Doubly autoparallel structure
on the probability simplex

Atsumi Ohara\textsuperscript{1} and Hideyuki Ishi\textsuperscript{2,3}

\textsuperscript{1} University of Fukui, Fukui 910-8507, Japan
ohara@fuee.u-fukui.ac.jp
\textsuperscript{2} Nagoya University, Furo-cho, Nagoya 464-8602, Japan
hideyuki@math.nagoya-u.ac.jp
\textsuperscript{3} JST, PRESTO, 4-1-8, Honcho, Kawaguchi 332-0012, Japan

Abstract. On the probability simplex, we can consider the standard
information geometric structure with the e- and m-affine connec-
tions mutually dual with respect to the Fisher metric. The geometry naturally
defines submanifolds simultaneously autoparallel for the both affine con-
nections, which we call doubly autoparallel submanifolds.
In this note we discuss their several interesting common properties. Fur-
ther, we algebraically characterize doubly autoparallel submanifolds on
the probability simplex and give their classification.

Keywords: statistical manifold, dual affine connections, doubly autopar-
allel submanifolds, mutation of Hadamard product

1 Introduction

Let us consider information geometric structure $\Pi (g, \nabla, \nabla^*)$ on a manifold $\mathcal{M}$,
where $g, \nabla, \nabla^*$ are, respectively, a Riemannian metric and a pair of torsion-free
affine connections satisfying
$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z), \quad \forall X, Y, Z \in \mathcal{X}(\mathcal{M}).$$

Here, $\mathcal{X}(\mathcal{M})$ denotes a set of vector fields on $\mathcal{M}$. Such a manifold with the
structure $(g, \nabla, \nabla^*)$ is called a statistical manifold and we say $\nabla$ and $\nabla^*$ are
mutually dual with respect to $g$. When curvature tensors of $\nabla$ and $\nabla^*$ vanish,
the statistical manifold is said dually flat. For a statistical manifold, we can
introduce a one-parameter family of affine connections called $\alpha$-connection:
$$\nabla^{(\alpha)} = \frac{1 + \alpha}{2} \nabla + \frac{1 - \alpha}{2} \nabla^*, \quad \alpha \in \mathbb{R}.$$ 

It is seen that $\nabla^{(\alpha)}$ and $\nabla^{-\alpha}$ are mutually dual with respect to $g$.

In a statistical manifold, we can naturally define a submanifold $\mathcal{N}$ that is
simultaneously autoparallel with respect to both $\nabla$ and $\nabla^*$.

Definition 1. Let $(\mathcal{M}, g, \nabla, \nabla^*)$ be a statistical manifold and $\mathcal{N}$ be its subman-
ifold. We call $\mathcal{N}$ doubly autoparallel in $\mathcal{M}$ when the followings hold:
$$\nabla_X Y \in \mathcal{X}(\mathcal{N}), \quad \nabla_X^* Y \in \mathcal{X}(\mathcal{N}), \quad \forall X, Y \in \mathcal{X}(\mathcal{N}).$$
We immediately see that doubly autoparallel submanifolds \( \mathcal{N} \) possess the following properties: (Note that 4) and 5) hold if \( \mathcal{M} \) is dually flat.)

**Proposition 1.** The following statements are equivalent:

1) a submanifold \( \mathcal{N} \) is doubly autoparallel (DA),
2) a submanifold \( \mathcal{N} \) is autoparallel w.r.t. to \( \nabla^{(\alpha)} \) for two different \( \alpha \)'s,
3) a submanifold \( \mathcal{N} \) is autoparallel w.r.t. to \( \nabla^{(\alpha)} \) for all \( \alpha \)'s,
4) the \( \alpha \)-geodesics connecting two points on \( \mathcal{N} \) (if they exist) lay in \( \mathcal{N} \) for all \( \alpha \)'s,
5) a submanifold \( \mathcal{N} \) is affinely constrained in both \( \nabla \)- and \( \nabla^* \)-affine coordinates of \( \mathcal{M} \).

Furthermore, when \( \mathcal{M} \) is dually flat and \( \mathcal{N} \) is DA, the \( \alpha \)-projections to \( \mathcal{N} \) (if they exist) are unique for all \( \alpha \)'s.

The concept of doubly autoparallelism has sometimes appeared but played important roles in several applications of information geometry [6,7,8,9]. However, the literature mostly treat information geometry of positive definite matrices or symmetric cones, and the study for statistical models has not been exploited yet.

