Autoparallelity of Quantum Statistical Manifolds in The Light of Quantum Estimation Theory

Hiroshi Nagaoka[1*], Akio Fujiwara[2]

[1]The University of Electro-Communications, Japan, [2]Osaka University, Japan

Submitted to Journal:
Frontiers in Physics

Specialty Section:
Statistical and Computational Physics

Article type:
Original Research Article

Manuscript ID:
1253743

Received on:
06 Jul 2023

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

CRedIT Author Statement

Akio Fujiwara: Writing – review & editing. Hiroshi Nagaoka: Writing – original draft, Writing – review & editing.

Keywords

Quantum estimation theory, information geometry, Autoparallel submanifold, Dual connection, Torsion, SLD (symmetric logarithmic derivative)

Abstract

Word count: 137

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 eautoparallelity, some of which are valid for the autoparallelity w.r.t. an affine connection in a more general geometrical situation.

Contribution to the field

The field of this paper is quantum information geometry, whose main interest lies in the relationship between geometric structures of quantum statistical manifolds consisting of density operators (quantum states) and physical and/or engineering aspects of quantum systems. This paper focuses on the relationship between the autoparallelity and estimation problems for quantum states. The autoparallelity is a differential-geometrical concept representing how a submanifold is flatly embedded in the whole manifold equipped with an affine connection, although the word "flatness" as a mathematical terminology is used for a different notion. In this paper, we consider the autoparallelity for a submanifold of the space of strictly positive density operators on which an affine connection is introduced by an information-geometric method, and show that a submanifold is autoparallel if and only if it has certain good properties from a viewpoint of quantum estimation theory. On the purely geometrical side, the paper sheds new light on the importance of autoparallelity for an affine connection having non-vanishing torsion.

Funding information

JSPS KAKENHI Grant Number 23H05492

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Autoparallelity of Quantum Statistical Manifolds in The Light of Quantum Estimation Theory

Hiroshi Nagaoka*[1] and Akio Fujiwara[2]

[1]The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan. nagaoka@is.uec.ac.jp
[2]Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan. fujiwara@math.sci.osaka-u.ac.jp
Correspondence : Corresponding Author*

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 geometrical situation.

Keywords— quantum estimation theory, information geometry, autoparallel submanifold, dual connection, torsion, SLD (symmetric logarithmic derivative)

1 Introduction

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. In the classical information geometry for manifolds of probability distributions, where the triple $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ of the Fisher metric g, the e-connection $\nabla^{(\mathrm{e})}$ and the m-connection $\nabla^{(\mathrm{m})}$ plays the leading role, the autoparallelity w.r.t. (with respect to) the e-connection, which is called the e-autoparallelity, is particularly important. This is because the e-autoparallel submanifolds of the space $\mathcal{P}$ consisting of all strictly positive probability distributions on a finite set are the exponential families, which is one of the key concepts in probability theory and statistics. For quantum statistical manifolds, which are submanifolds of the space $\mathcal{S}$ consisting of all strictly positive density operators on a finite-dimensional Hilbert space, we may introduce an analogous notion of exponential families as autoparallel submanifolds w.r.t. an affine connection on $\mathcal{S}$ analogous to the classical e-connection. However, the notion of quantum exponential families introduced in this way does not necessarily have statistical and/or physical importance. One of the main achievements of the present paper is that the e-autoparallelity for the SLD structure, which is among a family of information-geometric structures introduced on $\mathcal{S}$ in a unified way, has been shown to possess estimation-theoretical characterizations.

In order to clarify our motivation, 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 $\mathcal{M} = \{p_\xi \,|\, \xi = (\xi^1, \ldots, \xi^n) \in \Xi\} \subset \mathcal{P}$ 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, whose existence 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.

Condition (ii) in the above theorem means that the elements of $\mathcal{M}$ are represented as

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

and that the parameter ξ satisfies

$$\xi^i = E_\xi(F^i), \tag{1.2}$$

where E_ξ denotes the expectation w.r.t. the distribution p_ξ. This condition is expressed in the language of geometry as follows.

(ii)' $\mathcal{M}$ is autoparallel in $\mathcal{P}$ w.r.t. the e-connection of $\mathcal{P}$, and ξ forms an affine coordinate system w.r.t. the m-connection of $\mathcal{M}$.

Remark 1.2. In the notation of [3] as well as of many references on 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.

A quantum version of Theorem 1.1 is known, which we state 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 $V_\xi \geq G_\xi^{-1}$ [6, 7], where G_ξ is the SLD Fisher information matrix (see Section 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

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

and that

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

When a model $\mathcal{M} = \{\rho_\xi\}$ is represented as (1.3), we call it a *quasi-classical exponential family*. As is pointed out in [3] and will be verified in Section 4 of this paper, a quasi-classical exponential family is e-autoparallel in $\mathcal{S}$ w.r.t. the SLD structure. Note, however, that this is merely a special case of e-autoparallel submanifolds. Namely, the existence of efficient estimator is too strong as a characterization of the e-autoparallelity. Is it then possible to characterize the e-autoparallelity by an estimation-theoretic condition which is weaker than the existence of efficient estimator? We give an affirmative answer to this question in Section 5. We also give another characterization of the e-autoparallelity in Section 7 by considering estimation for scalar-valued functions instead of estimation for vector-valued parameters.

As mentioned above, the e-autoparallelity in the SLD structure has estimation-theoretical significance and is therefore a concept worth studying further. It should be noted here that the e-connection in the SLD structure is curvature-free but not torsion-free, so that the e-connection is not flat. This also means that $\mathcal{S}$ is not a dually flat space w.r.t. the SLD structure. 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 obvious. For a non-flat connection, on the other hand, we cannot see the whole picture of autoparallel submanifolds and are faced with the new problem of what kind of condition ensures the existence of autoparallel submanifold. Therefore, it is also important to study the autoparallelity from a purely geometrical point of view, away from estimation problems. This is another concern in this paper, along with the estimation-theoretical consideration.

The paper is organized as follows. In Section 2, we explain the basic issues about the autoparallelity for an affine connection ∇ on a general differential manifold, focusing in particular on the situation where the dual connection of ∇ w.r.t. a Riemannian metric is flat. In Section 3, we introduce a family of information-geometric structures on the space $\mathcal{S}(\mathcal{H})$, and derive the basic issues concerning the e-autoparallelity by applying the results of Section 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 descriptions are as self-contained as possible. Section 4 also consists of basically known results, where Theorem 1.3 is revisited, and it is clarified that the existence of efficient estimator only partially characterizes the e-autoparallelity for the SLD structure. This observation motivates Section 5, where a sequence of estimators, which is called a filtration of estimators, is treated instead of a single estimator, and the existence of efficient filtration is shown to characterize the e-autoparallelity. In Section 6, it is shown that a quantum Gaussian shift model has an efficient filtration 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 geometrical argument there. Section 7 treats an estimation problem for scalar-valued functions, where it is shown that a quantum statistical manifold is e-autoparallel in $\mathcal{S}$ if and only if the linear space formed by functions having efficient estimators is of maximal dimension. In Section 8, we move away from estimation theory and consider the condition for existence of e-autoparallel submanifolds from a purely geometrical point of view, where the involutivity of a parallel distribution of tangent spaces is studied in relation to the torsion tensor. In Section 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. It is shown there that $\mathcal{S}(\mathcal{H})$ in this case has a characteristic property that every e-parallel distribution is involutive. Section 10 is devoted to concluding remarks. Some proofs and additional results are included in Appendix 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 $\xi = (\xi^1, \ldots, \xi^n)$ is a coordinate system of a manifold M in the subsequent sections, it basically means that ξ is a map : $M \to \mathbb{R}^n$ (a global chart of M) which represents each point $p \in M$ by an n-dimensional vector $\xi(p) = (\xi^1(p), \ldots, \xi^n(p)) \in \mathbb{R}^n$, 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 Section 6 for quantum Gaussian states and in Section 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 $\mathcal{S}, \mathcal{M}, \mathcal{N}, \ldots$, while general manifolds are denoted by $S, M, N, \ldots$.

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 $\mathscr{X}(S)$, respectively. Suppose that S is provided with an affine connection ∇, which is a map $\mathscr{X}(S) \times \mathscr{X}(S) \to \mathscr{X}(S)$, $(X, Y) \mapsto \nabla_X Y$. Given a submanifold M of S, let $\mathscr{X}(S/M)$ 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 $\mathscr{X}(M) \times \mathscr{X}(S/M) \to \mathscr{X}(S/M)$ so that for any $X \in \mathscr{X}(M)$ and any $Y \in \mathscr{X}(S/M)$, $\nabla_X Y$ is defined as an element of $\mathscr{X}(S/M)$. Since $\mathscr{X}(M) =$

$\mathscr{X}(M/M) \subset \mathscr{X}(S/M)$, $\nabla_X Y \in \mathscr{X}(S/M)$ is defined for any $X, Y \in \mathscr{X}(M)$, although it does not necessarily belong to $\mathscr{X}(M)$.

When $\nabla_X Y$ belongs to $\mathscr{X}(M)$ for every $X, Y \in \mathscr{X}(M)$, M is said to be *autoparallel* w.r.t. ∇ or ∇-*autoparallel* in S (e.g., Sec. 8 in Chap. VII of [4]). 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 merely for simplicity of descriptions.

Remark 2.2. A similar but different notion to autoparallelity is total geodesicness. 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 geodesicness, but the converse is not true in general except when ∇ is torsion-free ([4], Theorem 8.4 in Chap. VII). We will revisit this topic in Remark 8.6.

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

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. When (S, ∇) is curvature-free, the parallel transport $\Phi^{(\nabla)}_{p,q} : T_p(S) \to T_q(S)$ is defined for arbitrary two points $p, q \in S$ so that a vector field $X \in \mathscr{X}(S)$ is ∇-parallel iff $\forall p, q \in S$, $\Phi^{(\nabla)}_{p,q}(X_p) = X_q$.

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

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

(i) X is ∇-parallel.

