k \cap \mathcal{H}_l$. Thus, the statement follows. □

Figures 1 and 2 taken from Ref. 27 show examples of α-Voronoi partitions for four common probability distributions on $\mathcal{S}^2$: (0.2, 0.7, 0.1), (0.3, 0.3, 0.4), (0.4, 0.4, 0.2), (0.6, 0.1, 0.3) with $\alpha = -0.6$ and 2. While the left ones are represented with usual probabilities on $\mathcal{S}^2$ (the axis p_3 is omitted), right ones are the corresponding partitions represented with escort probabilities on $\mathcal{E}^2$. In right ones of the both figures, the bisectors are straight line segments on $\mathcal{E}^2$ because they are simultaneously $\nabla^{(-\alpha)}$- and ∇^*-geodesics as is proved in (i) of Theorem 1.

Remark 1. Voronoi partitions for broader class of divergences that are not necessarily associated with any convex potentials are theoretically studied[34] from more general affine differential geometric points of views.

On the other hand, the α-divergence can be expressed as a Bregman divergence if the domain is extended from $\mathcal{S}^n$ to the positive orthant $\mathbf{R}_+^{n+1}$.[5–7] Hence, the α-geometry on $\mathbf{R}_+^{n+1}$ is dually flat. Using this property, α-Voronoi partitions on $\mathbf{R}_+^{n+1}$ is discussed by Nielsen and Nock.[35]

However, while both of the above mentioned methods require constructions of the polyhedrons in the space of dimension $d = n+2$, the new one proposed in this paper does in the space of dimension $d = n+1$. Since it is known[36] that the optimal computational time of polyhedrons depends on the dimension d by $O(m \log m + m^{\lfloor d/2 \rfloor})$, the new one is better when n is even and m is large.

The next proposition is a simple and relevant application of escort probabilities. Define the *α-centroid* $\boldsymbol{c}^{(\alpha)}$ for given m points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ by the minimizer of the following problem:

$$\min_{\boldsymbol{p} \in \mathcal{S}^n} \sum_{k=1}^{m} D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_k).$$

Proposition 2. *The α-centroid $\boldsymbol{c}^{(\alpha)}$ for given m points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ is represented in escort probabilities by the weighted average of conformal factors*

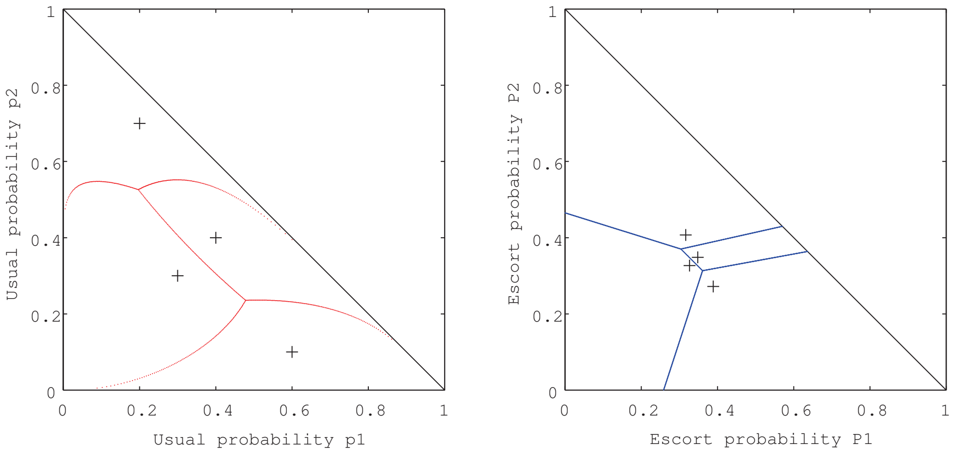

Fig. 1. An example of α-Voronoi partition on $\mathcal{S}^2$ (left) for $\alpha = -0.6$ (or $q = 0.2$) and the corresponding one on $\mathcal{E}^2$ (right).