In this note, we consider doubly autoparallel structure on the probability simplex, which can be identified with probability distributions on discrete and finite sample spaces. As a result, we give an algebraic characterization and classification of doubly autoparallel submanifolds in the probability simplex.

Such manifolds commonly possess the above interesting properties. Hence, the obtained results might be expected to give a useful insight into constructing statistical models for wide area of applications in information science, mathematics and statistical physics and so on [2,3,4]. Further, it should be mentioned that Nagaoka has recently reported the significance of this concept in study of Markov equivalence for statistical models [5].

2 Preliminaries

2.1 Information geometry of \( \mathcal{S}^n \) and \( \mathbb{R}^{n+1}_+ \)

Let us represent an element \( p \in \mathbb{R}^{n+1}_+ \) with its components \( p_i, \ i = 1, \ldots, n+1 \) as \( p = (p_i) \in \mathbb{R}^{n+1}_+ \). Denote, respectively, the positive orthant by

\[
\mathbb{R}^{n+1}_+ := \{ p = (p_i) \in \mathbb{R}^{n+1}_+ | p_i > 0, \ i = 1, \ldots, n+1 \},
\]

and the relative interior of the probability simplex by

\[
\mathcal{S}^n := \left\{ p \in \mathbb{R}^{n+1}_+ \mid \sum_{i=1}^{n+1} p_i = 1 \right\}.
\]

For a subset \( \mathcal{Q} \subset \mathbb{R}^{n+1}_+ \) and an element \( p \in \mathcal{Q} \), we simply write

\[
\log \mathcal{Q} := \{ \log p | p \in \mathcal{Q} \}, \quad \log p := (\log p_i) \in \mathbb{R}^{n+1}_+.
\]
Each element $p$ in the closure of $S^n$ denoted by $\text{cl}S^n$ can be identified with a discrete probability distribution for the sample space $\Omega = \{1, 2, \cdots, n, n + 1\}$. However, we only consider distributions $p(X)$ with positive probabilities, i.e., $p(i) = p_i > 0$, $i = 1, \cdots, n + 1$, defined by

$$p(X) = \sum_{i=1}^{n+1} p_i \delta_i(X), \quad \delta_i(j) = \delta^i_j \text{ (the Kronecker’s delta)},$$

which is identified with $S^n$.

A statistical model in $S^n$ is represented with parameters $\xi = (\xi_j), j = 1, \cdots, d \leq n$ by

$$p(X; \xi) = \sum_{i=1}^{n+1} p_i(\xi) \delta_i(X),$$

where each $p_i$ is a function of $\xi$. For example, $p_i = \xi_i$, $i = 1, \cdots, n$ with the condition $\sum_{i=1}^{n} \xi_i < 1$ is the full model, i.e.,

$$p(X; \xi) = \sum_{i=1}^{n} \xi_i \delta_i(X) + \left(1 - \sum_{i=1}^{n} \xi_i\right) \delta_{n+1}(X)$$

For the submodel, $\xi^j$, $j = 1, \cdots, d < n$ can be also regarded as coordinates of the corresponding submanifold in $S^n$.

The standard information geometric structure on $S^n$ [1] denoted by $(g, \nabla^{(e)}, \nabla^{(m)})$ are composed of the pair of flat affine connections $\nabla^{(e)}$ and $\nabla^{(m)}$. The affine connections $\nabla^{(e)} = \nabla^{(1)}$ and $\nabla^{(m)} = \nabla^{(-1)}$ are respectively called the exponential connection and the mixture connection. They are mutually dual with respect to the Fisher metric $g$.

By writing $\partial_i := \partial/\partial \xi_i$, $i = 1, \cdots, n$, they are explicitly represented as follows:

$$g_{ij}(p) = \sum_{X \in \Omega} p(X)(\partial_i \log p(X))(\partial_j \log p(X)), \quad i, j = 1, \cdots, n,$$

$$I^{(m)}_{ij,k}(p) = \sum_{X \in \Omega} p(X)(\partial_i \partial_j p(X))(\partial_k \log p(X)) \quad i, j, k = 1, \cdots, n, \quad (1)$$

$$I^{(e)}_{ij,k}(p) = \sum_{X \in \Omega} p(X)(\partial_i \partial_j \log p(X))(\partial_k \log p(X)), \quad i, j, k = 1, \cdots, n. \quad (2)$$

There exist two special coordinate systems. The one is the expectation coordinate $\eta_i := \sum_{X \in \Omega} p(X) \delta_i(X) = p_i$, $i = 1, \cdots, n$, which is $\nabla^{(m)}$-affine from [1]. It implies that if each $\eta_i$ is an affine function of all the model parameters $\xi_i$’s, then the statistical model is $\nabla^{(m)}$-autoparallel (or sometimes called $m$-flat).