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

Proposition 2.4. *For an n-dimensional submanifold M of S on which a curvature-free connection ∇ is given, 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* $\tilde{X}^1, \ldots, \tilde{X}^n$ *on* S *such that* $\forall i,\ \tilde{X}^i|_M \in \mathscr{X}(\mathcal{M})$*.*

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 flat. Namely, the triple (g, ∇, ∇^*) satisfies the duality [5] [3]:

$$\forall X, Y, Z \in \mathscr{X}(S),\ Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z), \tag{2.1}$$

and ∇^* is flat in the sense that it is curvature-free and torsion free. The flatness is known to be equivalent to the existence of a coordinate system $\xi = (\xi^i)$, which is called an *affine coordinate system w.r.t.* ∇^*, such that $\partial_i = \frac{\partial}{\partial \xi^i}, i \in \{1, \ldots, \dim S\}$, are all ∇^*-parallel. In this case, ∇ turns out to be curvature-free, since the curvature-freeness is preserved by the duality of connections (Theorem 3.3 of [3]), but is not necessarily torsion-free, and hence (S, g, ∇, ∇^*) is not necessarily dually flat.

Proposition 2.5. *In the above situation, we have:*

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

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

Proof The proposition relies not on the torsion-freeness of ∇^* but only on the curvature-freeness of ∇ and ∇^*, so that it suffices to show (1). For an $X \in \mathscr{X}(S)$, we have

$$\begin{aligned} X \text{ is } \nabla\text{-parallel} &\Leftrightarrow \forall Y, \forall Z,\ g(\nabla_Z X, Y) = 0 \\ &\overset{\text{a}}{\Leftrightarrow} \forall Y : \nabla^*\text{-parallel}, \forall Z,\ g(\nabla_Z X, Y) = 0 \\ &\overset{\text{b}}{\Leftrightarrow} \forall Y : \nabla^*\text{-parallel}, \forall Z,\ Zg(X, Y) = 0 \\ &\Leftrightarrow \forall Y : \nabla^*\text{-parallel},\ g(X, Y) \text{ is constant}, \end{aligned} \tag{2.2}$$

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

$$Zg(X,Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y) = g(\nabla_Z X, Y). \tag{2.3}$$

$\square$

Suppose that M is ∇-autoparallel in S. Then M is a Riemannian submanifold of (S,g) equipped with the affine connection $\nabla|_M$. Hence 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.,

$$\forall X, Y, Z \in \mathscr{X}(M),\ Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, (\nabla^*_M)_X Z), \tag{2.4}$$

where we have applied $(\nabla|_M)_X Y = \nabla_X Y$. From (2.1) and (2.4), we have

$$\forall X, Y, Z \in \mathscr{X}(M),\ g(Y, (\nabla^*_M)_X Z) = g(Y, \nabla^*_X Z), \tag{2.5}$$

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

$$\begin{tikzcd}
\nabla \arrow[r, "\text{restriction}"] \arrow[d, "g\text{-dual}"'] & \nabla|_M \arrow[d, "g|_M\text{-dual}"] \\
\nabla^* \arrow[r, "g\text{-projection}"] & \nabla^*_M
\end{tikzcd}$$

Since the curvature-freeness and the torsion-freeness are respectively preserved by the duality and the projection, ∇^*_M is flat as in the case with ∇^*, which ensures the existence of ∇^*_M-affine coordinate system of M.

The following proposition will be of fundamental importance for later arguments.

Proposition 2.6. *For an n-dimensional submanifold M of S and a coordinate system ξ of M, the following conditions are equivalent.*

(i) M is ∇-autoparallel in S, and ξ is a ∇^*_M-affine coordinate system.

(ii) For every $i \in \{1, \ldots, n\}$, the vector field

$$X^i := \sum_j g^{ij} \partial_j \in \mathscr{X}(M) \tag{2.6}$$

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 $X^i \in \mathscr{X}(M)$ by (2.6). Then we have $g(X^i, \partial_j) = \delta^i_j$ for every i, j, which is constant on M. Noting that $n(= \dim M)$ vector fields $\{\partial_j\}$ are ∇^*_M-parallel and that Prop. 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 Prop. 2.5 that X^i is $\nabla|_M$-parallel, and hence it is ∇-parallel.

We next show (ii) $\Rightarrow$ (i). Assume (ii). Then according to Prop. 2.4, M is ∇-autoparallel in S. Noting that $g(X^i, \partial_j)$ is constant on M and applying item (2) of Prop. 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. $\square$

3 Information geometric structures on quantum statistical manifolds

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

Let $\mathcal{L} = \mathcal{L}(\mathcal{H})$, $\mathcal{L}_\mathrm{h} = \mathcal{L}_\mathrm{h}(\mathcal{H})$ 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 $\mathcal{L}_{\mathrm{h},1} := \{A \in \mathcal{L}_\mathrm{h} \mid \mathrm{Tr}A = 1\}$, so that a flat affine connection is naturally introduced on $\mathcal{S}$, which we call the *m-connection* and denote by $\nabla^{(\mathrm{m})}$. In order to express $\nabla^{(\mathrm{m})}$ more explicitly, we introduce the embedding map $\iota : \mathcal{S} \to \mathcal{L}_{\mathrm{h},1}$ so that $\rho \in \mathcal{S}$ is denoted by $\iota(\rho)$ when treating it as an element of $\mathcal{L}_{\mathrm{h},1}$. Since ι is a smooth map, it has the differential at every point $\rho \in \mathcal{S}$, which we denote by $\iota_* = (d\iota)_\rho : T_\rho(\mathcal{S}) \to \mathcal{L}_{\mathrm{h},0} := \{A \in \mathcal{L}_\mathrm{h} \mid \mathrm{Tr}A = 0\}$. For a vector field $X \in \mathscr{X}(\mathcal{S})$, the map $\iota_*(X) : \mathcal{S} \to \mathcal{L}_{\mathrm{h},0}$ 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:

$$\forall X, Y \in \mathscr{X}(\mathcal{S}),\ \iota_*(\nabla^{(\mathrm{m})}_X Y) = X\iota_*(Y), \tag{3.1}$$

where $X\iota_*(Y) : \mathcal{S} \to \mathcal{L}_{\mathrm{h},0}$ 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

$$\iota_*(\partial_i) = \partial_i \rho_\xi, \tag{3.2}$$

and

$$\iota_*\left(\nabla^{(\mathrm{m})}_{\partial_i} \partial_j\right) = \partial_i\, \iota_*(\partial_j) = \partial_i \partial_j \rho_\xi, \tag{3.3}$$

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 $\mathcal{L}_\mathrm{h}$, where the correspondence $\rho \mapsto \langle \cdot, \cdot \rangle_\rho$ is smooth, and assume that

$$\forall \rho \in \mathcal{S}, \forall A \in \mathcal{L}_\mathrm{h},\ \langle A, I \rangle_\rho = \langle A \rangle_\rho := \mathrm{Tr}(\rho A). \tag{3.4}$$

The inner products are represented as

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

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

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

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

$$\forall A \in \mathcal{L}_\mathrm{h},\ X_\rho \langle A \rangle = \left\langle L_{X_\rho}, A \right\rangle_\rho, \tag{3.7}$$

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

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

From (3.4) and (3.7), we have

$$\left\langle L_{X_\rho} \right\rangle_\rho = \left\langle L_{X_\rho}, I \right\rangle_\rho = X_\rho \langle I \rangle = X_\rho 1 = 0. \tag{3.9}$$

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

$$\{L_{X_\rho} \,|\, X_\rho \in T_\rho(\mathcal{S})\} = \{A \in \mathcal{L}_\mathrm{h} \,|\, \langle A \rangle_\rho = 0\}. \tag{3.10}$$

In the following, we often express (3.4) as

$$\forall A \in \mathcal{L}_\mathrm{h},\ \langle A, I \rangle = \langle A \rangle \tag{3.11}$$

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

$$\forall A \in \mathcal{L}_\mathrm{h},\ X \langle A \rangle = \langle L_X, A \rangle, \tag{3.12}$$

where X is a vector field on $\mathcal{S}$, L_X denotes the map $\mathcal{S} \to \mathcal{L}_\mathrm{h},\ \rho \mapsto L_{X_\rho}$, and $\langle L_X, A \rangle$ denotes the function $\rho \mapsto \left\langle L_{X_\rho}, A \right\rangle_\rho$.

For a submanifold $\mathcal{M}$ of $\mathcal{S}$ and a vector field $X \in \mathscr{X}(\mathcal{M})$ on it, the map $L_X : \mathcal{M} \to \mathcal{L}_\mathrm{h}, \rho \mapsto L_{X_\rho}$ 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

$$\forall A \in \mathcal{L}_\mathrm{h},\ \partial_i \langle A \rangle = \partial_i \langle A \rangle |_\mathcal{M} = \langle L_i, A \rangle, \tag{3.13}$$

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 $\iota_*(X_\rho) = X_\rho^{(\mathrm{m})}$ and $L_{X_\rho} = X_\rho^{(\mathrm{e})}$, respectively. We do not use the symbols $X_\rho^{(\mathrm{m})}$ and $X_\rho^{(\mathrm{e})}$ in this paper, but may use the following notation when convenient:

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

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

Note that

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

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

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

which is equivalently written as

$$\forall X, Y \in \mathscr{X}(\mathcal{S}),\ g(X, Y) := \langle L_X, L_Y \rangle = \mathrm{Tr}\left(\iota_*(X) L_Y\right). \tag{3.18}$$

We also have the expression

$$\forall X, Y \in \mathscr{X}(\mathcal{S}),\ g(X, Y) = -\langle X L_Y \rangle = -\langle Y L_X \rangle, \tag{3.19}$$

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

$$\begin{aligned} 0 = X_\rho \langle L_Y \rangle = X_\rho \langle L_{Y.} \rangle_{.} &= X_\rho \left\langle L_{Y_\rho} \right\rangle_{.} + X_\rho \langle L_{Y.} \rangle_\rho \\ &= \left\langle L_{X_\rho}, L_{Y_\rho} \right\rangle_\rho + \langle X_\rho L_Y \rangle_\rho, \end{aligned} \tag{3.20}$$

