



Autoparallelity of quantum statistical manifolds in light of quantum estimation theory

Hiroshi Nagaoka¹ · Akio Fujiwara^{2,3}

Received: 2 October 2025 / Revised: 8 August 2026 / Accepted: 11 August 2026
© The Author(s) 2026

Abstract

In this paper we study the autoparallelity w.r.t. the e-connection for an information-geometric structure called the SLD structure, which consists of a Riemannian metric and mutually dual e- and m-connections, induced on the manifold of strictly positive density operators. Unlike the classical information geometry, the e-connection has non-vanishing torsion, which brings various mathematical difficulties. The notion of e-autoparallel submanifolds is regarded as a quantum version of exponential families in classical statistics, which is known to be characterized as statistical models having efficient estimators (unbiased estimators uniformly achieving the equality in the Cramér-Rao inequality). As quantum extensions of this classical result, we present two different forms of estimation-theoretical characterizations of the e-autoparallel submanifolds. We also give several results on the e-autoparallelity, some of which are valid for the autoparallelity w.r.t. an affine connection in a more general geometric situation.

Keywords Quantum estimation theory · Information geometry · Autoparallel submanifold · Dual connection · Torsion · SLD (symmetric logarithmic derivative)

Communicated by Jan Naudts.

✉ Hiroshi Nagaoka
nagaoka@is.uec.ac.jp

Akio Fujiwara
fujiwara@math.sci.osaka-u.ac.jp

¹ The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan

² Department of Mathematics, The University of Osaka, Toyonaka, Osaka 560-0043, Japan

³ Center for Quantum Information and Quantum Biology, The University of Osaka, Toyonaka, Osaka 560-0043, Japan

1 Introduction

Information geometry has revealed that many fundamental structures in classical statistics, most notably exponential families and efficient estimation, are naturally encoded in the geometry of the statistical manifold, with certain “straight” submanifolds singled out by their optimal estimation properties. In quantum statistics, the space of density operators carries a rich family of information-geometric structures, among which the SLD (symmetric logarithmic derivative) structure is distinguished by its close connection to quantum Cramér–Rao bounds, but the corresponding geometric picture is far less understood than in the classical, dually flat setting. The present paper investigates how this geometric notion of “straightness” extends to quantum statistical models via autoparallel submanifolds with respect to the SLD structure, and how such submanifolds can be characterized by estimation-theoretic conditions weaker than global efficiency as well as by purely geometric existence conditions for a non-flat, torsionful connection.

To clarify the motivation for our study, we begin with an overview of a result in the classical estimation theory. Let $\mathcal{P} = \mathcal{P}(\Omega)$ be the totality of strictly positive probability distributions (probability mass functions) on a finite set Ω , and let $\mathcal{M} = \{p_\xi \mid \xi = (\xi^1, \dots, \xi^n) \in \Xi\} \subset \mathcal{P}$ be a statistical model whose elements p_ξ are smoothly and injectively parametrized by an n -dimensional parameter ξ ranging over an open subset Ξ of $\mathbb{R}^n$. As is well known, the variance matrix $V_\xi \in \mathbb{R}^{n \times n}$ of an arbitrary unbiased estimator for the parameter ξ satisfies the Cramér–Rao inequality $V_\xi \geq G_\xi^{-1}$, where $G_\xi \in \mathbb{R}^{n \times n}$ denotes the Fisher information matrix. An unbiased estimator achieving the equality $V_\xi = G_\xi^{-1}$ for every $\xi \in \Xi$ is called an *efficient estimator* for the parameter ξ . A statistical model does not have an efficient estimator in general. Indeed, the existence of efficient estimator is known to impose strong restrictions on both the set $\mathcal{M}$ and the parametrization $\xi \mapsto p_\xi$. Namely, we have the following theorem (e.g. § 5a.2 (p.324) of [1], Eq. (7.14) in § 7.2 of [2], Theorem 3.12 of [3]).

Theorem 1.1 *For a statistical model $\mathcal{M} = \{p_\xi\}$, the following conditions are equivalent.*

- (i) *There exists an efficient estimator for the parameter ξ .*
- (ii) *$\mathcal{M}$ is an exponential family, and ξ is an expectation parameter; i.e., there exist functions $C(\omega)$, $F^i(\omega)$, $\theta_i(\xi)$, $\psi(\xi)$ ($i = 1, \dots, n$) such that the elements of $\mathcal{M}$ are represented as*

$$p_\xi(\omega) = \exp[C(\omega) + \sum_{i=1}^n \theta_i(\xi) F^i(\omega) - \psi(\xi)] \quad (1.1)$$

and that the parameter ξ satisfies

$$\xi^i = E_\xi(F^i), \quad (1.2)$$

where E_ξ denotes the expectation w.r.t. the distribution p_ξ .

Remark 1.2 In the notation of [3] as well as of many references in information geometry, θ_i , F^i and ξ^i in (1.1) and (1.2) are expressed as θ^i , F_i and η_i , respectively. The reason why the upper and lower indices are reversed here (and throughout the paper) is that we first treat ξ as an arbitrary coordinate system, which has an upper index (superscript) as in the standard notation of differential geometry, and then consider the condition for ξ to become an m-affine coordinate system such as the expectation coordinate system.

What is important about Theorem 1.1 for the main concern of this paper is that condition (ii) is expressed in the language of geometry as follows:

- (ii)' $\mathcal{M}$ is an e-autoparallel submanifold of $\mathcal{P}$, and ξ forms an m-affine coordinate system of $\mathcal{M}$.

We refer the reader to Theorem 3.12 of [3] and the later sections of this paper for details. For the present, it suffices to understand only the following points.

- The autoparallelity is a multi-dimensional version (including the 1-dimensional case in particular) of the notion of geodesic for a manifold equipped with an affine connection, which expresses the condition that a submanifold is “straight” according to the criterion defined by the connection.
- The e-autoparallelity means the autoparallelity w.r.t. the e-connection of $\mathcal{P}$ denoted by $\nabla^{(e)}$, which is one of the most important differential-geometric structures in information geometry.
- The manifold $\mathcal{P}$ has two other important structures: an affine connection called the m-connection $\nabla^{(m)}$ and a Riemannian metric called the Fisher metric g . The triple $(g, \nabla^{(e)}, \nabla^{(m)})$ makes $\mathcal{P}$ into what is called a dually flat space.
- The parameter ξ is said to be m-affine when it is regarded as an affine coordinate system of $\mathcal{M}$ w.r.t. the affine structure induced from the m-connection of $\mathcal{M}$ that is induced from $\nabla^{(m)}$ and g .

The equivalence (i) $\Leftrightarrow$ (ii)' is regarded as one of the most fundamental results in information geometry, revealing a beautiful harmony between estimation theory and geometry.

Let us now turn to the quantum case and observe how the harmony that prevailed in the classical case breaks down. We first note that a quantum version of Theorem 1.1 is actually known, which is stated below. Let $\mathcal{H}$ be a finite-dimensional Hilbert space, $\mathcal{S} = \mathcal{S}(\mathcal{H})$ be the totality of strictly positive density operators on $\mathcal{H}$, and $\mathcal{M} = \{\rho_\xi \mid \xi \in \Xi\}$ be an arbitrary quantum statistical model consisting of states ρ_ξ in $\mathcal{S}$. It is well known that the variance matrix V_ξ of an arbitrary unbiased estimator for the parameter ξ satisfies the SLD Cramér-Rao inequality $V_\xi \geq G_\xi^{-1}$ [4, 5], where G_ξ is the SLD Fisher information matrix (see Sect. 4 for details). When an unbiased estimator satisfies $V_\xi = G_\xi^{-1}$ for every $\xi \in \Xi$, we call it an efficient estimator for the parameter ξ as in the classical case. Then we have the following theorem (Theorem 7.6 of [3]).

Theorem 1.3 *For a quantum statistical model $\mathcal{M} = \{\rho_\xi\}$, the following conditions are equivalent.*

- (i) *There exists an efficient estimator for the parameter ξ .*
- (ii) *There exist mutually commuting Hermitian operators $F^1, \dots, F^n$ and a strictly positive operator P such that the elements of $\mathcal{M}$ are represented as*

$$\rho_\xi = \exp\left[\frac{1}{2}\left\{\sum_{i=1}^n \theta_i(\xi) F^i - \psi(\xi)\right\}\right] P \exp\left[\frac{1}{2}\left\{\sum_{i=1}^n \theta_i(\xi) F^i - \psi(\xi)\right\}\right] \quad (1.3)$$

and that

$$\forall \xi \in \Xi, \forall i, \quad \xi^i = \text{Tr}(\rho_\xi F^i). \quad (1.4)$$

(When a model $\mathcal{M} = \{\rho_\xi\}$ is represented as (1.3), we call it a quasi-classical exponential family.)

Compared with Theorem 1.1, this result is insufficient in several points except when $n = 1$.

From an estimation-theoretical point of view, although the primary interest in the theory of multi-parameter quantum estimation lies in the mathematical difficulties arising from the noncommutativity of operators, which often limit the significance of the SLD Cramér-Rao inequality itself, the above definition of efficient estimator considers only the highly special case in which these difficulties do not arise.

From a geometric point of view, on the other hand, condition (ii) is somewhat unsatisfactory in that it cannot be expressed as a simple geometric condition like condition (ii)' in the classical case. More specifically, in terms of the triple $(g, \nabla^{(e)}, \nabla^{(m)})$ of the Fisher metric and the e, m-connections defined by the SLD structure, which is among a family of information-geometric structures introduced on $\mathcal{S}$, conditions (ii) implies that $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$ and that ξ forms an m-affine coordinate system, but the converse does not hold unless $n = 1$ (see Section 7.4 of [3] and Sect. 4 of this paper). Namely, a quasi-classical exponential family is only a special case of an e-autoparallel submanifold, which means that the existence of an efficient estimator is too strong as a characterization of the e-autoparallelity.

Noting that Theorem 1.3 may be extended in several ways, we formulate in this paper the problem of determining an estimation-theoretic condition, which should be a weaker version of condition (i), equivalent to the geometric condition obtained by replacing (ii) with one of the same form as condition (ii)' in the classical case. In other words, the problem asks how the notion of efficient estimator should be extended in order for the e-autoparallelity to be characterized by its existence. We give two answers to the problem in Sects. 5 and 7, which constitute the highlights of this paper. In addition, we provide a detailed discussion of the properties of autoparallel submanifolds and the conditions for their existence. This is because, unlike in the classical case where the e-connection is flat, the e-connection in the quantum case is not generally flat due to the non-vanishing torsion, making new insights into autoparallel submanifolds essential.

Let us outline the paper by following its organization.

In Sect. 2, we explain the basic issues about the autoparallelity for an affine connection ∇ on a general differential manifold, focusing in particular on a curvature-free ∇ with its relation to the dual connection ∇^* w.r.t. a Riemannian metric g . Note that a connection is flat if and only if it is curvature-free and torsion-free and that, even though both ∇ and ∇^* are curvature-free in this case, they may have non-vanishing torsions and hence (g, ∇, ∇^*) may not define a dually flat structure. In Sect. 3, we introduce a family of information-geometric structures on the space $\mathcal{S}(\mathcal{H})$, where the e-connection has mostly a non-vanishing torsion, and derive the basic issues concerning the e-autoparallelity by applying the results of Sect. 2. These two sections are preliminaries for later sections. Although the results shown there are mostly known, we present them together with their derivations so that the description is as self-contained as possible.

In Sect. 4, preliminaries concerning quantum estimation are given. We give the definitions of (local-)unbiasedness and several kinds of (local-)efficiency with or without weight matrix, where the word “local” means the versions of these notions weakened by localizing them at a given point of the model. Essential for later arguments is Proposition 4.1, which states that the equality in the scalar version of SLD Cramér-Rao inequality $\mathbf{u}^T V_\xi \mathbf{u} \geq \mathbf{u}^T G_\xi^{-1} \mathbf{u}$ is attained as an infimum for each point ξ and each direction represented by a vector $\mathbf{u}$. Although the result itself is not new (e.g. Eq. (7.93) of [3] and Theorem 5.2 of [6]), considering its importance we provide it with a proof and a remark that the infimum is not generally replaced with the minimum even locally at each point. We introduce a family of randomized estimators, which is used for the proof of this proposition as well as for a main result of the paper presented in the next section. We revisit Theorem 1.3 and clarify that the existence of efficient estimator only partially characterizes the e-autoparallelity for the SLD structure.

In Sect. 5, we treat a one-parameter family of estimators $\tilde{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$, which we call a *filtration* of estimators, instead of a single estimator Π . We show Theorem 5.1, which gives the first answer to our problem, stating that a quantum statistical model is e-autoparallel in $\mathcal{S}$ if and only if for every direction $\mathbf{u}$, there exists a filtration $\tilde{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$ such that Π_ε is an unbiased estimator for every $\varepsilon > 0$ and that the equality in $\mathbf{u}^T V_\xi \mathbf{u} \geq \mathbf{u}^T G_\xi^{-1} \mathbf{u}$ is attained for every point ξ as the limit of $\varepsilon \downarrow 0$. (The condition imposed on the parametrization is omitted here for brevity.) This theorem can be regarded as an extension of Theorem 1.3 since an efficient estimator Π defines such a filtration by $\Pi_\varepsilon = \Pi$ ($\forall \varepsilon > 0$), independently of $\mathbf{u}$. In Sect. 6, it is shown that a quantum Gaussian shift model has such a filtration for every direction $\mathbf{u}$ and that the model in fact exhibits an analogous property to the e-autoparallelity in $\mathcal{S}(\mathcal{H})$, although the Hilbert space $\mathcal{H}$ is infinite-dimensional in this case, so that we cannot fully develop a differential geometric argument there.

While our concern in the previous sections lies in estimation of a multi-dimensional parameter $\xi = (\xi^i)$ that injectively represents the states ρ_ξ in a model even though the estimation error is measured only in one direction expressed by $\mathbf{u}$, Sect. 7 studies estimation of a scalar-valued function on the model, for which the efficiency of an estimator is defined in a similar way. This includes estimation of a single component ξ^i of the parameter ξ as a special case, which can be considered equivalent to a problem of estimation with nuisance parameters (e.g. [6]). Theorems 7.9 and 7.10 give the second answer to our question, where it is shown that an n -dimensional quantum statistical

model is e-autoparallel in $\mathcal{S}$ if and only if every component ξ^i of the parameter ξ has an efficient estimator. Since an efficient estimator Π for ξ induces an efficient estimator for each component ξ^i in a trivial way, this result can also be regarded as an extension of Theorem 1.3, although the set of efficient estimators for components induced from Π is not conceptually identified with Π itself. In contrast to the discussion in Sect. 5, the approach of this section does not involve the technical difficulties inherent in estimation of a multi-dimensional parameter of quantum states. We show that some of the estimation-theoretic results obtained in this section can be refined within a purely differential-geometric framework, where the use of differential forms proves to be particularly useful.

In Sect. 8, we move away from estimation theory and address the condition for existence of e-autoparallel submanifolds from a purely geometrical point of view. It should be noted that the e-connection in the SLD structure is curvature-free but is not flat due to the non-vanishing torsion. In the case of a flat connection, an autoparallel submanifold corresponds to an affine subspace in the coordinate space of an affine coordinate system, so that the existence condition for autoparallel submanifolds is trivial. For a non-flat connection, on the other hand, we cannot see the whole picture of autoparallel submanifolds in general, and are faced with the new problem of what kind of condition ensures the existence of autoparallel submanifold. At present, our understanding of this problem remains quite limited even for a curvature-free connection. We focus on the fact that an autoparallel submanifold is an integral manifold of a parallel distribution of tangent vectors, and investigate the involutivity of parallel distributions in relation to the torsion tensor. We show that, even when ∇ has a non-vanishing torsion, there is a case in which there exists a pair of dual foliations consisting of mutually orthogonal ∇ -autoparallel submanifolds and ∇^* -autoparallel submanifolds, as in the case of dually flat spaces.

In Sect. 9, we treat the case when $\dim \mathcal{H} = 2$ and study the SLD structure of the space $\mathcal{S}(\mathcal{H})$ of qubit states in detail. We show that $\mathcal{S}(\mathcal{H})$ in this case has a characteristic property that every e-parallel distribution is involutive, which means that the existence condition for autoparallel submanifolds is trivial as in the classical case. We also give several remarks on our randomized estimator for the qubit case, where its relation to some related results in quantum estimation theory is clarified.

Section 10 is devoted to concluding remarks. Some proofs and additional results are included in Appendices for the sake of readability of the main text.

Remark 1.4 We make some remarks on the nomenclature and the notation of the paper.

1. Throughout the paper, when we refer to a manifold, say M , it means that M is a manifold with a trivial global structure, so that we need not worry about the difference between global properties and local properties of M . For instance, M is always supposed to have a global coordinate system, and every closed differential form on M is considered to be exact.
2. When we say that $\xi = (\xi^1, \dots, \xi^n)$ is a coordinate system of a manifold M in the subsequent sections, it basically means that ξ is a map $: M \rightarrow \mathbb{R}^n$ (a global chart of M) which represents each point $p \in M$ by an n -dimensional vector $\xi(p) = (\xi^1(p), \dots, \xi^n(p)) \in \mathbb{R}^n$, although the same symbol ξ has appeared

above as a parameter to specify a point in the manifold. They are equivalent by $\xi(p) = \xi' \Leftrightarrow p = p_{\xi'}$. A parametrization is often more convenient than a coordinate system when dealing with concrete examples. In fact, we will use parametrizations in Sect. 6 for quantum Gaussian states and in Sect. 9 for qubit states.

3. This paper contains both arguments on quantum statistical manifolds (manifolds consisting of density operators) and those on general manifolds. We denote quantum statistical manifolds by $\mathcal{S}, \mathcal{M}, \mathcal{N}, \dots$, while general manifolds are denoted by $S, M, N, \dots$.

2 Basic issues about autoparallelity

In this section we summarize basic issues related to autoparallelity from the perspective of general differential geometry, which will be necessary for later discussions.

Let S be an arbitrary manifold, and denote the totality of smooth functions and that of smooth vector fields on S by $\mathcal{F}(S)$ and $\mathcal{X}(S)$, respectively. Suppose that S is provided with an affine connection ∇ , which is a map $\mathcal{X}(S) \times \mathcal{X}(S) \rightarrow \mathcal{X}(S)$, $(X, Y) \mapsto \nabla_X Y$. Given a submanifold M of S , let $\mathcal{X}(S/M)$ denote the totality of smooth mappings which map each point $p \in M$ to a tangent vector in $T_p(S)$, i.e., sections of the vector bundle $\bigsqcup_{p \in M} T_p(S)$. Then ∇ naturally induces a map $\mathcal{X}(M) \times \mathcal{X}(S/M) \rightarrow \mathcal{X}(S/M)$ so that for any $X \in \mathcal{X}(M)$ and any $Y \in \mathcal{X}(S/M)$, $\nabla_X Y$ is defined as an element of $\mathcal{X}(S/M)$. Since $\mathcal{X}(M) = \mathcal{X}(M/M) \subset \mathcal{X}(S/M)$, $\nabla_X Y \in \mathcal{X}(S/M)$ is defined for any $X, Y \in \mathcal{X}(M)$, although it does not necessarily belong to $\mathcal{X}(M)$.

When $\nabla_X Y$ belongs to $\mathcal{X}(M)$ for every $X, Y \in \mathcal{X}(M)$, M is said to be *autoparallel* w.r.t. ∇ or ∇ -*autoparallel* in S (e.g., Sect. 8 in Chap. VII of [7]). In particular, S itself is ∇ -autoparallel in S . An autoparallel curve is usually called a *geodesic* (or *pregeodesic* when we wish to clarify that our interest lies only in the image of the curve), so that the autoparallelity is a multi-dimensional extension of the notion of geodesic. When M is ∇ -autoparallel in S , ∇ defines an affine connection on M . We denote this connection by $\nabla|_M$ when we wish to distinguish it from the original connection ∇ on S . The autoparallelity is transitive in the sense that if M is ∇ -autoparallel in S and N is $\nabla|_M$ -autoparallel in M , then N is ∇ -autoparallel in S .

Remark 2.1 If M is ∇ -autoparallel in S and N is a nonempty open set of M , then N is also ∇ -autoparallel in S having the same dimension as M . In this paper, we restrict ourselves to maximal autoparallel submanifolds to avoid this ambiguity. In particular, an autoparallel submanifold of S having the same dimension as S is considered to be only S . This restriction is solely for simplicity of exposition.

Remark 2.2 A notion similar to, but different from, autoparallelity is total geodesicity. A submanifold M is said to be *totally geodesic* w.r.t. ∇ or ∇ -*totally geodesic* in S when for any point $p \in M$ and any tangent vector $X_p \in T_p(M)$ of M , the ∇ -geodesic passing through p in direction X_p lies in M . It is obvious that the autoparallelity implies the total geodesicity, but the converse is not true in general except when ∇ is torsion-free ([7], Theorem 8.4 in Chap. VII). We will revisit this topic in Remark 8.4.

A vector field $X \in \mathcal{X}(S)$ is said to be *parallel w.r.t. ∇* or ∇ -parallel when $\forall Y \in \mathcal{X}(S)$, $\nabla_Y X = 0$. More generally, $X \in \mathcal{X}(S/M)$ (including the case $X \in \mathcal{X}(M)$) is said to be ∇ -parallel when $\forall Y \in \mathcal{X}(M)$, $\nabla_Y X = 0$.

When there exist on S the same number of linearly independent ∇ -parallel vector fields as $\dim S$, we say that (S, ∇) is *curvature-free*. This condition is known to be equivalent to the curvature tensor of ∇ vanishing on S (: recall that S is assumed to be globally trivial). When (S, ∇) is curvature-free, the parallel transport $\Phi_{p,q}^{(\nabla)} : T_p(S) \rightarrow T_q(S)$ is defined for arbitrary two points $p, q \in S$ so that a vector field $X \in \mathcal{X}(S)$ is ∇ -parallel iff $\forall p, q \in S$, $\Phi_{p,q}^{(\nabla)}(X_p) = X_q$.

The following two propositions are straightforward, where the curvature-freeness is essential.

Proposition 2.3 *Let S be a manifold on which a curvature-free connection ∇ is given and M be a submanifold of S . For $X \in \mathcal{X}(S/M)$ (including the case when $X \in \mathcal{X}(M)$), the following conditions are equivalent.*

- (i) X is ∇ -parallel.
- (ii) $\exists \tilde{X} \in \mathcal{X}(S)$, $\tilde{X}$ is ∇ -parallel and $X = \tilde{X}|_M$.

Proposition 2.4 *Let S be a manifold on which a curvature-free connection ∇ is given and M be an n -dimensional submanifold of S . Then the following conditions are equivalent.*

- (i) M is ∇ -autoparallel in S .
- (ii) $\forall p, q \in M$, $\Phi_{p,q}^{(\nabla)}(T_p(M)) = T_q(M)$.
- (iii) There exist n linearly independent ∇ -parallel vector fields on M .
- (iv) There exist n linearly independent ∇ -parallel vector fields $\tilde{X}^1, \dots, \tilde{X}^n$ on S such that $\forall i$, $\tilde{X}^i|_M \in \mathcal{X}(M)$.

When these conditions hold, a vector field X on M is $\nabla|_M$ -parallel iff it is ∇ -parallel, and $\nabla|_M$ is curvature-free.

In the following, we consider the case when (S, ∇) is additionally provided with a Riemannian metric g for which the dual connection ∇^* of ∇ is defined. Namely, the triple (g, ∇, ∇^*) satisfies the duality [3, 8]:

$$\nabla_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X^* Z \rangle. \quad (2.1)$$

As is well known (see Theorems 3.2 and 3.3 of [3]), the curvature-freeness of ∇ and that of ∇^* are equivalent, and when they are curvature-free, their parallel transports satisfy

$$\nabla_{X_p} Y_p \in T_p(S), \quad g_p(X_p, Y_p) = g_q(\Phi_{p,q}^{(\nabla)}(X_p), \Phi_{p,q}^{(\nabla^*)}(Y_p)). \quad (2.2)$$

Proposition 2.5 *When ∇ and ∇^* are curvature-free in (S, g, ∇, ∇^*) , we have:*

- (1) For a vector field $X \in \mathcal{X}(S)$, X is ∇ -parallel iff $g(X, Y)$ is constant on S for every ∇^* -parallel vector field $Y \in \mathcal{X}(S)$.
- (2) For a vector field $X \in \mathcal{X}(S)$, X is ∇^* -parallel iff $g(X, Y)$ is constant on S for every ∇ -parallel vector field $Y \in \mathcal{X}(S)$.

Proof It suffices to show (1). For an $X \in \mathcal{X}(S)$, we have

$$\begin{aligned} X \text{ is } \nabla\text{-parallel} &\Leftrightarrow \forall Y, \forall Z, \quad g(\nabla_Z X, Y) = 0 \\ &\stackrel{a}{\Leftrightarrow} \forall Y : \nabla^*\text{-parallel}, \forall Z, \quad g(\nabla_Z X, Y) = 0 \\ &\stackrel{b}{\Leftrightarrow} \forall Y : \nabla^*\text{-parallel}, \forall Z, \quad Zg(X, Y) = 0 \\ &\Leftrightarrow \forall Y : \nabla^*\text{-parallel}, \quad g(X, Y) \text{ is constant,} \end{aligned} \quad (2.3)$$

where $\stackrel{a}{\Leftrightarrow}$ follows since the set of ∇^* -parallel vector fields has the same dimension as $\dim S$ due to the curvature-freeness of ∇^* , and $\stackrel{b}{\Leftrightarrow}$ follows since the duality (2.1) and the ∇^* -parallelity of Y imply

$$Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z^* Y) = g(\nabla_Z X, Y). \quad (2.4)$$

□

In order to state the next proposition, which will be of fundamental importance for later arguments, we need some preliminaries. First, suppose that M is a ∇ -autoparallel submanifold of S , so that M is a Riemannian submanifold of (S, g) equipped with the affine connection $\nabla|_M$. Then the dual connection of $\nabla|_M$ w.r.t. g (more precisely, w.r.t. the induced metric $g|_M$ on M) is defined, which we denote by $\nabla_M^* := (\nabla|_M)^*$; i.e.,

$$\forall X, Y, Z \in \mathcal{X}(M), \quad Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, (\nabla_M^*)_X Z), \quad (2.5)$$

where we have applied $(\nabla|_M)_X Y = \nabla_X Y$. From (2.1) and (2.5), we have

$$\forall X, Y, Z \in \mathcal{X}(M), \quad g(Y, (\nabla_M^*)_X Z) = g(Y, \nabla_X^* Z), \quad (2.6)$$

which means that $(\nabla_M^*)_X Z$ is the g -projection of $\nabla_X^* Z \in \mathcal{X}(S/M)$ onto $\mathcal{X}(M)$.

$$\begin{array}{ccc} \nabla & \xrightarrow{\text{restriction}} & \nabla|_M \\ g\text{-dual} \downarrow & & \downarrow g|_M\text{-dual} \\ \nabla^* & \xrightarrow{g\text{-projection}} & \nabla_M^* \end{array}$$

Next, we introduce the notions of flat connection and affine coordinate system. An affine connection ∇ is said to be *flat* when it is curvature-free and torsion-free. The flatness is known to be equivalent to the existence of a (generally local) coordinate system $\xi = (\xi^i)$, which is called an *affine coordinate system w.r.t. ∇* or a *∇ -affine coordinate system*, such that the vector fields $\partial_i = \frac{\partial}{\partial \xi^i}$, $i \in \{1, \dots, \dim S\}$, are all ∇ -parallel.

Proposition 2.6 *Suppose that ∇ and ∇^* are curvature-free in (S, g, ∇, ∇^*) , and let M be an n -dimensional submanifold of S with a coordinate system $\xi = (\xi^i)$. Then the following conditions are equivalent.*

- (i) M is ∇ -autoparallel in S , and ξ is a ∇_M^* -affine coordinate system (so that ∇_M^* is flat).
(ii) For every $i \in \{1, \dots, n\}$, the vector field

$$X^i := \sum_j g^{ij} \partial_j \in \mathcal{X}(M) \quad (2.7)$$

is ∇ -parallel, where $\partial_i := \frac{\partial}{\partial \xi^i}$ and $[g^{ij}] := [g_{ij} := g(\partial_i, \partial_j)]^{-1}$.

Proof We first show (i) $\Rightarrow$ (ii). Assume (i) and define $X^i \in \mathcal{X}(M)$ by (2.7). Then we have $g(X^i, \partial_j) = \delta_j^i$ for every i, j , which is constant on M . Noting that $n(= \dim M)$ vector fields $\{\partial_j\}$ are ∇_M^* -parallel and that Proposition 2.5 can be applied to $(M, g|_M, \nabla|_M, \nabla_M^*)$ due to the ∇ -autoparallelity of M , it follows from item (1) of Proposition 2.5 that X^i is $\nabla|_M$ -parallel, and hence it is ∇ -parallel (: recall the last sentence of Proposition 2.4).

We next show (ii) $\Rightarrow$ (i). Assume (ii). Then according to Proposition 2.4, M is ∇ -autoparallel in S . Noting that $g(X^i, \partial_j) = \delta_j^i$ is constant on M and applying item (2) of Proposition 2.5 to $(M, g|_M, \nabla|_M, \nabla_M^*)$, we have that $\{\partial_j\}$ are ∇_M^* -parallel, which means that ξ is ∇_M^* -affine. $\square$

Remark 2.7 In applying this proposition to later arguments, ∇^* will appear as the flat connection $\nabla^{(m)}$. When ∇^* is flat, its torsion-freeness is inherited by ∇_M^* , so that ∇_M^* turns out to be flat for every ∇ -autoparallel M . That is, the existence of ∇_M^* -affine coordinate system ξ follows from the ∇ -autoparallelity of M in condition (i) in this case.

3 Information-geometric structures on quantum statistical manifolds

In this section, we introduce a family of information-geometric structures on quantum statistical manifolds and apply the results of the previous section to them. The geometric structures treated here are essentially the same as those studied in § 7.3 of [3].

Let $\mathcal{L} = \mathcal{L}(\mathcal{H})$, $\mathcal{L}_h = \mathcal{L}_h(\mathcal{H})$ and $\mathcal{S} = \mathcal{S}(\mathcal{H})$ be the totality of linear operators on $\mathcal{H}$, that of Hermitian operators on $\mathcal{H}$ and that of strictly positive density operators on $\mathcal{H}$, respectively. Then $\mathcal{S}$ is an open subset of the affine space $\mathcal{L}_{h,1} := \{A \in \mathcal{L}_h \mid \text{Tr} A = 1\}$, so that a flat affine connection is naturally introduced on $\mathcal{S}$, which we call the *m-connection* and denote by $\nabla^{(m)}$. In order to express $\nabla^{(m)}$ more explicitly, we introduce the embedding map $\iota : \mathcal{S} \rightarrow \mathcal{L}_{h,1}$ so that $\rho \in \mathcal{S}$ is denoted by $\iota(\rho)$ when treating it as an element of $\mathcal{L}_{h,1}$. Since ι is a smooth map, it has the differential at every point $\rho \in \mathcal{S}$, which we denote by $\iota_* = (d\iota)_\rho : T_\rho(\mathcal{S}) \rightarrow \mathcal{L}_{h,0} := \{A \in \mathcal{L}_h \mid \text{Tr} A = 0\}$. For a vector field $X \in \mathcal{X}(\mathcal{S})$, the map $\iota_*(X) : \mathcal{S} \rightarrow \mathcal{L}_{h,0}$ is defined to be $\rho \mapsto \iota_*(X_\rho) = (d\iota)_\rho(X_\rho)$. Then the definition of the m-connection is represented as follows:

$$\forall X, Y \in \mathcal{X}(\mathcal{S}), \quad \iota_*(\nabla_X^{(m)} Y) = X \iota_*(Y), \quad (3.1)$$

where $X\iota_*(Y) : \mathcal{S} \rightarrow \mathcal{L}_{h,0}$ is the derivative of $\iota_*(Y)$ w.r.t. X . When a coordinate system $\xi = (\xi^i)$ is arbitrarily given and the elements of $\mathcal{S}$ is parametrized by it as ρ_ξ , we have

$$\iota_*(\partial_i) = \partial_i \rho_\xi, \quad (3.2)$$

and

$$\iota_* \left(\nabla_{\partial_i}^{(m)} \partial_j \right) = \partial_i \iota_*(\partial_j) = \partial_i \partial_j \rho_\xi, \quad (3.3)$$

where $\partial_i := \frac{\partial}{\partial \xi^i}$.