The other is the canonical coordinate $\theta^i$, which is defined by

$$\theta^i := \log \left( \frac{p_i}{1 - \sum_{i=1}^{n} p_i} \right), \quad i = 1, \cdots, n. \quad (3)$$
Thus, we find that \( p \) coordinates, respectively, from (4), (5) and 
\[
\log p = v
\]
For the simplicity we take the following 
\[
d \text{define a subspace of dimension } \eta \text{ the expectation coordinates}
\]
\[
\text{Example:}
\]
Take a vector \( v \) ambient structure. For arbitrary coordinates \( \tilde{\eta} \) take \( \tilde{\eta} \)
\[
\text{certain subspace} \ g, \text{The structure } (g, \nabla^{(e)}, \nabla^{(m)}) \text{ on } R^{n+1}_+	ext{.}
\]
The structure \( (g, \nabla^{(e)}, \nabla^{(m)}) \) on \( S^n \) is a submanifold geometry induced from this ambient structure. For arbitrary coordinates \( \xi, i = 1, \ldots, n+1 \text{ of } R^{n+1}_+ \), let us 
\[
\text{take } \tilde{\partial}_i := \partial/\partial \xi_i. \text{ Then their components are given by }
\]
\[
\tilde{g}_{ij}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \log p(X))(\tilde{\partial}_j \log p(X)), \quad i, j = 1, \ldots, n + 1,
\]
\[
\tilde{R}^{(m)}_{ij,k}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \tilde{\partial}_j p(X))(\tilde{\partial}_k \log p(X)), \quad i, j, k = 1, \ldots, n + 1, (4)
\]
\[
\tilde{R}^{(e)}_{ij,k}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \tilde{\partial}_j \log p(X))(\tilde{\partial}_k \log p(X)), \quad i, j, k = 1, \ldots, n + 1. (5)
\]
Thus, we find that \( p_i \)’s are \( \tilde{\nabla}^{(m)} \)-affine coordinates and \( \log p_i \)’s are \( \tilde{\nabla}^{(e)} \)-affine coordinates, respectively, from (4), (5) and 
\[
\text{log } p(X) = \sum_{X \in \Omega} (\log p_i) \delta_i(X).
\]

### 2.2 An example

**Example:** Let \( v^{(k)} = (\delta_i^k) \in R^{n+1}, k = 1, \ldots, n + 1 \) represent vertices on cl\( S^n \).
Take a vector \( v^{(0)} \in R^{n+1}_+ \) that is linearly independent of \( \{v^{(k)}\}_{k=1}^d \) \( d < n \) and 
define a subspace of dimension \( d + 1 \) by \( W = \text{span}\{v^{(0)}, v^{(1)}, \ldots, v^{(d)}\} \) Then 
\( M = S^n \cap W \) is doubly autoparallel.

We show this for the case \( d = 2 \) but similar arguments hold for general \( d \). For the simplicity we take the following \( v^{(i)}, i = 0, 1, 2 : \)
\[
\begin{align*}
v^{(0)} &= \begin{pmatrix} 0 & 0 & p_3 & \cdots & p_{n+1} \end{pmatrix}^T, & \sum_{i=3}^{n+1} p_i &= 1, & p_i > 0, & i = 3, \ldots, n + 1, \\
v^{(1)} &= \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}^T, & v^{(2)} &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^T, & (\cdot^T \text{ denotes the transpose}).
\end{align*}
\]
Since for \( p \in M \) we have a convex combination by parameters \( \xi_i \) as 
\[
p = \xi_1 v^{(1)} + \xi_2 v^{(2)} + (1 - \xi_1 - \xi_2) v^{(0)},
\]
the expectation coordinates \( \eta_i \)’s are 
\[
\eta_1 = \xi_1, \quad \eta_2 = \xi_2, \quad \eta_i = (1 - \xi_1 - \xi_2) p_i, \quad i = 3, \ldots, n + 1, \\
(\xi_1 > 0, \xi_2 > 0, \xi_1 + \xi_2 < 1).
\]
Thus, each \( \eta_i \) is affine in \( \xi_i \), \( i = 1, 2 \).