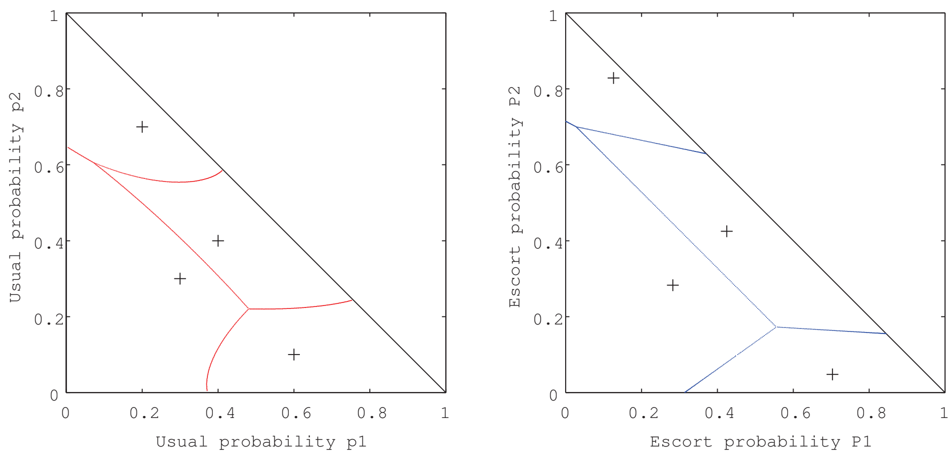

Fig. 2. An example of α-Voronoi partition on $\mathcal{S}^2$ (left) for $\alpha = 2$ (or $q = 1.5$) and the corresponding one on $\mathcal{E}^2$ (right).

$\lambda(\boldsymbol{p}_k) = 1/Z_q(\boldsymbol{p}_k)$, *i.e.*

$$P_i(\boldsymbol{c}^{(\alpha)}) = \frac{1}{\sum_{k=1}^m Z_q(\boldsymbol{p}_k)} \sum_{k=1}^m Z_q(\boldsymbol{p}_k) P_i(\boldsymbol{p}_k), \quad i = 1, \ldots, n+1 .$$

Proof. Let $\theta^i = \theta^i(\boldsymbol{p})$. Using (6), we have

$$\sum_{k=1}^m D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_k) = \sum_{k=1}^m Z_q(\boldsymbol{p}_k)\rho(\boldsymbol{p}, \boldsymbol{p}_k) = \sum_{k=1}^m Z_q(\boldsymbol{p}_k)\left\{\psi(\boldsymbol{\theta}) + \psi^*(\boldsymbol{\eta}(\boldsymbol{p}_k)) - \sum_{i=1}^n \theta^i \eta_i(\boldsymbol{p}_k)\right\}.$$

Then the optimality condition is

$$\frac{\partial}{\partial\theta^i}\sum_{k=1}^{m} D^{(\alpha)}(\boldsymbol{p},\boldsymbol{p}_k) = \sum_{k=1}^{m} Z_q(\boldsymbol{p}_k)(\eta_i - \eta_i(\boldsymbol{p}_k)) = 0, \quad i = 1,\ldots,n\,,$$

where $\eta_i = \eta_i(\boldsymbol{p})$. Thus, the statements for $i = 1,\ldots,n$ follow from Proposition 1. For $i = n+1$, it holds since the sum of the weights is equal to one. □

4. Related Dynamical Systems on the Simplex

In this section, we study properties of several dynamical systems naturally associated with the escort transformation, the conformal flattening and the resultant geometric structure.

4.1. *Conformal replicator equation*

Recall the *replicator system* on the simplex $\mathcal{S}^n$ for given functions $f_i(\boldsymbol{p})$ defined by

$$\dot{p}_i = p_i(f_i(\boldsymbol{p}) - \bar{f}(\boldsymbol{p})), \quad i = 1,\ldots,n+1, \quad \bar{f}(\boldsymbol{p}) := \sum_{i=1}^{n+1} p_i f_i(\boldsymbol{p})\,, \tag{10}$$

which is extensively studied in evolutionary game theory. It is known[37] that

(i) the solution of (10) is the gradient flow of a function $V(\boldsymbol{p})$ satisfying

$$f_i = \frac{\partial V}{\partial p_i}, \quad i = 1,\ldots,n+1\,,$$

with respect to the *Shahshahani metric*,[38]

(ii) the KL divergence is a local Lyapunov function for an equilibrium called the *evolutionary stable state* (*ESS*).