Suppose that we are given a family of inner products $\{\langle \cdot, \cdot \rangle_\rho \mid \rho \in \mathcal{S}(\mathcal{H})\}$ on the $\mathbb{R}$ -linear space $\mathcal{L}_h$, where the correspondence $\rho \mapsto \langle \cdot, \cdot \rangle_\rho$ is smooth, and assume that

$$\forall \rho \in \mathcal{S}, \forall A \in \mathcal{L}_h, \langle A, I \rangle_\rho = \langle A \rangle_\rho := \text{Tr}(\rho A), \quad (3.4)$$

where I denotes the identity operator on $\mathcal{H}$. (As usual, cI for $c \in \mathbb{C}$ is often denoted as c throughout the paper.) The inner products are represented as

$$\langle A, B \rangle_\rho = \langle A, \Omega_\rho(B) \rangle_{\text{HS}} = \text{Tr}(A \Omega_\rho(B)) \quad (3.5)$$

by a family of super-operators $\{\Omega_\rho : \mathcal{L}_h \rightarrow \mathcal{L}_h\}_{\rho \in \mathcal{S}}$, where $\langle \cdot, \cdot \rangle_{\text{HS}}$ denotes the Hilbert-Schmidt inner product. Note that the assumption (3.4) is equivalent to

$$\forall \rho \in \mathcal{S}, \Omega_\rho(I) = \rho. \quad (3.6)$$

For an arbitrary tangent vector $X_\rho \in T_\rho(\mathcal{S})$, a Hermitian operator $L_{X_\rho} \in \mathcal{L}_h$ is defined by the relation

$$\forall A \in \mathcal{L}_h, X_\rho \langle A \rangle = \langle L_{X_\rho}, A \rangle_\rho, \quad (3.7)$$

where the LHS denotes the derivative of the function $\langle A \rangle : \mathcal{S} \rightarrow \mathbb{R}, \rho \mapsto \langle A \rangle_\rho$ w.r.t. X_ρ . Noting that the LHS and the RHS are represented as $\text{Tr}(\iota_*(X_\rho)A)$ and $\text{Tr}(\Omega_\rho(L_{X_\rho})A)$, respectively, we can rewrite (3.7) into

$$\iota_*(X_\rho) = \Omega_\rho(L_{X_\rho}). \quad (3.8)$$

From (3.4) and (3.7), we have

$$\langle L_{X_\rho} \rangle_\rho = \langle L_{X_\rho}, I \rangle_\rho = X_\rho \langle I \rangle = X_\rho 1 = 0. \quad (3.9)$$

Since $X_\rho \leftrightarrow \iota_*(X_\rho) \leftrightarrow L_{X_\rho}$ are one-to-one correspondences, we obtain the following identity:

$$\{L_{X_\rho} \mid X_\rho \in T_\rho(\mathcal{S})\} = \{A \in \mathcal{L}_h \mid \langle A \rangle_\rho = 0\}. \quad (3.10)$$

In the following, we often express (3.4) as

$$\forall A \in \mathcal{L}_h, \langle A, I \rangle = \langle A \rangle \quad (3.11)$$

as an identity for functions on $\mathcal{S}$. Similarly, (3.7) is expressed as

$$\forall A \in \mathcal{L}_h, \quad X \langle A \rangle = \langle L_X, A \rangle, \quad (3.12)$$

where X is a vector field on $\mathcal{S}$, L_X denotes the map $\mathcal{S} \rightarrow \mathcal{L}_h$, $\rho \mapsto L_{X_\rho}$, and $\langle L_X, A \rangle$ denotes the function $\rho \mapsto \langle L_{X_\rho}, A \rangle_\rho$.

For a submanifold $\mathcal{M}$ of $\mathcal{S}$ and a vector field $X \in \mathcal{X}(\mathcal{M})$ on it, the map $L_X : \mathcal{M} \rightarrow \mathcal{L}_h$, $\rho \mapsto L_{X_\rho}$ is defined, for which (3.12) holds as an identity for functions on $\mathcal{M}$. We may write $X \langle A \rangle$ as $X \langle A \rangle|_{\mathcal{M}}$ in this case. In particular, given a coordinate system $\xi = (\xi^i)$ of $\mathcal{M}$, we have

$$\forall A \in \mathcal{L}_h, \quad \partial_i \langle A \rangle = \partial_i \langle A \rangle|_{\mathcal{M}} = \langle L_i, A \rangle, \quad (3.13)$$

where $\partial_i := \frac{\partial}{\partial \xi^i}$ and $L_i := L_{\partial_i}$.

Remark 3.1 In the terminology of [3], $\iota_*(X_\rho)$ and L_{X_ρ} are called the m-representation and e-representation, and are denoted by $\iota_*(X_\rho) = X_\rho^{(m)}$ and $L_{X_\rho} = X_\rho^{(e)}$, respectively. We do not use the symbols $X_\rho^{(m)}$ and $X_\rho^{(e)}$ in this paper, but may use the following notation when convenient:

$$T_\rho^{(m)}(\mathcal{M}) := \{\iota_*(X_\rho) \mid X_\rho \in T_\rho(\mathcal{M})\}, \quad (3.14)$$

$$T_\rho^{(e)}(\mathcal{M}) := \{L_{X_\rho} \mid X_\rho \in T_\rho(\mathcal{M})\}. \quad (3.15)$$

Note that

$$T_\rho^{(m)}(\mathcal{S}) = \mathcal{L}_{h,0} \quad \text{and} \quad T_\rho^{(e)}(\mathcal{S}) = \{A \in \mathcal{L}_h \mid \langle A \rangle_\rho = 0\}. \quad (3.16)$$

Now, we define a Riemannian metric g on $\mathcal{S}$ by

$$\forall \rho \in \mathcal{S}, \forall X_\rho, Y_\rho \in T_\rho(\mathcal{S}), \quad g_\rho(X_\rho, Y_\rho) := \langle L_{X_\rho}, L_{Y_\rho} \rangle_\rho = \text{Tr}(\iota_*(X_\rho)L_{Y_\rho}), \quad (3.17)$$

which is equivalently written as

$$\forall X, Y \in \mathcal{X}(\mathcal{S}), \quad g(X, Y) := \langle L_X, L_Y \rangle = \text{Tr}(\iota_*(X)L_Y). \quad (3.18)$$

We also have the expression

$$\forall X, Y \in \mathcal{X}(\mathcal{S}), \quad g(X, Y) = -\langle XL_Y \rangle = -\langle YL_X \rangle, \quad (3.19)$$

where $XL_Y : \mathcal{S} \rightarrow \mathcal{L}_h$, $\rho \mapsto X_\rho L_Y$, denotes the derivative of the map $L_Y : \mathcal{S} \rightarrow \mathcal{L}_h$, $\rho \mapsto L_{Y_\rho}$, w.r.t. X , and $\langle XL_Y \rangle$ denotes the function $\rho \mapsto \langle X_\rho L_Y \rangle_\rho$. This expression is derived as

$$0 = X_\rho \langle L_Y \rangle = X_\rho \langle L_Y \rangle_\rho = X_\rho \langle L_{Y_\rho} \rangle_\rho + X_\rho \langle L_Y \rangle_\rho$$

$$= \langle L_{X_\rho}, L_{Y_\rho} \rangle_\rho + \langle X_\rho L_Y \rangle_\rho, \quad (3.20)$$

where the first equality follows from (3.9) and the last from (3.7), with dots $\cdot$ added to clarify the positions of variables of maps.

An important class of such Riemannian metrics is that of *monotone metrics* [9] for which Ω_ρ is represented as

$$\Omega_\rho(A) = f(\Delta_\rho)(A\rho) = (f(\Delta_\rho)(A))\rho, \quad (3.21)$$

where $f : (0, \infty) \rightarrow (0, \infty)$ is an operator monotone function satisfying $\forall x > 0$, $xf(1/x) = f(x)$ and $f(1) = 1$, and $\Delta_\rho : \mathcal{L} \rightarrow \mathcal{L}$ is the modular operator defined by $\Delta_\rho(A) = \rho A \rho^{-1}$. The class contains the SLD metric, which plays the main role in this paper, defined by $f(x) = (x + 1)/2$ and

$$\langle A, B \rangle_\rho = \text{Re Tr}(\rho AB) = \text{Tr}(\rho(A \circ B)) = \text{Tr}((\rho \circ A)B), \quad A, B \in \mathcal{L}_h, \quad (3.22)$$

where $\circ$ denotes the symmetrized product: $A \circ B = \frac{1}{2}(AB + BA)$. In this case, L_{X_ρ} (L_X , resp.) is called the *SLD (symmetric logarithmic derivative)* of a tangent vector X_ρ (a vector field X , resp.). In particular, $L_i := L_{\partial_i}$ obeys the equation

$$\partial_i \rho_\xi = \rho_\xi \circ L_{i,\xi}, \quad (3.23)$$

which is a popular expression for the SLD.

Given a family of inner products $\{\langle \cdot, \cdot \rangle_\rho \mid \rho \in \mathcal{S}\}$ which determines a Riemannian metric g , let the *e-connection* $\nabla^{(e)}$ be defined as the dual connection of $\nabla^{(m)}$ w.r.t. g ; i.e.,

$$\forall X, Y, Z \in \mathcal{X}(\mathcal{S}), \quad Xg(Y, Z) = g(\nabla_X^{(e)} Y, Z) + g(Y, \nabla_X^{(m)} Z). \quad (3.24)$$

We have thus obtained $(\mathcal{S}, g, \nabla^{(e)}, \nabla^{(m)})$ as an example of $(\mathcal{S}, g, \nabla, \nabla^*)$ treated in the previous section, where $\nabla^{(e)}$ and $\nabla^{(m)}$ are dual w.r.t. g , $\nabla^{(e)}$ is curvature-free and $\nabla^{(m)}$ is flat (see Remark 2.7). The triple $(\mathcal{S}, g, \nabla^{(e)}, \nabla^{(m)})$ is called the *information-geometric structure* on $\mathcal{S}$ induced from a family of inner products $\{\langle \cdot, \cdot \rangle_\rho \mid \rho \in \mathcal{S}\}$. In particular, the information-geometric structure induced from the symmetrized inner product (3.22) is called the *SLD structure*. It is the SLD structure that will play a leading role in subsequent sections in relation to estimation theory, but this section will continue the discussion on general information-geometric structures.

Remark 3.2 As is shown in Theorem 7.1 of [3], there is only one information-geometric structure defined in the manner described above for which the e-connection is torsion-free (so that $(\mathcal{S}, g, \nabla^{(e)}, \nabla^{(m)})$ is dually flat). That is the structure induced from the BKM (Bogoliubov-Kubo-Mori) inner product

$$\langle A, B \rangle_\rho = \int_0^1 \text{Tr}(\rho^s A \rho^{1-s} B) ds, \quad A, B \in \mathcal{L}_h. \quad (3.25)$$

The induced Riemannian metric is a monotone metric corresponding to $f(x) = \frac{x-1}{\log x} = \int_0^1 x^s ds$. In the other cases, the torsion $\mathcal{T}^{(e)}(X, Y) = \nabla_X^{(e)} Y - \nabla_Y^{(e)} X - [X, Y]$ does not vanish, where $[\cdot, \cdot]$ denotes the Lie bracket for vector fields.

Remark 3.3 For the SLD structure, it is known ([3], Eq. (7.80)) that the torsion has the following representation: for each point $\rho \in \mathcal{S}$ and each tangent vectors $X_\rho, Y_\rho \in T_\rho(\mathcal{S})$, we have

$$\iota_*(\mathcal{T}^{(e)}(X_\rho, Y_\rho)) = \frac{1}{4}[[L_{X_\rho}, L_{Y_\rho}], \rho], \quad (3.26)$$

where $[\cdot, \cdot]$ in the RHS denotes the commutator for operators on $\mathcal{H}$. Since this representation will be used in Sects. 8 and 9, we show its proof in Appendix A for the reader's convenience.

Henceforth, we use the prefixes e- and m- for notions concerning the e-connection and m-connection; e.g., e-parallel, e-autoparallel, m-affine, etc.

Proposition 3.4 For any vector fields $X, Y, W \in \mathcal{X}(\mathcal{S})$, we have

$$W = \nabla_X^{(e)} Y \Leftrightarrow L_W = XL_Y - \langle XL_Y \rangle = XL_Y + g(X, Y). \quad (3.27)$$

Proof Differentiating $g(Y, Z) = \text{Tr}(L_Y \iota_*(Z))$ (see (3.18)) by X , we have

$$\begin{aligned} Xg(Y, Z) &= \text{Tr}((XL_Y) \iota_*(Z)) + \text{Tr}(L_Y (X \iota_*(Z))) \\ &= \text{Tr}((XL_Y) \iota_*(Z)) + g(Y, \nabla_X^{(m)} Z), \end{aligned}$$

where the second equality follows from (3.1). Letting $W' \in \mathcal{X}(\mathcal{S})$ be defined by $L_{W'} = XL_Y - \langle XL_Y \rangle$, whose existence is ensured by (3.10), the above equation is represented as

$$Xg(Y, Z) = g(W', Z) + g(Y, \nabla_X^{(m)} Z).$$

This means that $W' = \nabla_X^{(e)} Y$, and proves the proposition. $\square$

Proposition 3.5 For a vector field $X \in \mathcal{X}(\mathcal{S})$, we have

$$\begin{aligned} X \text{ is e-parallel} &\Leftrightarrow \exists A \in \mathcal{L}_h, L_X = A - \langle A \rangle \\ &\Leftrightarrow \exists A \in \mathcal{L}_h, \forall \rho \in \mathcal{S}, L_{X_\rho} = A - \langle A \rangle_\rho. \end{aligned} \quad (3.28)$$

Proof We may use Proposition 3.4 to prove this, but here we show an alternative proof. We first note that, according to (3.1), a vector field $Y \in \mathcal{X}(\mathcal{S})$ is m-parallel if and only if $\iota_*(Y) : \mathcal{S} \rightarrow \mathcal{L}_{h,0}$ is a constant map represented by an operator $B \in \mathcal{L}_{h,0}$ as $\iota_*(Y) = B$. Invoking Proposition 2.5 and (3.18), we have

$$\begin{aligned} X \text{ is e-parallel} &\Leftrightarrow \forall Y : \text{m-parallel}, g(X, Y) \text{ is constant on } \mathcal{S} \\ &\Leftrightarrow \forall B \in \mathcal{L}_{h,0}, \text{Tr}(L_X B) \text{ is constant on } \mathcal{S} \end{aligned}$$

$$\Leftrightarrow \exists A \in \mathcal{L}_h, L_X = A - \langle A \rangle.$$

□

Recalling Proposition 2.3, the following corollary is immediate.

Corollary 3.6 *For a submanifold $\mathcal{M}$ of $\mathcal{S}$ and for $X \in \mathcal{X}(\mathcal{S}/\mathcal{M})$ (including the case when $X \in \mathcal{X}(\mathcal{M})$), we have*

$$\begin{aligned} &X \text{ is } e\text{-parallel (i.e., parallel w.r.t. the } e\text{-connection on } \mathcal{S}) \\ &\Leftrightarrow \exists A \in \mathcal{L}_h, L_X = A - \langle A \rangle|_{\mathcal{M}} \\ &\Leftrightarrow \exists A \in \mathcal{L}_h, \forall \rho \in \mathcal{M}, L_{X_\rho} = A - \langle A \rangle_\rho. \end{aligned} \quad (3.29)$$

The following corollary is also immediate from Proposition 3.5.

Corollary 3.7 *The e -parallel transport $\Phi_{\rho,\sigma}^{(e)}: T_\rho(\mathcal{S}) \rightarrow T_\sigma(\mathcal{S})$ for arbitrary two points $\rho, \sigma \in \mathcal{S}$ is represented as follows: $\forall X_\rho \in T_\rho(\mathcal{S}), \forall X_\sigma \in T_\sigma(\mathcal{S})$,*

$$X_\sigma = \Phi_{\rho,\sigma}^{(e)}(X_\rho) \Leftrightarrow L_{X_\sigma} = L_{X_\rho} - \langle L_{X_\rho} \rangle_\sigma. \quad (3.30)$$

In the following, a pair $(\mathcal{M}, \xi)$ of a submanifold $\mathcal{M}$ of $\mathcal{S}$ and a coordinate system ξ of $\mathcal{M}$ is called a *model*.

Proposition 3.8 *For an n -dimensional model $(\mathcal{M}, \xi)$, the following conditions are equivalent.*

- (i) $\mathcal{M}$ is e -autoparallel in $\mathcal{S}$, and ξ is an m -affine coordinate system.
(Note: “ m -affine” means “affine w.r.t. the m -connection $\nabla_{\mathcal{M}}^{(m)}$ on $\mathcal{M}$ ”.)
- (ii) $\exists \{F^1, \dots, F^n\} \subset \mathcal{L}_h$ such that for every $i \in \{1, \dots, n\}$,

$$\sum_j g^{ij} L_j = F^i - \langle F^i \rangle|_{\mathcal{M}}, \quad (3.31)$$

where $\partial_i := \frac{\partial}{\partial \xi^i}$, $L_j := L_{\partial_j}$ and $[g^{ij}] := [g_{ij} := g(\partial_i, \partial_j)]^{-1}$.

- (iii) $\exists \{F^1, \dots, F^n\} \subset \mathcal{L}_h$ such that for every $i \in \{1, \dots, n\}$,

$$\sum_j g^{ij} L_j = F^i - \xi^i. \quad (3.32)$$

(Note: (3.32) implies $\xi^i = \langle F^i \rangle|_{\mathcal{M}}$.)

- (iv) $\exists \{F^1, \dots, F^n\} \subset \mathcal{L}_h$ such that for every $i \in \{1, \dots, n\}$,

$$\forall \rho \in \mathcal{M}, L_{i,\rho} \in \text{span} \{F^j\}_{j=1}^n \oplus \mathbb{R} \text{ and } \xi^i(\rho) = \langle F^i \rangle_\rho, \quad (3.33)$$

where $\mathbb{R}$ is identified with $\{cI \mid c \in \mathbb{R}\}$ (see Remark 3.10 below).

Proof The equivalence (i) $\Leftrightarrow$ (ii) is immediate from Proposition 2.6 and Cor. 3.6, and (iii) $\Rightarrow$ (ii) is obvious since $\langle L_i \rangle = 0$.

To show (ii) $\Rightarrow$ (iii), assume (ii). Then we have

$$\sum_j g^{ij} \langle L_j, L_k \rangle = \langle F^i, L_k \rangle.$$

Here, the LHS is $\sum_j g^{ij} g_{jk} = \delta_k^i = \partial_k \xi^i$ and the RHS is $\partial_k \langle F^i \rangle$ due to (3.13). Hence, there exists a constant vector $(c^i) \in \mathbb{R}^n$ such that $\langle F^i \rangle|_{\mathcal{M}} = \xi^i + c^i$. Redefining $F^i := F^i - c^i$, (3.31) is rewritten as (3.32), and (iii) is verified.

Since (3.32) implies that

$$L_{i,\rho} = \sum_j g_{ij}(\rho) F^j - \left(\sum_j g_{ij}(\rho) \xi^j(\rho) \right) \in \text{span} \{F^j\}_{j=1}^n \oplus \mathbb{R},$$

we have (iii) $\Rightarrow$ (iv). To show the converse, we assume the existence of $\{F^1, \dots, F^n\}$ in (iv). Then we have $\xi^i = \langle F^i \rangle|_{\mathcal{M}}$, and for each $\rho \in \mathcal{M}$ there exist $[a_{ij}] \in \mathbb{R}^{n \times n}$ and $[b_i] \in \mathbb{R}^n$ such that for any i

$$L_{i,\rho} = \sum_j a_{ij} F^j + b_i.$$

This implies that

$$\begin{aligned} g_{ik}(\rho) &= \langle L_{k,\rho}, L_{i,\rho} \rangle_\rho = \sum_j a_{ij} \langle L_{k,\rho}, F^j \rangle_\rho \\ &= \sum_j a_{ij} (\partial_k \langle F^j \rangle)_\rho = \sum_j a_{ij} (\partial_k \xi^j)_\rho = a_{ik}. \end{aligned}$$

Hence we have

$$\sum_j g^{ij}(\rho) L_{j,\rho} = F^i + \sum_j g^{ij}(\rho) b_j.$$

Here, the constant $\sum_j g^{ij}(\rho) b_j$ should be equal to $-\langle F^i \rangle_\rho = -\xi^i(\rho)$ due to $\langle L_{j,\rho} \rangle_\rho = 0$. Thus (iii) is concluded. $\square$

Remark 3.9 As noted in Remark 2.7, the flatness of the m-connection on $\mathcal{S}$ ensures that every e-autoparallel submanifold is m-flat and has an m-affine coordinate system. This fact can be seen directly from Proposition 3.8 as follows. The e-autoparallelity of $\mathcal{M}$ is equivalent to the existence of $\{F^1, \dots, F^n\} \subset \mathcal{L}_h$ s.t. $\forall \rho \in \mathcal{M}$, $T_\rho^{(e)}(\mathcal{M}) = \text{span}\{F^i - \langle F^i \rangle_\rho\}_{i=1}^n$ (see (3.15) for $T_\rho^{(e)}(\mathcal{M})$), which is equivalent to the first condition that $\forall i, \forall \rho \in \mathcal{M}$, $L_{i,\rho} \in \text{span}\{F^j\}_{j=1}^n \oplus \mathbb{R}$ of (iv) in the proposition. When $\mathcal{M}$ satisfies this condition, letting $\xi = (\xi^i)$ be defined by $\xi^i = \langle F^i \rangle|_{\mathcal{M}}$, we have that $(\mathcal{M}, \xi)$ satisfies the two conditions of (iv), so that ξ is m-affine due to (iv) $\Rightarrow$ (i).

Remark 3.10 In (3.31) and (3.32), the operators $\{F^i - \langle F^i \rangle_\rho\}_{i=1}^n$ turn out to be linearly independent for each $\rho \in \mathcal{M}$, which implies that $\{F^1, \dots, F^n, I\}$ are linearly independent, or equivalently that $\{F^1, \dots, F^n\}$ are linearly independent in the quotient space $\mathcal{L}_h/\mathbb{R}$ with identification $\mathbb{R} = \{cI \mid c \in \mathbb{R}\}$.

At the end of this section, we present a proposition which claims that i.i.d. extensions of a model preserves the e-autoparallelity. For the proposition, we assume the following condition on the family of inner products from which the information-geometric structure is defined:

$$\forall \{A_t\}_{t=1}^k, \forall \{B_t\}_{t=1}^k \subset \mathcal{L}_h(\mathcal{H}), \left\langle \bigotimes_{t=1}^k A_t, \bigotimes_{t=1}^k B_t \right\rangle_{\rho^{\otimes k}} = \prod_{t=1}^k \langle A_t, B_t \rangle_\rho, \quad (3.34)$$

which is equivalent to

$$\forall \{A_t\}_{t=1}^k \subset \mathcal{L}_h(\mathcal{H}), \Omega_{\rho^{\otimes k}}\left(\bigotimes_{t=1}^k A_t\right) = \bigotimes_{t=1}^k \Omega_\rho(A_t). \quad (3.35)$$

The assumption is satisfied when the inner products are defined from a function f by (3.5) and (3.21) for which $\Omega_\rho = f(\Delta_\rho)$ and $\Omega_{\rho^{\otimes k}} = f(\Delta_{\rho^{\otimes k}})$ hold. In particular, the proposition holds for the SLD structure.

Proposition 3.11 *Given a model $(\mathcal{M}, \xi)$ in $\mathcal{S}(\mathcal{H})$ and a natural number $k \geq 2$, define the model $(\tilde{\mathcal{M}}, \tilde{\xi})$ in $\mathcal{S}(\mathcal{H}^{\otimes k})$ by $\tilde{\mathcal{M}} := \{\rho^{\otimes k} \mid \rho \in \mathcal{M}\}$ and $\tilde{\xi}^i(\rho^{\otimes k}) = \xi^i(\rho)$ for $\rho \in \mathcal{M}$. Under the assumption (3.34)-(3.35), the following conditions are equivalent.*

- (i) $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$, and ξ is m-affine.
- (ii) $\tilde{\mathcal{M}}$ is e-autoparallel in $\mathcal{S}(\mathcal{H}^{\otimes k})$, and $\tilde{\xi}$ is m-affine.

Proof Let $\partial_i := \frac{\partial}{\partial \xi^i}$, $\tilde{\partial}_i := \frac{\partial}{\partial \tilde{\xi}^i}$, and $L_i = L_{\partial_i}$, $\tilde{L}_i = L_{\tilde{\partial}_i}$, which are determined by

$$\iota_*((\partial_i)_\rho) = \Omega_\rho(L_{i,\rho})$$

and

$$\tilde{\iota}_*((\tilde{\partial}_i)_{\rho^{\otimes k}}) = \Omega_{\rho^{\otimes k}}(\tilde{L}_{i,\rho^{\otimes k}}),$$

where $\tilde{\iota}$ denotes the natural embedding $\mathcal{S}(\mathcal{H}^{\otimes k}) \rightarrow \mathcal{L}_{h,1}(\mathcal{H}^{\otimes k})$. With the aid of the parametric representation (3.2), we see that

$$\begin{aligned} \tilde{\iota}_*((\tilde{\partial}_i)_{\rho^{\otimes k}}) &= \tilde{\partial}_i(\rho^{\otimes k})_{\tilde{\xi}} = \partial_i \rho_{\xi}^{\otimes k} = \sum_{t=1}^k \rho_{\xi}^{\otimes(t-1)} \otimes \partial_i \rho_{\xi} \otimes \rho_{\xi}^{\otimes(k-t)} \\ &= \sum_{t=1}^k \rho^{\otimes(t-1)} \otimes \iota_*((\partial_i)_\rho) \otimes \rho^{\otimes(k-t)} \end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^k \rho^{\otimes(t-1)} \otimes \Omega_\rho(L_{i,\rho}) \otimes \rho^{\otimes(k-t)} \\
&\stackrel{*}{=} \sum_{t=1}^k \Omega_{\rho^{\otimes k}}(I^{\otimes(t-1)} \otimes L_{i,\rho} \otimes I^{\otimes(k-t)}) = \Omega_{\rho^{\otimes k}}(L_{i,\rho}^{(k)}),
\end{aligned}$$

where $\stackrel{*}{=}$ follows from (3.6) and (3.35), and we have used the notation

$$A^{(k)} := \sum_{t=1}^k I^{\otimes(t-1)} \otimes A \otimes I^{\otimes(k-t)} \quad \text{for } A \in \mathcal{L}_h(\mathcal{H}).$$

This implies that $\tilde{L}_{i,\rho^{\otimes k}} = (L_{i,\rho})^{(k)}$ and that $\tilde{g}_{ij} = k g_{ij}$ for $g_{ij}(\rho) = \langle L_{i,\rho}, L_{j,\rho} \rangle_\rho$ and $\tilde{g}_{ij}(\rho^{\otimes k}) = \langle \tilde{L}_{i,\rho^{\otimes k}}, \tilde{L}_{j,\rho^{\otimes k}} \rangle_{\rho^{\otimes k}}$, which leads to $\tilde{g}^{ij} = \frac{1}{k} g^{ij}$. Hence, $L^i := \sum_j g^{ij} L_j$ and $\tilde{L}^i := \sum_j \tilde{g}^{ij} \tilde{L}_j$ are linked by $\tilde{L}_{\rho^{\otimes k}}^i = \frac{1}{k} (L_\rho^i)^{(k)}$. Now, according to Proposition 3.8, conditions (i), (ii) of the present proposition are respectively expressed as

- (i)' $\exists \{F^i\} \subset \mathcal{L}(\mathcal{H}), \forall i, \forall \rho \in \mathcal{M}, L_\rho^i = F^i - \xi^i(\rho).$
- (ii)' $\exists \{\tilde{F}^i\} \subset \mathcal{L}(\mathcal{H}^{\otimes k}), \forall i, \forall \rho \in \mathcal{M}, \frac{1}{k} (L_\rho^i)^{(k)} = \tilde{F}^i - \xi^i(\rho).$

They are obviously equivalent with relation $\tilde{F}^i = \frac{1}{k} (F^i)^{(k)}$. $\square$

4 Preliminaries from quantum estimation theory

From this section, we investigate the relationship between estimation problems and geometric properties for quantum statistical models. Henceforth, we will consider only the SLD structure as an information-geometric structure unless otherwise stated. As a first step, we introduce several notions of quantum estimation theory and discuss their basic properties in this section.

Given a model $(\mathcal{M}, \xi)$ in $\mathcal{S}(\mathcal{H})$, an *estimator* for coordinates ξ is generally represented by a POVM $\Pi = \Pi(d\hat{\xi})$ on $\mathcal{H}$, where $\hat{\xi}$ is a variable representing an estimate. A representative case is when Π is expressed as $\Pi(d\hat{\xi}) = \sum_{\omega \in \Omega} \pi_\omega \delta_{f(\omega)}(d\hat{\xi})$ by a POVM $\pi = (\pi_\omega)_{\omega \in \Omega}$ on a discrete set Ω and a function $f : \Omega \rightarrow \mathbb{R}^n$, where $\delta_{f(\omega)}$ denotes the δ -measure concentrated on the point $f(\omega) \in \mathbb{R}^n$. This estimator, which is denoted by $\Pi = (\pi, f)$, represents the estimation procedure in which the estimate is determined as $\xi = f(\omega)$ from the outcome ω of the measurement π .