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

$$\Omega_\rho(A) = f(\Delta_\rho)(A\rho) = f(\Delta_\rho)(A)\rho, \tag{3.21}$$

where $f : (0, \infty) \to (0, \infty)$ is an operator monotone function satisfying $\forall x > 0$, $xf(1/x) = f(x)$ and $f(1) = 1$, and $\Delta_\rho : \mathcal{L} \to \mathcal{L}$ 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

$$\langle A, B \rangle_\rho = \mathrm{Re}\,\mathrm{Tr}(\rho AB) = \mathrm{Tr}(\rho(A \circ B)) = \mathrm{Tr}((\rho \circ A)B), \quad A, B \in \mathcal{L}_\mathrm{h}, \tag{3.22}$$

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 the tangent vector X_ρ (the vector field X, resp.). In particular, $L_i := L_{\partial_i}$ obeys the equation

$$\partial_i \rho_\xi = \rho_\xi \circ L_{i,\xi}, \tag{3.23}$$

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* $\nabla^{(\mathrm{e})}$ be defined as the dual connection of $\nabla^{(\mathrm{m})}$ w.r.t. g; i.e.,

$$\forall X, Y, Z \in \mathscr{X}(\mathcal{S}),\ Xg(Y, Z) = g(\nabla^{(\mathrm{e})}_X Y, Z) + g(Y, \nabla^{(\mathrm{m})}_X Z). \tag{3.24}$$

We have thus obtained $(\mathcal{S}, g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ as an example of (S, g, ∇, ∇^*) treated in the previous section, where $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ are dual w.r.t. g, $\nabla^{(\mathrm{e})}$ is curvature-free and $\nabla^{(\mathrm{m})}$ is flat. The triple $(\mathcal{S}, g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ 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 $(\mathcal{S}, g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ is dually flat). That is the structure induced from the BKM (Bogoliubov-Kubo-Mori) inner product

$$\langle A, B \rangle_\rho = \int_0^1 \mathrm{Tr}\left(\rho^s A \rho^{1-s} B\right) ds, \quad A, B \in \mathcal{L}_{\mathrm{h}}. \tag{3.25}$$

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 $\mathcal{T}^{(\mathrm{e})}(X, Y) = \nabla^{(\mathrm{e})}_X Y - \nabla^{(\mathrm{e})}_Y X - [X, Y]$ 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

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

where $[\cdot, \cdot]$ in the RHS denotes the commutator for operators on $\mathcal{H}$. Since this representation will be used in Sections 8 and 9, we show its proof in A1 of appendix 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 $X, Y, W \in \mathscr{X}(\mathcal{S})$, we have*

$$W = \nabla^{(\mathrm{e})}_X Y \;\Leftrightarrow\; L_W = XL_Y - \langle XL_Y \rangle = XL_Y + g(X, Y). \tag{3.27}$$

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

$$\begin{aligned} Xg(Y, Z) &= \mathrm{Tr}\left((XL_Y)\, \iota_*(Z)\right) + \mathrm{Tr}\left(L_Y\, (X\iota_*(Z))\right) \\ &= \mathrm{Tr}\left((XL_Y)\, \iota_*(Z)\right) + g(Y, \nabla^{(\mathrm{m})}_X Z), \end{aligned}$$

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

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

This means that $W' = \nabla^{(\mathrm{e})}_X Y$, and proves the proposition. □

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

$$\begin{aligned} X \text{ is e-parallel } &\Leftrightarrow\; \exists A \in \mathcal{L}_{\mathrm{h}},\; L_X = A - \langle A \rangle \\ &\Leftrightarrow\; \exists A \in \mathcal{L}_{\mathrm{h}}, \forall \rho \in \mathcal{S},\; L_{X_\rho} = A - \langle A \rangle_\rho . \end{aligned} \tag{3.28}$$

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

$$\begin{aligned} X \text{ is e-parallel } &\Leftrightarrow\; \forall Y : \text{m-parallel},\; g(X, Y) \text{ is constant on } \mathcal{S} \\ &\Leftrightarrow\; \forall B \in \mathcal{L}_{\mathrm{h},0},\; \langle L_X, B \rangle \text{ is constant on } \mathcal{S} \\ &\Leftrightarrow\; \exists A \in \mathcal{L}_{\mathrm{h}},\; L_X = A - \langle A \rangle . \end{aligned}$$

□

Recalling Prop. 2.3, the following corollary is immediate.

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

$$\begin{aligned} X \text{ is e-parallel } &\Leftrightarrow\; \exists A \in \mathcal{L}_{\mathrm{h}},\; L_X = A - \langle A \rangle |_{\mathcal{M}} \\ &\Leftrightarrow\; \exists A \in \mathcal{L}_{\mathrm{h}}, \forall \rho \in \mathcal{M},\; L_{X_\rho} = A - \langle A \rangle_\rho . \end{aligned} \tag{3.29}$$

The following corollary is also immediate from Prop. 3.5.

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

$$X_\sigma = \Phi^{(\mathrm{e})}_{\rho,\sigma}(X_\rho) \;\Leftrightarrow\; L_{X_\sigma} = L_{X_\rho} - \left\langle L_{X_\rho} \right\rangle_\sigma . \tag{3.30}$$

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 $\nabla^{(\mathrm{m})}_{\mathcal{M}}$ on $\mathcal{M}$".)

(ii) $\exists\{F^1, \ldots, F^n\} \subset \mathcal{L}_{\mathrm{h}}$ such that for every $i \in \{1, \ldots, n\}$

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

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) $\exists\{F^1, \ldots, F^n\} \subset \mathcal{L}_{\mathrm{h}}$ such that for every $i \in \{1, \ldots, n\}$

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

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

(iv) $\exists\{F^1, \ldots, F^n\} \subset \mathcal{L}_{\mathrm{h}}$ such that for every $i \in \{1, \ldots, n\}$

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

where $\mathbb{R}$ is identified with $\{cI \,|\, c \in \mathbb{R}\}$ (see Remark 3.9 below).

Proof The equivalence (i) $\Leftrightarrow$ (ii) is immediate from Prop. 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 $\sum_j g^{ij} g_{jk} = \delta^i_k = \partial_k \xi^i$ 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

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

we have (iii) $\Rightarrow$ (iv). To show the converse, we assume the existence of $\{F^1, \ldots, F^n\}$ 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

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

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 $-\left\langle F^i \right\rangle_\rho = -\xi^i(\rho)$ due to $\langle L_{j,\rho} \rangle_\rho = 0$. Thus (iii) is concluded. $\square$

Remark 3.9. In (3.31) and (3.32), the operators $\{F^i - \left\langle F^i \right\rangle_\rho\}_{i=1}^n$ turn out to be linearly independent for each $\rho \in \mathcal{M}$, which implies that $\{F^1, \ldots, F^n, I\}$ are linearly independent, or equivalently that $\{F^1, \ldots, F^n\}$ are linearly independent in the quotient space $\mathcal{L}_\mathrm{h}/\mathbb{R}$ 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:

$$
\forall \{A_t\}_{t=1}^k, \forall \{B_t\}_{t=1}^k \subset \mathcal{L}_\mathrm{h}(\mathcal{H}),\ \Big\langle \bigotimes_{t=1}^k A_t, \bigotimes_{t=1}^k B_t \Big\rangle_{\rho^{\otimes k}} = \prod_{t=1}^k \langle A_t, B_t \rangle_\rho, \tag{3.34}
$$

which is equivalent to

$$
\forall \{A_t\}_{t=1}^k \subset \mathcal{L}_\mathrm{h}(\mathcal{H}),\ \Omega_{\rho^{\otimes k}}\Big(\bigotimes_{t=1}^k A_t\Big) = \bigotimes_{t=1}^k \Omega_\rho(A_t). \tag{3.35}
$$

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 $\Omega_{\rho^{\otimes k}} = f\big(\Delta_{\rho^{\otimes k}}\big)$ hold. In particular, the proposition hols for the SLD structure.

Proposition 3.10. *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 $\mathcal{S}(\mathcal{H}^{\otimes k}) \to \mathcal{L}_{\mathrm{h},1}(\mathcal{H}^{\otimes k})$. With the aid of the parametric representation (3.2), we see that

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

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

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

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 $\tilde{g}_{ij}(\rho^{\otimes k}) = \left\langle \tilde{L}_{i,\rho^{\otimes k}}, \tilde{L}_{j,\rho^{\otimes k}} \right\rangle_{\rho^{\otimes k}}$, 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 $\tilde{L}^i_{\rho^{\otimes k}} = \frac{1}{k}\,(L^i_\rho)^{(k)}$. Now, according to Prop. 3.8, conditions (i), (ii) of the present proposition are respectively expressed as

(i)' $\exists \{F^i\} \subset \mathcal{L}(\mathcal{H}),\ \forall i, \forall \rho \in \mathcal{M},\ L^i_\rho = F^i - \xi^i(\rho)$.

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

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

4 Efficient estimators

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.

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 finite set Ω and a function $f : \Omega \to \mathbb{R}^n$, 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

$$E_\rho(\Pi) := \int \hat{\xi} \, \mathrm{Tr}\left(\rho \, \Pi(d\hat{\xi})\right), \tag{4.1}$$

$$V_\rho(\Pi) := \int (\hat{\xi} - \xi(\rho))(\hat{\xi} - \xi(\rho))^T \, \mathrm{Tr}\left(\rho \, \Pi(d\hat{\xi})\right), \tag{4.2}$$

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

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

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

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

$$E^i_\rho(\Pi) = \xi^i(\rho) \quad \text{and} \quad \partial_j E^i_\rho(\Pi) = \delta^i_j, \tag{4.5}$$

where $\partial_j E^i_\rho(\Pi)$ denotes the derivative of the function $E^i(\Pi) : \sigma \mapsto E^i_\sigma(\Pi)$ 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 $L_i = L_{\partial_i}$, $i \in \{1, \ldots, n\}$, we have

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

where

$$A^i := \int \hat{\xi}^i \Pi(d\hat{\xi}) \in \mathcal{L}_{\mathrm{h}}(\mathcal{H}). \tag{4.7}$$

The well-known SLD Cramér-Rao inequality [6, 7] states that every $\Pi \in \mathcal{U}(\rho, \xi)$ obeys

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

where $G_\rho = [g_{ij}(\rho)]$ denotes the SLD Fisher information matrix defined by $g_{ij} = \langle L_i, L_j \rangle = g(\partial_i, \partial_j)$, where g is the SLD metric. Furthermore, the following proposition holds.

