Chapter 2
Finite Information Geometry

The considerations of this chapter are restricted to the situation of probability distributions on a finite number of symbols, and are hence of a more elementary nature. We pay particular attention to this case for two reasons. On the one hand, many applications of information geometry are based on statistical models associated with finite sets, and, on the other hand, the finite case will guide our intuition within the study of the infinite-dimensional setting. Some of the definitions in this chapter can and will be directly extended to more general settings.

2.1 Manifolds of Finite Measures

Basic Geometric Objects We consider a non-empty and finite set I.[1] The real algebra of functions $I \to \mathbb{R}$ is denoted by $\mathcal{F}(I)$, and its unity $\mathbb{1}_I$ or simply $\mathbb{1}$ is given by $\mathbb{1}(i) = 1$, $i \in I$. This vector spans the space $\mathbb{R} \cdot \mathbb{1} := \{c \cdot \mathbb{1} \in \mathcal{F}(I) : c \in \mathbb{R}\}$ of constant functions which we also abbreviate by $\mathbb{R}$. Given a function $g : \mathbb{R} \to \mathbb{R}$ and an $f \in \mathcal{F}(I)$, by $g(f)$ we denote the composition $i \mapsto g(f)(i) := g(f(i))$.

The space $\mathcal{F}(I)$ has the canonical basis e_i, $i \in I$, with

$$e_i(j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases}$$

and every function $f \in \mathcal{F}(I)$ can be written as

$$f = \sum_{i \in I} f^i \, e_i,$$

where the coordinates f^i are given by the corresponding values $f(i)$. We naturally interpret linear forms $\sigma : \mathcal{F}(I) \to \mathbb{R}$ as signed measures on I and denote the

[1]This set I will play the role of the no longer necessarily finite space Ω in Chap. 3.

N. Ay et al., *Information Geometry*, Ergebnisse der Mathematik und ihrer Grenzgebiete.
3. Folge / A Series of Modern Surveys in Mathematics 64,
DOI 10.1007/978-3-319-56478-4_2

corresponding dual space $\mathcal{F}(I)^*$, the space of $\mathbb{R}$-valued linear forms on $\mathcal{F}(I)$, by $\mathcal{S}(I)$. In a more general context, this interpretation is justified by the Riesz representation theorem. Here, it allows us to highlight a particular geometric perspective, which makes it easier to introduce natural information-geometric objects. Later in the book, we will treat general signed measures, and thereby have to carefully distinguish between various function spaces and their dual spaces.

The space $\mathcal{S}(I)$ has the dual basis δ^i, $i \in I$, defined by

$$\delta^i(e_j) := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Each element δ^i of the dual basis corresponds, interpreted as a measure, to the Dirac measure concentrated in i. A linear form $\mu \in \mathcal{S}(I)$, with $\mu_i := \mu(e_i)$, then has the representation

$$\mu = \sum_{i \in I} \mu_i \, \delta^i$$

with respect to the dual basis. This representation highlights the fact that μ can be interpreted as a signed measure, given by a linear combination of Dirac measures. The basis e_i, $i \in I$, allows us to consider the natural isomorphism between $\mathcal{F}(I)$ and $\mathcal{S}(I)$ defined by $e_i \mapsto \delta^i$. Note that this isomorphism is based on the existence of a distinguished basis of $\mathcal{F}(I)$. Information geometry, on the other hand, aims at identifying structures that are independent of such a particular choice of a basis. Therefore, the canonical basis will be used only for convenience, and all relevant information-geometric structures will be independent of this choice.

In what follows, we introduce several submanifolds of $\mathcal{S}(I)$ which play an important role in this chapter and which will be generalized and studied later in the book:

$$\begin{aligned}
\mathcal{S}_a(I) &:= \Big\{ \sum_{i \in I} \mu_i \, \delta^i : \sum_{i \in I} \mu_i = a \Big\} \quad (\text{for } a \in \mathbb{R}), \\
\mathcal{M}(I) &:= \big\{ \mu \in \mathcal{S}(I) : \mu_i \geq 0 \text{ for all } i \in I \big\}, \\
\mathcal{M}_+(I) &:= \big\{ \mu \in \mathcal{M}(I) : \mu_i > 0 \text{ for all } i \in I \big\}, \\
\mathcal{P}(I) &:= \Big\{ \mu \in \mathcal{M}(I) : \mu_i \geq 0 \text{ for all } i \in I, \text{ and } \sum_{i \in I} \mu_i = 1 \Big\}, \\
\mathcal{P}_+(I) &:= \Big\{ \mu \in \mathcal{P}(I) : \mu_i > 0 \text{ for all } i \in I, \text{ and } \sum_{i \in I} \mu_i = 1 \Big\}.
\end{aligned} \tag{2.1}$$

Obviously, we have the following inclusion chain of submanifolds of $\mathcal{S}(I)$:

$$\mathcal{P}_+(I) \subseteq \mathcal{M}_+(I) \subseteq \mathcal{S}(I).$$

In Sect. 3.1, we shall also alternatively interpret $\mathcal{P}_+(I)$ as the set of measures in $\mathcal{M}_+(I)$ that are defined up to a scaling factor, that is, as the projective space associated with $\mathcal{M}_+(I)$. From that point of view, $\mathcal{P}_+(I)$ is a positive spherical sector rather than a simplex.

Tangent and Cotangent Bundles We start with the vector space $\mathcal{S}(I)$. Given a point $\mu \in \mathcal{S}(I)$, clearly the tangent space is given by

$$T_\mu \mathcal{S}(I) = \{\mu\} \times \mathcal{S}(I). \tag{2.2}$$

Consider the natural identification of $\mathcal{S}(I)^* = \mathcal{F}(I)^{**}$ with $\mathcal{F}(I)$:

$$\mathcal{F}(I) \longrightarrow \mathcal{S}(I)^*, \qquad f \longmapsto \big(\mathcal{S}(I) \to \mathbb{R}, \mu \mapsto \mu(f)\big). \tag{2.3}$$

With this identification, the cotangent space of $\mathcal{S}(I)$ in μ is given by

$$T_\mu^* \mathcal{S}(I) = \{\mu\} \times \mathcal{F}(I). \tag{2.4}$$

As an open submanifold of $\mathcal{S}(I)$, $\mathcal{M}_+(I)$ has the same tangent and cotangent space at a point $\mu \in \mathcal{M}_+(I)$:

$$T_\mu \mathcal{M}_+(I) = \{\mu\} \times \mathcal{S}(I), \qquad T_\mu^* \mathcal{M}_+(I) = \{\mu\} \times \mathcal{F}(I). \tag{2.5}$$

Finally, we consider the manifold $\mathcal{P}_+(I)$. Obviously, for $\mu \in \mathcal{P}_+(I)$ we have

$$T_\mu \mathcal{P}_+(I) = \{\mu\} \times \mathcal{S}_0(I). \tag{2.6}$$

In order to specify the cotangent space, we consider the natural identification map (2.3). In terms of this identification, each $f \in \mathcal{F}(I)$ defines a linear form on $\mathcal{S}(I)$, which we now restrict to $\mathcal{S}_0(I)$. We obtain the map $\rho : \mathcal{F}(I) \to \mathcal{S}_0(I)^*$ that assigns to each f the linear form $\mathcal{S}_0(I) \to \mathbb{R}$, $\mu \mapsto \mu(f)$. Obviously, the kernel of ρ consists of the space of constant functions, and we obtain the natural isomorphism $\overline{\rho} : \mathcal{F}(I)/\mathbb{R} \to \mathcal{S}_0(I)^*$, $f + \mathbb{R} \mapsto \overline{\rho}(f + \mathbb{R}) := \rho(f)$. It will be useful to express the inverse $\overline{\rho}^{-1}$ in terms of the basis δ^i, $i \in I$, of $\mathcal{S}(I)$. In order to do so, assume $f \in \mathcal{S}_0(I)^*$, and consider an extension $\widetilde{f} \in \mathcal{S}(I)^*$. One can easily see that, with $f^i := \widetilde{f}(\delta^i)$, $i \in I$, the following holds:

$$\overline{\rho}^{-1}(f) = \left(\sum_{i \in I} f^i \, e_i\right) + \mathbb{R}. \tag{2.7}$$

Summarizing, we obtain

$$T_\mu^* \mathcal{P}_+(I) = \{\mu\} \times \big(\mathcal{F}(I)/\mathbb{R}\big) \tag{2.8}$$

as the cotangent space of $\mathcal{P}_+(I)$ at μ.