The expectation $E_\rho(\Pi) \in \mathbb{R}^n$ and the mean squared error (the variance in the unbiased case) $V_\rho(\Pi) \in \mathbb{R}^{n \times n}$ of Π for a state ρ are defined by

$$E_\rho(\Pi) := \int \hat{\xi} \operatorname{Tr}(\rho \Pi(d\hat{\xi})), \quad (4.1)$$

$$V_\rho(\Pi) := \int (\hat{\xi} - \xi(\rho))(\hat{\xi} - \xi(\rho))^T \operatorname{Tr}(\rho \Pi(d\hat{\xi})), \quad (4.2)$$

where $\mathbb{R}^n$ is regarded as the space of column vectors $\mathbb{R}^{n \times 1}$ and T denotes the transpose. For $\Pi = (\pi, f)$, we have

$$E_\rho(\Pi) = \sum_{\omega} f(\omega) \operatorname{Tr}(\rho \pi_{\omega}), \quad (4.3)$$

$$V_\rho(\Pi) = \sum_{\omega} (f(\omega) - \xi(\rho))(f(\omega) - \xi(\rho))^T \operatorname{Tr}(\rho \pi_{\omega}). \quad (4.4)$$

An estimator Π is said to be *locally unbiased* for a coordinate system ξ at $\rho \in \mathcal{M}$ when the elements $E_\rho^i(\Pi)$, $i \in \{1, \dots, n\}$, of $E_\rho(\Pi)$ satisfy

$$E_\rho^i(\Pi) = \xi^i(\rho) \quad \text{and} \quad \partial_j E_\rho^i(\Pi) = \delta_j^i, \quad (4.5)$$

where $\partial_j E_\rho^i(\Pi)$ denotes the derivative of the function $E^i(\Pi) : \sigma \mapsto E_\sigma^i(\Pi)$ by $\partial_i = \frac{\partial}{\partial \xi^i}$ evaluated at the point ρ . We denote by $\mathcal{U}(\rho, \xi)$ the totality of locally unbiased estimators for ξ at ρ . Using the symmetrized inner product $\langle \cdot, \cdot \rangle$ of (3.22) and the SLDs $L_i = L_{\partial_i}$, $i \in \{1, \dots, n\}$, we have

$$\Pi \in \mathcal{U}(\rho, \xi) \Leftrightarrow \forall i, j, \left\langle A^i \right\rangle_{\rho} = \xi^i(\rho) \quad \text{and} \quad \left\langle A^i, L_{j, \rho} \right\rangle_{\rho} = \delta_j^i, \quad (4.6)$$

where

$$A^i := \int \hat{\xi}^i \Pi(d\hat{\xi}) \in \mathcal{L}_h(\mathcal{H}). \quad (4.7)$$

The well-known SLD Cramér-Rao inequality [4, 5] states that every $\Pi \in \mathcal{U}(\rho, \xi)$ obeys

$$V_\rho(\Pi) \geq G_\rho^{-1}, \quad (4.8)$$

where $G_\rho = [g_{ij}(\rho)]$ denotes the SLD Fisher information matrix defined by $g_{ij} = \langle L_i, L_j \rangle = g(\partial_i, \partial_j)$ with g being the SLD metric. Furthermore, we have the following proposition, whose proof will be given after Lemma 4.3.

Proposition 4.1 *For every column vector $\mathbf{u} \in \mathbb{R}^n = \mathbb{R}^{n \times 1}$, we have*

$$\inf_{\Pi \in \mathcal{U}(\rho, \xi)} \mathbf{u}^T V_\rho(\Pi) \mathbf{u} = \mathbf{u}^T G_\rho^{-1} \mathbf{u}. \quad (4.9)$$

Remark 4.2 The infimum on the LHS of (4.9) cannot, in general, be attained as a minimum. Using the terminology that will be introduced later, this can be rephrased as saying that a locally $\mathbf{u}\mathbf{u}^T$ -efficient estimator at a point ρ for ξ does not generally exist. A necessary condition for the attainability is given in Appendix B.

Let us introduce a class of randomized procedures for estimation that will be useful in the proofs of both Proposition 4.1 above and Theorem 5.1 later. Suppose that a point $\rho \in \mathcal{M}$, a basis $\{\mathbf{u}^1, \dots, \mathbf{u}^n\}$ of $\mathbb{R}^n$, a positive probability vector $(p_1, \dots, p_n)$ s.t. $\forall i$,

$p_i > 0$, $\sum_i p_i = 1$ and n^2 real numbers $\{\gamma_k^i\}$ satisfying

$$\sum_k p_k \gamma_k^i = \xi^i(\rho) \quad (4.10)$$

are arbitrarily given. Let

$$X^k := \sum_i u_i^k L_\rho^i \in \mathcal{L}_h(\mathcal{H}), \quad (4.11)$$

where $\mathbf{u}^k = (u_i^k)$, $L_\rho^i := \sum_j g^{ij}(\rho) L_{j,\rho}$, $G_\rho^{-1} = [g^{ij}(\rho)]$, and consider the random measurement such that $k \in \{1, \dots, n\}$ is randomly chosen according to the probability distribution $(p_1, \dots, p_n)$ and then the observable X^k is measured. This measurement is represented by the POVM $\pi = \{\pi_{k,r}\} = \{p_k \pi_r^k\}$, where $\{\pi_r^k\}$ are the projectors in the spectral decomposition $X^k = \sum_r x_r^k \pi_r^k$. (Do not confuse X^k , x_r^k and π_r^k with the n th powers of X , x_r and π_r .) When an eigenvalue x_r^k is observed by measuring X^k , we estimate ξ by $\hat{\xi} = f(k, x_r^k)$, where $f = (f^1, \dots, f^n)^T$ is defined by

$$f^i(k, x) := \gamma_k^i + \frac{w_k^i}{p_k} x, \quad (4.12)$$

and $[w_k^i] = [u_i^k]^{-1}$, i.e., $\sum_k u_i^k w_k^j = \delta_i^j$ and $\sum_i u_i^k w_l^i = \delta_l^k$. This estimation procedure defines the estimator $\Pi := (\pi, f)$, which is characterized by the following property: for any polynomial function $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\int \varphi(\hat{\xi}) \Pi(d\hat{\xi}) = \sum_k p_k \varphi(f(k, X^k)). \quad (4.13)$$

The procedure for obtaining the estimate $\hat{\xi}$ is illustrated below:

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sum_i u_i^k L_\rho^i} \xrightarrow{\text{meas.}} x \\ \phantom{\boxed{\sum_i u_i^k L_\rho^i}} \phantom{\xrightarrow{\text{meas.}}} \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \boldsymbol{\gamma}_k + \frac{x}{p_k} \mathbf{w}_k, \quad (4.14)$$

where $\boldsymbol{\gamma}_k := (\gamma_k^i)$, $\mathbf{w}_k := (w_k^i) \in \mathbb{R}^n$ and “meas.” stands for “measure”.

In this situation we have the following lemma, whose proof is given in Appendix C.

Lemma 4.3 *The estimator $\Pi = (\pi, f)$ satisfies:*

- (1) $\Pi \in \mathcal{U}(\rho, \xi)$.
- (2) $V_\rho(\Pi) = \sum_k \frac{1}{p_k} \left((\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k \right) \mathbf{w}_k (\mathbf{w}_k)^T + \sum_k p_k (\gamma_k - \xi(\rho)) (\gamma_k - \xi(\rho))^T$,
where $\mathbf{w}_k := (w_k^i)$, $\gamma_k := (\gamma_k^i) \in \mathbb{R}^n = \mathbb{R}^{n \times 1}$.
- (3) $\forall k, (\mathbf{u}^k)^T V_\rho(\Pi) \mathbf{u}^k = \frac{1}{p_k} (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k + \sum_l p_l (a_l^k)^2$,
where $a_l^k := \sum_i u_i^k (\gamma_l^i - \xi^i(\rho))$.

Proof of Proposition 4.1 Let $\gamma_k^i := \xi^i(\rho)$, which satisfies (4.10) so that Lemma 4.3 is applicable:

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sum_i u_i^k L_\rho^i} \xrightarrow{\text{meas.}} x \\ \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \xi(\rho) + \frac{x}{p_k} \mathbf{w}_k. \quad (4.15)$$

Since $a_l^k = 0$ in this case, we have

$$(\mathbf{u}^k)^T V_\rho(\Pi) \mathbf{u}^k = \frac{1}{p_k} (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k. \quad (4.16)$$

The proposition is obvious when $\mathbf{u} = 0$, so we assume $\mathbf{u} \neq 0$ and choose $p \in (0, 1)$ arbitrarily. Taking $(\mathbf{u}^1, \dots, \mathbf{u}^n)$ and $(p_1, \dots, p_n)$ in the above construction such that $\mathbf{u}^1 = \mathbf{u}$ and $p_1 = p$, the resulting Π satisfies $\mathbf{u}^T V_\rho(\Pi) \mathbf{u} = \frac{1}{p} \mathbf{u}^T G_\rho^{-1} \mathbf{u}$. Since p can be arbitrarily close to 1, we have the proposition. $\square$

In the following, we introduce several notions of efficiency for estimators for a model $(\mathcal{M}, \xi)$. We begin with local notions which are properties of estimators at each point ρ of the model.

- A locally unbiased estimator $\Pi \in \mathcal{U}(\rho, \xi)$ is said to be *locally efficient* for ξ at ρ if

$$\forall \Pi' \in \mathcal{U}(\rho, \xi), \quad V_\rho(\Pi) \leq V_\rho(\Pi'). \quad (4.17)$$

- Given a positive-semidefinite matrix $W \in \mathbb{R}^{n \times n}$ as a weight, an estimator $\Pi \in \mathcal{U}(\rho, \xi)$ is said to be *locally W-efficient* for ξ at ρ if

$$\forall \Pi' \in \mathcal{U}(\rho, \xi), \quad \text{tr}(W V_\rho(\Pi)) \leq \text{tr}(W V_\rho(\Pi')), \quad (4.18)$$

or equivalently if

$$\text{tr}(W V_\rho(\Pi)) = \min_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W V_\rho(\Pi')). \quad (4.19)$$

Proposition 4.4 Given a model $(\mathcal{M}, \xi)$, a point $\rho \in \mathcal{M}$ and an estimator $\Pi \in \mathcal{U}(\rho, \xi)$, the following conditions are equivalent.

- Π is locally efficient for ξ at ρ .
- $V_\rho(\Pi) = G_\rho^{-1}$.
- Π is locally $\mathbf{u}\mathbf{u}^T$ -efficient for ξ at ρ for every column vector $\mathbf{u} \in \mathbb{R}^n$.
- Π is locally W -efficient for ξ at ρ for every positive weight $W > 0$.

Proof The equivalence (i) $\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv) is obvious since

$$\begin{aligned} V_\rho(\Pi) \leq V_\rho(\Pi') &\Leftrightarrow \forall \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{u}^T V_\rho(\Pi) \mathbf{u} \leq \mathbf{u}^T V_\rho(\Pi') \mathbf{u} \\ &\Leftrightarrow \forall W > 0, \quad \text{tr}(W V_\rho(\Pi)) \leq \text{tr}(W V_\rho(\Pi')), \end{aligned}$$

and (ii) $\Leftrightarrow$ (iii) follows from Proposition 4.1. $\square$

Remark 4.5 As is well known, there exists a locally efficient estimator at ρ iff the SLDs $L_{1,\rho}, \dots, L_{n,\rho}$ mutually commute (e.g. § 7.4 of [3]).

We next extend the notions of local efficiency introduced above to those that pertain to the model as a whole.

- An estimator Π is said to be *efficient* for a coordinate system ξ if Π is locally efficient for ξ at every point $\rho \in \mathcal{M}$.
- Given a weight field $\mathcal{W} = \{W_\rho \mid \rho \in \mathcal{M}\}$, Π is said to be *$\mathcal{W}$ -efficient* for ξ if Π is locally W_ρ -efficient for ξ at every $\rho \in \mathcal{M}$.
- When $W_\rho = W$ for all ρ , we simply call it *W -efficient* for ξ .

According to Proposition 4.4, Π is efficient $\Leftrightarrow \Pi$ is $\mathbf{u}\mathbf{u}^T$ -efficient for every $\mathbf{u} \in \mathbb{R}^n$
 $\Leftrightarrow \Pi$ is W -efficient for every $W > 0$.

The condition for existence of efficient estimator was mentioned in the introduction as Theorem 1.3. Suppose that $(\mathcal{M}, \xi)$ is represented as (1.3) and (1.4) in the theorem; namely, every $\rho \in \mathcal{M}$ is represented in the form

$$\rho = \exp\left[\frac{1}{2}\left\{\sum_{i=1}^n \theta_i(\rho) F^i - \psi(\rho)\right\}\right] P \exp\left[\frac{1}{2}\left\{\sum_{i=1}^n \theta_i(\rho) F^i - \psi(\rho)\right\}\right] \quad (4.20)$$

and satisfies $\xi^i(\rho) = \langle F^i \rangle_\rho$, where $F^1, \dots, F^n$ are mutually commuting observables. Note that P can be chosen to be an arbitrary element of $\mathcal{M}$ if we wish. The SLDs of $\mathcal{M}$ w.r.t. ξ are represented as

$$L_i = \partial_i \left(\sum_j \theta_j F^j - \psi \right) = \sum_j (\partial_i \theta_j) (F^j - \partial^j \psi), \quad (4.21)$$

where $\partial_i := \frac{\partial}{\partial \xi^i}$ and $\partial^i := \frac{\partial}{\partial \theta_i}$. Noting that the positions of upper/lower indices (superscripts and subscripts) are reversed from the standard notation of information geometry as in [3] (see Remark 1.2), we have $\partial_i \theta_j = g_{ij}$, $\partial^i \psi = \xi^i$ and

$$L^i = \sum_j g^{ij} L_j = F^i - \xi^i. \quad (4.22)$$

According to Proposition 3.8, this means that $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$ and that ξ is an m-affine coordinate system w.r.t. the SLD structure. Furthermore, the induced e-connection $\nabla^{(e)}|_{\mathcal{M}}$ on M turns out to be torsion-free and hence flat, for which $\theta = (\theta_i)$ forms an affine coordinate system. We have thus seen that $\mathcal{M}$ is dually flat just as classical exponential families.

When $n = 1$, (4.20) is written as

$$\rho = \exp\left[\frac{1}{2}\{\theta(\rho)F - \psi(\rho)\}\right] P \exp\left[\frac{1}{2}\{\theta(\rho)F - \psi(\rho)\}\right]. \quad (4.23)$$

This is a general form of e-geodesic in the sense that every e-geodesic is represented in this form by some P and F . In the multi-dimensional case $n \geq 2$, on the other hand, (4.20) provides merely a special case of e-autoparallel submanifolds. In order to characterize the e-autoparallelity by an estimation-theoretical notion, the existence of efficient estimator is too strong, and we need a new variant of the notion of efficient estimators, which will be introduced in the next section.

5 Characterization of e-autoparallelity in terms of filtrations of estimators

We now consider a one-parameter family of estimators $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon \in (0, \varepsilon_1)}$ instead of a single estimator, and call it a *filtration of estimators* or simply a *filtration*. The upper limit ε_1 can be an arbitrary positive number or ∞ , but our interest lies only in the limiting property of $\varepsilon \downarrow 0$ and the value of ε_1 plays no role. So we simply write $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon > 0}$. The notions of efficiency for estimators introduced in the previous section are extended to those for filtrations as follows:

- Given a nonnegative (i.e., positive-semidefinite) real matrix $W \in \mathbb{R}^{n \times n}$ as a weight, a filtration $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon > 0}$ is said to be *locally W -efficient* for ξ at ρ if $\Pi_\varepsilon \in \mathcal{U}(\rho, \xi)$ for every $\varepsilon > 0$ and $\lim_{\varepsilon \downarrow 0} \text{tr}(W V_\rho(\Pi_\varepsilon)) \leq \text{tr}(W V_\rho(\Pi'))$ for every $\Pi' \in \mathcal{U}(\rho, \xi)$, which is equivalent to

$$\lim_{\varepsilon \downarrow 0} \text{tr}(W V_\rho(\Pi_\varepsilon)) = \inf_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W V_\rho(\Pi')). \quad (5.1)$$

When $W = \mathbf{u}\mathbf{u}^T$ with $\mathbf{u} \in \mathbb{R}^n$, in particular, this is represented as

$$\lim_{\varepsilon \downarrow 0} \mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} = \mathbf{u}^T G_\rho^{-1} \mathbf{u} \quad (5.2)$$

due to Proposition 4.1.

- Given a weight field $\mathcal{W} = \{W_\rho \mid \rho \in \mathcal{M}\}$, a filtration $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon > 0}$ is said to be *$\mathcal{W}$ -efficient* for ξ if it is W_ρ -locally efficient for ξ at every $\rho \in \mathcal{M}$.
- When $W_\rho = W$ for all ρ , we simply call it *W -efficient* for ξ .

Compare (5.1) with (4.19), and note that a locally W -efficient filtration at ρ always exists, even when a locally W -efficient estimator does not exist; see Remark 4.2.

Now, we have the following theorem, which gives an estimation-theoretical characterization of the e-autoparallelity w.r.t. the SLD structure.

Theorem 5.1 *For a model $(\mathcal{M}, \xi)$, the following conditions are equivalent.*

- $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$, and ξ is an m -affine coordinate system.
- For every $\mathbf{u} \in \mathbb{R}^n$, there exists a $\mathbf{u}\mathbf{u}^T$ -efficient filtration for ξ .

Proof According to Proposition 3.8, it suffices to show the equivalence of (ii) and the condition that

$$\exists \{F^i\}, \forall i, L^i = F^i - \xi^i, \quad (5.3)$$

where $L^i := \sum_j g^{ij} L_j$.

We first show (ii) $\Rightarrow$ (5.3). Fix $i \in \{1, \dots, n\}$ arbitrarily, and let $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$ be an $\mathbf{e}^i(\mathbf{e}^i)^T$ -efficient filtration for ξ , where $\mathbf{e}^i = (\delta_j^i)$ denotes the i th vector of the natural basis of $\mathbb{R}^n$. For each $\rho \in \mathcal{M}$, we have

$$\begin{aligned} (\mathbf{e}^i)^T V_\rho(\Pi_\varepsilon) \mathbf{e}^i &= \int (\hat{\xi}^i - \xi^i(\rho))^2 \text{Tr}(\rho \Pi_\varepsilon(d\hat{\xi})) \\ &\geq \text{Tr}(\rho(F_\varepsilon^i - \xi^i(\rho))^2) = \|F_\varepsilon^i - \xi^i(\rho)\|_\rho^2, \end{aligned} \quad (5.4)$$

where

$$F_\varepsilon^i := \int \hat{\xi}^i \Pi_\varepsilon(d\hat{\xi}) \in \mathcal{L}_h(\mathcal{H}),$$

and $\|\cdot\|_\rho$ denotes the norm for the symmetrized inner product $\langle \cdot, \cdot \rangle_\rho$. (We also denote the norm for the metric g_ρ by the same symbol.) Note that the inequality in (5.4) follows from

$$\int (\hat{\xi}^i - \xi^i(\rho))^2 \Pi_\varepsilon(d\hat{\xi}) - (F_\varepsilon^i - \xi^i(\rho))^2 = \int (\hat{\xi}^i - F_\varepsilon^i) \Pi_\varepsilon(d\hat{\xi}) (\hat{\xi}^i - F_\varepsilon^i) \geq 0.$$

From the local unbiasedness condition (4.6) applied to Π_ε , we have

$$\left\langle F_\varepsilon^i - \xi^i(\rho), L_{j,\rho} \right\rangle_\rho = \left\langle F_\varepsilon^i, L_{j,\rho} \right\rangle_\rho = \delta_j^i = \left\langle L_\rho^i, L_{j,\rho} \right\rangle_\rho,$$

where the first equality follows from (3.9). This means that L_ρ^i is the orthogonal projection of $F_\varepsilon^i - \xi^i(\rho)$ onto $\text{span}\{L_{j,\rho}\}_{j=1}^n$. Hence we have

$$\|F_\varepsilon^i - \xi^i(\rho)\|_\rho^2 = \|L_\rho^i\|_\rho^2 + \|F_\varepsilon^i - \xi^i(\rho) - L_\rho^i\|_\rho^2. \quad (5.5)$$

The $\mathbf{e}^i(\mathbf{e}^i)^T$ -efficiency of $\vec{\Pi}$ is represented as

$$\lim_{\varepsilon \downarrow 0} (\mathbf{e}^i)^T V_\rho(\Pi_\varepsilon) \mathbf{e}^i = (\mathbf{e}^i)^T G_\rho^{-1} \mathbf{e}^i = g^{ii}(\rho) = \|L^i\|_\rho^2,$$

which, combined with (5.4) and (5.5), yields

$$\lim_{\varepsilon \downarrow 0} \|F_\varepsilon^i - \xi^i(\rho) - L_\rho^i\|_\rho^2 = 0.$$

This implies that a ρ -independent Hermitian operator $F^i := \lim_{\varepsilon \downarrow 0} F_\varepsilon^i$ exists and satisfies $L_\rho^i = F^i - \xi^i(\rho)$ for every $\rho \in \mathcal{M}$, which concludes (5.3).

We next show (5.3) $\Rightarrow$ (ii). Let $\mathbf{u} = (u_i) \in \mathbb{R}^n$ be arbitrarily given, for which we will construct $\mathbf{u}\mathbf{u}^T$ -efficient filtration by assuming the existence of $\{F^i\}$ of (5.3). We can assume $\mathbf{u} \neq 0$, and take a basis $\{\mathbf{u}^1, \dots, \mathbf{u}^n\}$, $\mathbf{u}^k = (u_i^k)$, of $\mathbb{R}^n$ such that $\mathbf{u}^1 = \mathbf{u}$, whereby for each k we define

$$\mathbf{Y}^k := \sum_i u_i^k F^i. \quad (5.6)$$

Let $\varepsilon \in (0, 1)$, and take a positive probability vector $(p_1, \dots, p_n)$ such that $p_1 = 1 - \varepsilon$. We define the estimator Π_ε by the following estimation procedure: randomly choose $k \in \{1, \dots, n\}$ according to the probability distribution $(p_1, \dots, p_n)$, measure the observable Y^k to get an outcome y , and then estimate ξ by $\hat{\xi} = g(k, y)$ using the function $g = (g^1, \dots, g^n)^T$ defined by

$$g^i(k, y) := \frac{w_k^i}{p_k} y, \quad (5.7)$$

where $[w_k^i] = [u_i^k]^{-1}$:

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sum_i u_i^k F^i} \xrightarrow{\text{meas.}} y \\ \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \frac{y}{p_k} \mathbf{w}_k. \quad (5.8)$$

Invoking (5.3) evaluated at an arbitrary point $\rho \in \mathcal{M}$, we have

$$\begin{aligned} g^i(k, Y^k) &= \frac{w_k^i}{p_k} \sum_j u_j^k F^j = \frac{w_k^i}{p_k} \sum_j u_j^k (L_\rho^j + \xi^j(\rho)) \\ &= \gamma_k^i + \frac{w_k^i}{p_k} X^k = f^i(k, X^k), \end{aligned}$$

where X^k and f^i are those defined by (4.11) and (4.12) with

$$\gamma_k^i := \frac{w_k^i}{p_k} \sum_j u_j^k \xi^j(\rho) = \frac{w_k^i}{p_k} \mathbf{u}^k \cdot \xi(\rho). \quad (5.9)$$

Namely, (5.8) is equivalent to the following:

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sum_i u_i^k L_\rho^i} \xrightarrow{\text{meas.}} x \\ \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \frac{\mathbf{u}^k \cdot \xi(\rho) + x}{p_k} \mathbf{w}_k. \quad (5.10)$$

Since this γ_k^i satisfies (4.10), Lemma 4.3 applies to conclude that Π_ε is locally unbiased at ρ and satisfies, for every k and every $\rho \in \mathcal{M}$,

$$(\mathbf{u}^k)^T V_\rho(\Pi_\varepsilon) \mathbf{u}^k = \frac{1}{p_k} (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k + \sum_l p_l (a_l^k)^2, \quad (5.11)$$

where $a_l^k := \sum_i u_i^k (\gamma_l^i - \xi^i(\rho))$. From (5.9), we have

$$a_l^k = \frac{1}{p_l} \sum_{i,j} u_i^k w_l^j u_j^l \xi^j(\rho) - \sum_i u_i^k \xi^i(\rho) = \left(\frac{\delta_l^k}{p_k} - 1 \right) \sum_i u_i^k \xi^i(\rho)$$

and hence

$$\sum_l p_l (a_l^k)^2 = \sum_l p_l \left(\frac{\delta_l^k}{p_k} - 1 \right)^2 \left(\sum_i u_i^k \xi^i(\rho) \right)^2 = \frac{1-p_k}{p_k} \left(\sum_i u_i^k \xi^i(\rho) \right)^2. \quad (5.12)$$

Thus, letting $k = 1$ in (5.11) with (5.12), we obtain

$$\mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} = \frac{1}{1-\varepsilon} \mathbf{u}^T G_\rho^{-1} \mathbf{u} + \frac{\varepsilon}{1-\varepsilon} \left(\sum_i u_i^k \xi^i(\rho) \right)^2.$$

This implies that $\lim_{\varepsilon \downarrow 0} \mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} = \mathbf{u}^T G_\rho^{-1} \mathbf{u}$ for every ρ and that $\vec{\Pi} := (\Pi_\varepsilon)_{\varepsilon \in (0,1)}$ is a $\mathbf{u}\mathbf{u}^T$ -efficient filtration. $\square$

Remark 5.2 The term $\frac{1-p_k}{p_k} \left(\sum_i u_i^k \xi^i(\rho) \right)^2$ in (5.11) with (5.12), which does not appear in (4.16), is an additional estimation error incurred to ensure that the estimation procedure does not depend on the state $\rho \in \mathcal{M}$.

The following proposition follows immediately from Theorem 5.1 and Proposition 3.11.

Proposition 5.3 *In the situation of Proposition 3.11 where $(\tilde{\mathcal{M}}, \tilde{\xi})$ is the k th i.i.d. extension of $(\mathcal{M}, \xi)$, $(\tilde{\mathcal{M}}, \tilde{\xi})$ has an efficient filtration if and only if $(\mathcal{M}, \xi)$ has an efficient filtration.*

6 A sufficient condition for the e-autoparallelity and its relation to the Gaussian states

Proposition 6.1 *For a model $(\mathcal{M}, \xi)$, the following condition is sufficient for (i) and (ii) of Theorem 5.1:*

- *For every positive weight $W > 0$, there exists a W -efficient estimator for ξ .*

Proof Given $\mathbf{u} \in \mathbb{R}^n$ and $\varepsilon > 0$, arbitrarily, let Π_ε be a $(\mathbf{u}\mathbf{u}^T + \varepsilon I)$ -efficient estimator. Then, for an arbitrary $\rho \in \mathcal{M}$ and an arbitrary $\Pi' \in \mathcal{U}(\rho, \xi)$, we have

$$\begin{aligned} \mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} &\leq \text{tr}((\mathbf{u}\mathbf{u}^T + \varepsilon I) V_\rho(\Pi_\varepsilon)) \\ &\leq \text{tr}((\mathbf{u}\mathbf{u}^T + \varepsilon I) V_\rho(\Pi')) = \mathbf{u}^T V_\rho(\Pi') \mathbf{u} + \varepsilon \text{tr}(V_\rho(\Pi')). \end{aligned}$$

This implies that $\lim_{\varepsilon \downarrow 0} \mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} \leq \mathbf{u}^T V_\rho(\Pi') \mathbf{u}$ for every $\Pi' \in \mathcal{U}(\rho, \xi)$, so that the filtration $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon > 0}$ is $\mathbf{u}\mathbf{u}^T$ -efficient. $\square$

An important example of a model satisfying the condition of Proposition 6.1 is given by a model consisting of quantum Gaussian states. Strictly speaking, the model

is mathematically out of our scope in that the underlying Hilbert space is infinite-dimensional. Nevertheless, it is worthwhile to consider the relationship between this important model and the e-autoparallelity, even at the expense of some rigor.

We begin by quickly reviewing the definition of quantum Gaussian states based on the description in Chapter 5 of [10]. Let Z be an even-dimensional real linear space on which a symplectic form $\Delta(\cdot, \cdot)$ is given, and $U(\cdot)$ be an irreducible projective representation of (Z, Δ) on a separable Hilbert space $\mathcal{H}$; i.e., $\{U(z) \mid z \in Z\}$ is a family of unitary operators on $\mathcal{H}$ satisfying $\forall z, z' \in Z, U(z)U(z') = e^{\sqrt{-1}\Delta(z, z')/2}U(z + z')$ and having no nontrivial invariant subspace. For each $z \in Z$, a self-adjoint operator $R(z)$ is defined by $U(tz) = e^{\sqrt{-1}tR(z)}, t \in \mathbb{R}$, and satisfies

$$\forall z, z' \in Z, [R(z), R(z')] = -\sqrt{-1}\Delta(z, z')I. \quad (6.1)$$

Given a symmetric bilinear form $\alpha(\cdot, \cdot)$ on Z satisfying $\forall z, z' \in Z, \alpha(z, z) + \alpha(z', z') \geq \Delta(z, z')$ and a linear functional $\mu(\cdot)$ on Z , there uniquely exists a density operator ρ on $\mathcal{H}$ such that

$$\forall z \in Z, \text{Tr}(\rho U(z)) = e^{\sqrt{-1}\mu(z) - \frac{1}{2}\alpha(z, z)}. \quad (6.2)$$

This ρ is called the *Gaussian state* determined by (μ, α) , and satisfies

$$\mu(z) = \langle I, R(z) \rangle_\rho = \langle R(z) \rangle_\rho, \quad (6.3)$$

$$\alpha(z, z') = \langle R(z) - \mu(z), R(z') - \mu(z') \rangle_\rho, \quad (6.4)$$

where $\langle \cdot, \cdot \rangle_\rho$ denotes the symmetrized inner product (3.22).

Assuming that α and linearly independent $\mu_1, \dots, \mu_n$ are arbitrarily given, consider the model $\mathcal{M} = \{\rho_\xi \mid \xi = (\xi^1, \dots, \xi^n) \in \mathbb{R}^n\}$, where ρ_ξ is the Gaussian state determined by (μ_ξ, α) with $\mu_\xi := \sum_i \xi^i \mu_i$. We call $\mathcal{M} = \{\rho_\xi\}$ a *quantum Gaussian shift model*. Holevo showed (§6.9 of [10]) that the model has W -efficient estimator for every positive weight W . Namely, the sufficient condition presented in Proposition 6.1 is fulfilled. Hence, if $\mathcal{H}$ were finite-dimensional, it would have been concluded that $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$ and that ξ is m-affine as a coordinate system of $\mathcal{M}$. In reality, however, $\mathcal{H}$ is infinite-dimensional, and the e-autoparallelity for a model in $\mathcal{S}$ is not given a mathematical definition in the framework of the present paper. Nevertheless, there is no essential difference from the finite-dimensional case. In fact, we can verify that the model $(\mathcal{M}, \xi)$ satisfies conditions (ii)-(iv) of Proposition 3.8 as follows. According to [10], the i th SLD $L_{i, \xi}$ is represented as

$$L_{i, \xi} = R(z_i) - \mu_\xi(z_i), \quad (6.5)$$

where z_i is the element of Z determined by the condition $\forall z \in Z, \mu_i(z) = \alpha(z_i, z)$. The SLD Fisher information matrix $G = [g_{ij}]$ is given by

$$g_{ij} = \langle L_{i, \xi}, L_{j, \xi} \rangle_\xi = \alpha(z_i, z_j), \quad (6.6)$$

which does not depend on the parameter ξ . Letting $F^i := \sum_j g^{ij} R(z_j)$, where $G^{-1} = [g^{ij}]$, we have

$$\sum_j g^{ij} L_{j,\xi} = F^i - \sum_j g^{ij} \mu_\xi(z_j) = F^i - \xi^i, \quad (6.7)$$

where the second equality follows from $\mu_\xi(z_j) = \sum_k \xi^k \mu_k(z_j)$ and $\mu_k(z_j) = \alpha(z_k, z_j) = g_{kj}$. We thus have verified that $(\mathcal{M}, \xi)$ satisfies condition (iii) of Proposition 3.8, which is evidently equivalent to (ii) and (iv) even in this case. Hence, we may consider that the model also satisfies condition (i) at least in a naive sense. In order to mathematically justify this consideration, we need a rigorous treatment of $\mathcal{S}(\mathcal{H})$ as an infinite-dimensional manifold equipped with an information-geometric structure, which is out of scope of the present paper.