On the other hand, the canonical coordinates \( \theta^i \)'s are

\[
\begin{align*}
\theta^1 &= \zeta_1, \\
\theta^2 &= \zeta_2, \\
\theta^i &= \log p_i + c, \quad i = 3, \ldots, n + 1, \\
(\zeta_i &= \log \{\xi_i/(1 - \xi_1 - \xi_2)\}, \quad i = 1, 2, \quad c = -\log p_{n+1}).
\end{align*}
\]

Thus, each \( \theta^i \) is affine in parameters \( \zeta_i \), \( i = 1, 2 \). Hence, \( M \) is doubly autoparallel.

### 2.3 Denormalization

**Definition 2.** Let \( M \) be a submanifold in \( S^n \). The submanifold \( \tilde{M} \) in \( \mathbb{R}^{n+1} \) defined by

\[
\tilde{M} = \{ \tau p \in \mathbb{R}^{n+1} | p \in M, \tau > 0 \}
\]

is called a denormalization of \( M \). [1]

**Lemma 1.** A submanifold \( M \) is \( \nabla^{(\pm 1)} \)-autoparallel in \( S^n \) if and only if the denormalization \( \tilde{M} \) is \( \tilde{\nabla}^{(\pm 1)} \)-autoparallel in \( \mathbb{R}^{n+1} \).

A key observation derived from the above lemma is as follows:

Since \( p_i, \ i = 1, \ldots, n+1 \) are \( \nabla^{(m)} \)-affine coordinates for \( S^n \), a submanifold \( M \subset S^n \) is \( \nabla^{(m)} \)-autoparallel if and only if it is represented as \( M = W \cap S^n \) for a subspace \( W \subset \mathbb{R}^{n+1} \). Hence, by definition \( \tilde{M} \) is nothing but

\[
\tilde{M} = W \cap \mathbb{R}^{n+1}.
\]

On the other hand, since the coordinates \( \log p_i, \ i = 1, \ldots, n+1 \) for \( \mathbb{R}^{n+1} \) are \( \tilde{\nabla}^{(e)} \)-affine, \( \tilde{M} \) is \( \tilde{\nabla}^{(e)} \)-autoparallel if and only if there exist a subspace \( V \subset \mathbb{R}^{n+1} \) and a constant element \( b \in \mathbb{R}^{n+1} \) satisfying

\[
\log \tilde{M} = b + V,
\]

where \( \dim W = \dim V \). If so, \( M \) is also \( \nabla^{(e)} \)-autoparallel from lemma [1]

Thus, we study conditions for the denormalization \( \tilde{M} \) to have simultaneously dualistic representations [9] and [10], which is equivalent to doubly autoparallelism of \( M \).

### 3 Main results

First we introduce an algebra \( (\mathbb{R}^{n+1}, \circ) \) via the Hadamard product \( \circ \), i.e.,

\[
x \circ y = (x_1) \circ (y_1) := (x_1 y_1), \quad x, y \in \mathbb{R}^{n+1},
\]

where the identity element \( e \) and an inverse \( x^{-1} \) are

\[
e = 1, \quad x^{-1} = \left( \frac{1}{x_i} \right),
\]
respectively. Here, 1 ∈ \( \mathbb{R}^{n+1}_+ \) is the element all the components of which are one. Note that the set of invertible elements

\[ I := \{ x = (x_i) \in \mathbb{R}^{n+1}_+ | x_i \neq 0, \ i = 1, \cdots, n+1 \} \]

contains \( \mathbb{R}^{n+1}_+ \). We simply write \( x^k \) for the powers recursively defined by \( x^k = x \circ x^{k-1} \).

For an arbitrarily fixed \( a \in \mathcal{I} \) the algebra \( (\mathbb{R}^{n+1}_+, \circ) \) induces another algebra called a \textit{mutation} \( (\mathbb{R}^{n+1}_+, \circ_{a-1}) \), the product of which is defined by

\[ x \circ_{a-1} y := x \circ a^{-1} \circ y = (x_i y_i / a_i), \ x, y \in \mathbb{R}^{n+1}_+, \]

with its identity element \( a \). We write \( x^{(\circ a^{-1})k} \) for the powers by \( \circ a^{-1} \).