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

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

Note that inf in (4.9) cannot be replaced with min in general.

Let us introduce a class of randomized procedures for estimation that will be useful in the proofs of both Prop. 4.1 above and Theorem 5.1 later. Suppose that a point $\rho \in \mathcal{M}$, a basis $\{\boldsymbol{u}^1, \ldots, \boldsymbol{u}^n\}$ of $\mathbb{R}^n$, a positive probability vector $(p_1, \ldots, p_n)$ s.t. $\forall i,\ p_i > 0, \sum_i p_i = 1$ and n^2 real numbers $\{\gamma_k^i\}$ satisfying

$$\sum_k p_k\, \gamma_k^i = \xi^i(\rho) \tag{4.10}$$

are arbitrarily given. Let

$$X^k := \sum_i u_i^k\, L_\rho^i \in \mathcal{L}_{\mathrm{h}}(\mathcal{H}), \tag{4.11}$$

where $\boldsymbol{u}^k = [u_i^k]$, $L_\rho^i := \sum_j g^{ij}(\rho) L_{j,\rho}$, $G_\rho^{-1} = [g^{ij}(\rho)]$, and consider the random measurement such that $k \in \{1, \ldots, n\}$ is randomly chosen according to the probability distribution $(p_1, \ldots, p_n)$ 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 nth 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 $f = (f^1, \ldots, f^n)^T$ is defined by

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

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 $\varphi : \mathbb{R}^n \to \mathbb{R}$,

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

In this situation we have the following lemma, whose proof is given in A2 of Appendix.

Lemma 4.2. *The estimator $\Pi = (\pi, f)$ satisfies:*

(1) $\Pi \in \mathcal{U}(\rho, \xi)$.

(2) $\forall k,\ (\boldsymbol{u}^k)^T V_\rho(\Pi) \boldsymbol{u}^k = \dfrac{1}{p_k} (\boldsymbol{u}^k)^T G_\rho^{-1} \boldsymbol{u}^k + \sum_l p_l (a_l^k)^2,$

where $a_l^k := \sum_i u_i^k (\gamma_l^i - \xi^i(\rho))$.

Proof of Prop. 4.1 Let $\gamma_k^i := \xi^i(\rho)$, which satisfies (4.10) so that Lemma 4.2 is applicable. Since $a_l^k = 0$ in this case, we have

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

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

A locally unbiased estimator $\Pi \in \mathcal{U}(\rho, \xi)$ is called *locally efficient* for ξ at ρ if $V_\rho(\Pi) \leq V_\rho(\Pi')$ for all $\Pi' \in (\rho, \xi)$. Given a positive-semidefinite matrix $W \in \mathbb{R}^{n\times n}$ as a weight, an estimator $\Pi \in \mathcal{U}(\rho, \xi)$ is called *locally W-efficient* for ξ at ρ if $\mathrm{tr}\,(WV_\rho(\Pi)) \leq \mathrm{tr}\,(WV_\rho(\Pi'))$ for all $\Pi' \in \mathcal{U}(\rho, \xi)$, or equivalently if

$$\mathrm{tr}\,(WV_\rho(\Pi)) = \min_{\Pi' \in \mathcal{U}(\rho,\xi)} \mathrm{tr}\,(WV_\rho(\Pi')). \tag{4.14}$$

Here, the symbol tr is used for the trace of $n \times n$ matrices to distinguish it from the trace Tr for operators on $\mathcal{H}$.

Proposition 4.3. *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.*

(i) Π *is locally efficient for ξ at ρ.*

(ii) $V_\rho(\Pi) = G_\rho^{-1}$.

(iii) Π *is locally $\boldsymbol{u}\boldsymbol{u}^T$-efficient for ξ at ρ for every column vector $\boldsymbol{u} \in \mathbb{R}^n$.*

(iv) 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

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

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

Remark 4.4. As is well known, there exists a locally efficient estimator at ρ iff the SLDs $L_{1,\rho}, \ldots, L_{n,\rho}$ mutually commute (e.g. § 7.4 of [3]).

An estimator Π is called *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 called *$\mathcal{W}$-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 Prop. 4.3, Π

is efficient $\Leftrightarrow$ Π is $\boldsymbol{u}\boldsymbol{u}^T$-efficient for every $\boldsymbol{u} \in \mathbb{R}^n$ $\Leftrightarrow$ Π 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

$$\rho = \exp\Big[\frac{1}{2}\Big\{\sum_{i=1}^{n}\theta_i(\rho)F^i - \psi(\rho)\Big\}\Big]\, P \exp\Big[\frac{1}{2}\Big\{\sum_{i=1}^{n}\theta_i(\rho)F^i - \psi(\rho)\Big\}\Big] \tag{4.15}$$

and satisfies $\xi^i(\rho) = \langle F^i \rangle_\rho$, where $F^1, \ldots, F^n$ 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

$$L_i = \partial_i\Big(\sum_j \theta_j F^j - \psi\Big) = \sum_j (\partial_i\theta_j)(F^j - \partial^j\psi), \tag{4.16}$$

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

$$L^i = \sum_j g^{ij}L_j = F^i - \xi^i. \tag{4.17}$$

According to Prop. 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 $\nabla^{(\mathrm{e})}|_{\mathcal{M}}$ on M turns out to 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 $n = 1$, (4.15) is written as

$$\rho = \exp\left[\frac{1}{2}\big\{\theta(\rho)F - \psi(\rho)\big\}\right] P \exp\left[\frac{1}{2}\big\{\theta(\rho)F - \psi(\rho)\big\}\right]. \tag{4.18}$$

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, (4.15) 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 Efficient filtration 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 $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$. Given a nonnegative real matrix $W \in \mathbb{R}^{n\times n}$ as a weight, a filtration $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$ is called *locally W-efficient* for ξ at ρ if $\Pi_\varepsilon \in \mathcal{U}(\rho,\xi)$ for every $\varepsilon > 0$ and $\lim_{\varepsilon\downarrow 0} \mathrm{tr}\,(WV_\rho(\Pi_\varepsilon)) \le \mathrm{tr}\,(WV_\rho(\Pi'))$ for every $\Pi' \in \mathcal{U}(\rho,\xi)$, which is equivalent to

$$\lim_{\varepsilon\downarrow 0} \mathrm{tr}\,(WV_\rho(\Pi_\varepsilon)) = \inf_{\Pi'\in\mathcal{U}(\rho,\xi)} \mathrm{tr}\,(WV_\rho(\Pi')). \tag{5.1}$$

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

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

due to Prop. 4.1. Compare (5.1) with (4.14), and note that a locally W-efficient filtration at ρ always exists, even when a locally W-efficient estimator does not exist.

Given a weight field $\mathcal{W} = \{W_\rho \,|\, \rho \in \mathcal{M}\}$, a filtration $\vec{\Pi} = (\Pi_\varepsilon)_{\varepsilon>0}$ is called *$\mathcal{W}$-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.

Now, we have the following theorem, which gives an estimation-theoretical characterization of the e-autoparallelity.

Theorem 5.1. *For a 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.

(ii) For every $\boldsymbol{u} \in \mathbb{R}^n$, there exists a $\boldsymbol{u}\boldsymbol{u}^T$-efficient filtration for ξ.

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

$$\exists\{F^i\},\ \forall i,\ \ L^i = F^i - \xi^i, \tag{5.3}$$

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

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

$$\begin{aligned}(\boldsymbol{e}^i)^T V_\rho(\Pi)\boldsymbol{e}^i &= \int (\hat{\xi}^i - \xi^i(\rho))^2 \mathrm{Tr}(\rho\Pi_\varepsilon(d\hat{\xi})) \\ &\ge \mathrm{Tr}(\rho(F^i_\varepsilon - \xi^i(\rho))^2) = \|F^i_\varepsilon - \xi^i(\rho)\|^2_\rho,\end{aligned} \tag{5.4}$$

where

$$F^i_\varepsilon := \int \hat{\xi}^i \Pi_\varepsilon(d\hat{\xi}) \in \mathcal{L}_{\mathrm{h}}(\mathcal{H}),$$

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

$$\left\langle F^i_\varepsilon - \xi^i(\rho), L_{j,\rho}\right\rangle_\rho = \delta^i_j = \left\langle L^i_\rho, L_{j,\rho}\right\rangle_\rho.$$

This means that L^i_ρ is the orthogonal projection of $F^i_\varepsilon - \xi^i(\rho)$ onto $\mathrm{span}\{L_{j,\rho}\}_{j=1}^n$. Hence we have

$$\|F^i_\varepsilon - \xi^i(\rho)\|^2_\rho = \|L^i_\rho\|^2_\rho + \|F^i_\varepsilon - \xi^i(\rho) - L^i_\rho\|^2_\rho. \tag{5.5}$$

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

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

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

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

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

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

$$Y^k := \sum_i u^k_i F^i. \tag{5.6}$$

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

$$g^i(k, y) := \frac{w^i_k}{p_k} y, \tag{5.7}$$

where $[w^i_k] = [u^k_i]^{-1}$. Invoking (5.3) evaluated at an arbitrary point $\rho \in \mathcal{M}$, we have

$$\begin{aligned} g^i(k, Y^k) &= \frac{w^i_k}{p_k} \sum_j u^k_j F^j = \frac{w^i_k}{p_k} \sum_j u^k_j (L^j_\rho + \xi^j(\rho)) \\ &= \gamma^i_k + \frac{w^i_k}{p_k} X^k = f^i(k, X^k), \end{aligned}$$

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

$$\gamma^i_k := \frac{w^i_k}{p_k} \sum_j u^k_j \xi^j(\rho). \tag{5.8}$$

Since this γ_k^i satisfies (4.10), Lemma 4.2 applies to conclude that Π_ε is locally unbiased at ρ and satisfies, for every k and every $\rho \in \mathcal{M}$,

$$(\boldsymbol{u}^k)^T V_\rho(\Pi_\varepsilon)\boldsymbol{u}^k = \frac{1}{p_k}(\boldsymbol{u}^k)^T G_\rho^{-1}\boldsymbol{u}^k + \sum_l p_l (a_l^k)^2, \tag{5.9}$$

where $a_l^k := \sum_i u_i^k(\gamma_l^i - \xi^i(\rho))$. From (5.8), we have

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

and hence

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

Thus, letting $k = 1$ in (5.9), we obtain

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

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

The following proposition follows immediately from Theorem 5.1 and Prop. 3.10.