The fact that $G = [g_{ij}]$ is constant on $\mathcal{M}$ means that $(\mathcal{M}, g)$ is a Euclidean manifold and the m-affine coordinate system ξ is also affine w.r.t. the flat Levi-Civita connection of g . This implies that $\mathcal{M}$ is dually flat and that the e, m-connections on $\mathcal{M}$ coincide with the Levi-Civita connection. Note also that the SLDs $\{L_{i,\xi}\}$ do not commute and hence $\mathcal{M}$ has no efficient estimator.

Remark 6.2 Let α be arbitrarily fixed, and let $\mathcal{N} := \{\rho_\mu \mid \mu \in Z^*\}$, where Z^* denotes the dual linear space of Z and ρ_μ denotes the Gaussian state determined by (μ, α) . This is a special (maximal) case of $\mathcal{M}$ treated above, so that $\mathcal{N}$ is “e-autoparallel in $\mathcal{S}$ ” in the same naive sense. The SLD structure of $\mathcal{N}$ is Euclidean, and the model $\mathcal{M} = \{\rho_\xi \mid \xi \in \mathbb{R}^n\}$ treated above, where $\mu_\xi = \sum_i \xi^i \mu_i$, forms an e,m-autoparallel submanifold of $\mathcal{N}$. Generally, a submanifold $\mathcal{M}$ of $\mathcal{N}$ is e,m-autoparallel in $\mathcal{N}$ iff there exists an affine subspace $\mathcal{A}$ of Z^* such that $\mathcal{M} = \{\rho_\mu \mid \mu \in \mathcal{A}\}$, which is represented as $\mathcal{M} = \{\rho_\xi \mid \xi \in \mathbb{R}^n\}$ with $\mu_\xi = \mu_0 + \sum_{i=1}^n \xi^i \mu_i$. Note that the construction of W -efficient estimator for the case $\mu_0 = 0$ which was established by Holevo is immediately applied to this extended model, so that it satisfies the sufficient condition of Proposition 6.1.

7 Characterization of e-autoparallelity in terms of estimation of scalar functions

In this section we give another characterization to the e-autoparallelity w.r.t. the SLD structure by considering a different type of estimation problem. Before we get into the main discussion, some preliminaries on geometrical language are in order.

On a general Riemannian manifold (M, g) , a one-to-one correspondence between a tangent vector $X_p \in T_p(M)$ and a cotangent vector $\omega_p \in T_p^*(M)$ at a point $p \in M$ is naturally defined; denoting the correspondence by $\overset{g_p}{\longleftrightarrow}$, we have

$$X_p \overset{g_p}{\longleftrightarrow} \omega_p \Leftrightarrow \forall Y_p \in T_p(M), \quad \omega_p(Y_p) = g_p(X_p, Y_p). \quad (7.1)$$

In terms of the Dirac notation, $X_p \overset{g_p}{\longleftrightarrow} \omega_p$ means that $X_p = |X_p\rangle$ and $\omega_p = \langle X_p|$ for the inner product $\langle \cdot | \cdot \rangle = g_p$. This is extended to the correspondence $\overset{g}{\longleftrightarrow}$ between a

vector field $X \in \mathcal{X}(M)$ and a differential 1-form $\omega \in \mathcal{D}(M)$, where $\mathcal{D}(M)$ denotes the totality of 1-forms on M , such that

$$\begin{aligned} X \xleftrightarrow{g} \omega &\Leftrightarrow \forall p \in M, X_p \xleftrightarrow{g_p} \omega_p \\ &\Leftrightarrow \forall Y \in \mathcal{X}(M), \omega(Y) = g(X, Y). \end{aligned} \quad (7.2)$$

When a coordinate system $\xi = (\xi^i)$ is given on M , and $X \in \mathcal{X}(M)$ and $\omega \in \mathcal{D}(M)$ are represented as $X = \sum_i X^i \partial_i$ and $\omega = \sum_j \omega_j d\xi^j$ by functions $\{X^i\}, \{\omega_j\} \subset \mathcal{F}(M)$, we have

$$X \xleftrightarrow{g} \omega \Leftrightarrow \forall j, \omega_j = \sum_i X^i g_{ij} \Leftrightarrow \forall i, X^i = \sum_j \omega_j g^{ij}, \quad (7.3)$$

where $g_{ij} = g(\partial_i, \partial_j)$ and $[g^{ij}] = [g_{ij}]^{-1}$.

For a function $f \in \mathcal{F}(M)$, its *gradient* w.r.t. g is defined as the vector field $X \in \mathcal{X}(M)$ such that $X \xleftrightarrow{g} df$, which we denote by $X = \text{grad } f$. This is represented as

$$\text{grad } f = \sum_{i,j} (\partial_i f) g^{ij} \partial_j. \quad (7.4)$$

The correspondence $\xleftrightarrow{g_p}$ induces an inner product and a norm on the cotangent space $T_p^*(M)$ such that $\xleftrightarrow{g_p}$ is an isometry; i.e., $X_p \xleftrightarrow{g_p} \omega_p \Rightarrow \|X_p\|_p = \|\omega_p\|_p$. In particular, we have

$$\|(\text{grad } f)_p\|_p^2 = \|(df)_p\|_p^2 = \sum_{i,j} g^{ij}(p) \partial_i f(p) \partial_j f(p). \quad (7.5)$$

Now, we are ready to start the main discussion of this section. Let $\mathcal{M}$ be an n -dimensional submanifold of $\mathcal{S} = \mathcal{S}(\mathcal{H})$, and $f \in \mathcal{F}(\mathcal{M})$ be a smooth function on it. We consider the problem of estimating the scalar value $f(\rho)$ for unknown $\rho \in \mathcal{M}$. An estimator is generally represented by a POVM $\Lambda = \Lambda(d\hat{t})$, where $\hat{t}$ is a scalar variable representing an estimate for $t = f(\rho)$. The expectation $E_\rho(\Lambda)$ and the mean squared error (the variance in the unbiased case) $V_\rho(\Lambda)$ of Λ for a state ρ are defined by

$$E_\rho(\Lambda) := \int \hat{t} \text{Tr}(\rho \Lambda(d\hat{t})), \quad (7.6)$$

$$V_\rho(\Lambda) := \int (\hat{t} - f(\rho))^2 \text{Tr}(\rho \Lambda(d\hat{t})). \quad (7.7)$$

Remark 7.1 When f is a component, say ξ^1 , of a coordinate system $\xi = (\xi^i)$ of $\mathcal{M}$, our problem treated here can be regarded as an example of estimation with nuisance parameters (see e.g. [6]), with ξ^1 being the parameter of interest and $\xi^2, \dots, \xi^n$ being the nuisance parameters. Note, however, that the problem itself depends only on $f =$

ξ^1 , not on the nuisance parameters, and that in our coordinate-free description, the nuisance parameters as well as their parameter orthogonalization [6] do not appear explicitly. See Remark 7.4 below.

Localizing the unbiasedness condition $E(\Lambda) = f$, where the LHS denotes the function $\mathcal{M} \rightarrow \mathbb{R}$, $\rho \mapsto E_\rho(\Lambda)$, we say that Λ is *locally unbiased* for f at $\rho \in \mathcal{M}$ when

$$E_\rho(\Lambda) = f(\rho) \quad \text{and} \quad (dE(\Lambda))_\rho = (df)_\rho. \quad (7.8)$$

When a coordinate system $\xi = (\xi^i)$ is arbitrarily given on $\mathcal{M}$, the second condition in (7.8) is expressed as

$$\forall i \in \{1, \dots, n\}, \quad \partial_i E_\rho(\Lambda) = \partial_i f(\rho), \quad (7.9)$$

where $\partial_i E_\rho(\Lambda)$ and $\partial_i f(\rho)$ denote the derivatives of the functions $E(\Lambda)$ and f by $\partial_i = \frac{\partial}{\partial \xi^i}$ evaluated at ρ . We denote by $\mathcal{U}(\rho, f)$ the totality of locally unbiased estimators for f at ρ .

Proposition 7.2 *For any $f \in \mathcal{F}(\mathcal{M})$ and any $\rho \in \mathcal{M}$, we have*

$$\min_{\Lambda \in \mathcal{U}(\rho, f)} V_\rho(\Lambda) = \|(df)_\rho\|_\rho^2, \quad (7.10)$$

where $\|\cdot\|_\rho$ denotes the norm for cotangent vectors w.r.t. the SLD metric on $\mathcal{M}$. The minimum of (7.10) is achieved by the spectral measure of the observable

$$F_\rho := f(\rho) + \sum_i \partial_i f(\rho) L_\rho^i, \quad (7.11)$$

where $L_\rho^i := \sum_j g^{ij}(\rho) L_{i,\rho}$ and $L_i := L_{\partial_i}$.

Proof Given an estimator Λ , let

$$A := \int \hat{t} \Lambda(d\hat{t}) \in \mathcal{L}_h.$$

Then the local unbiasedness of Λ at ρ is represented as

$$\langle A \rangle_\rho = f(\rho) \quad \text{and} \quad \forall i, \quad \partial_i \langle A \rangle_\rho = \partial_i f(\rho), \quad (7.12)$$

and we have

$$V_\rho(\Lambda) \geq \left\langle (A - f(\rho))^2 \right\rangle_\rho = \|A - f(\rho)\|_\rho^2, \quad (7.13)$$

where $\|\cdot\|_\rho$ denotes the norm w.r.t. the symmetrized inner product $\langle \cdot, \cdot \rangle_\rho$ on $\mathcal{L}_h$. Noting that the second condition of (7.12) is equivalent to

$$\forall i, \quad \langle L_{i,\rho}, A - f(\rho) \rangle_\rho = \langle L_{i,\rho}, A \rangle_\rho = \left\langle L_{i,\rho}, \sum_j \partial_j f(\rho) L_\rho^j \right\rangle_\rho,$$

where the first equality follows from (3.9), we see that $\sum_j \partial_j f(\rho) L_\rho^j$ is the orthogonal projection of $A - f(\rho)$ onto the space $\text{span}\{L_{j,\rho}\}_{j=1}^n$. Hence we have

$$\begin{aligned} \|A - f(\rho)\|_\rho^2 &= \left\| \sum_j \partial_j f(\rho) L_\rho^j \right\|_\rho^2 + \|A - f(\rho) - \sum_j \partial_j f(\rho) L_\rho^j\|_\rho^2 \\ &\geq \left\| \sum_j \partial_j f(\rho) L_\rho^j \right\|_\rho^2 \\ &= \sum_{i,j} g^{ij}(\rho) \partial_i f(\rho) \partial_j f(\rho) = \|(df)_\rho\|_\rho^2. \end{aligned} \quad (7.14)$$

The inequality in (7.13) holds with equality when Λ is the spectral measure of A , and the inequality in (7.14) holds with equality when $A = f(\rho) + \sum_j \partial_j f(\rho) L_\rho^j = F_\rho$. These observations prove the proposition. $\square$

Remark 7.3 Proposition 7.2 is essentially equivalent to Theorem 7.2 of [3] when the symmetrized inner product (3.22) is taken as what is called a generalized inner product in [3] (which is the same as a family of inner products $\{\langle \cdot, \cdot \rangle_\rho \mid \rho \in \mathcal{S}(\mathcal{H})\}$ introduced in Sect. 3 of the present paper).

Remark 7.4 Letting $f = \xi^1$ in (7.10), we have

$$\min_{\Lambda \in \mathcal{U}(\rho, \xi^1)} V_\rho(\Lambda) = \|(d\xi^1)_\rho\|_\rho^2 = g^{11}(\rho). \quad (7.15)$$

This formula is equivalent to Theorem 5.3 of [6], where $g^{11}(\rho)$ is expressed as the inverse of the Schur complement $\tilde{g}_{11}(\rho)$ of the matrix $G_\rho = [g_{ij}(\rho)]$. Letting $(\tilde{\partial}_1)_\rho$ be the g_ρ -projection of $(\partial_1)_\rho$ onto the orthogonal complement of $\text{span}\{(\partial_2)_\rho, \dots, (\partial_n)_\rho\}$, the Schur complement is represented as $\tilde{g}_{11}(\rho) = g_\rho((\tilde{\partial}_1)_\rho, (\tilde{\partial}_1)_\rho) = \left\langle \tilde{L}_{1,\rho}, \tilde{L}_{1,\rho} \right\rangle_\rho$, where $\tilde{L}_{1,\rho} := L_{(\tilde{\partial}_1)_\rho}$, and the conversion from $(\partial_1)_\rho$ to $(\tilde{\partial}_1)_\rho$, or from $L_{1,\rho}$ to $\tilde{L}_{1,\rho}$, is called the local parameter orthogonalization in [6].

Based on Proposition 7.2, we call an estimator Λ *locally efficient* for f at ρ when $\Lambda \in \mathcal{U}(\rho, f)$ and $V_\rho(\Lambda) = \|(df)_\rho\|_\rho^2$, and call it *efficient* for f when it is locally efficient for f at every $\rho \in \mathcal{M}$. Note that, unlike the case of estimation for multi-dimensional coordinates (ξ^i) where the infimum in (4.9) cannot be replaced with minimum in general, there always exists a locally efficient estimator for a scalar function f at each ρ . Furthermore, since a locally efficient estimator is obtained as the spectral measure of an observable as is shown in the proof of Proposition 7.2, it suffices to treat only estimators represented by Hermitian operators. Note that an estimator $F \in \mathcal{L}_h$ is efficient for f iff

$$\forall \rho \in \mathcal{M}, \langle F \rangle_\rho = f(\rho) \quad \text{and} \quad V_\rho(F) := \langle (F - \langle F \rangle_\rho)^2 \rangle_\rho = \|(df)_\rho\|_\rho^2. \quad (7.16)$$

We define

$$\mathcal{E}(\mathcal{M}) := \{f \in \mathcal{F}(\mathcal{M}) \mid \text{there exists an efficient estimator for } f\}. \quad (7.17)$$

Proposition 7.5 For a function $f \in \mathcal{F}(\mathcal{M})$, the following conditions are equivalent.

- (i) $f \in \mathcal{E}(\mathcal{M})$.
- (ii) $\exists F \in \mathcal{L}_h$, $F - f = \sum_i (\partial_i f) L^i$.
- (iii) $\exists F \in \mathcal{L}_h$, $F - \langle F \rangle|_{\mathcal{M}} = \sum_i (\partial_i f) L^i$.
- (iv) $\text{grad } f$ is e -parallel (i.e. parallel w.r.t. the e -connection on $\mathcal{S}$).

In (ii), the observable F gives an efficient estimator for f .

Proof From Proposition 7.2, it immediately follows that (i) $\Leftrightarrow$ (ii) and that F in (ii) gives an efficient estimator for f .

It is obvious that (ii) $\Rightarrow$ (iii). To show the converse, assume that an operator $F \in \mathcal{L}_h$ satisfies $F - \langle F \rangle|_{\mathcal{M}} = \sum_i (\partial_i f) L^i$. Then we have

$$\partial_i \langle F \rangle = \langle L_i, F \rangle = \langle L_i, F - \langle F \rangle|_{\mathcal{M}} \rangle = \left\langle L_i, \sum_j (\partial_j f) L^j \right\rangle = \partial_i f,$$

which implies that $\exists c \in \mathbb{R}$, $f = \langle F \rangle|_{\mathcal{M}} + c$. Redefining $F := F + c$, we have $F - f = \sum_i (\partial_i f) L^i$. This proves (iii) $\Rightarrow$ (ii).

Let $X := \text{grad } f$. Then (7.4) yields

$$L_X = \sum_{i,j} (\partial_i f) g^{ij} L_j = \sum_i (\partial_i f) L^i,$$

and Cor. 3.6 yields

$$X \text{ is } e\text{-parallel} \Leftrightarrow \exists F \in \mathcal{L}_h, L_X = F - \langle F \rangle|_{\mathcal{M}}.$$

Thus we obtain (iii) $\Leftrightarrow$ (iv). □

Corollary 7.6 $\mathcal{E}(\mathcal{M})$ is an $\mathbb{R}$ -linear space.

Proof Obvious from Proposition 7.5. □

Proposition 7.7 For a vector field $X \in \mathcal{X}(\mathcal{M})$, we have

$$X \text{ is } e\text{-parallel} \Leftrightarrow \exists f \in \mathcal{E}(\mathcal{M}), X = \text{grad } f. \quad (7.18)$$

Proof The implication $\Leftarrow$ follows from (i) $\Rightarrow$ (iv) in Proposition 7.5. To show the converse, assume that X is e -parallel. Then, according to Cor. 3.6, there exists $F \in \mathcal{L}_h$ such that $L_X = F - f$, where $f := \langle F \rangle|_{\mathcal{M}}$. For any $Y \in \mathcal{X}(\mathcal{M})$ we have

$$g(X, Y) = \langle L_Y, F - f \rangle = \langle L_Y, F \rangle \stackrel{*}{=} Y \langle F \rangle|_{\mathcal{M}} = Yf = df(Y), \quad (7.19)$$

where $\stackrel{*}{=}$ follows from (3.7). This implies that $X = \text{grad } f$. Since X is e -parallel, it follows from (iv) $\Rightarrow$ (i) in Proposition 7.5 that $f \in \mathcal{E}(\mathcal{M})$. Thus, $\Rightarrow$ in (7.18) has been verified. □

Define

$$d\mathcal{E}(\mathcal{M}) := \{df \mid f \in \mathcal{E}(\mathcal{M})\} \subset \mathcal{D}(\mathcal{M}). \quad (7.20)$$

Since $df = df' \Leftrightarrow f - f' = \text{const.}$, we have the natural identification $d\mathcal{E}(\mathcal{M}) \simeq \mathcal{E}(\mathcal{M})/\mathbb{R}$. We also define

$$\mathcal{X}_{\text{e-par}}(\mathcal{M}) := \{X \in \mathcal{X}(\mathcal{M}) \mid X \text{ is e-parallel}\}. \quad (7.21)$$

Then we have the following proposition.

Proposition 7.8 *The correspondence $\xleftrightarrow{g|_{\mathcal{M}}}$ establishes a linear isomorphism between $\mathcal{X}_{\text{e-par}}(\mathcal{M})$ and $d\mathcal{E}(\mathcal{M})$. As a consequence, we have*

$$\dim d\mathcal{E}(\mathcal{M}) = \dim \mathcal{E}(\mathcal{M}) - 1 = \dim \mathcal{X}_{\text{e-par}}(\mathcal{M}) \leq \dim \mathcal{M}. \quad (7.22)$$

Proof It suffices to show that for an arbitrary pair $(X, \omega) \in \mathcal{X}(\mathcal{M}) \times \mathcal{D}(\mathcal{M})$ satisfying $X \xleftrightarrow{g|_{\mathcal{M}}} \omega$, the following equivalence holds:

$$X \in \mathcal{X}_{\text{e-par}}(\mathcal{M}) \Leftrightarrow \exists f \in \mathcal{E}(\mathcal{M}), \omega = df. \quad (7.23)$$

Since $\omega = df \Leftrightarrow X = \text{grad } f$, this is just Proposition 7.7. $\square$

Now, we present two theorems for characterization of the e-autoparallelity in terms of $\mathcal{E}(\mathcal{M})$.

Theorem 7.9 *For an n -dimensional submanifold $\mathcal{M}$ of S , the following conditions are equivalent.*

- (i) $\mathcal{M}$ is e-autoparallel in S .
- (ii) $\dim \mathcal{E}(\mathcal{M}) = n + 1$.

Proof We have (i) $\Leftrightarrow \dim \mathcal{X}_{\text{e-par}}(\mathcal{M}) = n$ by Proposition 2.4, and $\dim \mathcal{X}_{\text{e-par}}(\mathcal{M}) = n \Leftrightarrow$ (ii) by Proposition 7.8. $\square$

Theorem 7.10 *For an n -dimensional model $(\mathcal{M}, \xi)$, the following conditions are equivalent.*

- (i) $\mathcal{M}$ is e-autoparallel in S , and ξ is an m -affine coordinate system.
- (ii) $\forall i \in \{1, \dots, n\}, \xi^i \in \mathcal{E}(\mathcal{M})$.
- (iii) $\mathcal{E}(\mathcal{M}) = \left\{ c + \sum_{i=1}^n u_i \xi^i \mid (c, u_1, \dots, u_n) \in \mathbb{R}^{n+1} \right\}$.

Proof Let $X^i := \text{grad } \xi^i = \sum_j g^{ij} \partial_j$. Then we have

$$(i) \Leftrightarrow \forall i, X^i \in \mathcal{X}_{\text{e-par}}(\mathcal{M}) \Leftrightarrow (ii),$$

where the first $\Leftrightarrow$ follows from Proposition 2.6 and the second $\Leftrightarrow$ follows from Proposition 7.5.

Next, noting that constant functions on $\mathcal{M}$ belong to $\mathcal{E}(\mathcal{M})$, we have

$$\begin{aligned} \text{(ii)} &\Leftrightarrow \left\{ c + \sum_{i=1}^n u_i \xi^i \mid (c, u_1, \dots, u_n) \in \mathbb{R}^{n+1} \right\} \subset \mathcal{E}(\mathcal{M}) \\ &\Leftrightarrow \text{(iii),} \end{aligned}$$

where the second $\Leftrightarrow$ follows since $\dim \mathcal{E}(\mathcal{M}) \leq n+1$ by (7.22) and $\{1, \xi^1, \dots, \xi^n\}$ are linearly independent. $\square$

Remark 7.11 Suppose that condition (ii) in Theorem 7.10 holds, and let $F^1, \dots, F^n$ be the efficient estimators of $\xi^1, \dots, \xi^n$, respectively, which means that $\langle F^i \rangle|_{\mathcal{M}} = \xi^i$ and $\langle (F^i - \xi^i)^2 \rangle|_{\mathcal{M}} = g^{ii}$ hold for every i . If, in addition, the operators $F^1, \dots, F^n$ commute, their simultaneous measurement $\Pi(d\hat{\xi})$ defines an efficient estimator of the coordinate system ξ . Conversely, if $\Pi(d\hat{\xi})$ is an efficient estimator of ξ , using a similar argument to the proof of Theorem 5.1, we can show that the operator $F^i := \int \hat{\xi}^i \Pi(d\hat{\xi})$ satisfies $F^i - \xi^i(\rho) = L_\rho^i$ for every $\rho \in \mathcal{M}$ and every i , which implies that F^i is an efficient estimator of ξ^i and that $F^1, \dots, F^n$ commute. In this sense, Theorem 7.10 can be regarded as an extension of Theorem 1.3.

Remark 7.12 If we replace $\mathcal{S}(\mathcal{H})$ by $\mathcal{P}(\Omega)$ in Theorems 7.9 and 7.10, these theorems hold as they are in the classical case. When the coordinate functions $\xi^1, \dots, \xi^n$ satisfy condition (ii) in Theorem 7.10, they have their efficient estimators $F^1, \dots, F^n$, which are functions $\Omega \rightarrow \mathbb{R}$ in this case, and the map $(F^1, \dots, F^n) : \Omega \rightarrow \mathbb{R}^n$ becomes an efficient estimator for $\xi = (\xi^1, \dots, \xi^n)$. Thus, we see that the equivalence (i) $\Leftrightarrow$ (ii) in the theorem becomes just Theorem 1.1.

Finally, we present three propositions that will aid in understanding the above results in a purely geometric context, whose proofs are given in Appendix D.

Proposition 7.13 $\forall F \in \mathcal{L}_h, \forall \rho \in \mathcal{S}, V_\rho(F) := \langle (F - \langle F \rangle_\rho)^2 \rangle_\rho = \|(d\langle F \rangle)_\rho\|_\rho^2$.

Proposition 7.14 We have

$$\begin{aligned} \mathcal{E}(\mathcal{S}) &= \{\langle F \rangle \mid F \in \mathcal{L}_h\} \\ &= \{f \in \mathcal{F}(\mathcal{S}) \mid \text{grad } f \text{ is } e\text{-parallel}\} \\ &= \{f \in \mathcal{F}(\mathcal{S}) \mid df \text{ is } m\text{-parallel}\}, \end{aligned} \quad (7.24)$$

where a 1-form $\omega \in \mathcal{D}(\mathcal{S})$ is said to be m -parallel when

$$\forall X, Y \in \mathcal{X}(\mathcal{S}), (\nabla_X^{(m)} \omega)(Y) := X\omega(Y) - \omega(\nabla_X^{(m)} Y) = 0. \quad (7.25)$$

Proposition 7.15 For an arbitrary submanifold $\mathcal{M}$ of $\mathcal{S}$, we have

$$\begin{aligned} \mathcal{E}(\mathcal{M}) &= \{f \in \mathcal{F}(\mathcal{M}) \mid \exists \tilde{f} \in \mathcal{E}(\mathcal{S}), \\ &\quad f = \tilde{f}|_{\mathcal{M}} \text{ and } \forall \rho \in \mathcal{M}, \|(df)_\rho\|_\rho = \|(d\tilde{f})_\rho\|_\rho\}. \end{aligned} \quad (7.26)$$

As these propositions suggest, the discussion for $\mathcal{E}(\mathcal{M})$ given in this section can be extended to a more general geometrical setting. Let us recall the situation treated in Sect. 2 where a manifold S is provided with a Riemannian metric g together with mutually dual curvature-free connections ∇ and ∇^* , and now assume the flatness of ∇^* additionally. We define

$$\begin{aligned}\mathcal{E}(S) &:= \{f \in \mathcal{F}(S) \mid \text{grad } f \text{ is } \nabla\text{-parallel}\} \\ &= \{f \in \mathcal{F}(S) \mid df \text{ is } \nabla^*\text{-parallel}\},\end{aligned}\quad (7.27)$$

where we have invoked the fact that the correspondence $\mathcal{X}(S) \ni X \xleftrightarrow{g} \omega \in \mathcal{D}(S)$ implies (cf. (D2) in Appendix D)

$$X \text{ is } \nabla\text{-parallel} \Leftrightarrow \omega \text{ is } \nabla^*\text{-parallel}.\quad (7.28)$$

Given a submanifold M of S , let

$$\begin{aligned}\mathcal{E}(M) &:= \{f \in \mathcal{F}(M) \mid \exists \tilde{f} \in \mathcal{E}(S), f = \tilde{f}|_M \text{ and} \\ &\quad \forall \rho \in M, \|(df)_\rho\|_\rho = \|(d\tilde{f})_\rho\|_\rho\}.\end{aligned}\quad (7.29)$$

Then it is not difficult to verify that Theorems 7.9 and 7.10 as well as their proofs are extended to this general situation almost as they are.

It should be noted that the flatness of ∇^* is essential in that it ensures $\dim \mathcal{E}(S) = \dim S + 1$. To clarify the role of the flatness, let us consider a more general situation by removing the assumption that ∇ is curvature-free and ∇^* is flat, assuming only that they are dual w.r.t. g . We start from the following general identity: for any 1-form $\omega \in \mathcal{D}(S)$ and any vector fields $X, Y \in \mathcal{X}(S)$, we have

$$\begin{aligned}(d\omega)(X, Y) &:= X\omega(Y) - Y\omega(X) - \omega([X, Y]) \\ &= (\nabla_X^* \omega)(Y) + \omega(\nabla_X^* Y) - (\nabla_Y^* \omega)(X) - \omega(\nabla_Y^* X) - \omega([X, Y]) \\ &= (\nabla_X^* \omega)(Y) - (\nabla_Y^* \omega)(X) + \omega(\mathcal{T}^{(\nabla^*)}(X, Y)),\end{aligned}\quad (7.30)$$

where $\mathcal{T}^{(\nabla^*)}$ denotes the torsion of ∇^* : $\mathcal{T}^{(\nabla^*)}(X, Y) = \nabla_X^* Y - \nabla_Y^* X - [X, Y]$. When ∇^* is torsion-free, this implies that for any $\omega \in \mathcal{D}(S)$

$$\omega \text{ is } \nabla^*\text{-parallel} \Rightarrow d\omega = 0 \Leftrightarrow \exists f \in \mathcal{F}(S), \omega = df\quad (7.31)$$

(see Remark 1.4) and that for any $X \in \mathcal{X}(S)$

$$X \text{ is } \nabla\text{-parallel} \Rightarrow \exists f \in \mathcal{F}(S), X = \text{grad } f.\quad (7.32)$$

This leads to

$$d\mathcal{E}(S) := \{df \mid f \in \mathcal{E}(S)\} = \{\omega \mid \omega \text{ is } \nabla^*\text{-parallel}\}\quad (7.33)$$

and hence

$$\dim \mathcal{E}(S) = \dim\{\omega \mid \omega \text{ is } \nabla^* \text{-parallel}\} + 1 = \dim \mathcal{X}_{\nabla\text{-par}}(S) + 1, \quad (7.34)$$

where $\mathcal{X}_{\nabla\text{-par}}(S)$ denotes the totality of ∇ -parallel vector fields on S . If, in addition, ∇ is curvature-free, which is equivalent to the flatness of ∇^* , then $\dim \mathcal{X}_{\nabla\text{-par}}(S) = \dim S$, and we obtain $\dim \mathcal{E}(S) = \dim S + 1$.

8 Integrability conditions

We begin by considering the general situation where an affine connection ∇ is given on a manifold S . For an arbitrary point $p \in S$ and an arbitrary 1-dimensional subspace V of the tangent space $T_p(S)$, there always exists a ∇ -autoparallel curve, i.e. a ∇ -geodesic, that passes through p in direction V . For the existence of multi-dimensional autoparallel submanifolds, the situation differs greatly depending on whether ∇ is flat or not. When ∇ is flat (curvature-free and torsion-free), the autoparallel submanifolds are those determined by arbitrary affine constraints on ∇ -affine coordinates. This ensures that, for an arbitrary point $p \in S$ and an arbitrary linear subspace V of the tangent space $T_p(S)$, there uniquely exists a ∇ -autoparallel submanifold M satisfying $p \in M$ and $T_p(M) = V$. This is the case with the e -connection on the space $\mathcal{P}$ of probability distributions, for which the autoparallel submanifolds are the exponential families. We meet a similar situation in the quantum case if we consider the BKM structure: see Remark 3.2. When ∇ is not flat, on the other hand, the existence of multi-dimensional autoparallel submanifolds is not ensured in general. In this section we investigate conditions for existence of autoparallel submanifolds.

Let us consider the case when ∇ is curvature-free as in the e -connection on $\mathcal{S}(\mathcal{H})$. According to (i) $\Leftrightarrow$ (iv) of Proposition 2.4, an n -dimensional submanifold M of S is ∇ -autoparallel iff there exist n linearly independent ∇ -parallel vector fields $X^1, \dots, X^n \in \mathcal{X}(S)$ such that their restrictions $X^1|_M, \dots, X^n|_M$ belong to $\mathcal{X}(M)$. This means that M is an integral manifold of $\{X^1, \dots, X^n\}$, or equivalently that M is an integral manifold of the n -dimensional distribution

$$\mathcal{V} : S \ni p \mapsto V_p := \text{span}\{X_p^1, \dots, X_p^n\} \subset T_p(S), \quad (8.1)$$

which is ∇ -parallel in the sense that $\forall p, q \in S, \Phi_{p,q}^{(\nabla)}(V_p) = V_q$, where $\Phi_{p,q}^{(\nabla)}$ denotes the parallel transport w.r.t. ∇ .