We give a basic result in terms of \( (\mathbb{R}^{n+1}_+, \circ) \).

**Theorem 1.** Assume that \( a \in \tilde{M} = W \cap \mathbb{R}^{n+1}_+ \). Then, there exists a subspace \( V \) satisfying

\[ \log \tilde{M} = \log \{ (a + W) \cap \mathbb{R}^{n+1}_+ \} = \log a + V \]

if and only if the following two conditions hold:

1) \( V = a^{-1} \circ W \), 2) \( \forall u, w \in W, u \circ a^{-1} \circ w \in W \).

**Proof.** ("only if" part): For all \( w \in W \) and small \( t \in \mathbb{R}_+ \), we have \( \log(a + tw) \in \log a + V \). Hence, it holds that

\[ \frac{d}{dt} \log(a + tw) \bigg|_{t=0} = a^{-1} \circ w \in V. \]

Thus, the condition 1) holds.

Similarly, for all \( u, w \in W \) and small \( t \in \mathbb{R}_+ \) and \( s \in \mathbb{R}_+ \), we have \( \log(a + su + tw) \in \log a + V \) and obtain

\[ \frac{\partial}{\partial s} \left( \frac{\partial}{\partial t} \log(a + su + tw) \bigg|_{t=0} \right) \bigg|_{s=0} = -a^{-1} \circ u \circ a^{-1} \circ w \in V. \]

Hence, we see that the condition 2) holds, using the condition 1).

("if" part): For \( w = (w_i) \in W \) satisfying \( a + w \in \mathbb{R}^{n+1}_+ \), take \( t \in \mathbb{R}_+ \) be larger than \( (1 + \max \{ |w_i| / a_i \} ) / 2 \). Then there exists \( u = (u_i) \in W \) satisfying

\[ a + w = ta + u, \ ta_i > |u_i|, \ i = 1, \cdots, n+1. \]

Hence, we have

\begin{align*}
\log(a + w) &= \log(ta + u) = \log\{(ta) \circ (e + (ta)^{-1} \circ u)\} \\
&= (\log t)e + \log a + \log\{e + (ta)^{-1} \circ u\}.
\end{align*}

Using the inequalities in (8) and the Taylor series

\[ \log(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k, \ |x| < 1, \]

\[ (e + (ta)^{-1} \circ u) \]

\[ (ta) \circ (e + (ta)^{-1} \circ u) \]

\[ \log(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k, \ |x| < 1, \]
we expand the right-hand side of (9) as
\[ (\log t)e + \log a + a^{-1} \circ \left( \frac{1}{l} \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{-1}{t} \right)^{k-1} u^{(a^{-1})k} \right). \]

Since each \( u^{(a^{-1})k} \) belongs to \( W \) from the condition 2), the third term is in \( V \) by the condition 1). Further the condition 1) implies that \( e \in V \), so is the first term. This completes the proof.

**Remark 1.**

i) The condition 2) claims that \( W \) is a subalgebra of \((\mathbb{R}^{n+1}, \circ_{a^{-1}})\).

ii) The affine subspace \( \log a + V \) is independent of the choice of \( a \in \tilde{M} = W \cap \mathbb{R}^{n+1} \). This follows from the proof of “if” part by taking \( a' = a + w \).

The following algebraic characterization of doubly autoparallel submanifold in \( S^n \) is immediate from the above theorem and lemma 1 in section 2.

**Corollary 1.** A \( \nabla^{(m)} \)-autoparallel submanifold \( M = W \cap S^n \) is doubly autoparallel if and only if the subspace \( W \) is a subalgebra of \((\mathbb{R}^{n+1}, \circ_{a^{-1}})\) with \( a \in \tilde{M} \).