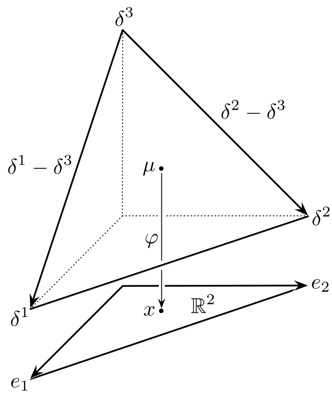

Fig. 2.1 Illustration of the chart φ for $n = 2$, with the two coordinate vectors $\delta^1 - \delta^3$ and $\delta^2 - \delta^3$

After having specified tangent and cotangent spaces at individual points μ, we finally list the corresponding tangent and cotangent bundles:

$$\begin{aligned} T\mathcal{S}(I) &= \mathcal{S}(I) \times \mathcal{S}(I), & T^*\mathcal{S}(I) &= \mathcal{S}(I) \times \mathcal{F}(I), \\ T\mathcal{M}_+(I) &= \mathcal{M}_+(I) \times \mathcal{S}(I), & T^*\mathcal{M}_+(I) &= \mathcal{M}_+(I) \times \mathcal{F}(I), \\ T\mathcal{P}_+(I) &= \mathcal{P}_+(I) \times \mathcal{S}_0(I), & T^*\mathcal{P}_+(I) &= \mathcal{P}_+(I) \times \big(\mathcal{F}(I)/\mathbb{R}\big). \end{aligned} \tag{2.9}$$

Example 2.1 (Local coordinates) In this example we consider a natural coordinate system of $\mathcal{P}_+(I)$ which is quite useful (see Fig. 2.1). We assume $I = [n+1] = \{1, \ldots, n, n+1\}$. With the open set

$$U := \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_i > 0 \text{ for all } i, \text{ and } \sum_{i=1}^n x_i < 1 \right\},$$

we consider the map

$$\varphi : \mathcal{P}_+(I) \to U, \qquad \mu = \sum_{i=1}^{n+1} \mu_i \, \delta^i \mapsto \varphi(\mu) := (\mu_1, \ldots, \mu_n)$$

and its inverse

$$\varphi^{-1} : U \to \mathcal{P}_+(I), \qquad (x_1, \ldots, x_n) \mapsto \sum_{i=1}^n x_i \, \delta^i + \left(1 - \sum_{i=1}^n x_i\right) \delta^{n+1}.$$

We have the coordinate vectors

$$\left.\frac{\partial}{\partial x_i}\right|_{\mu} = \left.\frac{\partial \varphi^{-1}}{\partial x_i}\right|_{\varphi(\mu)} = \delta^i - \delta^{n+1}, \quad i = 1, \ldots, n,$$

which form a basis of $\mathcal{S}_0(I)$. Formula (2.7) allows us to identify its dual basis with the following basis of $\mathcal{F}(I)/\mathbb{R}$:

$$dx_i := e_i + \mathbb{R}, \quad i = 1, \ldots, n.$$

Each vector $f + \mathbb{R}$ in $\mathcal{F}(I)/\mathbb{R}$ can be expressed with respect to dx_i, $i = 1, \ldots, n$:

$$\begin{aligned}
f + \mathbb{R} &= \left(\sum_{i=1}^{n+1} f^i\, e_i\right) + \mathbb{R} \\
&= \sum_{i=1}^{n+1} f^i\,(e_i + \mathbb{R}) \\
&= \sum_{i=1}^{n} f^i\,(e_i + \mathbb{R}) + f^{n+1}(e_{n+1} + \mathbb{R}) \\
&= \sum_{i=1}^{n} f^i\,(e_i + \mathbb{R}) + f^{n+1}\left(\left(\mathbb{1} - \sum_{i=1}^{n} e_i\right) + \mathbb{R}\right) \\
&= \sum_{i=1}^{n} f^i (e_i + \mathbb{R}) - \sum_{i=1}^{n} f^{n+1} (e_i + \mathbb{R}) \\
&= \sum_{i=1}^{n} \bigl(f^i - f^{n+1}\bigr)(e_i + \mathbb{R}).
\end{aligned}$$

The coordinate system of this example will be useful for explicit calculations later on.

2.2 The Fisher Metric

The Definition Given a measure $\mu \in \mathcal{M}_+(I)$, we have the following natural L^2-product on $\mathcal{F}(I)$:

$$\langle f, g\rangle_\mu = \mu(f \cdot g), \quad f, g \in \mathcal{F}(I). \tag{2.10}$$

This product allows us to consider the vector space isomorphism

$$\mathcal{F}(I) \longrightarrow \mathcal{S}(I), \qquad f \longmapsto f\mu := \langle f, \cdot\rangle_\mu. \tag{2.11}$$

The notation $f\mu$ emphasizes that, via this isomorphism, functions define linear forms on $\mathcal{F}(I)$ in terms of densities with respect to μ. The inverse, which we denote by $\widetilde{\phi}_\mu$, maps linear forms to functions and represents a simple version of the

Radon–Nikodym derivative with respect to μ:

$$\widetilde{\phi}_\mu : \mathcal{S}(I) \longrightarrow \mathcal{F}(I) = \mathcal{S}(I)^*, \qquad a = \sum_i a_i \, \delta^i \longmapsto \frac{da}{d\mu} := \sum_i \frac{a_i}{\mu_i} \, e_i. \tag{2.12}$$

This gives rise to the formulation of (2.10) on the dual space of $\mathcal{F}(I)$:

$$\langle a, b \rangle_\mu = \mu\left(\frac{da}{d\mu} \cdot \frac{db}{d\mu}\right) = \sum_i \frac{1}{\mu_i} a_i b_i, \quad a, b \in \mathcal{S}(I). \tag{2.13}$$

With this metric, the orthogonal complement of $\mathcal{S}_0(I)$ in $\mathcal{S}(I)$ is given by $\mathbb{R} \cdot \mu = \{\lambda \cdot \mu : \lambda \in \mathbb{R}\}$, and we have the orthogonal decomposition $a = \Pi_\mu^\top a + \Pi_\mu^\perp a$ of vectors $a \in \mathcal{S}(I)$, where

$$\Pi_\mu^\top(a) = \sum_{i \in I} \left(a_i - \mu_i \sum_{j \in I} a_j\right) \delta^i, \qquad \Pi_\mu^\perp(a) = \sum_{i \in I} \left(\mu_i \sum_{j \in I} a_j\right) \delta^i. \tag{2.14}$$

If we restrict this metric to $\mathcal{S}_0(I) \subseteq \mathcal{S}(I)$, then we obtain the following identification of $\mathcal{S}_0(I)$ with the dual space:

$$\phi_\mu : \mathcal{S}_0(I) \longrightarrow \mathcal{F}(I)/\mathbb{R} = \mathcal{S}_0(I)^*, \qquad a \longmapsto \frac{da}{d\mu} + \mathbb{R}. \tag{2.15}$$

With the natural inclusion map $\imath : \mathcal{S}_0(I) \to \mathcal{S}(I)$, and $\imath_\mu := \widetilde{\phi}_\mu \circ \imath \circ {\phi_\mu}^{-1}$, the following diagram commutes:

$$\begin{array}{ccc} \mathcal{S}(I) & \xrightarrow{\widetilde{\phi}_\mu} & \mathcal{S}(I)^* \\ {\scriptstyle \imath}\uparrow & & \uparrow{\scriptstyle \imath_\mu} \\ \mathcal{S}_0(I) & \xrightarrow[\phi_\mu]{} & \mathcal{S}_0(I)^* \end{array} \tag{2.16}$$

This diagram defines linear maps between tangent and cotangent spaces in the individual points of $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I)$. Collecting all these maps to corresponding bundle maps, we obtain a commutative diagram between the tangent and cotangent bundles:

$$\begin{array}{ccc} T\mathcal{M}_+(I) & \xrightarrow{\widetilde{\phi}} & T^*\mathcal{M}_+(I) \\ {\scriptstyle \imath}\uparrow & & \uparrow{\scriptstyle \imath} \\ T\mathcal{P}_+(I) & \xrightarrow[\phi]{} & T^*\mathcal{P}_+(I) \end{array} \tag{2.17}$$

The inner product (2.13) defines a Riemannian metric on $\mathcal{M}_+(I)$, on which the maps $\widetilde{\phi}$ and ϕ are based.

Definition 2.1 (Fisher metric) Given two vectors $A = (\mu, a)$ and $B = (\mu, b)$ of the tangent space $T_\mu \mathcal{M}_+(I)$, we consider

$$\mathfrak{g}_\mu(A, B) := \langle a, b \rangle_\mu. \tag{2.18}$$

This Riemannian metric $\mathfrak{g}$ on $\mathcal{M}_+(I)$ is called the *Fisher metric*.

The Fisher metric was introduced as a Riemannian metric by Rao [219]. It is relevant for estimation theory within statistics and also appears in mathematical population genetics where it is known as the *Shahshahani metric* [123, 124]. We shall outline the biological perspective of this metric in Sect. 6.2.

We now express the Fisher metric with respect to the coordinates of Example 2.1, where we concentrate on the restriction of the Fisher metric to $\mathcal{P}_+(I)$. With respect to the chart φ of Example 2.1, the first fundamental form of the Fisher metric is given as

$$g_{ij}(\mu) = \sum_{k=1}^{n} \frac{1}{\mu_k} \delta_{ki}\, \delta_{kj} + \frac{1}{\mu_{n+1}} = \begin{cases} \frac{1}{\mu_i} + \frac{1}{\mu_{n+1}}, & \text{if } i = j, \\ \frac{1}{\mu_{n+1}}, & \text{otherwise.} \end{cases} \tag{2.19}$$

The inverse of this matrix is given as

$$g^{ij}(\mu) = \begin{cases} \mu_i\,(1 - \mu_i), & \text{if } i = j, \\ -\mu_i\, \mu_j, & \text{otherwise.} \end{cases} \tag{2.20}$$

Written as matrices, we have

$$G(\mu) := (g_{ij})(\mu) = \frac{1}{\mu_{n+1}} \begin{pmatrix} \frac{\mu_{n+1}}{\mu_1} + 1 & 1 & \cdots & 1 \\ 1 & \frac{\mu_{n+1}}{\mu_2} + 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{\mu_{n+1}}{\mu_n} + 1 \end{pmatrix}, \tag{2.21}$$

$$G^{-1}(\mu) = \big(g^{ij}\big)(\mu) = \begin{pmatrix} \mu_1\,(1 - \mu_1) & -\mu_1\, \mu_2 & \cdots & -\mu_1\, \mu_n \\ -\mu_2\, \mu_1 & \mu_2\,(1 - \mu_2) & \cdots & -\mu_2\, \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_n\, \mu_1 & -\mu_n\, \mu_2 & \cdots & \mu_n\,(1 - \mu_n) \end{pmatrix}. \tag{2.22}$$

This is nothing but the covariance matrix of the probability distribution μ, in the following sense. We draw the element $i \in \{1, \dots, n\}$ with probability μ_i, and we put $X_i = 1$ and $X_j = 0$ for $j \neq i$ when i happens to be drawn. We then have the expectation values

$$\mu_i = \mathbb{E}(X_i) = \mathbb{E}\big(X_i^2\big), \tag{2.23}$$

and hence, the variances and covariances are

$$\mathrm{Var}(X_i) = \mu_i(1-\mu_i), \qquad \mathrm{Cov}(X_i X_j) = -\mu_i \mu_j \quad \text{for } j \neq i, \tag{2.24}$$

that is, (2.22). In fact, this is the statistical origin of the Fisher metric as a covariance matrix [219].