Proposition 8.1 *Suppose that we are given a manifold S with a curvature-free connection ∇ and an n -dimensional ∇ -parallel distribution $\mathcal{V} : S \ni p \mapsto V_p$. Define*

$$\mathcal{X}(S : \mathcal{V}) := \{X \in \mathcal{X}(S) \mid \forall p \in S, X_p \in V_p\} \quad (8.2)$$

and

$$\mathcal{X}_{\nabla\text{-par}}(S : \mathcal{V}) := \{X \in \mathcal{X}(S : \mathcal{V}) \mid X \text{ is } \nabla\text{-parallel}\}. \quad (8.3)$$

Then the following conditions are equivalent.

- (i) For every $p \in S$, there exists a ∇ -autoparallel submanifold M of S satisfying $p \in M$ and $T_p(M) = V_p$.
- (ii) The distribution $\mathcal{V}$ is involutive in the sense that $\forall X, Y \in \mathcal{X}(S: \mathcal{V})$, $[X, Y] \in \mathcal{X}(S: \mathcal{V})$.
- (iii) $\forall X, Y \in \mathcal{X}_{\nabla\text{-par}}(S: \mathcal{V})$, $[X, Y] \in \mathcal{X}(S: \mathcal{V})$.
- (iv) The torsion $\mathcal{T}^{(\nabla)}$ of ∇ satisfies $\forall p \in S$, $\mathcal{T}_p^{(\nabla)}(V_p \times V_p) \subset V_p$.

Proof (i) is equivalent to the condition that for any point $p \in S$, there exists an integral manifold of $\mathcal{V}$ containing p , and is equivalent to (ii) by the famous Frobenius theorem for integrability.

(ii) $\Rightarrow$ (iii) is obvious, and (iii) $\Rightarrow$ (ii) follows since there exist n linearly independent ∇ -parallel vector fields $\{X_1, \dots, X_n\} \subset \mathcal{X}_{\nabla\text{-par}}(S: \mathcal{V})$, whereby every element of $\mathcal{X}(S: \mathcal{V})$ is expressed as $\sum_i f_i X_i$ by some functions $\{f_1, \dots, f_n\} \subset \mathcal{F}(M)$.

For any ∇ -parallel vector fields X and Y , we have

$$\mathcal{T}^{(\nabla)}(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y] = -[X, Y]. \quad (8.4)$$

Hence (iii) is equivalent to

$$\forall X, Y \in \mathcal{X}_{\nabla\text{-par}}(S: \mathcal{V}), \quad \mathcal{T}^{(\nabla)}(X, Y) \in \mathcal{X}(S: \mathcal{V}). \quad (8.5)$$

Since $\mathcal{T}^{(\nabla)}$ is a tensor field so that $(\mathcal{T}^{(\nabla)}(X, Y))_p = \mathcal{T}_p^{(\nabla)}(X_p, Y_p)$ holds at each point p , (8.5) is equivalent to (iv). $\square$

Remark 8.2 Condition (i) in Proposition 8.1 (and hence (ii)-(iv) also) means that there exists a foliation $S = \bigsqcup_{\alpha} M_{\alpha}$ such that each leaf M_{α} is ∇ -autoparallel in S and satisfies $T_p(M_{\alpha}) = V_p$ for every $p \in M_{\alpha}$.

The following proposition is an immediate consequence of (i) $\Leftrightarrow$ (iv) in Proposition 8.1.

Proposition 8.3 For a manifold S with a curvature-free connection ∇ , the following conditions are equivalent.

- (i) For every point $p \in S$ and every linear subspace V of $T_p(S)$, there exists a ∇ -autoparallel submanifold M satisfying $p \in M$ and $T_p(M) = V$.
- (ii) $\forall p \in S$, $\forall X_p, Y_p \in T_p(S)$, $\mathcal{T}_p^{(\nabla)}(X_p, Y_p) \in \text{span}\{X_p, Y_p\}$, or equivalently, $\forall X, Y \in \mathcal{X}(S)$, $\mathcal{T}^{(\nabla)}(X, Y) \in \text{span}_{\mathcal{F}(S)}\{X, Y\} := \{fX + gY \mid f, g \in \mathcal{F}(S)\}$.

Condition (i) in the above proposition means that the condition for existence of autoparallel submanifolds is trivial as in the flat case, which corresponds to the vanishing torsion $\mathcal{T}^{(\nabla)} = 0$ implying condition (ii) obviously.

Remark 8.4 Let us revisit the relationship between autoparallelity and total geodesicity described in Remark 2.2 in the context of Proposition 8.3. Suppose that (S, ∇) satisfies conditions (i)-(ii) of Proposition 8.3 and that a submanifold M of S is ∇ -totally geodesic. Given a point $p \in M$ arbitrarily, there exists a ∇ -autoparallel submanifold

N which satisfies $p \in N$ and $T_p(N) = T_p(M)$ by condition (i). Since N is also ∇ -totally geodesic, we have $M = N$, so that M is ∇ -autoparallel. Namely, condition (ii) together with the curvature-freeness of ∇ implies the equivalence between ∇ -autoparallelity and ∇ -total geodesicity. In fact, the curvature-freeness is unnecessary. We have the following proposition, whose proof is given in Appendix E.

Proposition 8.5 *Suppose that an affine connection ∇ is given on a manifold S whose torsion satisfies*

$$\forall X, Y \in \mathcal{X}(S), \quad T^{(\nabla)}(X, Y) \in \text{span}_{\mathcal{F}(S)}\{X, Y\}. \quad (8.6)$$

Then every ∇ -totally geodesic submanifold of S is ∇ -autoparallel.

Let us apply Proposition 8.1 to $\mathcal{S} = \mathcal{S}(\mathcal{H})$ with the SLD structure and its submanifolds. Let $\mathcal{A}$ be an arbitrary linear subspace of $\mathcal{L}_h$, and define for each point $\rho \in \mathcal{S}$

$$\begin{aligned} V_{\mathcal{A}, \rho} &:= \{X_\rho \in T_\rho(\mathcal{S}) \mid \exists A \in \mathcal{A}, L_{X_\rho} = A - \langle A \rangle_\rho\} \\ &= \{X_\rho \in T_\rho(\mathcal{S}) \mid L_{X_\rho} \in \mathcal{A} + \mathbb{R}\}, \end{aligned} \quad (8.7)$$

where $\mathbb{R}$ is identified with $\{cI \mid c \in \mathbb{R}\}$. Then $\mathcal{V}_{\mathcal{A}} : \mathcal{S} \ni \rho \mapsto V_{\mathcal{A}, \rho}$ defines an e-parallel distribution on $\mathcal{S}$, whose dimension $\dim V_{\mathcal{A}, \rho}$ is equal to $\dim \mathcal{A}$ when $I \notin \mathcal{A}$ and $\dim \mathcal{A} - 1$ otherwise. Every e-parallel distribution on $\mathcal{S}$ is represented as $\mathcal{V}_{\mathcal{A}}$ by some $\mathcal{A}$, and $\mathcal{V}_{\mathcal{A}} = \mathcal{V}_{\mathcal{A}'}$ iff $\mathcal{A} + \mathbb{R} = \mathcal{A}' + \mathbb{R}$. This means that $\mathcal{A} \mapsto \mathcal{V}_{\mathcal{A}}$ establishes a one-to-one correspondence between linear subspaces of the quotient space $\mathcal{L}_h/\mathbb{R}$ and e-parallel distributions on $\mathcal{S}$.

Theorem 8.6 *Given a subspace $\mathcal{A} \subset \mathcal{L}_h$, the following conditions are equivalent.*

- (i) *For every $\rho \in \mathcal{S}$, there exists an e-autoparallel (w.r.t. the SLD structure) submanifold $\mathcal{M}$ of $\mathcal{S}$ satisfying $\rho \in \mathcal{M}$ and $T_\rho(\mathcal{M}) = V_{\mathcal{A}, \rho}$.*
- (ii) *For every $\rho \in \mathcal{S}$,*

$$\{[A, B], \rho \mid A, B \in \mathcal{A}\} \subset \{\rho \circ C \mid C \in \mathcal{A} + \mathbb{R}\}. \quad (8.8)$$

Proof From (3.26) it follows that for any $X_\rho, Y_\rho, Z_\rho \in T_\rho(\mathcal{S})$

$$T_\rho^{(e)}(X_\rho, Y_\rho) = Z_\rho \Leftrightarrow \frac{1}{4}[[L_{X_\rho}, L_{Y_\rho}], \rho] = \iota_*(Z_\rho) = \rho \circ L_{Z_\rho}.$$

Hence, noting that $[A - \langle A \rangle_\rho, B - \langle B \rangle_\rho] = [A, B]$, we obtain (i) $\Leftrightarrow$ (ii) from (i) $\Leftrightarrow$ (iv) in Proposition 8.1. $\square$

We have mentioned in Remark 8.2 that the conditions in Proposition 8.1 imply the existence of a foliation of S with ∇ -autoparallel leaves. If, in addition, the dual connection of ∇ w.r.t. a Riemannian metric g is flat as in the case where $\nabla = \nabla^{(e)}$ and $\nabla^* = \nabla^{(m)}$ for an information-geometric structure $(g, \nabla^{(e)}, \nabla^{(m)})$, then S has also a

foliation with ∇^* -autoparallel leaves, which forms a pair of dual foliations as in the case of dually flat spaces (cf. § 3.7 of [3]). Namely, we have the following proposition, whose proof is given in Appendix F.

Proposition 8.7 *Given (S, g, ∇, ∇^*) , suppose that ∇ is curvature-free and satisfies conditions (i)–(iv) in Proposition 8.1 for a ∇ -parallel distribution $\mathcal{V} : S \ni p \mapsto V_p$ and that ∇^* is flat. Then:*

- (1) *S has two foliations $S = \bigsqcup_{\alpha} M_{\alpha} = \bigsqcup_{\beta} N_{\beta}$ such that:*
 - (a) *$\forall \alpha, M_{\alpha}$ is ∇ -autoparallel in S .*
 - (b) *$\forall \beta, N_{\beta}$ is ∇^* -autoparallel in S .*
 - (c) *The leaves of the two foliations intersect orthogonally; i.e., if $p \in M_{\alpha} \cap N_{\beta}$, then $T_p(M_{\alpha}) \perp T_p(N_{\beta})$ w.r.t. the inner product g_p .*
- (2) *Given ∇ -parallel vector fields $\{X^1, \dots, X^n\}$ spanning $\mathcal{X}_{\nabla\text{-par}}(S : \mathcal{V})$, there exist functions $\{\zeta^1, \dots, \zeta^n\} \subset \mathcal{F}(S)$ such that $\forall i, X^i \xrightarrow{g} d\zeta^i$. Letting $\zeta := (\zeta^i)$, we have:*
 - (a) *$\forall \alpha, \zeta|_{M_{\alpha}}$ forms a $\nabla_{M_{\alpha}}^*$ -affine coordinate system of M_{α} , where $\nabla_{M_{\alpha}}^*$ denotes the dual connection of $\nabla|_{M_{\alpha}}$ w.r.t. $g|_{M_{\alpha}}$ as defined in Sect. 2 for (2.5).*
 - (b) *$\forall \beta, \exists c \in \mathbb{R}^n, N_{\beta} = \{p \in S \mid \zeta(p) = c\}$.*

Corollary 8.8 *Let $\{F^1, \dots, F^n\}$ be Hermitian operators on $\mathcal{H}$ such that $\{F^1, \dots, F^n, I\}$ are linearly independent, and suppose that $\mathcal{A} := \text{span}\{F^i\}$ satisfies condition (ii) in Theorem 8.6, which means that there exist n^3 functions $\{c_k^{ij}\} \subset \mathcal{F}(S)$ such that*

$$\forall i, j, \forall \rho \in S, [[F^i, F^j], \rho] = \sum_k c_k^{ij}(\rho) \rho \circ (F^k - \langle F^k \rangle_{\rho}). \quad (8.9)$$

- (1) *The functions $\{c_k^{ij}\}$ satisfy*

$$T^{(e)}(X^i, X^j) = \frac{1}{4} \sum_k c_k^{ij} X^k, \quad (8.10)$$

where $\{X^i\}$ denote the e -parallel vector fields on S defined by $L_{X^i} = F^i - \langle F^i \rangle$.

- (2) *Given a state $\rho \in S$ arbitrarily, there exists an e -autoparallel submanifold $\mathcal{M}$ of S satisfying $\rho \in \mathcal{M}$ and $\forall i, X^i|_{\mathcal{M}} \in \mathcal{X}(\mathcal{M})$. Letting $\xi^i := \langle F^i \rangle|_{\mathcal{M}}$, $\xi = (\xi^i)$ forms an m -affine coordinate system of $\mathcal{M}$.*
- (3) *In the situation of (2), the manifold*

$$\mathcal{N} := \{\rho' \in S \mid \forall i, \langle F^i \rangle_{\rho'} = \xi^i(\rho)\} \quad (8.11)$$

is m -autoparallel in S and intersects $\mathcal{M}$ orthogonally at ρ .

- (4) *Letting the state ρ range over S in (2) and (3), a pair of dual foliations $\mathcal{S} = \bigsqcup_{\alpha} \mathcal{M}_{\alpha} = \bigsqcup_{\beta} \mathcal{N}_{\beta}$ is generated as in Proposition 8.7.*

Proof (1), (2) and (4) are immediate from the discussion so far. It is not difficult to show (3) directly, while it is also derived from Proposition 8.7 as follows. Let $\zeta^i := \langle F^i \rangle$. Invoking (3.9), (3.12) and (3.18) and noting that $L_{X^i} = F^i - \zeta^i$, we have

$$\forall Y \in \mathcal{X}(S), \quad g(X^i, Y) = \langle F^i - \zeta^i, L_Y \rangle = \langle F^i, L_Y \rangle = Y\zeta^i = d\zeta^i(Y),$$

which means that $X^i \xleftrightarrow{g} d\zeta^i$. Now (3) follows from Proposition 8.7.

Example 8.9 When the operators $F^1, \dots, F^n$ commute in Cor. 8.8, which corresponds to the case when (8.9) holds with vanishing $\{c_k^{ij}\}$, the manifolds $\mathcal{M}$ and $\mathcal{M}_\alpha$ are quasi-classical exponential families of the form (4.20). In particular, for an arbitrary Hermitian operator F such that $\{F, I\}$ are linearly independent, there always exists a pair of dual foliations $\mathcal{S} = \bigsqcup_\alpha \mathcal{M}_\alpha = \bigsqcup_\beta \mathcal{N}_\beta$ such that $\mathcal{M}_\alpha$ is an e-geodesic of the form (4.23) and $\mathcal{N}_\beta$ is an m-autoparallel submanifold of codimension 1 represented by a constraint of the form $\langle F \rangle_\rho = \text{const}$.

A more interesting example of a pair of dual foliations will be given in Sect. 9.

Remark 8.10 Recall the situation when we defined the quantum Gaussian shift model $\mathcal{M} = \{\rho_\xi \mid \xi = (\xi^1, \dots, \xi^n) \in \mathbb{R}^n\}$ in Sect. 6, and let $\mathcal{A} := \text{span}\{R(z_1), \dots, R(z_n)\}$. Then, for any $A, B \in \mathcal{A}$, we have $[A, B] = cI$ with a purely imaginary constant c and hence $[[A, B], \rho] = 0$. So, (ii) in Theorem 8.6 holds at least formally, and the Gaussian model may be regarded as an integral manifold of $\mathcal{V}_\mathcal{A}$. Note, however, that Theorem 8.6 is not valid in the infinite-dimensional case so that (ii) does not imply (i), because various mathematical problems arise that were not present in the finite dimensional case, such as the fact that a positive operator does not always have finite trace and hence is not always normalizable.

At the end of this section, we give an example of e-autoparallel submanifold that does not fall within the scope of Theorem 8.6. Let $\mathcal{B} = \{|1\rangle, \dots, |d\rangle\}$ with $d = \dim \mathcal{H}$ be an arbitrary orthonormal basis of $\mathcal{H}$, and let $\mathcal{L}^\mathcal{B} := \{\sum_{i,j} a_{ij} |i\rangle \langle j| \mid [a_{ij}] \in \mathbb{R}^{d \times d}\}$, $\mathcal{L}_h^\mathcal{B} := \mathcal{L}_h \cap \mathcal{L}^\mathcal{B}$ and $\mathcal{S}^\mathcal{B} := \mathcal{S} \cap \mathcal{L}^\mathcal{B}$. That is, $\mathcal{S}^\mathcal{B}$ is the set of density operators whose matrix representations w.r.t. the basis $\mathcal{B}$ are real matrices.

Proposition 8.11 $\mathcal{S}^\mathcal{B}$ is e-autoparallel in $\mathcal{S}$ w.r.t. the SLD structure. (See Remark 8.14 for a generalization.)

Proof It is easy to see that for each $\rho \in \mathcal{S}^\mathcal{B}$

$$T_\rho^{(m)}(\mathcal{S}^\mathcal{B}) := \{\iota_*(X_\rho) \mid X_\rho \in T_\rho(\mathcal{S}^\mathcal{B})\} = \{A \in \mathcal{L}_h^\mathcal{B} \mid \text{Tr} A = 0\} \quad (8.12)$$

and that

$$\begin{aligned} T_\rho^{(e)}(\mathcal{S}^\mathcal{B}) &:= \{L_{X_\rho} \mid X_\rho \in T_\rho(\mathcal{S}^\mathcal{B})\} \\ &= \{A \in \mathcal{L}_h \mid \exists B \in T_\rho^{(m)}(\mathcal{S}^\mathcal{B}), \quad B = \rho \circ A\} \\ &= \{A \in \mathcal{L}_h^\mathcal{B} \mid \langle A \rangle_\rho = 0\}. \end{aligned} \quad (8.13)$$

(See Remark 3.1 for the symbols $T^{(m)}$ and $T^{(e)}$.) It follows from (8.13) that $A \in T_{\rho}^{(e)}(\mathcal{S}^{\mathcal{B}}) \Leftrightarrow A - \langle A \rangle_{\sigma} \in T_{\sigma}^{(e)}(\mathcal{S}^{\mathcal{B}})$ for any $\rho, \sigma \in \mathcal{S}^{\mathcal{B}}$, and hence we have from (3.30) that $X_{\rho} \in T_{\rho}(\mathcal{S}^{\mathcal{B}}) \Leftrightarrow \Phi_{\rho, \sigma}^{(e)}(X_{\rho}) \in T_{\sigma}(\mathcal{S}^{\mathcal{B}})$. This proves the proposition by Proposition 2.4. $\square$

Let us examine whether the e-autoparallelity of $\mathcal{S}^{\mathcal{B}}$ can be understood as an example of Theorem 8.6. Namely, the problem is whether $\mathcal{S}^{\mathcal{B}}$ is an integral manifold of an e-parallel distribution $\mathcal{V}_{\mathcal{A}}$ for some $\mathcal{A}$ satisfying condition (ii) in Theorem 8.6. For each $\rho \in \mathcal{S}^{\mathcal{B}}$, we have $T_{\rho}^{(e)}(\mathcal{S}^{\mathcal{B}}) = \{A - \langle A \rangle_{\rho} \mid A \in \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}\}$ (see (8.13)), which means that $T_{\rho}(\mathcal{S}^{\mathcal{B}}) = \mathcal{V}_{\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}, \rho}$ and that $\mathcal{S}^{\mathcal{B}}$ is an integral manifold of the distribution $\mathcal{V}_{\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}}$. Noting that $\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}} + \mathbb{R} = \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}$, the problem comes down to whether

$$\{[[A, B], \rho] \mid A, B \in \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}\} \subset \{\rho \circ C \mid C \in \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}\} \quad (8.14)$$

holds for every $\rho \in \mathcal{S}$. The answer is no, except when $\dim \mathcal{H} = 2$.

Proposition 8.12 *When $\dim \mathcal{H} \geq 3$,*

$$\forall \rho \in \mathcal{S} \setminus \mathcal{S}^{\mathcal{B}}, \{[[A, B], \rho] \mid A, B \in \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}\} \not\subset \{\rho \circ C \mid C \in \mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}\}. \quad (8.15)$$

As a consequence, the distribution $\mathcal{V}_{\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}}$ is not involutive.

Proof We represent operators on a d -dimensional Hilbert space by $d \times d$ matrices, and show that if a strictly positive density matrix ρ satisfies

$$\{[[A, B], \rho] \mid A, B : \text{real symmetric}\} \subset \{\rho \circ C \mid C : \text{real symmetric}\},$$

which is equivalent to

$$\{[\rho, F] \mid F : \text{real skew-symmetric}\} \subset \{\rho \circ C \mid C : \text{real symmetric}\}, \quad (8.16)$$

then ρ is a real matrix. Denoting $\rho = Q + \sqrt{-1} R$ with Q : real symmetric and R : real skew-symmetric, (8.16) is rewritten as

$$\begin{aligned} \forall F : \text{real skew-symmetric}, \exists C : \text{real symmetric}, \\ [Q, F] = Q \circ C \text{ and } [R, F] = R \circ C. \end{aligned} \quad (8.17)$$

Now it suffices to show the following lemma, whose proof will be given in Appendix G. $\square$

Lemma 8.13 *Let Q be a $d \times d$ real symmetric positive-definite matrix and R be a $d \times d$ real skew-symmetric matrix. If they satisfy (8.17) and $d \geq 3$, then $R = 0$.*

The above result implies that the e-parallel distribution $\mathcal{V}_{\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}}$ does not induce a foliation with e-autoparallel leaves and that $\mathcal{S}^{\mathcal{B}}$ is an isolated integral manifold of $\mathcal{V}_{\mathcal{L}_{\mathfrak{h}}^{\mathcal{B}}}$ when $\dim \mathcal{H} \geq 3$. The exceptional case $\dim \mathcal{H} = 2$ will be discussed in the next section.

Remark 8.14 Proposition 8.11 holds for a wide class of information-geometric structures, not limited to the SLD structure. In fact, the proof of Proposition 8.11 given above relies only upon the fact that if $\rho \in \mathcal{S}^{\mathcal{B}}$ and $X_\rho \in T_\rho(\mathcal{S}^{\mathcal{B}})$, then $L_{X_\rho} \in \mathcal{L}_{\mathcal{h}}^{\mathcal{B}}$. Due to (3.8) stating that $\iota_*(X_\rho) = \Omega_\rho(L_{X_\rho})$, this fact is shared by the e-connection defined from an arbitrary family of inner products $\langle \cdot, \cdot \rangle_\rho = \langle \cdot, \Omega_\rho(\cdot) \rangle_{\text{HS}}$, $\rho \in \mathcal{S}$, such that

$$\forall \rho \in \mathcal{S}^{\mathcal{B}}, \quad \Omega_\rho(\mathcal{L}_{\mathcal{h}}^{\mathcal{B}}) = \mathcal{L}_{\mathcal{h}}^{\mathcal{B}}. \quad (8.18)$$

This means that Proposition 8.11 holds under this mild condition on $\{\Omega_\rho\}_{\rho \in \mathcal{S}}$. In particular, if Ω_ρ is represented in the form (3.21) by a function $f : (0, \infty) \rightarrow (0, \infty)$ such that $\forall x > 0$, $xf(1/x) = f(x)$ and $f(1) = 1$ as in the case of monotone metrics, condition (8.18) is satisfied. To verify this, we represent (3.21) as $\Omega_\rho = f(\Delta_\rho)\mathcal{R}_\rho$, where $\mathcal{R}_\rho : A \mapsto A\rho$, and consider Ω_ρ as a $\mathbb{C}$ -linear map $\mathcal{L} \rightarrow \mathcal{L}$. Then it is easy to see that if $\rho \in \mathcal{S}^{\mathcal{B}}$, then $\mathcal{R}_\rho(\mathcal{L}^{\mathcal{B}}) = \mathcal{L}^{\mathcal{B}}$ and $\Delta_\rho(\mathcal{L}^{\mathcal{B}}) = \mathcal{L}^{\mathcal{B}}$, which yields $f(\Delta_\rho)(\mathcal{L}^{\mathcal{B}}) = \mathcal{L}^{\mathcal{B}}$, and hence we have $\Omega_\rho(\mathcal{L}^{\mathcal{B}}) = \mathcal{L}^{\mathcal{B}}$. Combined with $\Omega_\rho(\mathcal{L}_{\mathcal{h}}) = \mathcal{L}_{\mathcal{h}}$, this proves (8.18).

Remark 8.15 Since (8.12) shows that the space $T_\rho^{(\text{m})}(\mathcal{S}^{\mathcal{B}}) = \iota_*(T_\rho(\mathcal{S}^{\mathcal{B}}))$ does not depend on ρ , $\mathcal{S}^{\mathcal{B}}$ is m-autoparallel in $\mathcal{S}$, so that $\mathcal{S}^{\mathcal{B}}$ is doubly autoparallel (e.g., [11]) w.r.t. the e, m-connections. In addition, the set $\mathcal{S}_{\text{diag}}^{\mathcal{B}} := \mathcal{S} \cap \mathcal{L}_{\mathcal{h}, \text{diag}}^{\mathcal{B}}$, where $\mathcal{L}_{\mathcal{h}, \text{diag}}^{\mathcal{B}} := \{\sum_i a_i |i\rangle\langle i| \mid (a_i) \in \mathbb{R}^d\}$, is doubly autoparallel in $\mathcal{S}^{\mathcal{B}}$ w.r.t. the e, m-connections of $\mathcal{S}^{\mathcal{B}}$, and therefore is doubly autoparallel in $\mathcal{S}$ w.r.t. the e, m-connections of $\mathcal{S}$. These examples exhibit a remarkable contrast to the following fact for the classical case [12]: if a submanifold $\mathcal{M}$ of $\mathcal{P}(\Omega)$, where Ω is an arbitrary finite set, is doubly autoparallel in $\mathcal{P}(\Omega)$ w.r.t. the e, m-connections, then $\mathcal{M}$ is statistically isomorphic to $\mathcal{P}(\Omega')$ for some finite set Ω' .

9 Qubit manifolds

Throughout this section, we assume $\mathcal{H}$ to be 2-dimensional. Our aim is to study the e-autoparallelity in $\mathcal{S}(\mathcal{H})$ w.r.t. the SLD structure. To begin with, we make some preparations. Let $\{\sigma_1, \sigma_2, \sigma_3\} \subset \mathcal{L}_{\mathcal{h}}$ be a triple of Pauli operators such that

$$\text{Tr } \sigma_i = 0, \quad \sigma_i^2 = I, \quad \text{and} \quad \sigma_i \sigma_{i+1} = \sqrt{-1} \sigma_{i+2} \quad (i : \text{mod } 3). \quad (9.1)$$

Then $\{\sigma_1, \sigma_2, \sigma_3\}$ form a basis of $\mathcal{L}_{\mathcal{h},0}$. For any $\vec{a} = (a_i) \in \mathbb{R}^3$, we write $\vec{a} \cdot \vec{\sigma} := \sum_i a_i \sigma_i$, so that we have

$$\mathcal{L}_{\mathcal{h},0} = \{\vec{a} \cdot \vec{\sigma} \mid \vec{a} \in \mathbb{R}^3\}. \quad (9.2)$$

It follows that

$$(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma}) = (\vec{a} \cdot \vec{b})I + \sqrt{-1}(\vec{a} \times \vec{b}) \cdot \vec{\sigma}, \quad (9.3)$$

$$(\vec{a} \cdot \vec{\sigma}) \circ (\vec{b} \cdot \vec{\sigma}) = (\vec{a} \cdot \vec{b})I, \quad (9.4)$$

$$[\vec{a} \cdot \vec{\sigma}, \vec{b} \cdot \vec{\sigma}] = 2\sqrt{-1}(\vec{a} \times \vec{b}) \cdot \vec{\sigma}, \quad (9.5)$$

where $\vec{a} \cdot \vec{b} = \sum_i a_i b_i$, and $\vec{a} \times \vec{b} = \vec{c} \Leftrightarrow \forall i : \text{mod } 3, a_i b_{i+1} - a_{i+1} b_i = c_{i+2}$. The manifold $\mathcal{S} = \mathcal{S}(\mathcal{H})$ is represented as

$$\mathcal{S} = \{\rho_{\vec{r}} \mid \vec{r} \in \mathcal{R}\}, \quad (9.6)$$

where

$$\rho_{\vec{r}} := \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}), \quad \mathcal{R} := \{\vec{r} \in \mathbb{R}^3 \mid \|\vec{r}\| := \sqrt{\vec{r} \cdot \vec{r}} < 1\}. \quad (9.7)$$

For $\rho = \rho_{\vec{r}}$ and $A = a_0 I + \vec{a} \cdot \vec{\sigma}$, we have $\langle A \rangle_{\rho} = a_0 + \vec{r} \cdot \vec{a}$.

A tangent vector $X_{\rho} \in T_{\rho}(\mathcal{S})$ at $\rho = \rho_{\vec{r}}$ is represented by a 3-dimensional vector $\vec{x} \in \mathbb{R}^3$ such that

$$\iota_*(X_{\rho}) = \frac{1}{2} \vec{x} \cdot \vec{\sigma}. \quad (9.8)$$

The SLD of X_{ρ} is then represented as

$$L_{X_{\rho}} = \ell_{\vec{r}}(\vec{x}) \cdot \vec{\sigma} - \lambda_{\vec{r}}(\vec{x}) I, \quad (9.9)$$

where

$$\lambda_{\vec{r}}(\vec{x}) := \frac{\vec{x} \cdot \vec{r}}{1 - \|\vec{r}\|^2} \quad \text{and} \quad \ell_{\vec{r}}(\vec{x}) := \vec{x} + \lambda_{\vec{r}}(\vec{x}) \vec{r}. \quad (9.10)$$

In fact, (9.9) is verified as follows: noting that (9.10) yields $\vec{r} \cdot \ell_{\vec{r}}(\vec{x}) = \lambda_{\vec{r}}(\vec{x})$, we have

$$\begin{aligned} \rho \circ L_{X_{\rho}} &= \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}) \circ (\ell_{\vec{r}}(\vec{x}) \cdot \vec{\sigma} - \lambda_{\vec{r}}(\vec{x}) I) \\ &= \frac{1}{2}(\ell_{\vec{r}}(\vec{x}) - \lambda_{\vec{r}}(\vec{x}) \vec{r}) \cdot \vec{\sigma} + \frac{1}{2}((\vec{r} \cdot \ell_{\vec{r}}(\vec{x})) - \lambda_{\vec{r}}(\vec{x})) I \\ &= \frac{1}{2} \vec{x} \cdot \vec{\sigma} = \iota_*(X_{\rho}). \end{aligned} \quad (9.11)$$

Let us investigate the e-autoparallel submanifolds of $\mathcal{S}$. We first consider the 1-dimensional case, i.e., the e-geodesics. We recall that the general form of e-geodesic is given by (4.23). Treating the coordinate θ as a parameter to specify states and choosing P in (4.23) to be a state ρ_0 , an arbitrary e-geodesic $\mathcal{M}$ is represented as the trajectory $\mathcal{M} = \{\rho_{\theta} \mid \theta \in \mathbb{R}\}$ of

$$\rho_{\theta} = \frac{1}{Z_{\theta}} \exp\left(\frac{\theta}{2} F\right) \rho_0 \exp\left(\frac{\theta}{2} F\right), \quad Z_{\theta} := \text{Tr}(\rho_0 \exp(\theta F)), \quad (9.12)$$

where F is a Hermitian operator such that $\{F, I\}$ are linearly independent. Since the transformation $F \rightarrow aF + b$ by $a, b \in \mathbb{R}, a \neq 0$, together with $\theta \rightarrow \frac{1}{a}\theta$ and $\psi \rightarrow \psi + \frac{b}{a}\theta$, keeps $\mathcal{M}$ invariant, we can assume that F is represented as $F = \vec{u} \cdot \vec{\sigma}$ by a unit vector $\vec{u}$.