Proposition 5.2. *In the situation of Prop. 3.10 where $(\tilde{\mathcal{M}}, \tilde{\xi})$ is the kth 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 $\boldsymbol{u} \in \mathbb{R}^n$ and $\varepsilon > 0$, arbitrarily, let Π_ε be a $(\boldsymbol{u}\boldsymbol{u}^T + \varepsilon I)$-efficient estimator. Then, for an arbitrary $\rho \in \mathcal{M}$ and an arbitrary $\Pi' \in \mathcal{U}(\rho, \xi)$, we have

$$\begin{aligned}\boldsymbol{u}^T V_\rho(\Pi_\varepsilon)\boldsymbol{u} &\le \mathrm{tr}\,((\boldsymbol{u}\boldsymbol{u}^T + \varepsilon I)V_\rho(\Pi_\varepsilon)) \\ &\le \mathrm{tr}\,((\boldsymbol{u}\boldsymbol{u}^T + \varepsilon I)V_\rho(\Pi')) = \boldsymbol{u}^T V_\rho(\Pi')\boldsymbol{u} + \varepsilon\,\mathrm{tr}\,(V_\rho(\Pi')).\end{aligned}$$

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

An important example of a model satisfying the condition of Prop. 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 underling 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 [8]. 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., $\{U(z) \,|\, z \in Z\}$ 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

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

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

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

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

$$\mu(z) = \langle I, R(z)\rangle_\rho = \langle R(z)\rangle_\rho, \tag{6.3}$$

$$\alpha(z,z') = \big\langle R(z) - \mu(z), R(z') - \mu(z')\big\rangle_\rho. \tag{6.4}$$

Assuming that α and linearly independent $\mu_1,\ldots,\mu_n$ are arbitrarily given, consider the model $\mathcal{M} = \{\rho_\xi \,|\, \xi = (\xi^1,\ldots,\xi^n) \in \mathbb{R}^n\}$, 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 [8]) that the model has W-efficient estimator for every positive weight W. Namely, the sufficient condition presented in Prop. 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 Prop. 3.8 as follows. According to [8], the ith SLD $L_{i,\xi}$ is represented as

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

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

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

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

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

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 Prop. 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 by Holevo is immediately applied to this extended model, so that it satisfies the sufficient condition of Prop. 6.1.

7 Another estimation-theoretical characterization of e-autoparallelity

In this section we give another characterization to the e-autoparallelity 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 $\xleftrightarrow{g_p}$, we have

$$X_p \xleftrightarrow{g_p} \omega_p \;\Leftrightarrow\; \forall Y_p \in T_p(M),\ \omega_p(Y_p) = g_p(X_p, Y_p). \tag{7.1}$$

This is extended to the correspondence $\xleftrightarrow{g}$ between a vector field $X \in \mathscr{X}(M)$ and a differential 1-form $\omega \in \mathcal{D}(M)$, where $\mathcal{D}(M)$ denotes the totality of 1-forms on M, such that

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

When a coordinate system $\xi = (\xi^i)$ is given on M, and $X \in \mathscr{X}(M)$ 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

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

where $g_{ij} = g(\partial_i, \partial_j)$ and $g^{ij} = g(d\xi^i, d\xi^j)$ which form the inverse matrices of each other.

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

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

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

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

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

$$E_\rho(\Lambda) := \int \hat{t}\, \mathrm{Tr}\left(\rho\, \Lambda(d\hat{t})\right), \tag{7.6}$$

$$V_\rho(\Lambda) := \int (\hat{t} - f(\rho))^2\, \mathrm{Tr}\left(\rho\, \Lambda(d\hat{t})\right). \tag{7.7}$$

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

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

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

$$\forall i \in \{1, \ldots, n\}, \;\; \partial_i E_\rho(\Lambda) = \partial_i f(\rho), \tag{7.9}$$

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

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

The minimum of (7.10) *is achieved by the spectral measure of the observable*

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

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

Proof Given an estimator Λ, let

$$A := \int \hat{t}\, \Lambda(d\hat{t}) \in \mathcal{L}_{\mathrm{h}}.$$

Then the local unbiasedness of Λ at ρ is represented as

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

and we have

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

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

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

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

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

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 $A = f(\rho) + \sum_j \partial_j f(\rho) L^j_\rho = F_\rho$. These observations prove the proposition. $\square$

Based on Proposition 7.1, we call an estimator Λ *locally efficient* for f at ρ when $\Lambda \in \mathcal{U}(\rho, f)$ and $V_\rho(\Lambda) = \|df\|^2_\rho$, 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 Prop. 7.1, it suffices to treat only estimators represented by Hermitian operators. Note that an estimator $F \in \mathcal{L}_{\mathrm{h}}$ is efficient for f iff

$$\forall \rho \in \mathcal{M},\ \langle F \rangle_\rho = f(\rho) \quad \text{and} \quad V_\rho(F) := \left\langle (F - \langle F \rangle_\rho)^2 \right\rangle_\rho = \|(df)_\rho\|^2_\rho. \tag{7.15}$$

We define

$$\mathcal{E}(\mathcal{M}) := \{ f \in \mathcal{F}(\mathcal{M}) \,|\, \text{there exists an efficient estimator for } f \}. \tag{7.16}$$

Proposition 7.2. *For a function $f \in \mathcal{F}(\mathcal{M})$, the following conditions are equivalent.*

(i) $f \in \mathcal{E}(\mathcal{M})$.

(ii) $\exists F \in \mathcal{L}_{\mathrm{h}},\ F - f = \sum_i (\partial_i f)\, L^i$.

(iii) $\exists F \in \mathcal{L}_{\mathrm{h}},\ F - \langle F \rangle |_{\mathcal{M}} = \sum_i (\partial_i f)\, L^i$.

(iv) $\operatorname{grad} f$ *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 Prop. 7.1, 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 $F \in \mathcal{L}_{\mathrm{h}}$ satisfies $F - \langle F \rangle |_{\mathcal{M}} = \sum_i (\partial_i f)\, L^i$. Then we have

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

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 $X := \operatorname{grad} f$. 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

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

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

Corollary 7.3. *$\mathcal{E}(\mathcal{M})$ is an $\mathbb{R}$-linear space.*

Proof Obvious from Prop. 7.2. $\square$

Proposition 7.4. *For a vector field $X \in \mathscr{X}(\mathcal{M})$, we have*

$$X \text{ is e-parallel} \Leftrightarrow \exists f \in \mathcal{E}(\mathcal{M}),\ X = \operatorname{grad} f. \tag{7.17}$$

Proof The implication $\Leftarrow$ follows from (i) $\Rightarrow$ (iv) in Prop. 7.2. To show the converse, assume that X is e-parallel. Then, according to Cor. 3.6, there exists $F \in \mathcal{L}_{\mathrm{h}}$ such that $L_X = F - f$, where $f := \langle F \rangle|_{\mathcal{M}}$. For any $Y \in \mathscr{X}(\mathcal{M})$ we have

$$g(X,Y) = \langle L_Y, F - f \rangle = \langle L_Y, F \rangle \overset{\star}{=} Y \langle F \rangle|_{\mathcal{M}} = Y f = df(Y), \tag{7.18}$$

where $\overset{\star}{=}$ follows from (3.7). This implies that $X = \operatorname{grad} f$. Since X is e-parallel, it follows from (iv) $\Rightarrow$ (i) in Prop. 7.2 that $f \in \mathcal{E}(\mathcal{M})$. Thus, $\Rightarrow$ in (7.17) has been verified. $\square$

Define

$$d\mathcal{E}(\mathcal{M}) := \{df \mid f \in \mathcal{E}(\mathcal{M})\} \subset \mathcal{D}(\mathcal{M}). \tag{7.19}$$

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

$$\mathscr{X}_{\text{e-par}}(\mathcal{M}) := \{X \in \mathscr{X}(\mathcal{M}) \mid X \text{ is e-parallel}\}. \tag{7.20}$$

Then we have the following proposition.

Proposition 7.5. *The correspondence $\overset{g|_{\mathcal{M}}}{\longleftrightarrow}$ establishes a linear isomorphism between $\mathscr{X}_{\text{e-par}}(\mathcal{M})$ and $d\mathcal{E}(\mathcal{M})$. As a consequence, we have*

$$\dim d\mathcal{E}(\mathcal{M}) = \dim \mathcal{E}(\mathcal{M}) - 1 = \dim \mathscr{X}_{\text{e-par}}(\mathcal{M}) \leq \dim \mathcal{M}. \tag{7.21}$$

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

$$X \in \mathscr{X}_{\text{e-par}}(\mathcal{M}) \Leftrightarrow \exists f \in \mathcal{E}(\mathcal{M}),\ \omega = df. \tag{7.22}$$

Since $\omega = df \Leftrightarrow X = \operatorname{grad} f$, this is just Prop. 7.4. $\square$

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

Theorem 7.6. *For an n-dimensional submanifold $\mathcal{M}$ of $\mathcal{S}$, the following conditions are equivalent.*

(i) $\mathcal{M}$ *is e-autoparallel in* $\mathcal{S}$.