The Fisher metric is an example of a covariant 2-tensor on M, in the sense of the following definition (see also (B.16) and (B.17) in Appendix B).

Definition 2.2 A *covariant n-tensor Θ on a manifold M* is a collection of n-multilinear maps

$$\Theta_\xi : \bigtimes^n T_\xi M \longrightarrow \mathbb{R}, \qquad (V_1, \dots, V_n) \longmapsto \Theta_\xi(V_1, \dots, V_n)$$

which is continuous in the sense that for continuous vector fields V_i the function $\xi \mapsto \Theta_\xi(V_1, \dots, V_n)$ is continuous.

If $f : M_1 \to M_2$ is a differentiable map between the manifolds M_1 and M_2, then it can be used to pull back covariant n-tensors. That is, if Θ is a covariant n-tensor on M_2, then its pullback to M_1 by f is defined to be the tensor $f^*(\Theta)$ on M_1 given as

$$f^*(\Theta)_\xi(V_1, \dots, V_n) := \Theta_{f(\xi)}\left(\frac{\partial f}{\partial V_1}, \dots, \frac{\partial f}{\partial V_n}\right). \tag{2.25}$$

Information geometry deals with statistical models, that is, submanifolds of $\mathcal{P}_+(I)$. Instead of considering submanifolds directly, we take a slightly different perspective here. We consider statistical models as a manifold together with an embedding into $\mathcal{P}_+(I)$, or more generally, into $\mathcal{M}_+(I)$. To be more precise, let be M an n-dimensional (differentiable) manifold and $M \hookrightarrow \mathcal{M}_+(I)$, $\xi \mapsto p(\xi) = \sum_{i \in I} p_i(\xi)\,\delta^i$, an embedding. The pullback (2.25) of the Fisher metric $\mathfrak{g}$ defines a metric on M. More precisely, for $A, B \in T_\xi M$ we set

$$\begin{aligned} g_\xi(A,B) &:= p^*(\mathfrak{g})_\xi(A,B) \\ &\overset{(2.25)}{=} \mathfrak{g}_{p(\xi)}\left(\frac{\partial p}{\partial A}, \frac{\partial p}{\partial B}\right) \\ &= \sum_i \frac{1}{p_i(\xi)} \frac{\partial p_i}{\partial A}(\xi) \frac{\partial p_i}{\partial B}(\xi) \\ &= \sum_i p_i(\xi) \frac{\partial \log p_i}{\partial A}(\xi) \frac{\partial \log p_i}{\partial B}(\xi). \end{aligned} \tag{2.26}$$

This representation of the Fisher metric is more familiar within the standard information-geometric treatment.

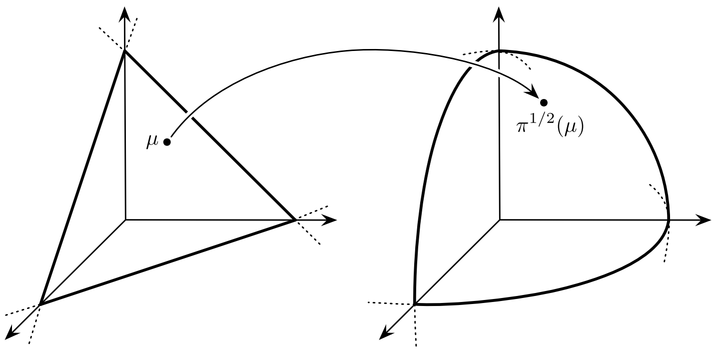

Fig. 2.2 Mapping from simplex to sphere

Extending the Fisher Metric to the Boundary As is obvious from (2.13) and also from the first fundamental form (2.19), the Fisher metric is not defined at the boundary of the simplex. It is, however, possible to extend the Fisher metric to the boundary by identifying the simplex with a sector of a sphere in $\mathbb{R}^I = \mathcal{F}(I)$. In order to be more precise, we consider the following sector of the sphere with radius 2 (or, equivalently (up to the factor 2, of course), the positive part of the projective space, according to the interpretation of the set of probability measures as a projective version of the space of positive measures alluded to above and taken up in Sect. 3.1):

$$S_{2,+}(I) := \left\{ f \in \mathcal{F}(I) : f(i) > 0 \text{ for all } i \in I, \text{ and } \sum_i f^2(i) = 4 \right\}.$$

As a submanifold of $\mathcal{F}(I)$ it carries the induced standard metric $\langle \cdot, \cdot \rangle$ of $\mathcal{F}(I)$. We consider the following diffeomorphism (see Fig. 2.2):

$$\pi^{1/2} : \mathcal{P}_+(I) \to S_{2,+}(I), \qquad \mu = \sum_i \mu_i \, \delta^i \mapsto 2 \sum_i \sqrt{\mu_i} e_i .$$

Note that $\|\pi^{1/2}(\mu)\| = \sqrt{\sum_i (2\sqrt{\mu_i})^2} = 2\sqrt{\sum_i \mu_i} = 2$.

Proposition 2.1 *The map $\pi^{1/2}$ is an isometry between $\mathcal{P}_+(I)$ with the Fisher metric $\mathfrak{g}$ and $S_{2,+}(I)$ with the induced canonical scalar product of $\mathcal{F}(I)$:*

$$\left\langle \frac{\partial \pi^{1/2}}{\partial A}(\mu), \frac{\partial \pi^{1/2}}{\partial B}(\mu) \right\rangle = \mathfrak{g}_\mu(A, B), \qquad A, B \in T_\mu \mathcal{P}_+(I).$$

That is, the Fisher metric coincides with the pullback of the standard metric on $\mathcal{F}(I)$ by the map $\pi^{1/2}$. In particular, the extension of the standard metric on $S_{2,+}(I)$ to the boundary can be considered as an extension of the Fisher metric.

Proof With $a, b \in \mathcal{S}_0(I)$ such that $A = (\mu, a)$ and $B = (\mu, b)$, we have

$$\left\langle \frac{\partial \pi^{1/2}}{\partial A}(\mu), \frac{\partial \pi^{1/2}}{\partial B}(\mu) \right\rangle = \left\langle \frac{d}{dt}\pi^{1/2}(\mu + ta)\Big|_{t=0}, \frac{d}{dt}\pi^{1/2}(\mu + tb)\Big|_{t=0} \right\rangle$$
$$= \sum_i \frac{1}{\sqrt{\mu_i}} a_i \cdot \frac{1}{\sqrt{\mu_i}} b_i$$
$$= \mathfrak{g}_\mu(A, B).$$

□

Fisher and Hellinger Distance Proposition 2.1 allows us to give an explicit formula for the Riemannian distance between two points $\mu, \nu \in \mathcal{P}_+(I)$ which is defined as

$$d^F(\mu, \nu) := \inf_{\substack{\gamma:[r,s]\to\mathcal{P}_+(I) \\ \gamma(r)=\mu,\, \gamma(s)=\nu}} L(\gamma),$$

where $L(\gamma)$ denotes the length of a curve $\gamma : [r, s] \to \mathcal{P}_+(I)$ given by

$$L(\gamma) = \int_r^s \|\dot{\gamma}(t)\|_{\gamma(t)}\, dt = \int_r^s \sqrt{\mathfrak{g}_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big)}\, dt.$$

We refer to d^F as the *Fisher distance*. With Proposition 2.1 we directly obtain the following corollary.

Corollary 2.1 *Let $d : S_{2,+}(I) \times S_{2,+}(I) \to \mathbb{R}$ denote the metric that is induced from the standard metric on $\mathcal{F}(I)$. Then*

$$d^F(\mu, \nu) = d\big(\pi^{1/2}(\mu), \pi^{1/2}(\nu)\big) = 2 \arccos\left(\sum_i \sqrt{\mu_i\, \nu_i}\right). \tag{2.27}$$

Proof We have

$$\cos\alpha = \frac{\langle \pi^{1/2}(\mu), \pi^{1/2}(\nu)\rangle}{\|\pi^{1/2}(\mu)\| \cdot \|\pi^{1/2}(\nu)\|} = \frac{\sum_i (2\sqrt{\mu_i})(2\sqrt{\nu_i})}{2 \cdot 2} = \sum_i \sqrt{\mu_i\, \nu_i}.$$

This implies

$$\frac{d^F(\mu, \nu)}{2} = \alpha = \arccos\left(\sum_i \sqrt{\mu_i\, \nu_i}\right).$$

□

A distance measure that is closely related to the Fisher distance is the *Hellinger distance*. It is not induced from $\mathcal{F}(I)$ onto $S_{2,+}(I)$ but restricted to $S_{2,+}(I)$:

$$d^H(\mu, \nu) := \sqrt{\sum_i \big(\sqrt{\mu_i} - \sqrt{\nu_i}\big)^2}. \tag{2.28}$$