Proposition 9.1 *Let $\rho_0 = \rho_{\vec{r}_0}$ and $F = \vec{u} \cdot \vec{\sigma}$ with $\|\vec{u}\| = 1$ in (9.12). Letting $\vec{v}$ be a unit vector such that $\vec{u} \cdot \vec{v} = 0$ and that $\vec{r}_0 \in \text{span}\{\vec{u}, \vec{v}\}$, the e-geodesic $\mathcal{M} = \{\rho_{\theta} \mid \theta \in \mathbb{R}\}$*

is represented as

$$\mathcal{M} = \{\rho_{\vec{r}} \mid \vec{r} \in \mathcal{Q}\} \quad \text{with} \quad \mathcal{Q} := \{\vec{r}(\xi) \mid -1 < \xi < 1\}, \quad (9.13)$$

where

$$\vec{r}(\xi) := \xi \vec{u} + c\sqrt{1 - \xi^2} \vec{v}, \quad (9.14)$$

$$c := \frac{b}{\sqrt{1 - a^2}}, \quad a := \vec{r}_0 \cdot \vec{u}, \quad b := \vec{r}_0 \cdot \vec{v}. \quad (9.15)$$

(Here ξ^2 denotes the square of ξ , while the same symbol will appear as the second component of $\xi = (\xi^i)$ later.) The parameter ξ is m -affine as a coordinate system of $\mathcal{M}$ and in one-to-one correspondence with the e -affine parameter θ by

$$\xi = \frac{(1+a)e^{2\theta} - (1-a)}{(1+a)e^{2\theta} + (1-a)} \quad \text{and} \quad \theta = \frac{1}{2} \log \frac{(1-a)(1+\xi)}{(1+a)(1-\xi)}. \quad (9.16)$$

Proof Noting that $F = \vec{u} \cdot \vec{\sigma}$ is represented as

$$F = \rho_{\vec{u}} - \rho_{-\vec{u}} = 1\rho_{\vec{u}} + (-1)\rho_{-\vec{u}}$$

and that this is the spectral decomposition of F with projectors $\{\rho_{\vec{u}}, \rho_{-\vec{u}}\}$, we have

$$\exp\left(\frac{\theta}{2}F\right) = e^{\theta/2}\rho_{\vec{u}} + e^{-\theta/2}\rho_{-\vec{u}} = \cosh(\theta/2)I + \sinh(\theta/2)\vec{u} \cdot \vec{\sigma}.$$

Using this expression and representing $\vec{r}_0$ as $\vec{r}_0 = a\vec{u} + b\vec{v}$ by $a := \vec{r}_0 \cdot \vec{u}$ and $b := \vec{r}_0 \cdot \vec{v}$, a direct calculation shows that

$$\exp\left(\frac{\theta}{2}F\right)\rho_{\vec{r}_0}\exp\left(\frac{\theta}{2}F\right) = \frac{Z_\theta}{2}I + \frac{1}{2}\{(a \cosh \theta + \sinh \theta)\vec{u} + b\vec{v}\} \cdot \vec{\sigma}$$

and $Z_\theta = \cosh \theta + a \sinh \theta$, which yields

$$\rho_\theta = \frac{1}{2}(I + \vec{s}(\theta) \cdot \vec{\sigma}),$$

where

$$\vec{s}(\theta) := \frac{a \cosh \theta + \sinh \theta}{\cosh \theta + a \sinh \theta} \vec{u} + \frac{b}{\cosh \theta + a \sinh \theta} \vec{v}.$$

If we define ξ from θ by (9.16), we have

$$\frac{a \cosh \theta + \sinh \theta}{\cosh \theta + a \sinh \theta} = \xi \quad \text{and} \quad \frac{b}{\cosh \theta + a \sinh \theta} = c\sqrt{1 - \xi^2},$$

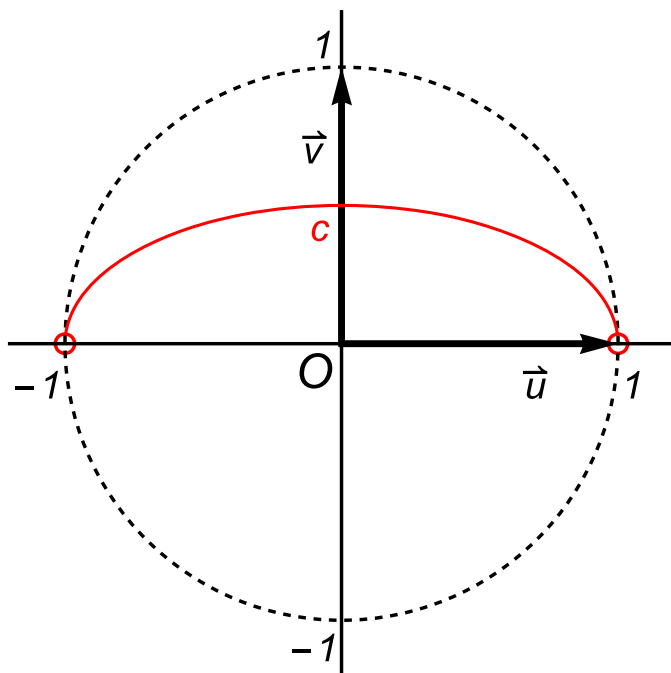


Fig. 1 The semi-ellipse representing an e-geodesic

so that $\vec{s}(\theta) = \vec{r}(\xi)$. It is easy to see that the range of ξ is $(-1, 1)$, and we obtain (9.13). In addition, since

$$\langle F \rangle_{\rho_{\vec{r}(\xi)}} = \vec{r}(\xi) \cdot \vec{u} = \xi,$$

the parameter ξ is m-affine. $\square$

Note that $\mathcal{Q}$ in the above proposition forms a semi-ellipse in the open unit ball $\mathcal{R}$ obtained by cutting an ellipse in half on the major axis: see Fig. 1. In the special case of $c = 0$, the semi-ellipse becomes a straight line.

Next, let us proceed to considering the 2-dimensional case. In searching for 2-dimensional e-autoparallel submanifolds, the previously obtained knowledge of e-geodesics provides an important clue. If a 2-dimensional submanifold $\mathcal{M} = \{\rho_{\vec{r}} \mid \vec{r} \in \mathcal{Q}\}$ is e-autoparallel, it must be e-totally geodesic, and hence the surface $\mathcal{Q}$ should be a union of semi-ellipses. The following proposition claims that a 2-dimensional e-autoparallel submanifold is obtained as a semi-ellipsoid formed by rotating a semi-ellipse representing an e-geodesic around its minor axis.

Proposition 9.2 *Given an orthonormal basis $\{\vec{u}_1, \vec{u}_2, \vec{v}\}$ of $\mathbb{R}^3$ and a real constant c satisfying $|c| < 1$, let*

$$\mathcal{Q} := \{\vec{r}(\xi) \mid \xi = (\xi^1, \xi^2) \in \mathbb{R}^2, (\xi^1)^2 + (\xi^2)^2 < 1\}, \quad (9.17)$$

where

$$\vec{r}(\xi) := \xi^1 \vec{u}_1 + \xi^2 \vec{u}_2 + c \sqrt{1 - (\xi^1)^2 - (\xi^2)^2} \vec{v}. \quad (9.18)$$

Then $\mathcal{M} := \{\rho_{\vec{r}} \mid \vec{r} \in \mathcal{Q}\}$ is e -autoparallel in $\mathcal{S}$, and the parameter $\xi = (\xi^1, \xi^2)$ is m -affine as a coordinate system of $\mathcal{M}$. More specifically, letting $F^i := \vec{u}_i \cdot \vec{\sigma}$, $\mathcal{A} := \text{span}\{F^1, F^2\}$, and $\mathcal{V}_{\mathcal{A}} : \mathcal{S} \ni \rho \mapsto V_{\mathcal{A}, \rho}$ be the e -parallel distribution defined from $\mathcal{A}$ by (8.7), $\mathcal{M}$ is an integral manifold of $\mathcal{V}_{\mathcal{A}}$ and $\xi^i = \langle F^i \rangle_{\rho_{\vec{r}(\xi)}}$.

Proof For $i \in \{1, 2\}$, let

$$\vec{x}_i := \partial_i \vec{r}(\xi) = \vec{u}_i - \frac{c \xi^i}{\alpha(\xi)} \vec{v},$$

where $\partial_i := \frac{\partial}{\partial \xi^i}$ and $\alpha(\xi) := \sqrt{1 - (\xi^1)^2 - (\xi^2)^2}$. Noting that

$$\|\vec{r}(\xi)\|^2 = 1 - (1 - c^2)\alpha(\xi)^2 \quad \text{and} \quad \vec{x}_i \cdot \vec{r}(\xi) = (1 - c^2)\xi^i,$$

we have

$$\begin{aligned} \ell_{\vec{r}(\xi)}(\vec{x}_i) &= \vec{x}_i + \frac{\xi^i}{\alpha(\xi)^2} \vec{r}(\xi) \\ &= \vec{u}_i + \frac{\xi^i \xi^1}{\alpha(\xi)^2} \vec{u}_1 + \frac{\xi^i \xi^2}{\alpha(\xi)^2} \vec{u}_2 \in \text{span}\{\vec{u}_1, \vec{u}_2\}, \end{aligned}$$

where the terms proportional to $\vec{v}$ included in $\vec{x}_i$ and $\vec{r}(\xi)$ cancel, yielding the last line. Owing to (9.9) this implies that the SLDs satisfy $L_{i, \xi} \in \text{span}\{F^1, F^2\} \oplus \mathbb{R}$ for $i \in \{1, 2\}$, which means that the first condition in (3.33) is satisfied. The second condition is also satisfied since $\langle F^i \rangle_{\rho_{\vec{r}(\xi)}} = \vec{u}_i \cdot \vec{r}(\xi) = \xi^i$. Thus the claim of the proposition follows from Proposition 3.8. $\square$

As can be seen from naive geometric intuition, for any point $\vec{r}$ in $\mathcal{R}$ and any plane $P = \vec{r} + V$ containing $\vec{r}$, where V is a 2-dimensional linear subspace of $\mathbb{R}^3$, there always exist an orthonormal basis $\{\vec{u}_1, \vec{u}_2, \vec{v}\}$ and a constant $c \in (-1, 1)$ such that the semi-ellipsoid $\mathcal{Q}$ defined from them by (9.17) and (9.18) contains $\vec{r}$ and has P as the tangent plane at $\vec{r}$. In fact, such $\{\vec{u}_1, \vec{u}_2, \vec{v}\}$ and c are obtained as follows: take an orthonormal basis $\{\vec{u}_1, \vec{u}_2, \vec{v}\}$ so that $\{\vec{u}_1, \vec{u}_2\} \subset \ell_{\vec{r}}(V) = \{\ell_{\vec{r}}(\vec{x}) \mid \vec{x} \in V\}$, and then let $\beta^2 := (\vec{r} \cdot \vec{u}_1)^2 + (\vec{r} \cdot \vec{u}_2)^2$ (i.e. the squared norm of the orthogonal projection of $\vec{r}$ onto $\ell_{\vec{r}}(V)$), $\gamma := \vec{r} \cdot \vec{v}$, and $c := \gamma / \sqrt{1 - \beta^2}$. Since $\dim \mathcal{S} = 3$, this fact means that $(\mathcal{S}, \nabla^{(e)})$ satisfies condition (i) of Proposition 8.3, and necessarily satisfies condition (ii) as well. Invoking (3.26), condition (ii) is expressed as follows.

Proposition 9.3 *When $\dim \mathcal{H} = 2$, for any $\rho \in \mathcal{S}$ and any $A, B \in \mathcal{L}_{\mathcal{H}}$ satisfying $\langle A \rangle_{\rho} = \langle B \rangle_{\rho} = 0$ we have*

$$[[A, B], \rho] \in \text{span}\{\rho \circ A, \rho \circ B\}. \quad (9.19)$$

This proposition can also be proved directly by the use of the following lemma, whose proof is given in Appendix H.

Lemma 9.4 When $\dim \mathcal{H} = 2$, for any $\rho \in \mathcal{S}$ and any $A, B \in \mathcal{L}_h$, we have

$$\begin{aligned} \frac{1}{2}[[A, B], \rho] = & (\text{Tr} A - 2 \langle A \rangle_\rho)(\rho \circ B) - (\text{Tr} B - 2 \langle B \rangle_\rho)(\rho \circ A) \\ & + \{(\text{Tr} B) \langle A \rangle_\rho - (\text{Tr} A) \langle B \rangle_\rho\} \rho. \end{aligned} \quad (9.20)$$

Letting $\langle A \rangle_\rho = \langle B \rangle_\rho = 0$ in (9.20), we obtain

$$[[A, B], \rho] = 2(\text{Tr} A)(\rho \circ B) - 2(\text{Tr} B)(\rho \circ A), \quad (9.21)$$

which proves Proposition 9.3.

The following proposition immediately follows from Proposition 9.3, which presents a remarkable contrast to Proposition 8.12 for the case $\dim \mathcal{H} \geq 3$.

Proposition 9.5 When $\dim \mathcal{H} = 2$, for any orthonormal basis $\mathcal{B}$ of $\mathcal{H}$ it holds that

$$\forall \rho \in \mathcal{S}, \{[[A, B], \rho] \mid A, B \in \mathcal{L}_h^\mathcal{B}\} \subset \{\rho \circ C \mid C \in \mathcal{L}_h^\mathcal{B}\}. \quad (9.22)$$

Proof Obvious from Proposition 9.3 since $\mathcal{L}_h^\mathcal{B}$ is an $\mathbb{R}$ -linear space with $I \in \mathcal{L}_h^\mathcal{B}$. $\square$

Thus, the distribution $\mathcal{V}_{\mathcal{L}_h^\mathcal{B}}$ is involutive, and induces a foliation $\mathcal{S} = \bigsqcup_\alpha \mathcal{M}_\alpha$ whose leaves $\{\mathcal{M}_\alpha\}$ are 2-dimensional e-autoparallel submanifolds that are integral manifolds of $\mathcal{V}_{\mathcal{L}_h^\mathcal{B}}$. Furthermore, we can see from the following lemma that every 2-dimensional e-autoparallel submanifold of $\mathcal{S}$ is an integral manifold of $\mathcal{V}_{\mathcal{L}_h^\mathcal{B}}$ for some $\mathcal{B}$.

Lemma 9.6 When $\dim \mathcal{H} = 2$, for any $A, B \in \mathcal{L}_h$ there exists an orthonormal basis $\mathcal{B}$ such that $\{A, B\} \subset \mathcal{L}_h^\mathcal{B}$.

Proof Let $\{|1\rangle, |2\rangle\}$ be an orthonormal basis that diagonalizes A , and choose $\beta \in \mathbb{C}$ so that $|\beta| = 1$ and $\beta \langle 1|B|2\rangle \in \mathbb{R}$. Then $\mathcal{B} := \{|1\rangle, \beta|2\rangle\}$ satisfies the desired condition. $\square$

According to Proposition 9.3, every e-parallel distribution is involutive in the qubit case, and should generate a pair of dual foliations due to Cor. 8.8. Let us observe these foliations based on the representations of e-autoparallel manifolds given in Proposition 9.1 and Proposition 9.2.

First, fix a unit vector $\vec{u} \in \mathbb{R}^3$ arbitrarily and let $F := \vec{u} \cdot \vec{\sigma}$ as in Proposition 9.1. The distribution $\mathcal{V}_\mathcal{A}$ with $\mathcal{A} = \text{span}\{F\}$ is involutive, and an integral manifold $\mathcal{M}$, which is an e-geodesic, is generally represented as (9.13) and (9.14). The representation has two parameters, a unit vector $\vec{v}$ orthogonal to $\vec{u}$ and a real number c lying in the interval $(-1, 1)$, but the true parameter that determines $\mathcal{M}$ injectively is $\vec{v}' := c\vec{v}$ lying in the open unit disc orthogonal to $\vec{u}$. Denoting the manifold $\mathcal{M}$ corresponding to $\vec{v}'$ by $\mathcal{M}_{\vec{v}'}$, we obtain the foliation $\mathcal{S} = \bigsqcup_{\vec{v}'} \mathcal{M}_{\vec{v}'}$ consisting of 1-dimensional e-autoparallel leaves. On the other hand, letting

$$\mathcal{N}_b := \{\rho \in \mathcal{S} \mid \langle F \rangle_\rho = b\} \quad (9.23)$$

for $b \in (-1, 1)$, we obtain the dual foliation $\mathcal{S} = \bigsqcup_b \mathcal{N}_b$ consisting of 2-dimensional m-autoparallel leaves which intersect $\mathcal{M}_{\vec{v}'}$ orthogonally for any $\vec{v}'$.

Next, fix a pair of orthonormal vectors $\{\vec{u}_1, \vec{u}_2\}$ in $\mathbb{R}^3$ arbitrarily and let $F^i := \vec{u}_i \cdot \vec{\sigma}$ as in Proposition 9.2. The distribution $\mathcal{V}_{\mathcal{A}}$ with $\mathcal{A} = \text{span}\{F^1, F^2\}$ is involutive, and an integral manifold $\mathcal{M}$ is generally represented as (9.17) and (9.18). Here the true parameter is $\vec{v}' := c\vec{v}$, whose range is the open line segment given by the intersection of the open unit ball and the line through the origin orthogonal to both $\vec{u}_1$ and $\vec{u}_2$. Denoting the manifold $\mathcal{M}$ corresponding to $\vec{v}'$ by $\mathcal{M}_{\vec{v}'}$, we obtain the foliation $\mathcal{S} = \bigsqcup_{\vec{v}'} \mathcal{M}_{\vec{v}'}$ consisting of 2-dimensional e-autoparallel leaves. The same foliation is also expressed as $\mathcal{S} = \bigsqcup_{c \in (-1, 1)} \mathcal{M}_c$ when $\vec{v}$ is chosen as one of the two unit vectors orthogonal to $\vec{u}_1$ and $\vec{u}_2$. On the other hand, letting

$$\mathcal{N}_b := \{\rho \in \mathcal{S} \mid \langle F^1 \rangle_\rho = b^1 \text{ and } \langle F^2 \rangle_\rho = b^2\} \quad (9.24)$$

for $b = (b^1, b^2) \in \mathbb{R}^2$ satisfying $(b^1)^2 + (b^2)^2 < 1$, we obtain the dual foliation $\mathcal{S} = \bigsqcup_b \mathcal{N}_b$ consisting of 1-dimensional m-autoparallel leaves which intersect $\mathcal{M}_{\vec{v}'}$ orthogonally for any $\vec{v}'$.

Remark 9.7 In a context of quantum parameter estimation with nuisance parameters, Suzuki et al. introduced in Section 6.3 of [6] the coordinate system $\tilde{\xi} = (\tilde{\xi}^i)$ on $\mathcal{S}$ defined by

$$\tilde{\xi}^i = \langle \sigma_i \rangle, \quad i \in \{1, 2\}, \quad \text{and} \quad \tilde{\xi}^3 = \langle \sigma_3 \rangle / \sqrt{1 - (\tilde{\xi}^1)^2 - (\tilde{\xi}^2)^2}. \quad (9.25)$$

It is shown there that

$$g(\tilde{\partial}_1, \tilde{\partial}_3) = g(\tilde{\partial}_2, \tilde{\partial}_3) = 0, \quad \tilde{\partial}_i := \frac{\partial}{\partial \tilde{\xi}^i}, \quad (9.26)$$

while the usual coordinate system $\xi = (\xi^i)$ with $\xi^i = \langle \sigma_i \rangle$, $i \in \{1, 2, 3\}$, has non-vanishing $g(\partial_i, \partial_j)$ for any i, j . In the situation where ξ^1 and ξ^2 are parameters of interest to be estimated and ξ^3 is a nuisance parameter, the parameter change from ξ to $\tilde{\xi}$, which is called the global orthogonalization in [6], is considered to remove the influence of nuisance parameter in some sense, while the parameters of interest are kept invariant as $\tilde{\xi}^i = \xi^i$, $i \in \{1, 2\}$. The orthogonality in (9.26) can be viewed as a reflection of the orthogonality of the pair of dual foliations $\mathcal{S} = \bigsqcup_c \mathcal{M}_c = \bigsqcup_b \mathcal{N}_b$ given above. To see this, let $\vec{u}_1, \vec{u}_2, \vec{v}$ in Proposition 9.2 be the natural basis $\vec{e}_1, \vec{e}_2, \vec{e}_3$ of $\mathbb{R}^3$. Then the 2-dimensional e-autoparallel leaf $\mathcal{M}_c$ is represented by the constraint $\xi^3 = c\sqrt{1 - (\xi^1)^2 - (\xi^2)^2}$ on the coordinates $\xi^i = \langle \sigma_i \rangle$, $i \in \{1, 2, 3\}$, which means that $\tilde{\xi}^3$ is fixed to c for the coordinates $(\tilde{\xi}^i)$ defined by (9.25). Therefore, $(\tilde{\partial}_1)_\rho$ and $(\tilde{\partial}_2)_\rho$ belong to $T_\rho(\mathcal{M}_c)$ when $\rho \in \mathcal{M}_c$. On the other hand, the 1-dimensional m-autoparallel leaf $\mathcal{N}_b$ is represented by the constraint $(\xi^1, \xi^2) = (b^1, b^2)$, or equivalently $(\tilde{\xi}^1, \tilde{\xi}^2) = (b^1, b^2)$, so that both $(\partial_3)_\rho$ and $(\tilde{\partial}_3)_\rho$ belongs to $T_\rho(\mathcal{N}_b)$ when $\rho \in \mathcal{N}_b$. Since every point $\rho \in \mathcal{S}$ belongs to $\mathcal{M}_c \cap \mathcal{N}_b$ for some (c, b) for which $T_\rho(\mathcal{M}_c) \perp T_\rho(\mathcal{N}_b)$ holds, we obtain (9.26).

Finally, let us make several remarks on estimators for qubit manifolds.

Remark 9.8 The following result is known for qubit models: for any model $(\mathcal{M}, \xi)$, any state $\rho \in \mathcal{M}$ and any positive-definite weight matrix W , it holds that

$$\min_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W V_\rho(\Pi')) = \left(\text{tr} \left(G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} \right)^{\frac{1}{2}} \right)^2 =: \tau_\rho(W) \quad (9.27)$$

and that a locally W -efficient estimator (see (4.19)) achieving the minimum in (9.27) is given by a randomized procedure similar to the one introduced in Sect. 4. Historically, apart from the 1-parameter case which is trivial, the 2-parameter case was established in [13, 14], and the general case including $\mathcal{M} = \mathcal{S}(\mathcal{H})$ was established in [15, 16]. A comprehensive and transparent exposition on this result can be found in [17]. While our estimator was chosen solely for the purpose of proving Proposition 4.1 and Theorem 5.1, and was not expected to be optimal in any sense, it is meaningful to clarify its relation to (9.27). Recall that our estimator $\Pi = (\pi, f)$ is determined by $\{\mathbf{u}^1, \dots, \mathbf{u}^n\}$, $(p_1, \dots, p_n)$ and $\{\gamma_k^i\}$ with $n = \dim \mathcal{M}$, and its variance matrix is represented as shown in item (2) of Lemma 4.3. The variance is minimized when we set $\gamma_k^i = \xi^i(\rho)$, and in this case we have

$$\text{tr}(W V_\rho(\Pi)) = \sum_k \frac{1}{p_k} \left((\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k \right) \left((\mathbf{w}_k)^T W \mathbf{w}_k \right). \quad (9.28)$$

Since Π belongs to $\mathcal{U}(\rho, \xi)$, it should satisfy

$$\text{tr}(W V_\rho(\Pi)) \geq \tau_\rho(W). \quad (9.29)$$

A direct proof of this inequality is given in Appendix I. There, it is shown that the equality holds iff there exists a positive-definite matrix K such that

$$W = K G_\rho^{-1} K, \quad (9.30)$$

and

$$K \propto \sum_k \frac{p_k}{(\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k} \mathbf{u}^k (\mathbf{u}^k)^T. \quad (9.31)$$

When a weight W and a point ρ of a model $(\mathcal{M}, \xi)$ are given, a matrix $K > 0$ satisfying (9.30) is determined from them and is represented as

$$K = G_\rho^{\frac{1}{2}} \left(G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} \right)^{\frac{1}{2}} G_\rho^{\frac{1}{2}}. \quad (9.32)$$

Then (9.31) tells us that an estimator Π achieving $\text{tr}(W V_\rho(\Pi)) = \tau_\rho(W)$ is obtained by choosing (p_k) and $\{\mathbf{u}^k\}$ so that (9.31) holds. There are infinitely many choices of such (p_k) and $\{\mathbf{u}^k\}$. An example is given by using the spectral decomposition (diagonalization) of K , i.e., by letting $\{\mathbf{u}^k\}$ be an orthonormal eigenvectors of K and letting

(p_k) be a probability vector such that $p_k/(\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k$ is proportional to the corresponding eigenvalue. Another example is given by using the spectral decomposition of $G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}}$ instead of K as follows. Let $\{\lambda_k\}$ and $\{\mathbf{v}_k\}$ be the eigenvalues and the orthonormal eigenvectors of $G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}}$ so that

$$G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} = \sum_k \lambda_k \mathbf{v}_k (\mathbf{v}_k)^T \quad \text{with} \quad (\mathbf{v}_k)^T \mathbf{v}_l = \delta_{kl}. \quad (9.33)$$

Then we have

$$K = G_\rho^{\frac{1}{2}} \left(\sum_k \sqrt{\lambda_k} \mathbf{v}_k (\mathbf{v}_k)^T \right) G_\rho^{\frac{1}{2}} = \sum_k \frac{\sqrt{\lambda_k}}{(\mathbf{v}_k)^T \mathbf{v}_k} (G_\rho^{\frac{1}{2}} \mathbf{v}_k) (G_\rho^{\frac{1}{2}} \mathbf{v}_k)^T. \quad (9.34)$$

This shows that (9.31) holds if we let

$$\mathbf{u}^k = G_\rho^{\frac{1}{2}} \mathbf{v}_k \quad \text{and} \quad p_k \propto \sqrt{\lambda_k}. \quad (9.35)$$

The estimator defined by (9.35) corresponds to that of [14, 16, 17]; see item (1) of Remark 9.13.

(End of Remark 9.8)

Example 9.9 Let us demonstrate the randomized estimators treated so far for a concrete example. Let $(\mathcal{M}, \xi)$ be a qubit model with $n = \dim \mathcal{M} (\leq 3)$ and ρ be a point of $\mathcal{M}$. With appropriate choices of a triple of Pauli operators $\{\sigma_i\}$ satisfying (9.1) and of a coordinate system up to affine transformation, we may assume without loss of generality that

$$\forall i \in \{1, \dots, n\}, \quad \xi^i(\rho) = \langle \sigma_i \rangle_\rho \quad \text{and} \quad L_\rho^i = \sigma_i - \xi^i(\rho). \quad (9.36)$$

In particular, when $\mathcal{M}$ is e-autoparallel in $\mathcal{S}$ and ξ is m-affine, we may assume that (9.36) holds uniformly at all points ρ of $\mathcal{M}$ due to (i) $\Leftrightarrow$ (iii) of Proposition 3.8. When $n = 1$ and when $n = 2$, this is equivalent to redefining the operators $F = \vec{u} \cdot \vec{\sigma}$ and $F^i = \vec{u}_i \cdot \vec{\sigma}$ in Proposition 9.1 and Proposition 9.2 as σ_1 and σ_i , respectively. Note that (9.36) implies that

$$g^{ij}(\rho) = \left\langle \sigma_i - \xi^i(\rho), \sigma_j - \xi^j(\rho) \right\rangle_\rho = \delta_{ij} - \xi^i(\rho) \xi^j(\rho), \quad (9.37)$$

which means that

$$G_\rho^{-1} = I - \xi(\rho) \xi(\rho)^T. \quad (9.38)$$

For simplicity, let us consider the special case when $\{\mathbf{u}^k\}$ is the natural basis $\{\mathbf{e}_k\}$ of $\mathbb{R}^n$ for our randomized estimators. Since $u_i^k = w_i^k = \delta_{ik}$ in this case, the estimation

procedures (4.15) and (5.8) are represented as

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sigma_k} \xrightarrow{\text{meas.}} y \\ \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \xi(\rho) + \frac{y - \xi^k(\rho)}{p_k} \mathbf{e}_k \quad (9.39)$$

and

$$(p_k) \xrightarrow{\text{rnd}} k \rightarrow \left\{ \begin{array}{l} \boxed{\sigma_k} \xrightarrow{\text{meas.}} y \\ \rightarrow k \end{array} \right\} \longrightarrow \hat{\xi} = \frac{y}{p_k} \mathbf{e}_k, \quad (9.40)$$

respectively. Letting $p_1 = 1 - \varepsilon$, both of these estimation procedures give a *locally* $\mathbf{e}_1(\mathbf{e}_1)^T$ -efficient filtration at ρ , say $\tilde{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$, such that

$$\lim_{\varepsilon \downarrow 0} v_\rho^{11}(\Pi_\varepsilon) = g^{11}(\rho) = 1 - \left(\xi^1(\rho) \right)^2, \quad (9.41)$$

where $V_\rho(\Pi_\varepsilon) = [v_\rho^{ij}(\Pi_\varepsilon)]$. As noted in Remark 5.2, $v_\rho^{11}(\Pi_\varepsilon)$ of (9.39) is smaller than that of (9.40) by $\frac{\varepsilon}{1-\varepsilon}(\xi^1(\rho))^2$, while (9.40) has the advantage that the procedure does not depend on ρ and gives an $\mathbf{e}_1(\mathbf{e}_1)^T$ -efficient filtration for an e-autoparallel model.

Remark 9.10 For a weight matrix $W = [w_{ij}]$, the estimator Π of the form (9.39) achieves $\text{tr}(W V_\rho(\Pi)) = \tau_\rho(W)$ iff the conditions (9.30) and (9.31) hold, which now reduce to

$$w_{ij} \propto p_i p_j \frac{\delta_{ij} - \xi^i(\rho) \xi^j(\rho)}{(1 - (\xi^i(\rho))^2)(1 - (\xi^j(\rho))^2)}. \quad (9.42)$$

When (p_k) is the uniform distribution $p_k = 1/n$, in particular, this turns out to be

$$w_{ij} \propto \frac{\delta_{ij} - \xi^i(\rho) \xi^j(\rho)}{(1 - (\xi^i(\rho))^2)(1 - (\xi^j(\rho))^2)}. \quad (9.43)$$

This form of weight matrix appears in a main result of [17], where a kind of efficiency of quantum state tomography is discussed. Further details and the relation to our formulation are given in Appendix J.