The Shahshahani metric is defined on the positive orthant $\mathbf{R}_+^{n+1}$ by

$$\tilde{g}_{ij} = \frac{\sum_{k=1}^{n+1} p_k}{p_i}\delta_{ij}, \quad i,j = 1,\ldots,n+1\,.$$

Note that a vector $X = \sum_{i=1}^{n} X^i\partial_i$ tangent to $\mathcal{S}^n$ is represented by a tangent vector $\tilde{X}$ on $\mathbf{R}_+^{n+1}$ by $\tilde{X} = \sum_{k=1}^{n+1}\tilde{X}^k\partial/\partial p_k$, where $\tilde{X}^i = X^i$, $i = 1,\ldots,n$ and $\tilde{X}^{n+1} = -\sum_{i=1}^{n} X^i$. Then we see that the Shahshahani metric induces the Fisher metric g in (3) on $\mathcal{S}^n$ because $\sum_{i,j}^{n} g_{ij}X^iX^j = \sum_{k,l}^{n+1}\tilde{g}_{kl}\tilde{X}^k\tilde{X}^l$ holds. Further, the KL divergence is a canonical divergence[7] of $(g,\nabla^{(1)},\nabla^{(-1)})$. Thus, the replicator dynamics (10) are closely related with the standard dually flat structure $(g,\nabla^{(1)},\nabla^{(-1)})$, which associates with exponential and mixture families of probability distributions.[39]

In this subsection, motivated by the above two features (i) and (ii), we define a modified replicator system compatible to the dually flat structure (h,∇,∇^*) and discuss their properties. See Harper[40] for another modification of the replicator system.

Consider a metric on $\mathbf{R}_+^{n+1}$ defined by $\tilde{h} := \lambda \tilde{g}$ and the following modified replicator system:

$$\dot{p}_i = Z_q(\boldsymbol{p}) p_i (f_i(\boldsymbol{p}) - \bar{f}(\boldsymbol{p})), \quad i = 1, \dots, n+1 \,. \tag{11}$$

It is easy to see the above right-hand sides define the vector that is tangent to $\mathcal{S}^n$ and the gradient of a function V with respect to $\tilde{h}$, since $\sum_{i=1}^{n+1} \dot{p}_i = 0$ and

$$\tilde{h}(\tilde{X}, \dot{\boldsymbol{p}}) = \sum_{i,j=1}^{n+1} \tilde{h}_{ij} \tilde{X}^i \dot{p}_j = \sum_{i=1}^{n+1} f_i \tilde{X}^i - \bar{f} \sum_{i=1}^{n+1} \tilde{X}^i = \sum_{i=1}^{n+1} \frac{\partial V}{\partial p_i} \tilde{X}^i \,,$$

respectively, hold for any tangent vector $\tilde{X}$ on $\mathcal{S}^n$. Thus, comparing (10) and (11), we can conclude as follows:

Proposition 3. *The gradient flow of a function V on $\mathcal{S}^n$ with respect to the conformal metric h is given by* (11). *Its trajectories coincide with those of* (10) *while velocities of time-evolutions are different by the factor $Z_q(\boldsymbol{p})$.*

We investigate properties of (11) in the case that $V(\boldsymbol{p}) = -\rho(\boldsymbol{r}, \boldsymbol{p})$ for a fixed distribution $\boldsymbol{r}$. Applying the result for gradient flows of divergences on dually flat spaces,[20] we see that the flow is explicitly given in the ∇-affine coordinates by

$$\theta^i(\boldsymbol{p}(t)) = \exp(-t)\{\theta^i(\boldsymbol{p}(0)) - \theta^i(\boldsymbol{r})\} + \theta^i(\boldsymbol{r}), \quad i = 1, \dots, n \,, \tag{12}$$

i.e. it converges to $\boldsymbol{r}$ along the ∇-geodesic (pregeodesic) curve.

On the other hand, consider the optimization problem maximizing $V(\boldsymbol{p}) = -\rho(\boldsymbol{r}, \boldsymbol{p})$ with m constraints of the escort expectations:

$$\begin{aligned} \langle\langle A_j \rangle\rangle_q &= \sum_{i=1}^{n+1} P_i(\boldsymbol{p}) A_j^i \\ &= \sum_{i=1}^{n} \eta_i(\boldsymbol{p}) A_j^i + \left(1 - \sum_{i=1}^{n} \eta_i(\boldsymbol{p})\right) A_j^{n+1} = \bar{A}_j, \quad j = 1, \dots, m \,, \end{aligned} \tag{13}$$

where A_j^i and $\bar{A}_j$ are prescribed values. Since the constraints (13) form a ∇^*-autoparallel submanifold in $\mathcal{S}^n$, the problem has the *unique* maximizer owing to the Pythagorean theorem[6,7] in a dually flat space. Defining the Lagrangian

$$L(\boldsymbol{p}) := \rho(\boldsymbol{r}, \boldsymbol{p}) + \sum_{j=1}^{m} \beta^j (\bar{A}_j - \langle\langle A_j \rangle\rangle_q) \,,$$