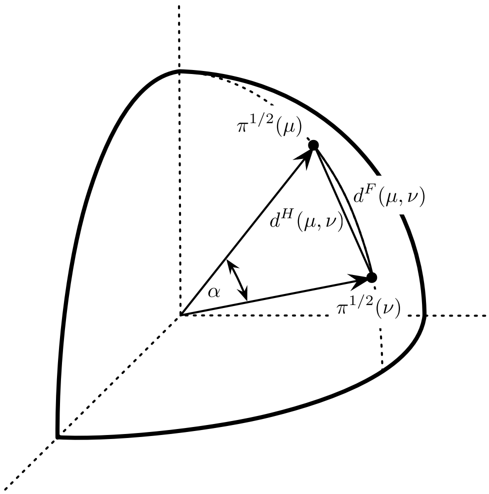

Fig. 2.3 Illustration of the relation between the Fisher distance $d^F(\mu,\nu)$ and the Hellinger distance $d^H(\mu,\nu)$ of two probability measures μ and ν, see Eq. (2.29)

We have the following relation between d^F and d^H (see Fig. 2.3):

$$
\begin{aligned}
d^H(\mu,\nu) &= \sqrt{\sum_i (\sqrt{\mu_i} - \sqrt{\nu_i})^2} \\
&= \sqrt{\sum_i (\mu_i - 2\sqrt{\mu_i\,\nu_i} + \nu_i)} \\
&= \sqrt{2\Bigl(1 - \sum_i \sqrt{\mu_i\,\nu_i}\Bigr)} \\
&= \sqrt{2\Bigl(1 - \cos\Bigl(\frac{1}{2}\,d^F(\mu,\nu)\Bigr)\Bigr)}.
\end{aligned} \tag{2.29}
$$

Chentsov's Characterization of the Fisher Metric In what follows, we present a classical characterization of the Fisher metric through invariance properties. This is due to Chentsov [64].

Consider two non-empty and finite sets I and I'. A *Markov kernel* is a map

$$K : I \to \mathcal{P}(I'), \qquad i \mapsto K^i := \sum_{i' \in I'} K^i_{i'}\,\delta^{i'}. \tag{2.30}$$

Particular examples of Markov kernels are given in terms of (deterministic) maps $f : I \to I'$. Given such a map, we simply define K^f by $i \mapsto \delta^{f(i)}$. Each Markov kernel K induces a corresponding map between probability distributions:

$$K_* : \mathcal{P}(I) \to \mathcal{P}(I'), \qquad \mu = \sum_{i \in I} \mu_i\,\delta^i \mapsto \sum_{i \in I} \mu_i\,K^i. \tag{2.31}$$

The map K_* is called the *Markov morphism induced by* K. Note that K_* may also be regarded as a linear map $K_* : \mathcal{S}(I) \to \mathcal{S}(I')$. Given a map $f : I \to I'$, we use $f_* := (K^f)_*$ as short-hand notation.

Now assume that $|I| \le |I'|$. We call a Markov kernel K *congruent* if there is a partition A_i, $i \in I$, of I', such that the following condition holds:

$$K^i_{i'} > 0 \quad \Leftrightarrow \quad i' \in A_i. \tag{2.32}$$

If K is congruent and $\mu \in \mathcal{P}_+(I)$ then $K_*(\mu) \in \mathcal{P}_+(I')$. This implies a differentiable map

$$K_* : \mathcal{P}_+(I) \to \mathcal{P}_+\big(I'\big),$$

and the differential in μ is given by

$$d_\mu K_* : T_\mu \mathcal{P}_+(I) \to T_{K_*(\mu)} \mathcal{P}_+\big(I'\big), \qquad (\mu, \nu - \mu) \mapsto \big(K_*(\mu), K_*(\nu) - K_*(\mu)\big).$$

The following theorem has been proven by Chentsov.

Theorem 2.1 (Cf. [65, Theorem 11.1]) *We assign to each non-empty and finite set I a metric h^I on $\mathcal{P}_+(I)$. If for each congruent Markov kernel $K : I \to \mathcal{P}(I')$ we have invariance in the sense*

$$h^I_p(A, B) = h^{I'}_{K_*(p)}\big(d_p K_*(A), d_p K_*(B)\big),$$

or for short $(K_*)^*(h^{I'}) = h^I$ in the notation of (2.25), then there is a constant $\alpha > 0$, such that $h^I = \alpha\, \mathfrak{g}^I$ for all I, where the latter is the Fisher metric on $\mathcal{P}_+(I)$.

Proof **Step 1:** First we consider permutations $\pi : I \to I$. The center $c_I := \frac{1}{|I|} \sum_{i \in I} \delta^i$ is left-invariant, that is, $\pi_*(c_I) = c_I$. With $E_i := (c_I, \delta^i - c_I) \in T_{c_I} \mathcal{P}_+(I)$, we also have

$$d_{c_I} \pi_*(E_i) = E_{\pi(i)}, \quad i \in I.$$

For each $i, j \in I$, $i \neq j$, choose a transposition π of i and j, that is, $\pi(i) = j$, $\pi(j) = i$, and $\pi(k) = k$ if $k \notin \{i, j\}$. This implies

$$\begin{aligned} h^I_{ii}(c_I) &= h^I_{c_I}(E_i, E_i) = h^I_{\pi_*(c_I)}\big(d_{c_I}\pi_*(E_i), d_{c_I}\pi_*(E_i)\big) = h^I_{c_I}(E_{\pi(i)}, E_{\pi(i)}) \\ &= h^I_{c_I}(E_j, E_j) = h^I_{jj}(c_I) =: f_1(n), \end{aligned}$$

where we set $n := |I|$. In a similar way, we obtain that all $h^I_{ij}(c_I)$ with $i \neq j$ coincide. We denote them by $f_2(n)$. The functions $f_1(n)$ and $f_2(n)$ have to satisfy the following:

$$\begin{aligned} f_1(n) + (n-1) f_2(n) &= \sum_{j \in I} h^I_{ij}(c_I) = \sum_{j \in I} h^I_{c_I}(E_i, E_j) \\ &= h^I_{c_I}\left(E_i, \sum_{j \in I} E_j\right) = h^I_{c_I}(E_i, 0) = 0. \end{aligned}$$

Consider two vectors

$$a = \sum_{i \in I} a_i \, \delta^i, \qquad b = \sum_{i \in I} b_i \, \delta^i.$$

Assuming $a, b \in \mathcal{S}_0(I)$, we have $\sum_{i \in I} a_i = 0$ and $\sum_{i \in I} b_i = 0$ and therefore

$$a = \sum_{i \in I} a_i \left(\delta^i - c_I\right), \qquad b = \sum_{i \in I} b_i \left(\delta^i - c_I\right).$$

This implies for $A = (c_I, a)$ and $B = (c_I, b)$

$$\begin{aligned}
h^I_{c_I}(A, B) &= \sum_{i,j \in I} a_i \, b_j \, h^I_{ij}(c_I) = \sum_{i \in I} a_i \, b_i \, h^I_{ii}(c_I) + \sum_{\substack{i,j \in I \\ i \neq j}} a_i \, b_j \, h^I_{ij}(c_I) \\
&= f_1(n) \sum_{i \in I} a_i \, b_i + f_2(n) \sum_{\substack{i,j \in I \\ i \neq j}} a_i \, b_j \\
&= -(n-1) f_2(n) \sum_{i \in I} a_i \, b_i + f_2(n) \sum_{\substack{i,j \in I \\ i \neq j}} a_i \, b_j \\
&= f_2(n) \left\{ -n \sum_{i \in I} a_i \, b_i + \sum_{i,j \in I} a_i \, b_j \right\} \\
&= -f_2(n) \sum_{i \in I} \frac{1}{\frac{1}{n}} a_i \, b_i \\
&= -f_2(n) \, \mathfrak{g}^I_{c_I}(A, B).
\end{aligned}$$

Step 2: In this step, we show that the function $f(n)$ is actually independent of n and therefore a constant. In order to do so, we divide each element $i \in I$ into k elements. More precisely, we set $I' := I \times \{1, \dots, k\}$. With the partition $A_i := \{(i, j) : 1 \leq j \leq k\}$, $i \in I$, we define the Markov kernel K by

$$i \mapsto K^i = \frac{1}{k} \sum_{j=1}^{k} \delta^{(i,j)}.$$

This kernel satisfies $K_*(c_I) = c_{I'}$, and we have

$$\begin{aligned}
d_{c_I} K_*(E_i) &= d_{c_I} K_*\bigl(c_I, \delta^i - c_I\bigr) \\
&= \left(c_{I'}, \frac{1}{k} \sum_{j=1}^{k} \left(\delta^{(i,j)} - \frac{1}{n} \sum_{i' \in I} \delta^{(i',j)} \right) \right)
\end{aligned}$$

$$= \left(c_{I'}, \frac{1}{k} \sum_{j=1}^{k} \left(\delta^{(i,j)} - \frac{1}{n\,k} \sum_{i' \in I} \sum_{j'=1}^{k} \delta^{(i',j')} \right) \right)$$

$$= \frac{1}{k} \sum_{j=1}^{k} E'_{(i,j)}.$$

With $r, s \in I$, $r \neq s$, this implies

$$f_2(n) = h^I_{c_I}(E_r, E_s) = h^{I'}_{c_{I'}} \left(\frac{1}{k} \sum_{j=1}^{k} E'_{r,j}, \frac{1}{k} \sum_{j=1}^{k} E'_{s,j} \right)$$

$$= \frac{1}{k^2} \sum_{j_1, j_2 = 1}^{k} h^{I'}_{c_{I'}} \big(E'_{r,j_1}, E'_{s,j_2} \big)$$

$$= \frac{1}{k^2} k^2 \, f_2(n \cdot k) = f_2(n \cdot k).$$

Exchanging the role of n and k, we obtain $f_2(k) = f_2(k \cdot n) = f_2(n)$ and therefore $-f_2(n)$ is a positive constant in n, which we denote by α. In the center c_I, we have shown that

$$h^I_{c_I}(A, B) = \alpha \, \mathfrak{g}^I_{c_I}(A, B) = 0, \quad A, B \in T_{c_I} \mathcal{P}_+(I).$$

It remains to show that this equality also holds for all other points. This is our next step.