Remark 9.11 In Sect. 6, we showed that if a model $(\mathcal{M}, \xi)$ has a W -efficient estimator for every positive weight W , then $\mathcal{M}$ is e-autoparallel in $\mathcal{S}(\mathcal{H})$ and ξ is an m-affine coordinate system (Proposition 6.1) and that a quantum Gaussian shift model gives an example of such a model. So it may be natural to ask whether an e-autoparallel model of qubit states gives another example. The answer is negative except for the 1-dimensional case. In fact, the following stronger proposition holds, whose proof is given in Appendix K

Proposition 9.12 *Let $\mathcal{M}$ be an e-autoparallel submanifold of $\mathcal{S}(\mathcal{H})$ with $\mathcal{H} \simeq \mathbb{C}^2$ and $n := \dim \mathcal{M} \in \{2, 3\}$, and ξ be an m-affine coordinate system of $\mathcal{M}$. Then the model $(\mathcal{M}, \xi)$ has no W -efficient estimator for any positive-definite $W \in \mathbb{R}^{n \times n}$.*

Remark 9.13 We add some supplementary remarks concerning Remark 9.8.

- (1) The construction of an estimator achieving the minimum of (9.29) given in [14] for the two-parameter case may appear to differ from that in [16, 17], but is in fact the same as shown below. In the notation of the present paper, the construction in [14] can be described as follows: defining a linear operator ϕ on $T_\rho^{(e)}(\mathcal{M}) = \text{span}\{L_{i,\rho}\}$ by

$$\phi := \sum_{i,j} w_{ij} \left| L_\rho^i \right\rangle \left\langle L_\rho^j \right| : A \mapsto \sum_{i,j} w_{ij} \left\langle L_\rho^j, A \right\rangle_\rho L_\rho^i, \quad (9.44)$$

where $W = [w_{ij}]$, and representing its spectral decomposition by

$$\phi = \sum_k \lambda_k \left| X^k \right\rangle \left\langle X^k \right| \quad \text{with} \quad \left\langle X^k, X^l \right\rangle_\rho = \delta^{kl}, \quad (9.45)$$

an estimator achieving the minimum is realized by the random choice of the observable X^k with probability $p_k = \sqrt{\lambda_k} / \sum_l \sqrt{\lambda_l}$. Representing X^k as $X^k = \sum_i u_i^k L_\rho^i$ by a vector $\mathbf{u}^k = (u_i^k)$ and invoking (9.44), the spectral decomposition (9.45) is rewritten as

$$W = \sum_k \lambda_k \mathbf{u}^k (\mathbf{u}^k)^T \quad \text{with} \quad (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^l = \delta^{kl}, \quad (9.46)$$

which is equivalent to (9.33) when $\mathbf{u}^k = G_\rho^{-\frac{1}{2}} \mathbf{v}_k$.

- (2) The proof of (9.29) with derivation of the equality condition given in Appendix I does not rely on the condition $\dim \mathcal{H} = 2$. Therefore, the following identity holds for an arbitrary model regardless of $\dim \mathcal{H}$:

$$\min_{\Pi' \in \hat{\mathcal{U}}(\rho, \xi)} \text{tr}(W V_\rho(\Pi')) = \tau_\rho(W), \quad (9.47)$$

where $\hat{\mathcal{U}}(\rho, \xi)$ denotes the totality of estimators in $\mathcal{U}(\rho, \xi)$ that are realized by the randomized procedure introduced in Sect. 4. This identity appears to be closely related to the discussion in [15], where it is claimed that the minimum in $\min_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W V_\rho(\Pi'))$ is achieved by a certain class of randomized estimator iff the minimum is equal to $\tau_\rho(W)$.

- (3) The identity (9.47) proves Proposition 4.1 as follows. Given a vector $\mathbf{u} \in \mathbb{R}^n$ and a positive number ε , define $W_\varepsilon := \mathbf{u} \mathbf{u}^T + \varepsilon I$ and let Π_ε be an estimator in $\hat{\mathcal{U}}(\rho, \xi)$ achieving the minimum for W_ε . Then we have

$$\begin{aligned} \mathbf{u}^T G_\rho^{-1} \mathbf{u} &\leq \mathbf{u}^T V_\rho(\Pi_\varepsilon) \mathbf{u} \leq \text{tr} W_\varepsilon V_\rho(\Pi_\varepsilon) \\ &= \left(\text{tr} \left(G_\rho^{-\frac{1}{2}} W_\varepsilon G_\rho^{-\frac{1}{2}} \right)^{\frac{1}{2}} \right)^2 \rightarrow \mathbf{u}^T G_\rho^{-1} \mathbf{u} \quad \text{as } \varepsilon \downarrow 0. \end{aligned}$$

10 Concluding remarks

In this paper we study the autoparallelity w.r.t. the e-connection for an information-geometric structure induced on $\mathcal{S}(\mathcal{H})$. In particular, we focus on the e-autoparallelity for the SLD structure, for which two different estimation-theoretical characterizations are given. We also investigate the existence conditions for e-autoparallel submanifolds by way of the involutivity of e-parallel distributions and its relation to the torsion tensor. As a result, a specialty of the qubit case has been revealed.

Since the obtained estimation-theoretical characterizations of the e-autoparallelity are complete in themselves, we do not see at this time what kind of development lies ahead. It is expected that the future development of quantum estimation theory and related fields may reveal new directions. The classical exponential family has a variety of important properties besides the existence of efficient estimator, some of which may present new material to characterize certain geometric notions.

For the autoparallelity w.r.t. non-flat connections, our understanding is still very limited. For example, we do not yet have the whole picture about e-autoparallel submanifolds of $\mathcal{S}(\mathcal{H})$ when $\dim \mathcal{H} \geq 3$. We look forward to further research on this topic in information geometry and/or general differential geometry.

It may also be a challenging problem to develop the infinite-dimensional quantum information geometry so that Theorem 5.1 is extended to the case when $\dim \mathcal{H} = \infty$ and that the naive geometric consideration on the quantum Gaussian shift model presented in Sect. 6 is mathematically justified.

Geometry of quantum statistical manifolds in an asymptotic framework would also be an important subject to be addressed. For example, consider a sequence $\mathcal{M}^{(n)} = \{\rho_{\xi}^{\otimes n}\}$, $n = 1, 2, \dots$, of i.i.d. extensions of a quantum statistical model $\mathcal{M} = \{\rho_{\xi}\}$. Recent progress in asymptotic quantum statistics has revealed that the sequence exhibits a desirable property called a quantum local asymptotic normality, which tells us that in a shrinking ($\sim 1/\sqrt{n}$) neighbourhood of a given point ξ_0 , the sequence converges to a quantum Gaussian shift model [18–23]. As pointed out in Sect. 6, the limiting quantum Gaussian shift model has a characteristic feature in view of quantum information geometry. It would, therefore, be an interesting future project to extend the geometrical idea presented in this paper to an asymptotic framework so that the convergence of quantum statistical manifolds can be discussed under a suitably chosen topology.

Appendix A Proof of (3.26)

We first consider the general situation where the e-connection $\nabla^{(e)}$ is determined by a family of inner products $\langle A, B \rangle_{\rho} = \langle A, \Omega_{\rho}(B) \rangle_{\text{HS}}$, and show that the torsion $\mathcal{T}^{(e)}$ of $\nabla^{(e)}$ is represented as follows: for any $X, Y \in \mathcal{X}(\mathcal{S})$,

$$\iota_*(\mathcal{T}^{(e)}(X, Y)) = (Y\Omega)(L_X) - (X\Omega)(L_Y), \quad (\text{A1})$$

where $Y\Omega : \rho \mapsto Y_{\rho}\Omega$ denotes the derivative of the super-operator-valued map $\Omega : \rho \mapsto \Omega_{\rho}$ w.r.t. Y , and $(Y\Omega)(L_X)$ denotes the map $\rho \mapsto (Y_{\rho}\Omega)(L_{X_{\rho}}) \in \mathcal{L}_{\text{h}}$. In

fact, invoking (3.8) and (3.27), we have for any $X, Y, Z \in \mathcal{X}(\mathcal{S})$,

$$\begin{aligned} Z = T^{(e)}(X, Y) &\Leftrightarrow Z = \nabla_X^{(e)} Y - \nabla_Y^{(e)} X - [X, Y] \\ &\Leftrightarrow L_Z = (XL_Y + g(X, Y)) - (YL_X + g(Y, X)) - L_{[X, Y]} \\ &\Leftrightarrow L_Z = XL_Y - YL_X - L_{[X, Y]} \\ &\Leftrightarrow \iota_*(Z) = \Omega(XL_Y - YL_X - L_{[X, Y]}) \\ &\Leftrightarrow \iota_*(Z) = \Omega(XL_Y) - \Omega(YL_X) - \iota_*([X, Y]). \end{aligned}$$

Noting that

$$\begin{aligned} \iota_*([X, Y]) &= X\iota_*(Y) - Y\iota_*(X) \\ &= X(\Omega(L_Y)) - Y(\Omega(L_X)) \\ &= (X\Omega)(L_Y) + \Omega(XL_Y) - (Y\Omega)(L_X) - \Omega(YL_X), \end{aligned}$$

we obtain (A1). For the SLD structure, we have $\Omega_\rho(A) = \frac{1}{2}(\iota(\rho)A + A\iota(\rho))$, which yields that for any point $\rho \in \mathcal{S}$ and any tangent vectors $X_\rho, Y_\rho \in T_\rho(\mathcal{S})$,

$$\begin{aligned} (Y_\rho\Omega)(L_{X_\rho}) &= \frac{1}{2} \{ \iota_*(Y_\rho)L_{X_\rho} + L_{X_\rho}\iota_*(Y_\rho) \} \\ &= \frac{1}{4} (\rho L_{Y_\rho} L_{X_\rho} + L_{Y_\rho} \rho L_{X_\rho} + L_{X_\rho} \rho L_{Y_\rho} + L_{X_\rho} L_{Y_\rho} \rho). \end{aligned}$$

Similarly, we have

$$(X_\rho\Omega)(L_{Y_\rho}) = \frac{1}{4} (\rho L_{X_\rho} L_{Y_\rho} + L_{X_\rho} \rho L_{Y_\rho} + L_{Y_\rho} \rho L_{X_\rho} + L_{Y_\rho} L_{X_\rho} \rho).$$

Substituting these into (A1), we obtain (3.26).

Appendix B On the attainability for the infimum in (4.9)

Let $(\mathcal{M}, \xi)$ be an n -dimensional model in $\mathcal{S}(\mathcal{H})$. Given a point $\rho \in \mathcal{M}$ and a vector $\mathbf{u} = (u_i) \in \mathbb{R}^n \setminus \{0\}$, we present some necessary conditions for the infimum in (4.9) to be attained by a minimum. Define

$$X_{\mathbf{u}, \rho} := \sum_{i,j} u_i g^{ij}(\rho) (\partial_i)_\rho \in T_\rho(\mathcal{M}), \quad (\text{B1})$$

where $\partial_i = \frac{\partial}{\partial \xi^i}$ and $G^{-1} = [g^{ij}]$ is the inverse matrix of the SLD Fisher information matrix $G = [g_{ij}]$. Then the SLD of the tangent vector $X_{\mathbf{u}, \rho}$ is given by

$$L_{\mathbf{u}, \rho} := \sum_i u_i L_\rho^i = \sum_{i,j} u_i g^{ij}(\rho) L_{i, \rho}, \quad (\text{B2})$$

where $\{L_i\}$ denotes the SLDs for $\partial_i = \frac{\partial}{\partial \xi^i}$. Let the spectral decomposition of $L_{\mathbf{u},\rho}$ be denoted by

$$L_{\mathbf{u},\rho} = \sum_k a_k v_k, \quad (\text{B3})$$

where $\{a_k\}$ are the distinct eigenvalues and $\{v_k\}$ are the projectors onto the eigenspaces. Then we have the following proposition.

Proposition B.1 *If the infimum of (4.9) is attained by a minimum so that*

$$\min_{\Pi \in \mathcal{U}(\rho, \xi)} \mathbf{u}^T V_\rho(\Pi) \mathbf{u} = \mathbf{u}^T G_\rho^{-1} \mathbf{u}, \quad (\text{B4})$$

then the operators $\{\sum_k v_k (\partial_i \rho) v_k\}_{i \in \{1, \dots, n\}}$ are linearly independent, or equivalently, the linear map

$$T_\rho(\mathcal{M}) \rightarrow \mathcal{L}_h(\mathcal{H}), \quad X_\rho \mapsto \sum_k v_k \iota_*(X_\rho) v_k \quad (\text{B5})$$

is injective, where ι_ denotes the differential map of the natural embedding $\iota : \mathcal{S} \rightarrow \mathcal{L}_{h,1}$ and $\partial_i \rho := \iota_*((\partial_i)_\rho)$ (cf. (3.2)).*

Before giving a proof, let us observe some consequences of Proposition B.1. We introduce the *commutation operator* (see [10]) $D_\rho : T_\rho(\mathcal{S}) \rightarrow T_\rho(\mathcal{S})$ defined by

$$\forall X_\rho \in T_\rho(\mathcal{S}), \quad \iota_*(D_\rho(X_\rho)) = \sqrt{-1} [\rho, L_{X_\rho}], \quad (\text{B6})$$

where L_{X_ρ} is the SLD for X_ρ and $[\cdot, \cdot]$ denotes the commutator of operators. Let $X'_{\mathbf{u},\rho}$ be a tangent vector of $\mathcal{S}$ (not necessarily of $\mathcal{M}$) whose SLD satisfies

$$L'_{\mathbf{u},\rho} \in \text{span}\{v_k\}_{k=1}^n, \quad (\text{B7})$$

including the case when $X'_{\mathbf{u},\rho} = X_{\mathbf{u},\rho}$. Then, as a corollary of Proposition B.1 we have:

Corollary B.2 *If $D_\rho(X'_{\mathbf{u},\rho}) \in T_\rho(\mathcal{M}) \setminus \{0\}$, then the infimum of (4.9) cannot be attained by a minimum.*

Proof For any k , we have

$$\begin{aligned} v_k \iota_*(D_\rho(X'_{\mathbf{u},\rho})) v_k &= v_k \sqrt{-1} [\rho, L'_{\mathbf{u},\rho}] v_k \\ &= \sqrt{-1} (v_k \rho L'_{\mathbf{u},\rho} v_k - v_k L'_{\mathbf{u},\rho} \rho v_k) = 0, \end{aligned}$$

where the last equality follows since $L'_{\mathbf{u},\rho} v_k = v_k L'_{\mathbf{u},\rho} \propto v_k$ due to (B7). Hence, if $D_\rho(X'_{\mathbf{u},\rho}) \in T_\rho(\mathcal{M}) \setminus \{0\}$, then the linear map in (B5) has a nontrivial kernel, so that the corollary follows from Proposition B.1. $\square$

Corollary B.3 *When $\mathcal{M} = \mathcal{S}(\mathcal{H})$, the infimum of (4.9) cannot be attained by a minimum if $[\rho, L_{\mathbf{u},\rho}] \neq 0$.*

Proof Obvious from Cor. B.2 since $[\rho, L_{\mathbf{u},\rho}] \neq 0 \Leftrightarrow D_\rho(X_{\mathbf{u},\rho}) \in T_\rho(\mathcal{S}) \setminus \{0\}$. $\square$

Let us proceed to proving Proposition B.1. Assume (B4), and let $\Pi = \Pi(d\hat{\xi})$ be an estimator achieving the minimum; i.e., Π is locally unbiased at ρ for ξ and satisfies

$$\mathbf{u}^T V_\rho(\Pi) \mathbf{u} = \mathbf{u}^T G_\rho^{-1} \mathbf{u}. \quad (\text{B8})$$

We need the following lemma.

Lemma B.4 *Letting $\{v_k\}$ be the projectors onto the eigenspaces of $L_{\mathbf{u},\rho}$ introduced in (B3), we have*

$$\Pi(d\hat{\xi}) = \sum_k v_k \Pi(d\hat{\xi}) v_k. \quad (\text{B9})$$

Proof Let

$$H := \sum_i u_i \int (\hat{\xi}^i - \xi^i(\rho)) \Pi(d\hat{\xi}) = \int h(\hat{\xi}) \Pi(d\hat{\xi}) \in \mathcal{L}_h(\mathcal{H}),$$

where

$$h(\hat{\xi}) := \sum_i u_i (\hat{\xi}^i - \xi^i(\rho)).$$

Then, just in parallel with the derivation of (ii) $\Rightarrow$ (5.3) in the proof of Theorem 5.1, we have

$$\mathbf{u}^T V_\rho(\Pi) \mathbf{u} \geq \|H\|_\rho^2 \geq \|L_{\mathbf{u},\rho}\|_\rho^2 = \mathbf{u}^T G_\rho^{-1} \mathbf{u}. \quad (\text{B10})$$

Here, the first inequality follows from the operator inequality

$$\int h(\hat{\xi})^2 \Pi(d\hat{\xi}) - H^2 = \int (h(\hat{\xi}) - H) \Pi(d\hat{\xi}) (h(\hat{\xi}) - H) \geq 0, \quad (\text{B11})$$

and the second inequality follows from

$$\forall j, \langle H, L_{j,\rho} \rangle_\rho = u_j = \langle L_{\mathbf{u},\rho}, L_{j,\rho} \rangle_\rho \quad (\text{B12})$$

implying

$$\|H\|_\rho^2 = \|L_{\mathbf{u},\rho}\|_\rho^2 + \|H - L_{\mathbf{u},\rho}\|_\rho^2.$$

Note that (B12) is a consequence of the local unbiasedness condition

$$\forall i, \forall j, \int (\hat{\xi}^i - \xi^i(\rho)) \text{Tr} \left((\partial_j \rho) \Pi(d\hat{\xi}) \right) = \delta_j^i, \quad (\text{B13})$$

where $\partial_j \rho$ denotes $\iota_*((\partial_j)_\rho)$ (cf. (3.2)). (Actually, the above proof of (B10) is a standard argument to prove the SLD Cramér-Rao inequality (4.8).) We see that (B8) holds iff the operator inequality (B11) holds with equality and $H = L_{\mathbf{u},\rho}$, which implies that

$$\int (h(\hat{\xi}) - L_{\mathbf{u},\rho}) \Pi(d\hat{\xi}) (h(\hat{\xi}) - L_{\mathbf{u},\rho}) = 0. \quad (\text{B14})$$

From this it follows that the support of the POVM (Positive Operator-Valued *Measure*) Π satisfies

$$\text{supp}(\Pi) \subset \Gamma := \bigcup_k h^{-1}(a_k) = \{\hat{\xi} \mid h(\hat{\xi}) \in \{a_1, a_2, \dots\}\}, \quad (\text{B15})$$

where $\{a_k\} = \{a_1, a_2, \dots\}$ denote the distinct eigenvalues of $L_{\mathbf{u}, \rho}$ as in (B3). To see this, let U be an arbitrary open set of $\mathbb{R}^n$ such that $U \cap \Gamma = \emptyset$. From (B11) and (B14), we have

$$\int_U (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \Pi(d\hat{\xi}) (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) = 0. \quad (\text{B16})$$

This means that the (unnormalized) POVM $\Lambda(d\hat{\xi}) = (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \Pi(d\hat{\xi}) (h(\hat{\xi}) - L_{\mathbf{u}, \rho})$ vanishes on U . Since $h(\hat{\xi}) - L_{\mathbf{u}, \rho}$ is invertible for all $\hat{\xi} \in U \subset \mathbb{R}^n \setminus \Gamma$, the POVM $\Pi(d\hat{\xi})$ also vanishes on U , i.e., $\Pi(U) = 0$. We thus have (B15).

Given an arbitrary Borel set $B \subset \mathbb{R}^n$, it follows from (B14) and (B15) that

$$\begin{aligned} 0 &= \int_B (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \Pi(d\hat{\xi}) (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \\ &= \sum_k \int_{B \cap h^{-1}(a_k)} (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \Pi(d\hat{\xi}) (h(\hat{\xi}) - L_{\mathbf{u}, \rho}) \\ &= \sum_k (a_k - L_{\mathbf{u}, \rho}) \Pi(B \cap h^{-1}(a_k)) (a_k - L_{\mathbf{u}, \rho}), \end{aligned}$$

which leads to

$$\forall k, \quad (a_k - L_{\mathbf{u}, \rho}) \Pi(B \cap h^{-1}(a_k)) = 0.$$

This implies that, for any k , the image of $\Pi(B \cap h^{-1}(a_k))$ is contained in the eigenspace of $L_{\mathbf{u}, \rho}$ corresponding to the eigenvalue a_k , so that

$$\forall k, \quad \Pi(B \cap h^{-1}(a_k)) = v_k \Pi(B \cap h^{-1}(a_k)) = v_k \Pi(B \cap h^{-1}(a_k)) v_k, \quad (\text{B17})$$

and hence

$$\forall k, \quad \Pi(B \cap h^{-1}(a_k)) = \sum_l v_l \Pi(B \cap h^{-1}(a_k)) v_l.$$

Then we have

$$\begin{aligned} \Pi(B) &= \sum_k \Pi(B \cap h^{-1}(a_k)) = \sum_{k,l} v_l \Pi(B \cap h^{-1}(a_k)) v_l \\ &= \sum_l v_l \Pi(B) v_l. \end{aligned}$$

Since B is arbitrary, the lemma has been proved. $\square$

Now, assume that

$$\sum_{i=1}^n c_i \sum_k v_k (\partial_i \rho) v_k = 0 \quad (\text{B18})$$

holds for some $(c_i) \in \mathbb{R}^n$. It follows from (B13) that for any i ,

$$\begin{aligned} c^i &= \sum_j c^j \int (\hat{\xi}^i - \xi^i(\rho)) \operatorname{Tr} \left((\partial_j \rho) \Pi(d\hat{\xi}) \right) \\ &= \int (\hat{\xi}^i - \xi^i(\rho)) \operatorname{Tr} \left(\left(\sum_j c^j \partial_j \rho \right) \Pi(d\hat{\xi}) \right). \end{aligned}$$

Since (B9) and (B18) yield

$$\begin{aligned} \operatorname{Tr} \left(\left(\sum_j c^j \partial_j \rho \right) \Pi(d\hat{\xi}) \right) &= \operatorname{Tr} \left(\left(\sum_j c^j \partial_j \rho \right) \sum_k v_k \Pi(d\hat{\xi}) v_k \right) \\ &= \operatorname{Tr} \left(\sum_k v_k \left(\sum_j c^j \partial_j \rho \right) v_k \Pi(d\hat{\xi}) \right) = 0, \end{aligned}$$

we obtain $\forall i, c_i = 0$. This means that the operators $\{\sum_k v_k (\partial_i \rho) v_k\}_{i \in \{1, \dots, n\}}$ are linearly independent, which completes the proof of Proposition B.1.

Remark B.5 Noting that (B17) implies $\Pi(h^{-1}(a_k)) = v_k \Pi(h^{-1}(a_k)) v_k \leq v_k$ and that $\sum_k \Pi(h^{-1}(a_k)) = I = \sum_k v_k$, we have

$$\forall k, \quad \Pi(h^{-1}(a_k)) = v_k. \quad (\text{B19})$$

This indicates that the measurement Π is a refinement of the projective measurement $\nu = (v_k)$. In particular, when Π is represented as $\Pi = (\pi, f)$ with a POVM $\pi = (\pi_\omega)_{\omega \in \Omega}$ on a discrete set Ω and a function $f = (f^i) : \Omega \rightarrow \mathbb{R}^n$, the LHS of (B19) is represented as

$$\Pi(h^{-1}(a_k)) = \sum_{\omega \in \hat{h}^{-1}(a_k)} \pi_\omega,$$

where $\hat{h}(\omega) := h(f(\omega)) = \sum_i u_i (f^i(\omega) - \xi^i(\rho))$, and (B19) implies that

$$\forall \omega, \exists k, \quad \pi_\omega \leq v_k.$$

According to Theorem 5 of [24], if the infimum of (4.9) is attained by a minimum, then an estimator attaining the minimum can be taken to have finite support with cardinality not exceeding $(\dim \mathcal{H})^2 + n(n+1)$. Therefore, Proposition B.1 could have been proved under the assumption that Π takes the form $\Pi = (\pi, f)$ with π having finite support.

Appendix C Proof of Lemma 4.3

(1) Let $A^i := \int \hat{\xi}^i \Pi(d\xi)$. Then we have

$$\begin{aligned} A^i &= \sum_k p_k f^i(k, X^k) = \sum_k p_k (\gamma_k^i + \frac{w_k^i}{p_k} X^k) \\ &= \xi^i(\rho) + \sum_k w_k^i X^k = \xi^i(\rho) + \sum_k \sum_j w_k^i u_j^k L_\rho^j = \xi^i(\rho) + L_\rho^i, \end{aligned}$$

where we have invoked (4.13), (4.10), (4.11) and (4.12). Now $\Pi \in \mathcal{U}(\rho, \xi)$ follows from (4.6).

(2) Invoking (4.13), we have for each i, j

$$\begin{aligned} B^{ij} &:= \int (\hat{\xi}^i - \xi^i(\rho))(\hat{\xi}^j - \xi^j(\rho)) \Pi(d\hat{\xi}) \\ &= \sum_k p_k \left(f^i(k, X^k) - \xi^i(\rho) \right) \left(f^j(k, X^k) - \xi^j(\rho) \right) \\ &= \sum_k p_k \left(\frac{w_k^i}{p_k} X^k + \gamma_k^i - \xi^i(\rho) \right) \left(\frac{w_k^j}{p_k} X^k + \gamma_k^j - \xi^j(\rho) \right). \end{aligned}$$

This leads to

$$\begin{aligned} (V_\rho(\Pi))^{ij} &= \text{Tr}(\rho B^{ij}) \\ &= \sum_k p_k \text{Tr} \left[\rho \left(\frac{w_k^i}{p_k} X^k + \gamma_k^i - \xi^i(\rho) \right) \left(\frac{w_k^j}{p_k} X^k + \gamma_k^j - \xi^j(\rho) \right) \right] \\ &= \sum_k \frac{w_k^i w_k^j}{p_k} \text{Tr}(\rho (X^k)^2) + \sum_k p_k (\gamma_k^i - \xi^i(\rho)) (\gamma_k^j - \xi^j(\rho)) \\ &= \sum_k \frac{w_k^i w_k^j}{p_k} (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^k + \sum_k p_k (\gamma_k^i - \xi^i(\rho)) (\gamma_k^j - \xi^j(\rho)), \end{aligned}$$

where we invoked $\text{Tr}(\rho X^k) = 0$ due to $X^k \in \text{span}\{L_{i,\rho}\}_{i=1}^n$.

(3) Obvious from (2) since $[w_k^i] = [u_k^i]^{-1}$.

Appendix D Proofs of Propositions 7.13, 7.14 and 7.15

Proof of Proposition 7.13 This proposition is essentially contained in Theorem 7.2 of [3]. Here we give a proof for the reader's convenience.

Given $F \in \mathcal{L}_h$ and $\rho \in \mathcal{S}$, there exists a tangent vector $X_\rho \in T_\rho(\mathcal{S})$ satisfying $L_{X_\rho} = F - \langle F \rangle_\rho$ by (3.10). Applying (7.19) to the case $\mathcal{M} = \mathcal{S}$, we have $X_\rho =$

$(\text{grad } \langle F \rangle)_\rho$, and hence

$$\|(d \langle F \rangle)_\rho\|_\rho^2 = \|X_\rho\|_\rho^2 = \langle L_{X_\rho}, L_{X_\rho} \rangle_\rho = \langle (F - \langle F \rangle_\rho)^2 \rangle_\rho = V_\rho(F).$$

□

Proof of Proposition 7.14 Recalling (7.16) and (7.17), we have

$$\mathcal{E}(\mathcal{S}) = \{f \in \mathcal{F}(\mathcal{S}) \mid \exists F \in \mathcal{L}_h, f = \langle F \rangle \text{ and } \forall \rho \in \mathcal{S}, V_\rho(F) = \|(df)_\rho\|_\rho^2\}. \quad (\text{D1})$$

Since the condition $V_\rho(F) = \|(df)_\rho\|_\rho^2$ is always satisfied by Proposition 7.13, we have the first equality in (7.14). The second equality follows from Proposition 7.5, and the third follows since under the relation $X \xrightarrow{g} \omega$ we have

$$\begin{aligned} X \text{ is e-parallel} &\Leftrightarrow \forall Y, Z \in \mathcal{X}(\mathcal{S}), g(\nabla_Y^{(\text{e})} X, Z) = 0 \\ &\Leftrightarrow \forall Y, Z \in \mathcal{X}(\mathcal{S}), Yg(X, Z) = g(X, \nabla_Y^{(\text{m})} Z) \\ &\Leftrightarrow \forall Y, Z \in \mathcal{X}(\mathcal{S}), Y\omega(Z) = \omega(\nabla_Y^{(\text{m})} Z) \\ &\Leftrightarrow \omega \text{ is m-parallel.} \end{aligned} \quad (\text{D2})$$

□

Proof of Proposition 7.15 By Propositions 7.13 and 7.14, the condition imposed on f in (7.26) is equivalent to the existence of $F \in \mathcal{L}_h$ satisfying (7.16). □

Appendix E Proof of Proposition 8.5

We present a proof of Proposition 8.5 below, which is almost parallel to the proof of Theorem 8.4 in Chap. VII of [7] cited as a result due to E. Cartan.