(ii) $\dim \mathcal{E}(\mathcal{M}) = n + 1$.

Proof We have (i) $\Leftrightarrow \dim \mathscr{X}_{\text{e-par}}(\mathcal{M}) = n$ by Prop. 2.4, and $\dim \mathscr{X}_{\text{e-par}}(\mathcal{M}) = n$ $\Leftrightarrow$ (ii) by Prop. 7.5. $\square$

Theorem 7.7. *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.*

(ii) $\forall i \in \{1, \ldots, n\}$, $\xi^i \in \mathcal{E}(\mathcal{M})$.

(iii) $\mathcal{E}(\mathcal{M}) = \big\{c + \sum_{i=1}^{n} u_i \xi^i \,\big|\, (c, u_1, \ldots, u_n) \in \mathbb{R}^{n+1}\big\}$.

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

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

where the first $\Leftrightarrow$ follows from Prop. 2.6 and the second $\Leftrightarrow$ follows from Prop. 7.2.

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

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

where the second $\Leftrightarrow$ follows since $\dim \mathcal{E}(\mathcal{M}) \le n + 1$ by (7.21) and $\{1, \xi^1, \ldots, \xi^n\}$ are linearly independent. $\square$

Remark 7.8. If we replace $\mathcal{S}(\mathcal{H})$ by $\mathcal{P}(\Omega)$ in Theorems 7.6 and 7.7, these theorems hold as they are in the classical case. When the coordinate functions $\xi^1, \ldots, \xi^n$ satisfy condition (ii) in Theorem 7.7, they have their efficient estimators $F^1, \ldots, F^n$, which are functions $\Omega \to \mathbb{R}$ in this case, and the map $(F^1, \ldots, F^n) : \Omega \to \mathbb{R}^n$ becomes an efficient estimator for $\xi = (\xi^1, \ldots, \xi^n)$. Thus, we see that the equivalence (i) $\Leftrightarrow$ (ii) in the theorem is 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 A3 of Appendix.

Proposition 7.9. $\forall F \in \mathcal{L}_{\mathrm{h}}, \forall \rho \in \mathcal{S},\ V_\rho(F) := \big\langle (F - \langle F \rangle_\rho)^2 \big\rangle_\rho = \|(d\, \langle F \rangle)_\rho\|_\rho^2$.

Proposition 7.10. *We have*

$$\begin{aligned}\mathcal{E}(\mathcal{S}) &= \{\langle F\rangle \mid F\in\mathcal{L}_{\mathrm{h}}\}\\ &= \{f\in\mathcal{F}(\mathcal{S}) \,|\, \mathrm{grad}\, f \ \textit{is e-parallel}\}\\ &= \{f\in\mathcal{F}(\mathcal{S}) \,|\, df \ \textit{is m-parallel}\},\end{aligned} \tag{7.23}$$

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

$$\forall X,Y\in\mathscr{X}(\mathcal{S}),\ (\nabla^{(\mathrm{m})}_X\omega)(Y) := X\omega(Y)-\omega\big(\nabla^{(\mathrm{m})}_X Y\big)=0. \tag{7.24}$$

Proposition 7.11. *For an arbitrary submanifold $\mathcal{M}$ of $\mathcal{S}$, we have*

$$\begin{aligned}\mathcal{E}(\mathcal{M}) = \big\{f\in\mathcal{F}(\mathcal{M}) \,\big|\, \exists\tilde f\in\mathcal{E}(\mathcal{S}),\ f=\tilde f|_{\mathcal{M}} \ \textit{and}\\ \forall\rho\in\mathcal{M},\ \|(df)_\rho\|_\rho = \|(d\tilde f)_\rho\|_\rho\big\}.\end{aligned} \tag{7.25}$$

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 teated in section 2 where a manifold S is provided with a Riemannian metric g together with mutually dual affine connections ∇ and ∇^* such that ∇ is curvature-free and ∇^* is flat. We define

$$\begin{aligned}\mathcal{E}(S) :=& \{f\in\mathcal{F}(S) \,|\, \mathrm{grad}\, f \text{ is } \nabla\text{-parallel}\}\\ =& \{f\in\mathcal{F}(S) \,|\, df \text{ is } \nabla^*\text{-parallel}\},\end{aligned} \tag{7.26}$$

where we have invoked the fact that the correspondence $\mathscr{X}(S)\ni X \xleftrightarrow{g} \omega\in\mathcal{D}(S)$ implies (cf. (A.7))

$$X \text{ is } \nabla\text{-parallel} \ \Leftrightarrow\ \omega \text{ is } \nabla^*\text{-parallel}. \tag{7.27}$$

Given a submanifold M of S, let

$$\begin{aligned}\mathcal{E}(M) := \big\{f\in\mathcal{F}(M) \,\big|\, \exists\tilde f\in\mathcal{E}(S),\ f=\tilde f|_{M} \text{ and}\\ \forall\rho\in M,\ \|(df)_\rho\|_\rho = \|(d\tilde f)_\rho\|_\rho\big\}.\end{aligned} \tag{7.28}$$

Then it is not difficult to verify that Theorems 7.6 and 7.7 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 $X,Y\in\mathscr{X}(S)$, we have

$$\begin{aligned}(d\omega)(X,Y) &:= X\omega(Y)-Y\omega(X)-\omega([X,Y])\\ &= (\nabla^*_X\omega)(Y)+\omega(\nabla^*_X Y)-(\nabla^*_Y\omega)(X)-\omega(\nabla^*_Y X)-\omega([X,Y])\\ &= (\nabla^*_X\omega)(Y)-(\nabla^*_Y\omega)(X)+\omega(\mathcal{T}^{(\nabla^*)}(X,Y)),\end{aligned} \tag{7.29}$$

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 $\omega \in \mathcal{D}(S)$

$$\omega \text{ is } \nabla^*\text{-parallel} \;\Rightarrow\; d\omega = 0 \;\Leftrightarrow\; \exists f \in \mathcal{F}(S),\ \omega = df \tag{7.30}$$

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

$$X \text{ is } \nabla\text{-parallel} \;\Rightarrow\; \exists f \in \mathcal{F}(S),\ X = \operatorname{grad} f. \tag{7.31}$$

This leads to

$$d\mathcal{E}(S) := \{df \mid f \in \mathcal{E}(S)\} = \{\omega \mid \omega \text{ is } \nabla^*\text{-parallel}\} \tag{7.32}$$

and hence

$$\dim \mathcal{E}(S) = \dim\{\omega \mid \omega \text{ is } \nabla^*\text{-parallel}\} + 1 = \dim \mathscr{X}_{\nabla\text{-par}}(S) + 1, \tag{7.33}$$

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

8 Integrability conditions

Consider 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. 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 Prop. 2.4, an n-dimensional submanifold M of S is ∇-autoparallel iff there exist n linearly independent ∇-parallel vector fields $X^1, \ldots, X^n \in \mathscr{X}(S)$ such that their restrictions $X^1|_M, \ldots, X^n|_M$ belong to $\mathscr{X}(M)$. This means that M is an integral manifold of $\{X^1, \ldots, X^n\}$, or equivalently that M is an integral manifold of the n-dimensional distribution

$$\mathcal{V} : S \ni p \mapsto V_p := \operatorname{span}\{X^1_p, \ldots, X^n_p\} \subset T_p(S), \tag{8.1}$$

which is ∇-parallel in the sense that $\forall p, q \in S,\ \Phi^{(\nabla)}_{p,q}(V_p) = V_q$, where $\Phi^{(\nabla)}_{p,q}$ 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*

$$\mathscr{X}(S\colon \mathcal{V}) := \{X \in \mathscr{X}(S) \,|\, \forall p \in S,\ X_p \in V_p\} \tag{8.2}$$

and

$$\mathscr{X}_{\nabla\text{-}par}(S\colon \mathcal{V}) := \{X \in \mathscr{X}(S\colon \mathcal{V}) \,|\, X \text{ is } \nabla\text{-parallel}\}. \tag{8.3}$$

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 $\forall X, Y \in \mathscr{X}(S\colon \mathcal{V}),\ [X, Y] \in \mathscr{X}(S\colon \mathcal{V})$.

(iii) $\forall X, Y \in \mathscr{X}_{\nabla\text{-}par}(S\colon \mathcal{V}),\ [X, Y] \in \mathscr{X}(S\colon \mathcal{V})$.

(iv) The torsion $\mathcal{T}^{(\nabla)}$ of ∇ satisfies $\forall p \in S,\ \mathcal{T}^{(\nabla)}_p(V_p \times V_p) \subset V_p$.

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 $\{X_1, \ldots, X_n\} \subset \mathscr{X}_{\nabla\text{-par}}(S\colon \mathcal{V})$, whereby every element of $\mathscr{X}(S\colon \mathcal{V})$ is expressed as $\sum_i f_i X_i$ by some functions $\{f_1, \ldots, f_n\} \subset \mathcal{F}(M)$.

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

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

Hence (iii) is equivalent to

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

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

Remark 8.2. Condition (i) in Prop. 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 Prop. 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) $\forall p \in S$, $\forall X_p, Y_p \in T_p(S)$, $\mathcal{T}_p^{(\nabla)}(X_p, Y_p) \in \mathrm{span}\{X_p, Y_p\}$, or equivalently, $\forall X, Y \in \mathscr{X}(S)$, $\mathcal{T}^{(\nabla)}(X, Y) \in \mathrm{span}_{\mathcal{F}(S)}\{X, Y\} := \{fX + gY \mid f, g \in \mathcal{F}(S)\}$.