Step 3: First consider a point $\mu \in \mathcal{P}_+(I)$ that has rational coordinates, that is,

$$\mu = \sum_{i \in I} \frac{k_i}{n} \, \delta^i,$$

with $\sum_i k_i = n$. We now define a set I' and a congruent Markov kernel $K : I \to \mathcal{P}(I')$ so that $K_*(\mu) = c_{I'}$. With

$$I' := \biguplus_{i \in I} \big(\{i\} \times \{1, \ldots, k_i\} \big),$$

("$\biguplus$" denotes the disjoint union) we consider the Markov kernel

$$K : i \mapsto \frac{1}{k_i} \sum_{j=1}^{k_i} \delta^{(i,j)}.$$

Obviously, we have

$$d_\mu K_* : A = \left(\mu, \sum_{i \in I} a_i \, \delta^i \right) \mapsto d_\mu K_*(A) = \left(c_{I'}, \sum_{i \in I} \sum_{j=1}^{k_i} \frac{a_i}{k_i} \, \delta^{(i,j)} \right).$$

This implies

$$\begin{aligned}
h^I_\mu(A,B) &= h^{I'}_{K_*(\mu)}\big(d_\mu K_*(A), d_\mu K_*(B)\big) = \alpha\, g^{I'}_{c_{I'}}\big(d_\mu K_*(A), d_\mu K_*(B)\big) \\
&= \alpha \sum_{i\in I}\sum_{j=1}^{k_i} \frac{1}{\frac{1}{n}}\,\frac{a_i}{k_i}\,\frac{b_i}{k_i} = \alpha \sum_{i\in I} k_i \frac{1}{\frac{1}{n}}\,\frac{a_i}{k_i}\,\frac{b_i}{k_i} = \alpha \sum_{i\in I} \frac{1}{\mu_i}\, a_i\, b_i \\
&= \alpha\, \mathfrak{g}^I_\mu(A,B).
\end{aligned}$$

We have this equality for all probability vectors μ with rational coordinates. As we assume continuity with respect to the base point μ, the equality of $h^I_\mu(A,B) = \alpha\, \mathfrak{g}^{I'}_\mu(A,B)$ holds for all $\mu \in \mathcal{P}_+(I)$. □

2.3 Gradient Fields

In this section, we are going to study vector and covector fields on $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I)$. We begin with the first case, which is the simpler one, and consider covector fields given by a differentiable function $V : \mathcal{M}_+(I) \to \mathbb{R}$. The differential in μ is defined as the linear form

$$d_\mu V : \mathcal{S}(I) \to \mathbb{R}, \qquad a \mapsto d_\mu V(a) = \frac{\partial V}{\partial a}(\mu).$$

In terms of the canonical basis, we have

$$d_\mu V = \sum_i \partial_i V(\mu)\, e_i \in \mathcal{F}(I), \tag{2.33}$$

where $\partial_i V(\mu) := \frac{\partial V}{\partial \mu_i}(\mu) := \frac{\partial V}{\partial \delta^i}(\mu)$. This defines the covector field

$$dV : \mathcal{M}_+(I) \to \mathcal{F}(I), \qquad \mu \mapsto d_\mu V.$$

The Fisher metric $\mathfrak{g}$ allows us to identify $d_\mu V$ with an element of $T_\mu \mathcal{M}_+(I)$ in terms of the map $\widetilde{\phi}_\mu$, the *gradient* of V in μ:

$$\operatorname{grad}_\mu V := \widetilde{\phi}_\mu^{-1}(d_\mu V) = \sum_i \mu_i\, \partial_i V(\mu)\, \delta^i. \tag{2.34}$$

Given a function $f : \mathcal{M}_+(I) \to \mathcal{F}(I)$, $\mu \mapsto f(\mu) = \sum_{i\in I} f^i(\mu)\, e_i$, we can ask whether there exists a differentiable function V such that $f(\mu) = d_\mu V$. In this case, f is called exact. It is easy to see that f is exact if and only if the condition

$$\frac{\partial f^i}{\partial \mu_j} = \frac{\partial f^j}{\partial \mu_i} \tag{2.35}$$

holds on $\mathcal{M}_+(I)$ for all $i, j \in I$.

Now we come to vector and covector fields on $\mathcal{P}_+(I)$. The commutative diagram (2.17) allows us to relate sections to each other. Of particular interest are sections in $T^*\mathcal{P}_+(I) = \mathcal{P}_+(I) \times (\mathcal{F}(I)/\mathbb{R})$ (*covector fields*) as well as sections in $T\mathcal{P}_+(I)$ (*vector fields*). As all bundles are of product form $\mathcal{P}_+(I) \times \mathcal{V}$, sections are given by functions $f : \mathcal{P}_+(I) \to \mathcal{V}$. We assume that f is a C^∞ function. We will also use C^∞ extensions $\widetilde{f} : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U}$ is an open subset of $\mathcal{S}(I)$ containing $\mathcal{P}_+(I)$, and $\widetilde{f}|_{\mathcal{U}} = f$. To simplify the notation, we will also use the same symbol f for the extension $\widetilde{f}$. Given a section $f : \mathcal{P}_+(I) \to \mathcal{F}(I)$, we assign various other sections to it:

$$\begin{aligned}
\overline{f} : \mathcal{P}_+(I) &\to \mathbb{R}, && \mu \mapsto \overline{f}(\mu) := \mu(f(\mu)) = \textstyle\sum_i \mu_i \, f^i(\mu), \\
[f] : \mathcal{P}_+(I) &\to (\mathcal{F}(I)/\mathbb{R}), && \mu \mapsto f(\mu) + \mathbb{R}, \\
\widetilde{\phi}(f) : \mathcal{P}_+(I) &\to \mathcal{S}(I), && \mu \mapsto f(\mu)\mu = \textstyle\sum_i \mu_i \, f^i(\mu)\, \delta^i, \\
\widehat{f} : \mathcal{P}_+(I) &\to \mathcal{S}_0(I), && \mu \mapsto (f(\mu) - \overline{f}(\mu))\mu = \textstyle\sum_i \mu_i (f^i(\mu) - \overline{f}(\mu))\delta^i.
\end{aligned}$$

In what follows, we consider covector fields given by a differentiable function $V : \mathcal{P}_+(I) \to \mathbb{R}$. The differential in μ is defined as the linear form

$$d_\mu V : T_\mu \mathcal{P}_+(I) \to \mathbb{R}, \qquad a \mapsto d_\mu V(a) = \frac{\partial V}{\partial a}(\mu),$$

which defines a covector field $dV : \mu \mapsto d_\mu V \in T^*_\mu \mathcal{P}_+(I)$. In order to interpret it as a vector in $\mathcal{F}(I)/\mathbb{R}$, consider an extension $\widetilde{V} : \mathcal{U} \to \mathbb{R}$ of V to an open neighborhood of $\mathcal{P}_+(I)$. This yields a corresponding extension $d_\mu \widetilde{V} : \mathcal{S}(I) \to \mathbb{R}$, and according to (2.7) we have

$$d_\mu V = \sum_i \partial_i \widetilde{V}(\mu)\, e_i + \mathbb{R}, \tag{2.36}$$

where $\partial_i \widetilde{V}(\mu) = \frac{\partial \widetilde{V}}{\partial \delta^i}(\mu)$. The Fisher metric $\mathfrak{g}$ allows us to identify $d_\mu V$ with an element of $T_\mu \mathcal{P}_+(I)$ via the map ϕ_μ, the *gradient* of V in μ:

$$\operatorname{grad}_\mu V := \phi_\mu^{-1}(d_\mu V). \tag{2.37}$$

(See (B.22) in Appendix B for the general construction.)

Proposition 2.2 *Let $V : \mathcal{P}_+(I) \to \mathbb{R}$ be a differentiable function, $\mathcal{U}$ an open subset of $\mathcal{S}(I)$ that contains $\mathcal{P}_+(I)$, and $\widetilde{V} : \mathcal{U} \to \mathbb{R}$ a differentiable continuation of V, that is, $\widetilde{V}|_{\mathcal{P}_+(I)} = V$. Then the coordinates of* $\operatorname{grad}_\mu V$ *with respect to δ^i are given as*

$$(\operatorname{grad}_\mu V)_i = \mu_i \Big(\partial_i \widetilde{V}(\mu) - \sum_j \mu_j \, \partial_j \widetilde{V}(\mu) \Big), \quad i \in I.$$

Proof This follows from (2.36), (2.37), and the definition of ϕ_μ. Alternatively, one can show this directly: We have to verify

$$\mathfrak{g}_\mu(\operatorname{grad}_\mu V, a) = d_\mu V(a), \quad a \in T_\mu \mathcal{P}_+(I).$$

$$\begin{aligned}
\mathfrak{g}_\mu(\operatorname{grad}_\mu V, a) &= \sum_i \frac{1}{\mu_i}\left(\mu_i\left(\partial_i \widetilde{V}(\mu) - \sum_j \mu_j \partial_j \widetilde{V}(\mu)\right)\right) a_i \\
&= \sum_i a_i\, \partial_i \widetilde{V}(\mu) - \underbrace{\sum_i a_i \sum_j \mu_j \partial_j \widetilde{V}(\mu)}_{=0} \\
&= \frac{\partial \widetilde{V}}{\partial a}(\mu) \\
&= \lim_{t\to 0} \frac{\widetilde{V}(\mu + ta) - \widetilde{V}(\mu)}{t} \\
&= \lim_{t\to 0} \frac{V(\mu + ta) - V(\mu)}{t} \\
&= \frac{\partial V}{\partial a}(\mu) \\
&= d_\mu V(a).
\end{aligned}$$

□

Proposition 2.3 *Consider a map $f : \mathcal{U} \to \mathcal{F}(I)$, $\mu \mapsto f(\mu) = \sum_{i\in I} f^i(\mu)\, e_i$, defined on a neighborhood of $\mathcal{P}_+(I)$. Then the following statements are equivalent*:

(1) *The vector field $\widehat{f}$ is a Fisher gradient field on $\mathcal{P}_+(I)$.*
(2) *The covector field $[f] : \mathcal{P}_+(I) \to \mathcal{F}(I)/\mathbb{R}$, $\mu \mapsto [f](\mu) := f(\mu) + \mathbb{R}$, is exact, that is, there exists a function $V : \mathcal{P}_+(I) \to \mathbb{R}$ satisfying $d_\mu V = [f](\mu)$.*
(3) *The relation*

$$\frac{\partial f^i}{\partial \mu_j} + \frac{\partial f^j}{\partial \mu_k} + \frac{\partial f^k}{\partial \mu_i} = \frac{\partial f^i}{\partial \mu_k} + \frac{\partial f^k}{\partial \mu_j} + \frac{\partial f^j}{\partial \mu_i} \tag{2.38}$$

holds on $\mathcal{P}_+(I)$ for all $i, j, k \in I$.