Let $\dim S = n + r$, and M be a ∇ -totally geodesic submanifold with $\dim M = n$. We take a coordinate system $\tilde{\xi} = (\tilde{\xi}^i)$ of S such that M is represented as

$$M = \{p \in S \mid \forall i \in \{n+1, \dots, n+r\}, \tilde{\xi}^i(p) = 0\}$$

and that $(\xi^1, \dots, \xi^n) := (\tilde{\xi}^1|_M, \dots, \tilde{\xi}^n|_M)$ forms a coordinate system of M . Let $\tilde{\partial}_i := \frac{\partial}{\partial \tilde{\xi}^i}$, $\partial_i := \frac{\partial}{\partial \xi^i}$, and denote the connection coefficients of ∇ w.r.t. $\tilde{\xi}$ by $\{\Gamma_{ij}^k\}$: $\nabla_{\tilde{\partial}_i} \tilde{\partial}_j = \sum_k \Gamma_{ij}^k \tilde{\partial}_k$ for $i, j \in \{1, \dots, n+r\}$. For arbitrary i, j , it follows from the assumption (8.6) that $\mathcal{T}^{(\nabla)}(\tilde{\partial}_i, \tilde{\partial}_j) = \sum_k (\Gamma_{ij}^k - \Gamma_{ji}^k) \tilde{\partial}_k \in \text{span}_{\mathcal{F}(S)}\{\tilde{\partial}_i, \tilde{\partial}_j\}$, which implies that $\Gamma_{ij}^k - \Gamma_{ji}^k = 0$ for any $k \notin \{i, j\}$. Hence we have

$$\forall i, j \in \{1, \dots, n\}, \forall k \in \{n+1, \dots, n+r\}, \Gamma_{ij}^k = \Gamma_{ji}^k. \quad (\text{E1})$$

Given a point $p \in M$ and a tangent vector $X_p = \sum_{i=1}^n x^i (\partial_i)_p \in T_p(M)$ arbitrarily, let $\gamma : t \mapsto \gamma(t)$ be a ∇ -geodesic with an affine parameter t satisfying $\gamma(0) = p$

and $\dot{\gamma}(0) := \frac{d}{dt}\gamma(t)|_{t=0} = X_p$. The geodesic should satisfy the differential equation $\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0$, which is represented as

$$\forall k \in \{1, \dots, n+r\}, \quad \frac{d^2}{dt^2} \tilde{\xi}^k(\gamma(t)) + \sum_{i,j=1}^{n+r} \frac{d}{dt} \tilde{\xi}^i(\gamma(t)) \frac{d}{dt} \tilde{\xi}^j(\gamma(t)) (\Gamma_{ij}^k)_{\gamma(t)} = 0.$$

Since M is assumed to be ∇ -totally geodesic, $\gamma(t)$ stays in M and hence $\tilde{\xi}^k(\gamma(t)) = 0$ for $k \in \{n+1, \dots, n+r\}$. Therefore, the above equation yields

$$\forall k \in \{n+1, \dots, n+r\}, \quad \sum_{i,j=1}^n \frac{d}{dt} \tilde{\xi}^i(\gamma(t)) \frac{d}{dt} \tilde{\xi}^j(\gamma(t)) (\Gamma_{ij}^k)_{\gamma(t)} = 0,$$

and letting $t = 0$, we obtain

$$\forall k \in \{n+1, \dots, n+r\}, \quad \sum_{i,j=1}^n x^i x^j (\Gamma_{ij}^k)_p = 0.$$

Since $p \in M$ and $X_p = \sum_{i=1}^n x^i (\partial_i)_p$ are arbitrary and Γ_{ij}^k is symmetric w.r.t. $i \leftrightarrow j$ by (E1), it follows that

$$\forall i, j \in \{1, \dots, n\}, \quad \forall k \in \{n+1, \dots, n+r\}, \quad \Gamma_{ij}^k|_M = 0. \quad (\text{E2})$$

Now, for arbitrary vector fields $X = \sum_{i=1}^n X^i \partial_i = \sum_{i=1}^n X^i \tilde{\partial}_i|_M$ and $Y = \sum_{j=1}^n Y^j \partial_j = \sum_{j=1}^n Y^j \tilde{\partial}_j|_M$ on M , where $\{X^i\}, \{Y^j\} \subset \mathcal{F}(M)$, we have

$$\begin{aligned} \nabla_X Y &= \sum_{i,j=1}^n \sum_{k=1}^{n+r} X^i Y^j \Gamma_{ij}^k|_M \tilde{\partial}_k|_M + \sum_{j=1}^n X(Y^j) \partial_j \\ &= \sum_{i,j=1}^n \sum_{k=1}^n X^i Y^j \Gamma_{ij}^k|_M \partial_k + \sum_{j=1}^n X(Y^j) \partial_j \in \mathcal{X}(M), \end{aligned}$$

which concludes that M is ∇ -autoparallel in S and completes the proof.

Note that the only difference from the proof of [7] is whether (E1) is derived from $\mathcal{T}^{(\nabla)} = 0$ or from the weaker assumption (8.6).

Appendix F Proof of Proposition 8.7

- (1) (a) is straightforward from condition (i) as already mentioned. To show (b) and (c), let us introduce the distribution $\mathcal{V}^\perp : S \ni p \mapsto V_p^\perp$, where $V_p^\perp$ denotes the orthogonal complement of V_p in $(T_p(S), g_p)$. Then $\mathcal{V}^\perp$ is ∇^* -parallel since

$$\forall p, q \in S, \quad \Phi_{p,q}^{(\nabla^*)}(V_p^\perp) = (\Phi_{p,q}^{(\nabla)}(V_p))^\perp = V_q^\perp,$$

where the first equality follows from (2.2). Since ∇^* is torsion-free, the ∇^* -parallel distribution $\mathcal{V}^\perp$ is integrable (cf. (iv) in Proposition 8.1) and generates a foliation $S = \bigsqcup_\beta N_\beta$ satisfying (b). Property (c) is obvious from the construction.

- (2) Define the 1-forms $\{\omega^i\} = \{\omega^1, \dots, \omega^n\}$ by $X^i \xleftarrow{g} \omega^i$. Then $\{\omega^i\}$ are ∇^* -parallel due to (7.28), and $\exists\{\zeta^i\}$, $\omega^i = d\zeta^i$ due to (7.31). Thus we have $X^i \xleftarrow{g} d\zeta^i$. For each point $p \in S$ and each i , the relation $X_p^i \xleftarrow{g_p} (d\zeta^i)_p$ implies that $\text{Ker}(d\zeta^i)_p = (X_p^i)^\perp$, so that $\text{Ker}(d\zeta)_p = V_p^\perp$. To show (a), let $M = M_\alpha$ for an arbitrary α , and define $\xi = (\xi^i)$ by $\xi^i := \zeta^i|_M$. Then we have

$$\forall p \in M, \quad \text{Ker}(d\xi)_p = \text{Ker}(d\zeta)_p \cap T_p(M) = V_p^\perp \cap V_p = \{0\},$$

which implies that ξ forms a coordinate system of M (at least locally). Noting that M is a ∇ -autoparallel submanifold of S generated by the distribution $\mathcal{V} : p \mapsto V_p = \text{span}\{X_p^i\}$ and that $X^i \xleftarrow{g} d\zeta^i$, we have

$$X^i|_M \text{ is a } \nabla|_M\text{-parallel vector field on } M \text{ and } X^i|_M \xleftarrow{g|_M} d\xi^i.$$

It then follows that $d\xi^i$ is ∇_M^* -parallel for every i (cf. (7.28)), which is equivalent to saying that the coordinate system $\xi = (\xi^i)$ is ∇_M^* -affine as

$$\begin{aligned} & \forall i, \forall Y \in \mathcal{X}(M), \quad (\nabla_M^*)_Y d\xi^i = 0 \\ \Leftrightarrow & \forall i, j, \forall Y \in \mathcal{X}(M), \quad ((\nabla_M^*)_Y d\xi^i)(\partial_j) = 0, \quad \text{where } \partial_j := \frac{\partial}{\partial \xi^j} \\ \Leftrightarrow & \forall i, j, \forall Y \in \mathcal{X}(M), \quad Y(d\xi^i(\partial_j)) - d\xi^i((\nabla_M^*)_Y \partial_j) = 0 \\ \Leftrightarrow & \forall j, \forall Y \in \mathcal{X}(M), \quad (\nabla_M^*)_Y \partial_j = 0, \end{aligned} \tag{F1}$$

where the last equivalence follows from $d\xi^i(\partial_j) = \delta_j^i = \text{const}$. This proves (a). Property (b) is derived from

$$\forall \beta, \forall p \in N_\beta, \quad T_p(N_\beta) = V_p^\perp = \text{Ker}(d\zeta)_p,$$

which means that each leaf N_β is represented by the constraint $\zeta = \text{const}$.

Appendix G Proof of Lemma 8.13

We can assume Q to be a diagonal matrix with no loss of generality, and let

$$Q = [q_i \delta_{ij}], \quad q_i > 0, \quad R = [r_{ij}], \quad r_{ij} = -r_{ji}.$$

Suppose that real matrices $F = [f_{ij}]$, $f_{ij} = -f_{ji}$ and $C = [c_{ij}]$, $c_{ij} = c_{ji}$ satisfy

$$[Q, F] = Q \circ C \quad \text{and} \quad [R, F] = R \circ C,$$

which are represented as

$$(q_i - q_j) f_{ij} = \frac{q_i + q_j}{2} c_{ij}$$

and

$$\sum_k (r_{ik} f_{kj} - f_{ik} r_{kj}) = \frac{1}{2} \sum_k (r_{ik} c_{kj} + c_{ik} r_{kj}).$$

Eliminating $\{c_{ij}\}$ from these equations, we have

$$\sum_k \frac{q_j}{q_j + q_k} r_{ik} f_{kj} = \sum_k \frac{q_i}{q_i + q_k} r_{kj} f_{ik}. \quad (\text{G1})$$

Condition (8.17) means that this equation holds identically for variables $\{f_{ij}\}$ satisfying $f_{ij} = -f_{ji}$. For each (i, j) with $i \neq j$, the variables $\{f_{kj}\}_{k \neq j}$ and $\{f_{ik}\}_{k \neq i}$ appearing in (G1) are all independent except for f_{ij} , so that the identity (G1) implies that

$$\forall k \in \{1, \dots, d\} \setminus \{i, j\}, \quad \frac{q_j}{q_j + q_k} r_{ik} = \frac{q_i}{q_i + q_k} r_{kj} = 0,$$

which leads to

$$\forall k \in \{1, \dots, d\} \setminus \{i, j\}, \quad r_{ik} = r_{kj} = 0. \quad (\text{G2})$$

When $d \geq 3$, the condition that (G2) holds for $\forall i \neq \forall j$ implies that $r_{ij} = 0$ for $\forall i \neq \forall j$, i.e., $R = 0$.

Appendix H Proof of Lemma 9.4

When $\text{Tr} A = \text{Tr} B = 0$, (9.20) is reduced to

$$\frac{1}{2} [[A, B], \rho] = -2 \langle A \rangle_\rho (\rho \circ B) + 2 \langle B \rangle_\rho (\rho \circ A), \quad (\text{H1})$$

which we prove first. Letting $A = \vec{a} \cdot \vec{\sigma}$, $B = \vec{b} \cdot \vec{\sigma}$ and $\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma})$, it immediately follows from (9.4) and (9.5) that

$$\frac{1}{2} [[A, B], \rho] = (\vec{r} \times (\vec{a} \times \vec{b})) \cdot \vec{\sigma}$$

and

$$-2 \langle A \rangle_\rho (\rho \circ B) + 2 \langle B \rangle_\rho (\rho \circ A) = \{(\vec{b} \cdot \vec{r})\vec{a} - (\vec{a} \cdot \vec{r})\vec{b}\} \cdot \vec{\sigma}.$$

Hence, the well-known formula for the vector triple product proves (H1).

Remove the assumption $\text{Tr} A = \text{Tr} B = 0$, and let $A' := A - \frac{\text{Tr} A}{2} I$ and $B' := B - \frac{\text{Tr} B}{2} I$. Then we have

$$\begin{aligned} \frac{1}{2} [[A, B], \rho] &= \frac{1}{2} [[A', B'], \rho] \\ &= -2 \langle A' \rangle_\rho (\rho \circ B') + 2 \langle B' \rangle_\rho (\rho \circ A') \\ &= (\text{Tr} A - 2 \langle A \rangle_\rho) (\rho \circ B) - (\text{Tr} B - 2 \langle B \rangle_\rho) (\rho \circ A) \end{aligned}$$

$$+ \{(\mathrm{Tr} B) \langle A \rangle_\rho - (\mathrm{Tr} A) \langle B \rangle_\rho\} \rho,$$

where the second equality follows from (H1). Thus we obtain (9.20).

Appendix I Proof of (9.29) and the condition for equality

Let $A = [a_{kl}]$ and $B = [b_{kl}]$ be $n \times n$ positive-definite real matrices defined by

$$a_{kl} := (\mathbf{u}^k)^T G_\rho^{-1} \mathbf{u}^l \quad \text{and} \quad b_{kl} := (\mathbf{w}_k)^T W \mathbf{w}_l.$$

Then we have

$$\begin{aligned} \mathrm{tr}(W V_\rho(\Pi)) &= \sum_k \frac{1}{p_k} a_{kk} b_{kk} = \sum_k \left(\frac{\sqrt{a_{kk} b_{kk}}}{\sqrt{p_k}} \right)^2 \cdot \sum_k (\sqrt{p_k})^2 \\ &\geq \left(\sum_k \sqrt{a_{kk} b_{kk}} \right)^2 \geq \left(\mathrm{tr} \left(A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)^2, \end{aligned} \quad (\text{I1})$$

where the last inequality can be regarded as a variant of the monotonicity of the fidelity or of the Bures distance [25]. Letting $U := (\mathbf{u}^1, \dots, \mathbf{u}^n) \in \mathbb{R}^{n \times n}$, we have $A = U^T G_\rho^{-1} U$, so that there exists an orthogonal matrix Q such that $A^{\frac{1}{2}} = Q G_\rho^{-\frac{1}{2}} U$. Noting that $U^{-1} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ and that $B = U^{-1} W (U^{-1})^T$, we have

$$A^{\frac{1}{2}} B A^{\frac{1}{2}} = Q G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} Q^T,$$

which yields

$$\left(A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} = Q \left(G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} \right)^{\frac{1}{2}} Q^T$$

and

$$\mathrm{tr} \left(A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} = \mathrm{tr} \left(G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}} \right)^{\frac{1}{2}}. \quad (\text{I2})$$

Now (9.29) follows from (I1) and (I2).

Equality in (9.29) holds iff both of the two inequalities in (I1) hold with equality, which occurs iff

$$p_k \propto \sqrt{a_{kk} b_{kk}} \quad (\text{I3})$$

and

$$\exists D = \mathrm{diag}\{d_1, \dots, d_n\}, \quad d_k > 0, \quad B = D A D. \quad (\text{I4})$$

Since $U B U^T = W$ and $(U^T)^{-1} A U^{-1} = G_\rho^{-1}$ we have

$$B = D A D \Leftrightarrow W = K G_\rho^{-1} K, \quad (\text{I5})$$

where

$$K := UDU^T = \sum_k d_k \mathbf{u}^k (\mathbf{u}^k)^T. \quad (I6)$$

In addition, it follows from $B = DAD$ that $\sqrt{a_{kk}b_{kk}} = d_k a_{kk}$. Hence, the equality conditions (I3) and (I4) are equivalent to the existence of positive numbers $\{d_k\}$ such that

$$d_k \propto \frac{p_k}{a_{kk}} \text{ and } W = KG_\rho^{-1}K \text{ with (I6),}$$

which is also equivalent to the existence of a positive-definite matrix K such that

$$W = KG_\rho^{-1}K \text{ and } K \propto \sum_k \frac{p_k}{a_{kk}} \mathbf{u}^k (\mathbf{u}^k)^T,$$

namely, (9.30) and (9.31).

Appendix J On a result of Yamagata [17] and its relation to our results

In the situation of Example 9.9, let π be the random measurement represented as

$$\pi : (p_k) \xrightarrow{\text{rnd}} k \rightarrow \boxed{\sigma_k} \xrightarrow{\text{meas.}} y. \quad (J1)$$

|
→ k

Note that the estimators defined in (9.39) and (9.40) are both represented in the form $\Pi = (\pi, f)$ by some functions f . Yamagata [17] treated the quantum tomography for qubit states as the MLE (maximum likelihood estimator) for the Stokes parameter $\xi = (\xi^1, \xi^2, \xi^3)$ of the full model $\mathcal{M} = \mathcal{S}(\mathcal{H})$ based on repetitive measurement of this π with the uniform distribution $p_k = 1/3$. According to the classical estimation theory, the asymptotic variance of the MLE for a state ρ is equal to $(G_\rho^{(\pi)})^{-1}$, where $G_\rho^{(\pi)}$ denotes the Fisher information matrix of the classical model induced from $(\mathcal{S}(\mathcal{H}), \xi)$ and the measurement π . In addition, owing to the achievability of the classical Cramér-Rao inequality, we can construct an estimator based on the measurement π which is locally unbiased at ρ and has the variance $(G_\rho^{(\pi)})^{-1}$. Hence, for an arbitrary weight matrix $W = [w_{ij}] > 0$, it follows from (9.27) that

$$\text{tr}(W(G_\rho^{(\pi)})^{-1}) \geq \tau_\rho(W). \quad (J2)$$

Yamagata showed that the necessary and sufficient condition for a weight field $\mathcal{W} = \{W_\rho = [w_{ij}(\rho)] \mid \rho \in \mathcal{S}\}$ to satisfy

$$\forall \rho \in \mathcal{S}(\mathcal{H}), \quad \text{tr}(W_\rho(G_\rho^{(\pi)})^{-1}) = \tau_\rho(W_\rho) \quad (J3)$$

is that (9.43) identically holds; i.e.,

$$\forall \rho \in \mathcal{S}(\mathcal{H}), \quad w_{ij}(\rho) \propto \frac{\delta_{ij} - \xi^i(\rho)\xi^j(\rho)}{(1 - (\xi^i(\rho))^2)(1 - (\xi^j(\rho))^2)}, \quad (J4)$$

where the proportionality constant may depend on ρ .

This result is readily extended to the situation of Example 9.9 where an e-autoparallel model $(\mathcal{M}, \xi)$ such that (9.36) holds for every $\rho \in \mathcal{M}$ is given and (p_k) in (J1) is an arbitrary positive probability vector, implying that for a weight field $\mathcal{W} = \{W_\rho = [w_{ij}(\rho)] \mid \rho \in \mathcal{M}\}$ on the model, we have

$$\begin{aligned} \forall \rho \in \mathcal{M}, \quad \text{tr}(W_\rho (G_\rho^{(\pi)})^{-1}) &= \tau_\rho(W_\rho) \\ \Leftrightarrow \forall \rho \in \mathcal{M}, \quad w_{ij}(\rho) &\propto p_i p_j \frac{\delta_{ij} - \xi^i(\rho) \xi^j(\rho)}{(1 - (\xi^i(\rho))^2)(1 - (\xi^j(\rho))^2)}. \end{aligned} \quad (\text{J5})$$

On the other hand, (9.42) means that

$$\begin{aligned} \forall \rho \in \mathcal{M}, \quad \text{tr}(W_\rho V_\rho(\Pi_\rho)) &= \tau_\rho(W_\rho) \\ \Leftrightarrow \forall \rho \in \mathcal{M}, \quad w_{ij}(\rho) &\propto p_i p_j \frac{\delta_{ij} - \xi^i(\rho) \xi^j(\rho)}{(1 - (\xi^i(\rho))^2)(1 - (\xi^j(\rho))^2)}, \end{aligned} \quad (\text{J6})$$

where Π_ρ denotes the estimation defined by (9.39). Here, Π_ρ depends on ρ in the form $\Pi_\rho = (\pi, f_\rho)$, where

$$f_\rho(k, y) := \xi(\rho) + \frac{y - \xi^k(\rho)}{p_k} \mathbf{e}_k. \quad (\text{J7})$$

The equivalence of (J5) and (J6) may be understood as a consequence of the identity

$$V_\rho(\Pi_\rho) = (G_\rho^{(\pi)})^{-1}, \quad (\text{J8})$$

which follows from the classical Cramér-Rao inequality $V_\rho(\Pi_\rho) \geq (G_\rho^{(\pi)})^{-1}$ and the existence of $W > 0$ such that $\text{tr}(W V_\rho(\Pi_\rho)) = \tau_\rho(W) \leq \text{tr}(W (G_\rho^{(\pi)})^{-1})$ (see (J2)).

The conditions in (J5) and (J6) are expressed as

$$\forall \rho \in (\mathcal{M}), \quad \text{tr}(W_\rho (G_\rho^{(\pi)})^{-1}) = \min_{\pi': \text{POVM}} \text{tr}(W_\rho (G_\rho^{(\pi')})^{-1}) \quad (\text{J9})$$

and

$$\forall \rho \in \mathcal{M}, \quad \text{tr}(W_\rho V_\rho(\Pi_\rho)) = \min_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W_\rho V_\rho(\Pi')). \quad (\text{J10})$$

It is worthwhile to compare these conditions with the definition of $\mathcal{W}$ -efficiency of an estimator given in Sect. 4; an estimator Π for the model $(\mathcal{M}, \xi)$ is said to be $\mathcal{W}$ -efficient when it satisfies $\forall \rho \in \mathcal{M}, \Pi \in \mathcal{U}(\rho, \xi)$, which is equivalent to the unbiasedness $\forall \rho \in \mathcal{M}, E_\rho(\Pi) = \xi(\rho)$, and

$$\forall \rho \in \mathcal{M}, \quad \text{tr}(W_\rho V_\rho(\Pi)) = \min_{\Pi' \in \mathcal{U}(\rho, \xi)} \text{tr}(W_\rho V_\rho(\Pi')). \quad (\text{J11})$$

Since f_ρ in $\Pi_\rho = (\pi, f_\rho)$ depends on ρ , (J10) does not mean the existence of a $\mathcal{W}$ -efficient estimator. On the other hand, if we take f to be $f_0(k, y) = \frac{y}{p_k} \mathbf{e}_k$ as in (9.40),

we obtain an unbiased estimator $\Pi_0 = (\pi, f_0)$, but its variance for a state $\rho \in \mathcal{M}$ does not achieve the minimum in (J11) except when $f_0 = f_\rho$, i.e. $\xi(\rho) = 0$.

Appendix K Proof of Proposition 9.12

Suppose that an n -dimensional e-autoparallel qubit model $(\mathcal{M}, \xi)$ and a positive weight $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ are arbitrary given as in the statement of the proposition. We assume the existence of a W -efficient estimator for $(\mathcal{M}, \xi)$ and derive a contradiction. As noted in Example 9.9, we may assume without loss of generality that (9.36) holds for any $\rho \in \mathcal{M}$; i.e.,

$$\forall \rho \in \mathcal{M}, \quad \forall i \in \{1, \dots, n\}, \quad \xi^i(\rho) = \langle \sigma_i \rangle_\rho \quad \text{and} \quad L_\rho^i = \sigma_i - \xi^i(\rho),$$

which implies (9.37); i.e.,

$$g_\rho^{ij} = \delta_{ij} - \xi^i(\rho)\xi^j(\rho). \quad (\text{K1})$$

Now, assume that an estimator $\Pi = \Pi(d\hat{\xi})$ for the model $(\mathcal{M}, \xi)$ is W -efficient. Namely, we assume that

$$\forall \rho \in \mathcal{M}, \quad E_\rho(\Pi) = \xi(\rho) \quad \text{and} \quad \text{tr}(WV_\rho(\Pi)) = \tau_\rho(W),$$

which is expressed as

$$\forall i \in \{1, \dots, n\}, \quad \forall \rho \in \mathcal{M}, \quad \left\langle A^i \right\rangle_\rho = \xi^i(\rho) \quad (\text{K2})$$

and

$$\forall \rho \in \mathcal{M}, \quad \sum_{i,j} w_{ij} \left\langle B^{ij} \right\rangle_\rho - \sum_{i,j} w_{ij} \xi^i(\rho)\xi^j(\rho) = \tau_\rho(W), \quad (\text{K3})$$

where

$$A^i := \int \hat{\xi}^i \Pi(d\hat{\xi}) \quad \text{and} \quad B^{ij} := \int \hat{\xi}^i \hat{\xi}^j \Pi(d\hat{\xi}),$$

Letting $\{\lambda_i(\rho)\}$ be the eigenvalues of $G_\rho^{-\frac{1}{2}} W G_\rho^{-\frac{1}{2}}$, we have

$$\tau_\rho(W) = \left(\sum_i \sqrt{\lambda_i(\rho)} \right)^2 = \sum_i \lambda_i(\rho) + 2 \sum_{i < j} \sqrt{\lambda_i(\rho)\lambda_j(\rho)}. \quad (\text{K4})$$

Since (K1) yields

$$\sum_i \lambda_i(\rho) = \text{tr}(W G_\rho^{-1}) = \sum_i w_{ii} - \sum_{i,j} w_{ij} \xi^i(\rho)\xi^j(\rho),$$

(K3) is expressed as

$$\forall \rho \in \mathcal{M}, \quad \langle B \rangle_\rho = \text{tr} W + 2 \sum_{i < j} \sqrt{\lambda_i(\rho)\lambda_j(\rho)}, \quad (\text{K5})$$

where

$$B := \sum_{i,j} w_{ij} B^{ij}.$$

Consider the case $n = 2$ first. Then there exists a real constant c for which the model $(\mathcal{M}, \xi)$ is represented as (9.17) and (9.18) in the Bloch ball. In the present situation, this means that

$$\forall \rho \in \mathcal{M}, \quad \langle \sigma_3 \rangle_\rho = c \sqrt{1 - \|\xi(\rho)\|^2}.$$

Representing B as

$$B = b_0 + \sum_{k=1}^3 b_k \sigma_k \quad (\text{K6})$$

and noting that

$$\lambda_1(\rho) \lambda_2(\rho) = \det(W G_\rho^{-1}) = (\det W)(1 - \|\xi(\rho)\|^2),$$

(K5) is expressed as

$$\begin{aligned} \forall \rho \in \mathcal{M}, \quad b_0 + \sum_{i=1}^2 b_i \xi^i(\rho) + b_3 c \sqrt{1 - \|\xi(\rho)\|^2} \\ = \operatorname{tr} W + 2\sqrt{\det W} \sqrt{1 - \|\xi(\rho)\|^2}, \end{aligned}$$

which implies that

$$b_0 = \operatorname{tr} W, \quad b_1 = b_2 = 0, \quad \text{and} \quad b_3 c = 2\sqrt{\det W}. \quad (\text{K7})$$

If $c = 0$, the third condition of (K7) leads to a contradiction. So we assume $c \neq 0$, which yields

$$B = \operatorname{tr} W + \frac{2\sqrt{\det W}}{c} \sigma_3. \quad (\text{K8})$$

On the other hand, letting A^i be represented as

$$A^i = a_0^i + \sum_{k=1}^3 a_k^i \sigma_k,$$

the unbiasedness condition (K2) is expressed as

$$\forall \rho \in \mathcal{M}, \quad a_0^i + \sum_{j=1}^2 a_j^i \xi^j(\rho) + a_3^i c \sqrt{1 - \|\xi(\rho)\|^2} = \xi^i(\rho),$$

which implies that

$$a_0^i = 0, \quad a_j^i = \delta_j^i, \quad \text{and} \quad a_3^i = 0,$$

or equivalently,

$$A^i = \sigma_i. \quad (\text{K9})$$

From the operator inequality

$$0 \leq \sum_{i,j} w_{ij} \int (\hat{\xi}^i - A^i) \Pi(d\hat{\xi}) (\hat{\xi}^j - A^j) = B - A,$$

where

$$A := \sum_{i,j} w_{ij} A^i A^j = \sum_{i,j} w_{ij} \sigma_i \sigma_j = \text{tr } W,$$

we have

$$B \geq \text{tr } W. \quad (\text{K10})$$

This contradicts with (K8).

Consider the case $n = 3$ next; i.e. $\mathcal{M} = \mathcal{S}$. Representing B as (K6), condition (K5) is expressed as

$$\forall \rho \in \mathcal{S}, \quad b_0 + \sum_{i=1}^3 b_i \xi^i(\rho) = \text{tr } W + 2 \sum_{i < j} \sqrt{\lambda_i(\rho) \lambda_j(\rho)}. \quad (\text{K11})$$

This implies that $\sum_{i < j} \sqrt{\lambda_i(\rho) \lambda_j(\rho)}$ is an affine function of the Stokes parameters (ξ^i) of ρ , which is clearly impossible. Hence, we arrive at a contradiction.

Author Contributions The content of the paper was developed through discussions between H.N. and A.F., based on H.N.'s original idea. The manuscript was primarily written by H.N., who incorporated A.F.'s feedback as needed. Both authors reviewed and approved the final version.

Funding This work is partly supported by JSPS KAKENHI Grant Numbers 23H05492 (HN), 25K24611 (HN), 17H02861 (HN, AF), 23H01090 (AF), 23K25787 (AF), and JST ERATO Grant Number JPM-JER2402 (AF).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors are current members of the Editorial Board of Information Geometry. The authors state that there is no other conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Rao, C.R.: Linear Statistical Inference and Its Applications, 2nd edn. Wiley, Amsterdam (1973)
2. Kiefer, J.C.: Introduction to Statistical Inference. Springer, New York (1987)
3. Amari, S., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Oxford University Press, Oxford (2000)
4. Helstrom, C.W.: Minimum mean-square error estimation in quantum statistics. *Phys. Lett.* **25A**, 101–102 (1967)
5. Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic Press, New York (1976)
6. Suzuki, J., Yang, Y., Hayashi, M.: Quantum state estimation with nuisance parameters. *J. Phys. A: Math. Theor.* **53**, 453001 (2020)
7. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, II. Wiley, New York (1969)
8. Nagaoka, H., Amari, S.: Differential geometry of smooth families of probability distributions. *Info. Geo.* <https://doi.org/10.1007/s41884-025-00187-y> (2026). (The original version appeared as METR 82-7, Dept. Math. Eng. and Instr. Phys., Univ. of Tokyo (1982).)
9. Petz, D.: Monotone metrics on matrix spaces. *Linear Algebra Appl.* **244**, 81–96 (1996)
10. Holevo, A.S.: Probabilistic and Statistical Aspects of Quantum Theory, Pisa: Edizioni della Normale (2011). (the previous edition. Amsterdam: North-Holland (1982).)
11. Ohara, A.: Geodesics for dual connections and means on symmetric cones. *Integr. Equ. Oper. Theory* **50**, 537–548 (2004)
12. Nagaoka, H.: Information-geometrical characterization of statistical models which are statistically equivalent to probability simplex. In: Proc. 2017 IEEE International Symposium on Information Theory, 1346–1350 (2017)
13. Nagaoka, H.: A new approach to Cramér-Rao bounds for quantum state estimation. In: Asymptotic Theory of Quantum Statistical Inference (ed. Hayashi M.), World Scientific, Chap. 8 (2005). (Originally appeared as IEICE Technical Report, 89, 228, IT 89-42:9–14 (1989).)
14. Nagaoka, H.: A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory. *Trans. Jap. Soc. Indust. Appl. Math.*, 1:43–56 (1991), in Japanese. (The English version translated by Tanaka F. and Tsuda Y. appeared in Asymptotic Theory of Quantum Statistical Inference (ed. Hayashi M.), World Scientific, Chap. 11 (2005).)
15. Hayashi, M.: A linear programming approach to attainable Cramér-Rao type bounds. In: Quantum Communication, Computing, and Measurement, Plenum Publishing, 99–108 (1997)
16. Gill, R.D., Massar, S.: State estimation for large ensembles. *Phys. Rev. A* **61**, 042312 (2000)
17. Yamagata, K.: Efficiency of quantum state tomography for qubits. *Int. J. Quant. Info.* **9**, 1167–1183 (2011)
18. Guță, M., Kahn, J.: Local asymptotic normality for qubit states. *Phys. Rev. A* **73**, 052108 (2006)
19. Kahn, J., Guță, M.: Local asymptotic normality for finite dimensional quantum systems. *Commun. Math. Phys.* **289**, 597–652 (2009)
20. Yamagata, K., Fujiwara, A., Gill, R.D.: Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Stat.* **41**, 2197–2217 (2013)
21. Fujiwara, A., Yamagata, K.: Noncommutative Lebesgue decomposition and contiguity with application to quantum local asymptotic normality. *Bernoulli* **26**, 2105–2142 (2020)
22. Fujiwara, A., Yamagata, K.: Efficiency of estimators for locally asymptotically normal quantum statistical models. *Ann. Stat.* **51**, 1159–1182 (2023)
23. Fujiwara, A.: Foundations of quantum local asymptotic normality via information geometry. *Info. Geo.* (2026). <https://doi.org/10.1007/s41884-026-00207-5>
24. Fujiwara, A.: Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *J. Phys. A: Math. Gen.* **39**, 12489–12504 (2006)
25. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)