we have the following optimality condition from (6) and (7):

$$\begin{aligned} \frac{\partial L}{\partial \eta_i} &= \theta^i - \theta^i_{\boldsymbol{r}} - \sum_{j=1}^{m} \beta^j (A_j^i - A_j^{n+1}) \\ &= \ln_q p_i + \psi(\boldsymbol{\theta}) - \theta^i_{\boldsymbol{r}} - \sum_{j=1}^{m} \beta^j (A_j^i - A_j^{n+1}) = 0, \quad i = 1, \dots, n \,, \end{aligned}$$

where θ^i and η_i are, respectively, the ∇- and the ∇^*-affine coordinates of $\boldsymbol{p}$ introduced in Theorem 1, and $\theta^i_{\boldsymbol{r}} := \theta^i(\boldsymbol{r})$. Hence, θ^i is affine with respect to β^j and the maximizer $\boldsymbol{p}$ is in the q-exponential family represented in (8). These facts imply that the set of maximizers forms a ∇-autoparallel submanifold parametrized by β^j, which are determined by the prescribed values $\bar{A}_j$.

Combining this consideration with (12), we see that the following holds:

Corollary 1. *Let $\boldsymbol{r}$ be any distribution, and suppose that $\boldsymbol{p}_0$ and $\boldsymbol{p}_\infty$ are in the q-exponential family* (8) *parametrized by β^j as $\theta^i = \sum_{j=1}^m (A^i_j - A^{n+1}_j)\beta^j + \theta^i_{\boldsymbol{r}}, i = 1,\ldots,n$ and $\theta^{n+1} \equiv 0$. The gradient flow* (11) *with $V(\boldsymbol{p}) = -\rho(\boldsymbol{p}_\infty, \boldsymbol{p})$ starting from $\boldsymbol{p}_0$ converges to $\boldsymbol{p}_\infty$ staying on the q-exponential family.*

In the above, $\boldsymbol{p}_0$ and $\boldsymbol{p}_\infty$ are respectively interpreted as maximizers of $-\rho(\boldsymbol{r}, \boldsymbol{p})$ under the constraints (13) with different values of $\bar{A}_j$'s. The corollary claims that the q-exponential family is an invariant manifold for the transition of distribution from $\boldsymbol{p}_0$ to $\boldsymbol{p}_\infty$ caused by the change of $\bar{A}_j$'s, if the transition dynamics are governed by the gradient flow.

4.2. *Flows of escort transformation*

Consider a dynamical system induced by the escort transformation from $\boldsymbol{p}$ to $\boldsymbol{P}$ defined by (5). When we identify the set of escort distributions $\mathcal{E}^n$ with $\mathcal{S}^n$, the transformation is regarded to define a flow $\boldsymbol{P}^{(t)}$ on $\mathcal{S}^n$ parametrized by $t \in \mathbf{R}$:

$$P_i^{(t)} = \frac{(p_i)^t}{\sum_{j=1}^{n+1}(p_j)^t}, \quad i = 1,\ldots,n+1, \quad \boldsymbol{P}^{(1)} = \boldsymbol{p} \in \mathcal{S}^n\,, \tag{14}$$

where $\boldsymbol{p}$ is a *fixed* probability distribution.

Recalling the standard dually flat structure, which is obtained by limiting $q \to 1$ (or $\alpha \to 1$) in Proposition 1, we have the corresponding coordinates[c] $\theta^i_{\boldsymbol{p}} := \theta^i(\boldsymbol{p}) = \ln(p_i) - \ln(p_{n+1}), i = 1,\ldots,n$. In this case, if a curve $(\theta^i(t))$ on $\mathcal{S}^n$ is affinely parametrized by $t \in \mathbf{R}$, we call it *e-geodesic.*[7]

Since it follows that

$$\theta^i(t) := \theta^i(\boldsymbol{P}^{(t)}) = \ln P_i^{(t)} - \ln P_{n+1}^{(t)} = t(\ln p_i - \ln p_{n+1}) = t\theta^i_{\boldsymbol{p}}, \quad i = 1,\ldots,n\,,$$

we conclude from a viewpoint of information geometry that the flow of the escort transformation (14) evolves along the e-geodesic curve that passes $\boldsymbol{p}$ at $t = 1$.