Proof (1) ⇔ (2) This is clear.

(2) ⇔ (3) The covector field $f + \mathbb{R}$ is exact if and only if it is closed. The latter property is expressed in local coordinates. Without restriction of generality we assume $I = \{1, \ldots, n, n+1\}$ and choose the coordinate system of Example 2.1.

$$\left.\frac{\partial \varphi^{-1}}{\partial x_i}\right|_{\varphi(p)} = \delta^i - \delta^{n+1}, \quad i = 1, \ldots, n.$$

This family is a basis of $\mathcal{S}_0(I)$. The dual basis in $\mathcal{F}(I)/\mathbb{R}$ is given as

$$e_i + \mathbb{R}, \quad i = 1, \ldots, n.$$

We now express the covector field $[f]$ in these coordinates:

$$\begin{aligned} f(\mu)+\mathbb{R} &= \left(\sum_{i=1}^{n+1} f^i(p)\, e_i\right)+\mathbb{R} \\ &= \sum_{i=1}^{n}\big(f^i(\mu)-f^{n+1}(\mu)\big)(e_i+\mathbb{R}). \end{aligned}$$

The covector field $f+\mathbb{R}$ is closed, if the coefficients $f^i - f^{n+1}$ satisfy the following integrability condition:

$$\frac{\partial(f^i - f^{n+1})}{\partial(\delta^j - \delta^{n+1})}(\mu) = \frac{\partial(f^j - f^{n+1})}{\partial(\delta^i - \delta^{n+1})}(\mu), \quad i, j = 1, \ldots, n.$$

This is equivalent to

$$\frac{\partial f^i}{\partial \delta^j} + \frac{\partial f^j}{\partial \delta^{n+1}} + \frac{\partial f^{n+1}}{\partial \delta^i} = \frac{\partial f^j}{\partial \delta^i} + \frac{\partial f^i}{\partial \delta^{n+1}} + \frac{\partial f^{n+1}}{\partial \delta^j}.$$

Replacing $n+1$ by k yields the integrability condition (2.38). □

2.4 The *m*- and *e*-Connections

The tangent bundle $T\mathcal{M}_+(I)$ and the cotangent bundle $T^*\mathcal{M}_+(I)$ are of product structure. Given two points μ and ν in $\mathcal{M}_+(I)$, this allows for the following natural identification of $T_\mu\mathcal{M}_+(I)$ with $T_\nu\mathcal{M}_+(I)$ and $T^*_\mu\mathcal{M}_+(I)$ with $T^*_\nu\mathcal{M}_+(I)$:

$$\widetilde{\Pi}^{(m)}_{\mu,\nu} : T_\mu\mathcal{M}_+(I) \longrightarrow T_\nu\mathcal{M}_+(I), \qquad (\mu, a) \longmapsto (\nu, a), \tag{2.39}$$

$$\widetilde{\Pi}^{(e)}_{\mu,\nu} : T^*_\mu\mathcal{M}_+(I) \longrightarrow T^*_\nu\mathcal{M}_+(I), \qquad (\mu, f) \longmapsto (\nu, f). \tag{2.40}$$

Note that these identifications of fibers is not a consequence of the triviality of the vector bundles only. In general, a trivial vector bundle has no distinguished trivialization. However, in our case the bundles have a natural product structure.

With the bundle isomorphism $\widetilde{\phi}$ (see diagram (2.17)) one can interpret $\widetilde{\Pi}^{(e)}_{\mu,\nu}$ as a parallel transport in $T\mathcal{M}_+(I)$, given by

$$\widetilde{\Pi}^{(e)}_{\mu,\nu} : T_\mu\mathcal{M}_+(I) \longrightarrow T_\nu\mathcal{M}_+(I), \qquad (\mu, a) \longmapsto \big(\nu, \big(\widetilde{\phi}_\nu^{-1} \circ \widetilde{\phi}_\mu\big)(a)\big).$$

Here, one has

$$\big(\widetilde{\phi}_\nu^{-1} \circ \widetilde{\phi}_\mu\big)(a) = \frac{da}{d\mu}\,\nu = \sum_i \nu_i\,\frac{a_i}{\mu_i}\,\delta^i.$$

One immediately observes the following duality of the two parallel transports with respect to the Fisher metric. With $A = (\mu, a)$ and $B = (\nu, b)$:

$$\mathfrak{g}_\nu\big(\widetilde{\Pi}^{(e)}_{\mu,\nu} A, \widetilde{\Pi}^{(m)}_{\mu,\nu} B\big) = \sum_i \frac{1}{\nu_i}\left(\nu_i \frac{a_i}{\mu_i}\right) b_i = \sum_i \frac{1}{\mu_i} a_i b_i = \mathfrak{g}_\mu(A, B). \tag{2.41}$$

The correspondence of tangent spaces can be encoded more effectively in terms of an affine connection, which is a differential version of the parallel transport that specifies the directional derivative of a vector field in the direction of another vector field. To be more precise, let A and B be two vector fields $\mathcal{M}_+(I) \to T\mathcal{M}_+(I)$. There exist maps $a, b : \mathcal{M}_+(I) \to \mathcal{S}(I)$ satisfying $B_\mu = (\mu, b_\mu)$ and $A_\mu = (\mu, a_\mu)$. With a curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}_+(I)$, $\gamma(0) = \mu$ and $\dot\gamma(0) = A_\mu$ the *covariant derivative of B in the direction of A* can be obtained from the parallel transports as follows (see Eq. (B.33) in Appendix B):

$$\widetilde{\nabla}^{(m,e)}_A B\big|_\mu := \lim_{t \to 0} \frac{1}{t}\big(\widetilde{\Pi}^{(m,e)}_{\gamma(t),\mu}(B_{\gamma(t)}) - B_\mu\big) \in T_\mu\mathcal{M}_+(I). \tag{2.42}$$

The pair (2.42) of affine connections $\widetilde{\nabla}^{(m)}$ and $\widetilde{\nabla}^{(e)}$ corresponds to two kinds of straight line, the so-called geodesic, and exponential maps which specify a natural way of locally identifying the tangent space in μ with a neighborhood of μ (in $\mathcal{M}_+(I)$).

Proposition 2.4

(1) *The affine connections $\widetilde{\nabla}^{(m)}$ and $\widetilde{\nabla}^{(e)}$, defined by* (2.42), *are given by*

$$\widetilde{\nabla}^{(m)}_A B\big|_\mu = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu)\right),$$

$$\widetilde{\nabla}^{(e)}_A B\big|_\mu = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu) - \left(\frac{da_\mu}{d\mu} \cdot \frac{db_\mu}{d\mu}\right)\mu\right).$$

(2) *As corresponding* (*maximal*) *m- and e-geodesic with initial point $\mu \in \mathcal{M}_+(I)$ and initial velocity $a \in T_\mu\mathcal{M}_+(I)$ we have*

$$\gamma^{(m)} : \,]t^-, t^+[\, \to \mathcal{M}_+(I), \quad t \mapsto \mu + ta,$$

with

$$t^- := -\min\left\{\frac{\mu_i}{a_i} : i \in I, a_i > 0\right\}, \qquad t^+ := \min\left\{\frac{\mu_i}{|a_i|} : i \in I, a_i < 0\right\}$$

(*we use the convention* $\min \emptyset = \infty$), *and*

$$\gamma^{(e)} : \mathbb{R} \to \mathcal{M}_+(I), \qquad t \mapsto \exp\left(t\,\frac{da}{d\mu}\right)\mu.$$

(3) *As corresponding exponential maps* $\widetilde{\exp}^{(m)}$ *and* $\widetilde{\exp}^{(e)}$, *we obtain*

$$\widetilde{\exp}^{(m)} : T \to \mathcal{M}_+(I), \qquad (\mu, a) \mapsto \mu + a, \tag{2.43}$$

with $T := \{(\mu, \nu - \mu) \in T\mathcal{M}_+(I) : \mu, \nu \in \mathcal{M}_+(I)\}$, *and*

$$\widetilde{\exp}^{(e)} : T\mathcal{M}_+(I) \to \mathcal{M}_+(I), \qquad (\mu, a) \mapsto \exp\left(\frac{da}{d\mu}\right)\mu. \tag{2.44}$$

Proof **The *m*-connection:**

$$\begin{aligned}
\widetilde{\nabla}_A^{(m)} B\big|_\mu &= \lim_{t\to 0} \frac{1}{t}\big(\widetilde{\Pi}^{(m)}_{\gamma(t),\mu}(B_{\gamma(t)}) - B_\mu\big) \\
&= \left(\mu, \lim_{t\to 0}\frac{1}{t}(b_{\gamma(t)} - b_\mu)\right) \\
&= \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu)\right).
\end{aligned}$$

In order to get the geodesic of the m-connection we consider the corresponding equation:

$$\ddot{\gamma} = 0 \quad \text{with } \gamma(0) = \mu,\ \dot{\gamma}(0) = a.$$

Its solution is given by

$$t \mapsto \mu + t\,a$$

which is defined for the maximal time interval $]t^-, t^+[$. Setting $t = 1$ gives us the corresponding exponential map $\widetilde{\exp}^{(m)}$.