Finally, in order to answer a natural question what structure is necessary and sufficient for \( W \), we classify subalgebras of \((\mathbb{R}^{n+1}, \circ_{a^{-1}})\). Let \( q \) and \( r \) be integers that meet \( q \geq 0, r > 0 \) and \( q + r = \dim W < n + 1 \). Define integers \( n_l, l = 1, \cdots, r \) satisfying
\[ q + \sum_{l=1}^{r} n_l = n + 1, \quad 2 \leq n_1 \leq \cdots \leq n_r. \]

Constructing subvectors \( a_l \in \mathbb{R}^{n_l}, l = 1, \cdots, r \) with components arbitrarily extracted from \( a \in W \cap \mathbb{R}^{n+1} \) without duplications, we denote by \( \Pi \) the permutation matrix that meets
\[ (a_0^T a_1^T \cdots a_r^T)^T = \Pi a, \quad (10) \]

where the subvector \( a_0 \in \mathbb{R}^q \) is composed of the remaining components in \( a \). We give the classification via the canonical form for \( W_0 = \Pi W = \{ w' \in \mathbb{R}^{n+1} | w' = \Pi w, w \in W \} \) based on this partition instead of the original form for \( W \).

**Theorem 2.** For the above setup, \( W \) is a subalgebra of \((\mathbb{R}^{n+1}, \circ_{a^{-1}})\) with \( a \in \tilde{M} \) if and only if \( W \) is isomorphic to \( \mathbb{R}^q \times \mathbb{R}^{a_1} \times \cdots \times \mathbb{R}^{a_r} \) and represented by \( \Pi^{-1} W_0 \), where
\[ W_0 = \{ (y^T t_1 a_1^T \cdots t_r a_r^T)^T \in \mathbb{R}^{n+1} | \forall y \in \mathbb{R}^q, a_l \in \mathbb{R}^{n_l}, \forall t_l \in \mathbb{R}, l = 1, \cdots, r \}. \]

**Proof.** ("only if" part) Let \( V \) be a subspace in \( \mathbb{R}^{n+1} \) defined by \( V = a^{-1} \circ W \). Then it is straightforward to check that \( W \) is a subalgebra of \((\mathbb{R}^{n+1}, \circ_{a^{-1}})\) if and only if \( V \) is a subalgebra of \((\mathbb{R}^{n+1}, \circ)\). Using this equivalence, we consider the necessity condition.
Since $e$ and $x^k$ are in $V$ for any $x = (x_i) \in V$ and positive integer $k$, the square matrix $\Xi$ defined by

$$\Xi := (e \ x \ \cdots \ x^n) = \begin{pmatrix} 1 & x_1 & \cdots & x_1^n \\ 1 & x_2 & \cdots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n+1} & \cdots & x_{n+1}^n \end{pmatrix}$$

is singular. The determinant of the Vandermonde's matrix $\Xi$ is calculated using the well-known formula, as

$$\det \Xi = (-1)^{(n+1)n/2} \prod_{i<j} (x_i - x_j).$$

Hence, it is necessary for $x$ to belong to $V$ that

$$\exists (i,j), \ x_i = x_j. \quad (11)$$

Denoting basis vectors of $V$ by $v^{(k)} = (v_i^{(k)}) \in \mathbb{R}^{n+1}$, $k = 1, \ldots, q+r (= \dim V)$, we can represent any $x$ as $x = \sum_{k=1}^{q+r} \alpha_k v^{(k)}$ using a coefficient vector $(\alpha_k) \in \mathbb{R}^{q+r}$. Hence, the necessary condition (11) is equivalent to

$$\forall (\alpha_k) \in \mathbb{R}^{q+r}, \ \exists (i,j), \ \sum_{k=1}^{q+r} \alpha_k (v_i^{(k)} - v_j^{(k)}) = 0. \quad (12)$$

It is easy to see, by contradiction, that (12) implies the following condition:

$$\exists (i,j), \ \forall k, \ v_i^{(k)} = v_j^{(k)}. \quad (13)$$

By normalization $v_i^{(k)} = v_j^{(k)} = 1$ for $(i, j)$ satisfying (13) and a proper permutation of $i = 1, \ldots, n+1$, we find that possible canonical form of subspace $V$, which we denote by $V_0$, is restricted to

$$V_0 = \{(z^T t_1 1^T \cdots t_r 1^T)^T \in \mathbb{R}^{n+1} | \forall z \in \mathbb{R}^q, \ t_l 1 \in \mathbb{R}^{n_l}, \ \forall t_l \in \mathbb{R}, \ l = 1, \ldots, r\}$$

for $q$, $r$ and $n_l$, $l = 1, \ldots, r$ given in the setup. Using the above permutation as $\Pi$ in the setup, i.e., $V = (\Pi^{-1} V_0)$, we have an isomorphic relation $W = \alpha \circ (\Pi^{-1} V_0)$. Thus, this means that $W_0 = \Pi W = (\Pi \alpha) \circ V_0$.