Note that the arbitrary flows (14) converge to the uniform distribution independently of $\boldsymbol{p}$, when $t \to 0$. On the other hand, when $t \to \pm\infty$, it converges to a distribution on the boundary of $\mathcal{S}^n$ depending on the maximum or minimum components of $\boldsymbol{p}$. See Ref. 41 as a relevant work. In several literature,[42,43] examples of physical models with a time-evolution of the power index of distribution functions are reported.

[c]These coordinates are called the *canonical parameters* in statistics literature.

The above result can be slightly generalized with a *projective transformation* $\Pi_{\boldsymbol{r}} : \mathcal{S}^n \to \mathcal{S}^n$ defined by

$$\boldsymbol{p} = (p_i) \mapsto \Pi_{\boldsymbol{r}}(\boldsymbol{p}) := \left(\frac{r_i p_i}{\sum_{i=1}^{n+1} r_i p_i} \right), \quad i = 1, \ldots, n+1 \,,$$

for a given vector $\boldsymbol{r} = (r_i) \in \mathbf{R}_+^{n+1}$, and the relation with the replicator equation is elucidated.

Proposition 4. *For arbitrary* $\boldsymbol{r}$ *the projective transformation of the escort flow given in* (14) *evolves along the e-geodesic curve that passes* $\tilde{\boldsymbol{r}} = \boldsymbol{r}/\|\boldsymbol{r}\|_1$ *at* $t = 0$ *and* $\Pi_{\boldsymbol{r}}(\boldsymbol{p})$ *at* $t = 1$. *This flow evolves along the trajectory of the replicator equation* (10) *with constants* $f_i = \ln(p_i), i = 1, \ldots, n+1$.

Proof. The first statement follows from direct calculation of coordinates θ^i for the standard dually flat structure when $q \to 1$ ($\alpha \to 1$):

$$\theta^i(\Pi_{\boldsymbol{r}}(\boldsymbol{P}^{(t)})) = \ln(r_i P_i^{(t)}) - \ln(r_{n+1} P_{n+1}^{(t)}) = t\theta^i_{\boldsymbol{p}} + \ln(r_i/r_{n+1}), \quad i = 1, \ldots, n \,.$$

To prove the second statement note that that the flow $\Pi_{\boldsymbol{r}}(\boldsymbol{P}^{(t)})$ is a normalization of a vector $\boldsymbol{y}(t)$, each component of which is $y_i(t) = r_i(p_i)^t$. Hence, $\boldsymbol{y}(t)$ satisfies the following linear differential equation:

$$\dot{y}_i = \ln(p_i) y_i, \quad y_i(0) = r_i, \quad i = 1, \ldots, n+1 \,.$$

By setting $x_i = y_i/\|\boldsymbol{y}\|_1$, we have

$$\frac{d}{dt}\ln(x_i) = \ln(p_i) - \frac{1}{\|\boldsymbol{y}\|_1}\sum_{j=1}^{n+1}\dot{y}_j = \ln(p_i) - \sum_{j=1}^{n+1} x_j \ln(p_j), \quad i = 1, \ldots, n+1 \,.$$

Thus, $\Pi_{\boldsymbol{r}}(\boldsymbol{P}^{(t)})$ is the solution of

$$\dot{x}_i = x_i \left(\ln(p_i) - \sum_{j=1}^{n+1} \ln(p_j) x_j \right), \quad x_i(0) = \frac{r_i}{\|\boldsymbol{r}\|_1}, \quad i = 1, \ldots, n+1 \,.$$

This proves the second statement. □

5. Concluding Remarks

We have discussed two applications of escort probabilities and the dually flat structure (h, ∇, ∇^*) on $\mathcal{S}^n$ induced by conformal transformations of the α-geometry. They are used to new directions except the studies of multifractal or nonextensive statistical physics.

We first demonstrate a direct application of the conformal flattening to computation of α-Voronoi partitions and α-centroids. Escort probabilities are found to work as a suitable coordinate system for the purpose. Further, conformal divergence and projective equivalence of affine connections also play important roles.

In behavioral analysis of dynamical systems we present the properties of gradient flows with respect to the conformal metric and discuss a relation with the replicator equation. Next, we show that the projective transformation of the escort flow is e-geodesic. This flow describes a time-evolution of the power index of distributions.

Physical interpretation of the obtained conformal structure is another future research direction.

Acknowledgments

We would like to thank Prof. Tatsuaki Wada for helpful comments. A. Ohara was partially supported by JSPS Grant-in-Aid (C) 23540134 and AOARD Grant No. 10-1-4060.