The *e*-connection: Now we consider the covariant derivative induced by the exponential parallel transport $\widetilde{\Pi}^{(e)}$:

$$\begin{aligned}
\widetilde{\nabla}_A^{(e)} B\big|_\mu &:= \lim_{t\to 0} \frac{1}{t}\big(\widetilde{\Pi}^{(e)}_{\gamma(t),\mu}(B_{\gamma(t)}) - B_\mu\big) \\
&= \left(\mu, \lim_{t\to 0}\frac{1}{t}\sum_i \left(\mu_i\,\frac{b_{\gamma(t),i}}{\gamma_i(t)} - b_{\mu,i}\right)\delta^i\right) \\
&= \left(\mu, \sum_i \frac{d}{dt}\left\{\mu_i\,\frac{b_{\gamma(t),i}}{\gamma_i(t)}\right\}\bigg|_{t=0}\delta^i\right) \\
&= \left(\mu, \sum_i \left(\frac{\partial b_i}{\partial a_\mu}(\mu) - \frac{1}{\mu_i}\,a_{\mu,i} b_{\mu,i}\right)\delta^i\right).
\end{aligned}$$

The equation for the corresponding e-geodesic is given as

$$\ddot{\gamma} - \frac{\dot{\gamma}^2}{\gamma} = 0 \quad \text{with } \gamma(0) = \mu,\ \dot{\gamma}(0) = a. \tag{2.45}$$

One can easily verify that the solution of (2.45) is given by the following curve γ:

$$t \mapsto \sum_i \mu_i \, e^{t \, \frac{a_i}{\mu_i}} \, \delta^i . \tag{2.46}$$

Setting $t = 1$ in (2.46), we obtain the corresponding exponential map $\widetilde{\exp}^{(e)}$ which is defined on the whole tangent bundle $T\mathcal{M}_+(I)$:

$$(\mu, a) \mapsto \exp\biggl(\frac{da}{d\mu}\biggr)\,\mu = \sum_i \mu_i \, e^{\frac{a_i}{\mu_i}} \, \delta^i .$$

□

In what follows, we restrict the m- and e-connections to the simplex $\mathcal{P}_+(I)$. First consider the m-connection. Given a point $\mu \in \mathcal{P}_+(I)$ and two vector fields $A, B : \mathcal{P}_+(I) \to T\mathcal{P}_+(I)$, we observe that the covariant derivative in μ is already in the tangent space of $\mathcal{P}_+(I)$ in μ, that is, $\widetilde{\nabla}^{(m)}_A B|_\mu \in T_\mu \mathcal{P}_+(I)$. This allows us to define the m-connection on $\mathcal{P}_+(I)$ simply by

$$\nabla^{(m)}_A B\big|_\mu := \widetilde{\nabla}^{(m)}_A B\big|_\mu . \tag{2.47}$$

The situation is different for the e-connection. There, we have in general $\widetilde{\nabla}^{(e)}_A B|_\mu \notin T_\mu \mathcal{P}_+(I)$. In order to obtain an e-connection on the simplex, we have to project $\widetilde{\nabla}^{(e)}_A B|_\mu$ onto $T_\mu \mathcal{P}_+(I)$ with respect to the Fisher metric $\mathfrak{g}_\mu$ in μ, which leads to the following covariant derivative on the simplex (see (2.14)):

$$\nabla^{(e)}_A B\big|_\mu = \biggl(\mu, \frac{\partial b}{\partial a_\mu}(\mu) - \biggl(\frac{da_\mu}{d\mu} \cdot \frac{db_\mu}{d\mu}\biggr)\mu + \mathfrak{g}_\mu(A_\mu, B_\mu)\,\mu\biggr). \tag{2.48}$$

Proposition 2.5 *Consider the affine connections $\nabla^{(m)}$ and $\nabla^{(e)}$ defined by (2.47) and (2.48), respectively. Then the following holds:*

(1) *The corresponding (maximal) m- and e-geodesic with initial point $\mu \in \mathcal{P}_+(I)$ and initial velocity $a \in T_\mu \mathcal{P}_+(I)$ are given by*

$$\gamma^{(m)} : \,]t^-, t^+[\, \to \mathcal{P}_+(I), \quad t \mapsto \mu + t a,$$

with

$$t^- := -\min\biggl\{\frac{\mu_i}{a_i} : i \in I, a_i > 0\biggr\}, \qquad t^+ := \min\biggl\{\frac{\mu_i}{|a_i|} : i \in I, a_i < 0\biggr\},$$

and

$$\gamma^{(e)} : \mathbb{R} \to \mathcal{P}_+(I), \qquad t \mapsto \frac{\exp(t \, \frac{da}{d\mu})}{\mu(\exp(t \, \frac{da}{d\mu}))} \, \mu .$$

(2) *As corresponding exponential maps* $\exp^{(m)}$ *and* $\exp^{(e)}$ *we have*

$$\exp^{(m)} : T \to \mathcal{P}_+(I), \qquad (\mu, a) \mapsto \mu + a,$$

with $T := \{(\mu, \nu - \mu) \in T\mathcal{P}_+(I) : \mu, \nu \in \mathcal{P}_+(I)\}$, *and*

$$\exp^{(e)} : T\mathcal{P}_+(I) \to \mathcal{P}_+(I), \qquad (\mu, a) \mapsto \frac{\exp(\frac{da}{d\mu})}{\mu(\exp(\frac{da}{d\mu}))}\,\mu.$$

Proof Clearly, we only have to prove the statements for the e-connection. From the definition (2.48), we obtain the equation for the corresponding e-geodesic:

$$\ddot{\gamma} - \frac{\dot{\gamma}^2}{\gamma} + \gamma \sum_i \frac{\dot{\gamma}_i^2}{\gamma_i} = 0 \quad \text{with } \gamma(0) = \mu,\ \dot{\gamma}(0) = a. \tag{2.49}$$

The solution of (2.49) is given by the following curve γ:

$$t \mapsto \sum_i \frac{\mu_i\, e^{t\, \frac{a_i}{\mu_i}}}{\sum_j \mu_j\, e^{t\, \frac{a_j}{\mu_j}}}\, \delta^i. \tag{2.50}$$

We now verify this: Obviously, we have $\gamma(0) = \mu$. Furthermore, a straightforward calculation gives us

$$\dot{\gamma}_i(t) = \gamma_i(t)\left(\frac{a_i}{\mu_i} - \sum_j \gamma_j(t)\,\frac{a_j}{\mu_j}\right)$$

and

$$\begin{aligned}\ddot{\gamma}_i(t) &= \dot{\gamma}_i(t)\left(\frac{a_i}{\mu_i} - \sum_j \gamma_j(t)\,\frac{a_j}{\mu_j}\right) - \gamma_i(t)\sum_j \dot{\gamma}_j(t)\,\frac{a_j}{\mu_j} \\ &= \gamma_i(t)\left(\frac{a_i}{\mu_i} - \sum_j \gamma_j(t)\,\frac{a_j}{\mu_j}\right)^2 - \gamma_i(t)\sum_j \dot{\gamma}_j(t)\,\frac{a_j}{\mu_j}.\end{aligned}$$

This implies $\dot{\gamma}(0) = a$ and

$$\begin{aligned}&\ddot{\gamma}_i(t) - \frac{\dot{\gamma}_i(t)^2}{\gamma_i} - \gamma_i(t)\sum_j\left(\ddot{\gamma}_j(t) - \frac{\dot{\gamma}_j(t)^2}{\gamma_j(t)}\right) \\ &\quad = -\gamma_i(t)\sum_j \dot{\gamma}_j(t)\,\frac{a_j}{\mu_j} + \gamma_i(t)\sum_j \gamma_j(t)\sum_k \dot{\gamma}_k(t)\,\frac{a_k}{\mu_k} \\ &\quad = 0,\end{aligned}$$

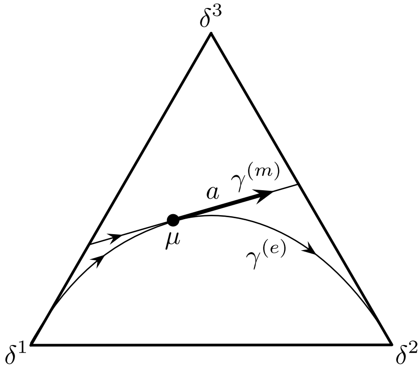

Fig. 2.4 m- and e-geodesic in $\mathcal{P}_+(\{1,2,3\})$ with initial point μ and velocity a

which proves that all conditions (2.49) are satisfied. Setting $t = 1$ in (2.50), we obtain the corresponding exponential map $\exp^{(e)}$ which is defined on the whole tangent bundle $T\mathcal{P}_+(I)$:

$$(\mu, a) \mapsto \frac{\exp(\frac{da}{d\mu})}{\mu(\exp(\frac{da}{d\mu}))}\,\mu = \sum_i \frac{\mu_i e^{\frac{a_i}{\mu_i}}}{\sum_j \mu_j e^{\frac{a_j}{\mu_j}}}\,\delta^i. \qquad \square$$

As an illustration of the m- and e-geodesic of Proposition 2.5(1), see Fig. 2.4.