("if" part) Conversely it is easy to confirm $V_0$ is a subalgebra of $(\mathbb{R}^{n+1}, \circ)$. We show that any other proper subspaces in $V_0$ cannot be a subalgebra with $e$, except for the trivial cases where several $t_l$'s or components of $z = (z_i)$ are fixed to be zero, or equal to each other.

4 These cases contradict the fact that $e \in V_0$.

5 These cases correspond to choosing smaller $q$ or $r$ in the setup.
Consider a subspace \( V' \subset V_0 \) with nontrivial linear constraints between \( t_l \)'s and \( z_i \)'s. If \( V' \) is a subalgebra, then for all \( x \in V' \) and integer \( m \) we have
\[
V' \ni x^m = (z^m)^T t^m_1 T \cdots t^m_r T, \quad z^m = (z^m_i) \in \mathbb{R}^q,
\]
where \( t^m_l \)'s and \( z^m_i \)'s should satisfy the same linear constraints. We, however, find this is impossible by the similar arguments with the Vandermonde's matrix in the “only if” part. This completes the proof.

Example (continued from section 2.2): As \( a \in \tilde{M} = W \cap \mathbb{R}^{n+1} \) we set
\[
a = (1 ~ 2 ~ p_3 \cdots p_{n+1})^T, \quad a_0 = (1 ~ 2)^T, \quad a_1 = (p_3 \cdots p_{n+1})^T.
\]
Then we have \( q = 2, \ r = 1, \ n_1 = n - 1 \) and need no permutation, i.e., \( W = W_0 \). Since every element in \( W \) can be represented by
\[
w = (\xi_1 \xi_2 \ t p_3 \cdots t p_{n+1})^T, \quad \xi_1, \xi_2, \ t \in \mathbb{R}
\]
we can confirm \( W \) is a subalgebra of \( (\mathbb{R}^{n+1}, \circ_{a^{-1}}) \) and
\[
V_0 = V = a^{-1} \circ W = \{(z^T \ t 1^T)^T \in \mathbb{R}^{n+1} | \forall z \in \mathbb{R}^2, \ t 1 \in \mathbb{R}^{n-1}, \forall t \in \mathbb{R}\}.
\]

4 Concluding remarks

We have studied doubly autoparallel structure of statistical models in the family of probability distributions on discrete and finite sample space. Identifying it by the probability simplex and using the mutation of Hadamard product, we give an algebraic characterization of doubly autoparallel submanifolds and their classification.

Acknowledgements

A. O. was partially supported by JSPS Grant-in-Aid (C) 15K04997.

References

1. S-I. Amari and H. Nagaoka, Methods of Information Geometry, Trans. Math. Monogr. vol 191 (American Mathematical Society and Oxford University Press) (2000).
2. S. Ikeda, T. Tanaka and S-I. Amari, Stochastic reasoning, free energy, and information geometry, Neural Computation, 16, 1779-1810 (2004).
3. F. Matus and N. Ay, On maximization of the information divergence form an exponential family, Proc. WUPES’03, 199-204 (2004).
4. G. F. Montúfar, Mixture decompositions of exponential families using a decomposition of their sample spaces, Kybernetika, 49, 1, 23-39 (2013).
5. H. Nagaoka, Information-geometrical characterization of statistical models which are statistically equivalent to probability simplex, arXiv:1701.07736v2 (2017).
6. A. Ohara, Information Geometric Analysis of an Interior Point Method for Semidefinite Programming, Proc. of Geometry in Present Day Science (O. Barndorf-Nielsen and V. Jensen eds.) World Scientific, 49-74 (1999).
7. A. Ohara, Geodesics for Dual Connections and Means on Symmetric Cones, Integral Equations and Operator Theory, Vol.50, 537-548 (2004).
8. A. Ohara and T. Wada, Information Geometry of q-Gaussian Densities and Behaviours of Solutions to Related Diffusion Equations, Journal of Physics A: Mathematical and Theoretical, Vol.43, 035002 (18pp.) (2010).
9. K. Uohashi and A. Ohara, Jordan Algebras and Dual Affine Connections on Symmetric Cones, Positivity, Vol. 8, No. 4, 369-378 (2004).