References

1. C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
2. C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer, New York, 2009).
3. J. Naudts, *Generalized Thermostatistics* (Springer, London, 2011).
4. A. Ohara, *Phys. Lett. A* **370** (2007) 184.
5. A. Ohara, *Eur. Phys. J. B* **70** (2009) 15.
6. S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Springer-Verlag, Berlin, 1985).
7. S.-I. Amari and H. Nagaoka, *Methods of Information Geometry* (AMS & Oxford, Rhode Island, 2000).
8. A. Ohara, H. Matsuzoe and S.-I. Amari, *J. Phys.: Conf. Ser.* **201** (2010) 012012.
9. S.-I. Amari and A. Ohara, *Entropy* **13** (2011) 1170.
10. H. Matsuzoe and A. Ohara, *Proc. 2nd Int. Colloquium on Differential Geometry and its Related Fields* (World Scientific, Singapore, 2011), p. 55.
11. K. Onishi and N. Takayama, *IEICE Trans. Fundamentals* **E79-A** (1996) 533.
12. K. Onishi and H. Imai, *Proc. 13th ACM-SIAM Symposium on Computational Geometry* (1997), p. 463.
13. J.-D. Boissonnat, F. Nielsen and N. Nock, *Discrete Comput. Geom.* **44** (2010) 281.
14. H. Imai and M. Inaba, *IEICE Trans. Inf. Syst.* **E83-D** (2000) 27.
15. G. Lebanon, *IEEE Trans. Pattern Anal. Mach. Intell.* **28** (2005) 497.
16. S.-I. Amari, *Neural Comput.* **19** (2007) 2780.
17. N. N. Čencov, *Statistical Decision Rules and Optimal Inference* (AMS, Rhode Island, 1982), [Originally published in Russian (Nauka, Moscow, 1972)].
18. H. Edelsbrunner and R. Seidel, *Discrete Comput. Geom.* **1** (1986) 25.
19. H. Edelsbrunner, *Algorithms in Combinatorial Geometry* (Springer-Verlag, 1987).
20. A. Fujiwara and S-I. Amari, *Physica D* **80** (1995) 317.
21. S. Eguchi, *Ann. Stat.* **11** (1983) 793.
22. T. Kurose, *Math. Z.* **203** (1990) 115.
23. T. Kurose, *Tôhoku Math. J.* **46** (1994) 427.
24. T. Kurose, *Interdiscip. Inform. Sci.* **8** (2002) 89.
25. H. Matsuzoe, *Hiroshima Math. J.* **29** (1999) 175.
26. C. Tsallis, *J. Stat. Phys.* **52** (1988) 479.
27. A. Ohara, H. Matsuzoe and S.-I. Amari, arXiv:1010.4965v1 (2010).

28. P. T. Landsberg and V. Vedral, *Phys. Lett. A* **247** (1998) 211.
29. A. K. Rajagopal and S. Abe, *Phys. Rev. Lett.* **83** (1999) 1711.
30. T. Wada and A. M. Scarfone, *Phys. Lett. A* **335** (2005) 351.
31. K. Nomizu and T. Sasaki, *Affine Differential Geometry* (Cambridge University Press, Cambridge, 1993).
32. J. Naudts, *J. Inequal Pure Appl. Math.* **5**(4) (2004) 102.
33. A. Rényi, *Proc. 4th Berkeley Symp. Math. Statist. Probability*, Vol. 1 (University of California Press, Berkeley, 1961), p. 547.
34. H. Matsuzoe, in *Emerging Trends in Visual Computing*, ed. F. Nielsen (Springer, Berlin, 2009), p. 103.
35. F. Nielsen and R. Nock, *Int. Symp. Voronoi Diagrams* ($ISVD$), DTU Lyngby, Denmark (IEEE Press, New York, 2009).
36. B. Chazelle, *Discrete Comput. Geom.* **10** (1993) 377.
37. J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection* (Cambridge University Press, 1988).
38. S. Shahshahani, *A New Mathematical Framework for the Study of Linkage and Selection*, Memoirs of the AMS, Vol. 17, No. 211 (AMS, 1979).
39. N. Ay and I. Erb, *J. Dyn. Differ. Equ.* **17** (2005) 427.
40. M. Harper, *Physica D* **240** (2011) 1411.
41. S. Abe, *Phys. Rev. E* **68** (2003) 031101.
42. A. Taruya and M. Sakagami, *Phys. Rev. Lett.* **90** (2003) 181101.
43. A. Campa, P. Chavanis, A. Giansanti and G. Morelli, *Phys. Rev. E* **78** (2008) 040102(R).