2.5 The Amari–Chentsov Tensor and the α-Connections

2.5.1 The Amari–Chentsov Tensor

We consider a covariant 3-tensor using the affine connections $\widetilde{\nabla}^{(m)}$ and $\widetilde{\nabla}^{(e)}$: For three vector fields $A : \mu \mapsto A_\mu = (\mu, a_\mu)$, $B : \mu \mapsto B_\mu = (\mu, b_\mu)$, and $C : \mu \mapsto C_\mu = (\mu, c_\mu)$ on $\mathcal{M}_+(I)$, we define

$$\begin{aligned}\mathbf{T}_\mu(A_\mu, B_\mu, C_\mu) &:= \mathfrak{g}_\mu\big(\widetilde{\nabla}^{(m)}_A B\big|_\mu - \widetilde{\nabla}^{(e)}_A B\big|_\mu, C_\mu\big) \\ &= \sum_{i \in I} \mu_i \frac{a_{\mu,i}}{\mu_i}\frac{b_{\mu,i}}{\mu_i}\frac{c_{\mu,i}}{\mu_i}. \end{aligned} \tag{2.51}$$

We refer to this tensor as the *Amari–Chentsov tensor*. Note that for vector fields A, B, C on $\mathcal{P}_+(I)$ and $\mu \in \mathcal{P}_+(I)$ we have

$$\mathbf{T}_\mu(A_\mu, B_\mu, C_\mu) = \mathfrak{g}_\mu\big(\nabla^{(m)}_A B\big|_\mu - \nabla^{(e)}_A B\big|_\mu, C_\mu\big).$$

We have seen that the Fisher metric $\mathfrak{g}$ on $\mathcal{P}_+(I)$ is uniquely characterized in terms of invariance (see Theorem 2.1). Following Chentsov, the same uniqueness property also holds for the tensor $\mathbf{T}$ on $\mathcal{P}_+(I)$, which is the content of the following theorem.

Theorem 2.2 *We assign to each non-empty and finite set I a (non-trivial) covariant 3-tensor S^I on $\mathcal{P}_+(I)$. If for each congruent Markov kernel $K : I \to \mathcal{P}(I')$ we have invariance in the sense that*

$$S^I_\mu(A, B, C) = S^{I'}_{K_*(\mu)}\big(d_\mu K_*(A), d_\mu K_*(B), d_\mu K_*(C)\big)$$

then there is a constant $\alpha > 0$ such that $S^I = \alpha\, \mathbf{T}^I$ for all I, where $\mathbf{T}^I$ denotes the Amari–Chentsov tensor on $\mathcal{P}_+(I)$.[2]

One can prove this theorem by following the same steps as in the proof of Theorem 2.1. Alternatively, it immediately follows from the more general result stated in Theorem 2.3.

By analogy, we can extend the definition (2.51) to a covariant n-tensor for all $n \geq 1$:

$$\begin{aligned} \tau^n_\mu\big(V^{(1)}, V^{(2)}, \dots, V^{(n)}\big) &:= \sum_{i\in I} \mu_i \, \frac{v^{(1)}_{\mu,i}}{\mu_i} \frac{v^{(2)}_{\mu,i}}{\mu_i} \cdots \frac{v^{(n)}_{\mu,i}}{\mu_i} \\ &= \sum_{i\in I} \frac{1}{{\mu_i}^{n-1}} \, v^{(1)}_{\mu,i}\, v^{(2)}_{\mu,i} \cdots v^{(n)}_{\mu,i}. \end{aligned} \tag{2.52}$$

Obviously, we have

$$\tau^2 = \mathfrak{g}, \quad \text{and} \quad \tau^3 = \mathbf{T}.$$

It is easy to extend the representation (2.26) of the Fisher metric $\mathfrak{g}$ to the covariant n-tensor τ^n. Given a differentiable manifold M and an embedding $p : M \hookrightarrow \mathcal{M}_+(I)$, one obtains as pullback of τ^n the following covariant n-tensor, defined on M:

$$\tau^n_\xi(V_1, \dots, V_n) := \sum_i p_i(\xi) \frac{\partial \log p_i}{\partial V_1}(\xi) \cdots \frac{\partial \log p_i}{\partial V_n}(\xi).$$

As suggested by (2.52), the tensor τ^n is closely related to the following multilinear form:

$$\begin{aligned} &L^n_I : \underbrace{\mathcal{F}(I) \times \cdots \times \mathcal{F}(I)}_{n \text{ times}} \to \mathbb{R}, \\ &(f_1, \dots, f_n) \mapsto L^n_I(f_1, \dots, f_n) := \sum_i {f_1}^i \cdots {f_n}^i. \end{aligned} \tag{2.53}$$

In order to see this, consider the map

$$\pi^{1/n} : \mathcal{M}_+(I) \to \mathcal{F}(I), \qquad \mu = \sum_i \mu_i\, \delta^i \mapsto \pi^{1/n}(\mu) := n \sum_i {\mu_i}^{\frac{1}{n}}\, e_i.$$

[2]Note that we use the abbreviation $\mathbf{T}$ if corresponding statements are clear without reference to the set I, which is usually the case throughout this book.

This implies

$$\begin{aligned} L_I^n\left(\frac{\partial \pi^{1/n}}{\partial v^{(1)}}(\mu), \ldots, \frac{\partial \pi^{1/n}}{\partial v^{(n)}}(\mu)\right) &= \sum_i \bigl(\mu_i^{-\frac{n-1}{n}} v_i^{(1)}\bigr) \cdots \bigl(\mu_i^{-\frac{n-1}{n}} v_i^{(n)}\bigr) \\ &= \tau_\mu^n\bigl(V^{(1)}, \ldots, V^{(n)}\bigr). \end{aligned}$$

This proves that the tensor τ^n is nothing but the $\pi^{1/n}$-pullback of the multi-linear form L_I^n. In this sense, it is a very natural tensor. Furthermore, for $n = 2$ and $n = 3$, we have seen that the restrictions of $\mathfrak{g}$ and $\mathbf{T}$ to the simplex $\mathcal{P}_+(I)$ are naturally characterized in terms of their invariance with respect to congruent Markov embeddings (see Theorem 2.1 and Theorem 2.2). This raises the question whether the tensors τ^n on $\mathcal{M}_+(I)$, or their restrictions to $\mathcal{P}_+(I)$, are also characterized by invariance properties. It is easy to see that for all n, τ^n is indeed invariant. However, τ^n are not the only invariant tensors. In fact, Chentsov's results treat the only nontrivial uniqueness cases. Already for $n = 2$, Campbell has shown that the metric $\mathfrak{g}$ is not the only one that is invariant if we consider tensors on $\mathcal{M}_+(I)$ rather than on $\mathcal{P}_+(I)$ [57]. Furthermore, for higher n, there are other possible invariant tensors, even when restricting to $\mathcal{P}_+(I)$. For instance, for $n = 4$ we can consider the tensors

$$\begin{aligned} \tau^{\{1,2\},\{3,4\}}(V_1, V_2, V_3, V_4) &:= \tau^2(V_1, V_2)\,\tau^2(V_3, V_4) = \mathfrak{g}(V_1, V_2)\,\mathfrak{g}(V_3, V_4), \\ \tau^{\{1,3\},\{2,4\}}(V_1, V_2, V_3, V_4) &:= \tau^2(V_1, V_3)\,\tau^2(V_2, V_4) = \mathfrak{g}(V_1, V_3)\,\mathfrak{g}(V_2, V_4), \\ \tau^{\{1,4\},\{2,3\}}(V_1, V_2, V_3, V_4) &:= \tau^2(V_1, V_4)\,\tau^2(V_2, V_3) = \mathfrak{g}(V_1, V_4)\,\mathfrak{g}(V_2, V_3). \end{aligned}$$

It is obvious that all of these invariant tensors are mutually different and also different from τ^4. Similarly, for $n = 5$ we have, for example,

$$\begin{aligned} \tau^{\{1,2\},\{3,4,5\}}(V_1, V_2, V_3, V_4, V_5) &:= \tau^2(V_1, V_2)\,\tau^3(V_3, V_4, V_5) \\ &= \mathfrak{g}(V_1, V_2)\,\mathbf{T}(V_3, V_4, V_5), \\ \tau^{\{1,4\},\{2,3,5\}}(V_1, V_2, V_3, V_4, V_5) &:= \tau^2(V_1, V_4)\,\tau^3(V_2, V_3, V_5) \\ &= \mathfrak{g}(V_1, V_4)\,\mathbf{T}(V_2, V_3, V_5). \end{aligned}$$

From these examples it becomes evident that for each partition

$$\mathbf{P} = \bigl\{\bigl\{i_1^1, \ldots, i_1^{n_1}\bigr\}, \ldots, \bigl\{i_l^1, \ldots, i_l^{n_l}\bigr\}\bigr\}$$

of the set $\{1, \ldots, n\}$ with $n = n_1 + \cdots + n_l$ one can define an invariant n-tensor $\tau^{\mathbf{P}}(V_1, \ldots, V_n)$ in a corresponding fashion, see Definition 2.6 below. Our generalization of Chentsov's uniqueness results, Theorem 2.3, will state that any invariant n-tensor will be a linear combination of these, i.e., the dimension of the space of invariant n-tensors depends on the number of partitions of the set $\{1, \ldots, n\}$. In fact, this result will even hold if we consider arbitrary (infinite) measure spaces (see Theorem 5.6).

http://www.springer.com/978-3-319-56477-7

Information Geometry
Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L.
2017, XI, 407 p. 15 illus., Hardcover
ISBN: 978-3-319-56477-7