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

$$\begin{aligned} V_{\mathcal{A},\rho} &:= \{X_\rho \in T_\rho(\mathcal{S}) \mid \exists A \in \mathcal{A},\ L_{X_\rho} = A - \langle A \rangle_\rho\} \\ &= \{X_\rho \in T_\rho(\mathcal{S}) \mid L_{X_\rho} \in \mathcal{A} + \mathbb{R}\}, \end{aligned} \tag{8.6}$$

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 $\mathcal{L}_{\mathrm{h}}/\mathbb{R}$ and e-parallel distributions on $\mathcal{S}$.

Theorem 8.4. *Given a subspace $\mathcal{A} \subset \mathcal{L}_{\mathrm{h}}$, the following conditions are equivalent.*

(i) For every $\rho \in \mathcal{S}$, there exists an e-autoparallel submanifold $\mathcal{M}$ of $\mathcal{S}$ satisfying $\rho \in \mathcal{M}$ and $T_\rho(\mathcal{M}) = V_{\mathcal{A},\rho}$.

(ii) For every $\rho \in \mathcal{S}$,

$$\{[[A, B], \rho] \mid A, B \in \mathcal{A}\} \subset \{\rho \circ C \mid C \in \mathcal{A} + \mathbb{R}\}. \tag{8.7}$$

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

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

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 Prop. 8.1. $\square$

Let $F^1, \ldots, F^n$ be Hermitian operators on $\mathcal{H}$ such that $\forall i, j,\ [F^i, F^j] = 0$ and that $\{F^1, \ldots, F^n, I\}$ are linearly independent (cf. Remark 3.9), and let $\mathcal{A} := \mathrm{span}\{F^1, \ldots, F^n\}$ ($\not\ni I$). Then for any $\rho \in \mathcal{S}$ and any $A, B \in \mathcal{A}$ we have $[[A, B], \rho] = 0$, so that (ii) in Theorem 8.4 trivially holds. Hence the distribution $\mathcal{V}_{\mathcal{A}}$ is integrable, and we obtain a foliation $\mathcal{S} = \bigsqcup_\alpha \mathcal{M}_\alpha$, whose leaves $\{\mathcal{M}_\alpha\}$ are n-dimensional quasi-exponential families of the form (4.15).

Remark 8.5. Recall the situation when we defined the quantum Gaussian shift model $\mathcal{M} = \{\rho_\xi \,|\, \xi = (\xi^1, \ldots, \xi^n) \in \mathbb{R}^n\}$ in Section 6, and let $\mathcal{A} := \mathrm{span}\{R(z_1), \ldots, R(z_n)\}$. 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 8.4 holds at least formally, and the Gaussian model may be regarded as an integral manifold of $\mathcal{V}_{\mathcal{A}}$. Note, however, that Theorem 8.4 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.

Remark 8.6. Let us revisit the relationship between autoparallelity and total geodesicness described in Remark 2.2 in the context of Prop. 8.3. Suppose that (S, ∇) satisfies conditions (i)-(ii) of Prop. 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 geodesicness. In fact, the curvature-freeness is unnecessary, and their equivalence follows from condition (ii) alone. See A4 of Appendix for details.

At the end of this section, we give an examples of e-autoparallel submanifold that does not fall within the scope of Theorem 8.4. Let $\mathcal{B} = \{|1\rangle, \ldots, |d\rangle\}$ with $d = \dim \mathcal{H}$ be an arbitrary orthonormal basis of $\mathcal{H}$, and let $\mathcal{L}^{\mathcal{B}} := \{\sum_{i,j} a_{ij} \,|i\rangle \langle j| \,|\, [a_{ij}] \in \mathbb{R}^{d \times d}\}$, $\mathcal{L}_{\mathrm{h}}^{\mathcal{B}} := \mathcal{L}_{\mathrm{h}} \cap \mathcal{L}^{\mathcal{B}}$ and $\mathcal{S}^{\mathcal{B}} := \mathcal{S} \cap \mathcal{L}^{\mathcal{B}}$.

Proposition 8.7. *$\mathcal{S}^{\mathcal{B}}$ is e-autoparallel in $\mathcal{S}$.*

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

$$T_\rho^{(\mathrm{m})}(\mathcal{S}^{\mathcal{B}}) := \{\iota_*(X_\rho) \,|\, X_\rho \in T_\rho(\mathcal{S}^{\mathcal{B}})\} = \{A \in \mathcal{L}_{\mathrm{h}}^{\mathcal{B}} \,|\, \mathrm{Tr} A = 0\} \tag{8.8}$$

and that

$$\begin{aligned} T_\rho^{(\mathrm{e})}(\mathcal{S}^{\mathcal{B}}) &:= \{L_{X_\rho} \,|\, X_\rho \in T_\rho(\mathcal{S}^{\mathcal{B}})\} \\ &= \{A \in \mathcal{L}_{\mathrm{h}} \,|\, \exists B \in T_\rho^{(\mathrm{m})}(\mathcal{S}^{\mathcal{B}}),\ B = \rho \circ A\} \\ &= \{A \in \mathcal{L}_{\mathrm{h}}^{\mathcal{B}} \,|\, \langle A \rangle_\rho = 0\}. \end{aligned} \tag{8.9}$$

(See Remark 3.1 for the symbols $T^{(\mathrm{m})}$ and $T^{(\mathrm{e})}$.) It follows from (8.9) that $A \in T_\rho^{(\mathrm{e})}(\mathcal{S}^{\mathcal{B}}) \;\Leftrightarrow\; A - \langle A \rangle_\sigma \in T_\sigma^{(\mathrm{e})}(\mathcal{S}^{\mathcal{B}})$, and hence we have from (3.30) that $X_\rho \in T_\rho(\mathcal{S}^{\mathcal{B}}) \;\Leftrightarrow\; \Phi_{\rho,\sigma}^{(\mathrm{e})}(X_\rho) \in T_\sigma(\mathcal{S}^{\mathcal{B}})$. This proves the proposition by Prop. 2.4. □

Let us examine whether the e-autoparallelity of $\mathcal{S}^{\mathcal{B}}$ can be understood as an example of Theorem 8.4. 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 8.4.

For each $\rho \in \mathcal{S}^{\mathcal{B}}$, we have $T^{(\mathrm{e})}_{\rho}(\mathcal{S}^{\mathcal{B}}) = \{A - \langle A \rangle_{\rho} \mid A \in \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}\}$ (see (8.9)), which means that $T_{\rho}(\mathcal{S}^{\mathcal{B}}) = V_{\mathcal{L}^{\mathcal{B}}_{\mathrm{h}},\rho}$ and that $\mathcal{S}^{\mathcal{B}}$ is an integral manifold of the distribution $\mathcal{V}_{\mathcal{L}^{\mathcal{B}}_{\mathrm{h}}}$. Noting that $\mathcal{L}^{\mathcal{B}}_{\mathrm{h}} + \mathbb{R} = \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}$, the problem comes down to whether

$$\{[[A, B], \rho] \mid A, B \in \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}\} \subset \{\rho \circ C \mid C \in \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}\} \tag{8.10}$$

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

Proposition 8.8. *When* $\dim \mathcal{H} \geq 3$*,*

$$\exists \rho \in \mathcal{S},\ \{[[A, B], \rho] \mid A, B \in \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}\} \not\subset \{\rho \circ C \mid C \in \mathcal{L}^{\mathcal{B}}_{\mathrm{h}}\}. \tag{8.11}$$

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

Proof We represent operators on a d-dimensional Hilbert space by $d \times d$ matrices, and show that there exist a strictly positive density matrix ρ and real symmetric matrices A, B such that $[[A, B], \rho]$ cannot be represented as $\rho \circ C$ by any real symmetric C when $d \geq 3$. Let

$$A_1 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ B_1 := \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},\ P_1 := \begin{pmatrix} 1 & i\varepsilon & i\varepsilon \\ -i\varepsilon & 1 & i\varepsilon \\ -i\varepsilon & -i\varepsilon & 1 \end{pmatrix},$$

where $i := \sqrt{-1}$ and ε is an arbitrary real number, and let A, B and ρ be $d \times d$ matrices with the block representations

$$A := \left(\begin{array}{c|c} A_1 & 0 \\ \hline 0 & 0 \end{array}\right),\ B := \left(\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array}\right),\ \rho := \frac{1}{d}\left(\begin{array}{c|c} P_1 & 0 \\ \hline 0 & I \end{array}\right).$$

Then A, B are real symmetric, and ρ is Hermitian with trace 1 and strictly positive when $|\varepsilon|$ is sufficiently small. A direct calculation shows that

$$[[A, B], \rho] = \frac{1}{d}\left(\begin{array}{c|c} Q_1 & 0 \\ \hline 0 & I \end{array}\right) \quad \text{with} \quad Q_1 := [[A_1, B_1], P_1] = \begin{pmatrix} 0 & 0 & -i\varepsilon \\ 0 & 0 & i\varepsilon \\ i\varepsilon & -i\varepsilon & 0 \end{pmatrix}.$$

Suppose that a $d \times d$ real symmetric matrix C satisfies $[[A, B], \rho] = \rho \circ C$. Letting C_1 be the 3×3 block of C, the 3×3 block of $\rho \circ C$ equals $\frac{1}{d} P_1 \circ C_1$. Hence we have $Q_1 = P_1 \circ C_1$, which is rewritten as

$$i\varepsilon \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} = C_1 + i\varepsilon \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \circ C_1.$$

Since $C_1 \in \mathbb{R}^{3\times 3}$ and $\varepsilon \in \mathbb{R}$, this implies that $C_1 = 0$ and $\varepsilon = 0$. Therefore, if we take $\varepsilon \neq 0$, no real symmetric C satisfies $[[A, B], \rho] = \rho \circ C$. □

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

Remark 8.9. Prop. 8.7 holds for a wide class of information geometric structures, not limited to the SLD structure. In fact, the proof of Prop. 8.7 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 $L_{X_\rho} \in \mathcal{L}_{\mathrm{h}}^{\mathcal{B}}$. 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 $\langle \cdot, \cdot \rangle_\rho = \langle \cdot, \Omega_\rho(\cdot) \rangle_{\mathrm{HS}}$, $\rho \in \mathcal{S}$, such that

$$\forall \rho \in \mathcal{S}^{\mathcal{B}},\ \Omega_\rho(\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}) = \mathcal{L}_{\mathrm{h}}^{\mathcal{B}}. \tag{8.12}$$

This means that Prop. 8.7 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 $f : (0, \infty) \to (0, \infty)$ such that $\forall x > 0$, $xf(1/x) = f(x)$ and $f(1) = 1$ as in the case of monotone metrics, condition (8.12) 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 $\mathcal{L} \to \mathcal{L}$. 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 $\Omega_\rho(\mathcal{L}_{\mathrm{h}}) = \mathcal{L}_{\mathrm{h}}$, this proves (8.12).

Remark 8.10. Since (8.8) shows that $T_\rho^{(\mathrm{m})}(\mathcal{S}^{\mathcal{B}}) = \iota_*(T_\rho(\mathcal{S}^{\mathcal{B}}))$ 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, [10]) w.r.t. the e, m-connections. This example exhibits a remarkable contrast to the following fact for the classical case [11]; 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. To begin with, we make some preparations. Let $\{\sigma_1, \sigma_2, \sigma_3\} \subset \mathcal{L}_{\mathrm{h}}$ be a triple of Pauli operators such that

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

Then $\{\sigma_1, \sigma_2, \sigma_3\}$ form a basis of $\mathcal{L}_{\mathrm{h},0}$. 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

$$\mathcal{L}_{\mathrm{h},0} = \{\vec{a} \cdot \vec{\sigma} \,|\, \vec{a} \in \mathbb{R}^3\}. \tag{9.2}$$

It follows that

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

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

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

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

$$\mathcal{S} = \{\rho_{\vec{r}} \,|\, \vec{r} \in \mathcal{R}\}, \tag{9.6}$$

where

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

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

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

The SLD of X_ρ is then represented as

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

where

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

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

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

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 (4.18). Treating the coordinate θ as a parameter to specify states and choosing P in (4.18) to be a state ρ_0, an arbitrary e-geodesic $\mathcal{M}$ is represented as the trajectory $\mathcal{M} = \{\rho_\theta \,|\, \theta \in \mathbb{R}\}$ of

$$\rho_\theta = \frac{1}{Z_\theta}\exp\Big(\frac{\theta}{2}F\Big)\,\rho_0 \exp\Big(\frac{\theta}{2}F\Big), \quad Z_\theta := \mathrm{Tr}(\rho_0 \exp(\theta F)), \tag{9.12}$$

where F is a Hermitian operator such that $\{F, I\}$ are linearly independent. Since the transformation $F \to aF + b$ by $a, b \in \mathbb{R}$, $a \neq 0$, together with $\theta \to \frac{1}{a}\theta$ and $\psi \to \psi + \frac{b}{a}\theta$, keeps 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 $\vec{r}_0 \in \mathrm{span}\{\vec{u}, \vec{v}\}$, the e-geodesic $\mathcal{M} = \{\rho_\theta \,|\, \theta \in \mathbb{R}\}$ is represented as*

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

where

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

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

(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

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

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

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

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

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

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

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

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. □

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 $\mathcal{M} = \{\rho_{\vec{r}}\,|\,\vec{r}\in\mathcal{Q}\}$ 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.

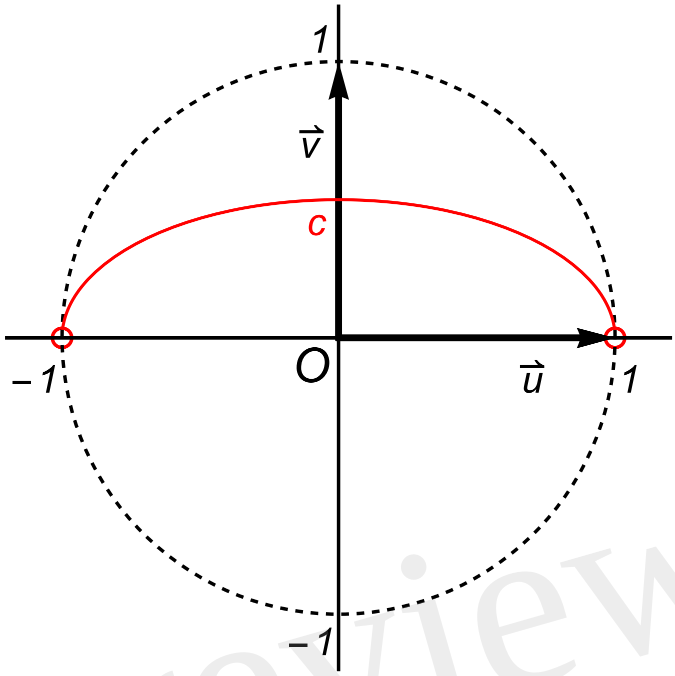

Figure 1: The semi-ellipse representing an e-geodesic

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*

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

where

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

Then $\mathcal{M} := \{\rho_{\vec{r}} \,|\, \vec{r} \in \mathcal{Q}\}$ 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 $F^i := \vec{u}_i \cdot \vec{\sigma}$, $\mathcal{A} := \mathrm{span}\{F^1, F^2\}$, and $\mathcal{V}_{\mathcal{A}} : \mathcal{S} \ni \rho \mapsto V_{\mathcal{A},\rho}$ be the e-parallel distribution defined from $\mathcal{A}$ by (8.6), $\mathcal{M}$ 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

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

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 $L_{i,\xi} \in \mathrm{span}\,\{F^1, F^2\} \oplus \mathbb{R}$ for $i \in \{1, 2\}$, which means that the first condition in (3.33) is satisfied. The second condition is also satisfied since $\left\langle F^i \right\rangle_{\rho_{\vec{r}(\xi)}} = \vec{u}_i \cdot \vec{r}(\xi) = \xi^i$. Thus the claim of the proposition follows from Prop. 3.8. □

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 $\{\vec{u}_1, \vec{u}_2\} \subset \ell_{\vec{r}}(V) = \{\ell_{\vec{r}}(\vec{x}) \,|\, \vec{x} \in V\}$, 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 Prop. 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* $A, B \in \mathcal{L}_{\mathrm{h}}$ *satisfying* $\langle A \rangle_\rho = \langle B \rangle_\rho = 0$ *we have*

$$[[A, B], \rho] \in \mathrm{span}\,\{\rho \circ A, \rho \circ B\}. \tag{9.19}$$

This proposition can also be proved directly by the use of the following lemma, whose proof is given in A5 of Appendix.

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

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

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

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

which proves Prop. 9.3.

The following proposition immediately follows from Prop. 9.3, which presents a remarkable contrast to Prop. 8.8 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*

$$\forall\rho\in\mathcal{S},\ \{[[A,B],\rho]\,|\,A,B\in\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}\}\subset\{\rho\circ C\,|\,C\in\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}\}. \tag{9.22}$$

Proof Obvious from Prop. 9.3 since $\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}$ is an $\mathbb{R}$-linear space with $I\in\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}$. □

Thus, the distribution $\mathcal{V}_{\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}}$ 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 $\mathcal{V}_{\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}}$. Furthermore, we can see from the following lemma that every 2-dimensional e-autoparallel submanifold of $\mathcal{S}$ is an integral manifold of $\mathcal{V}_{\mathcal{L}_{\mathrm{h}}^{\mathcal{B}}}$ for some $\mathcal{B}$.

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

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. □

10 Concluding remarks

In this paper we studied the autoparallelity w.r.t. the e-connection for an information-geometric structure induced on $\mathcal{S}(\mathcal{H})$. In particular, we focused on the e-autoparallelity for the SLD structure, for which two different estimation-theoretical characterizations were given. We also investigated 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 was 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 materical 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 Section 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

$\mathcal{M}^{(n)} = \{\rho_\xi^{\otimes n}\}$, $n = 1, 2, \ldots$, 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 [12, 13, 14, 15, 16]. As pointed out in Section 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.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Numbers 23H05492 (HN), 23H01090 (AF) and 17H02861 (HN, AF).

References

[1] Rao CR. Linear Statistical Inference and Its Applications (2nd ed). John Wiley & Sons (1973).

[2] Kiefer JC. Introduction to Statistical Inference. Springer-Verlag (1987).

[3] Amari S, Nagaoka H. Methods of Information Geometry. American Mathematical Society, Oxford University Press (2000).

[4] Kobayashi S, Nomizu K. Foundations of Differential Geometry, II. New York: John Wiley & Sons (1969).

[5] Nagaoka H., Amari S. Differential geometry of smooth families of probability distributions. METR 82-7, Dept Math Eng and Instr Phys, Univ. of Tokyo (1982). (https://www.keisu.t.u-tokyo.ac.jp/research/techrep/y1982/) (https://bsi-ni.brain.riken.jp/database/item/104)

[6] Helstrom CW. Minimum mean-square error estimation in quantum statistics. Phys Lett (1967) 25A:101–102 .

[7] Helstrom CW. Quantum Detection and Estimation Theory. New York: Academic Press (1976).

[8] Holevo AS. Probabilistic and Statistical Aspects of Quantum Theory. Pisa: Edizioni della Normale (2011). (the previous edition. Amsterdam: North-Holland (1982).)

[9] Petz D. Monotone metrics on matrix spaces. Linear Algebra Appl (1996) 244:81–96.

[10] Ohara A. Geodesics for dual connections and means on symmetric cones. Integr equ oper theory (2004) 50:537–548.

[11] Nagaoka H. Information-geometrical characterization of statistical models which are statistically equivalent to probability simplexes. Proc. 2017 IEEE International Symposium on Information (2017) 1346–1350.

[12] Guţă M, Kahn J. Local asymptotic normality for qubit states. Phys Rev A (2006) 73:052108.

[13] Kahn, J, Guţă M. Local asymptotic normality for finite dimensional quantum systems. Comm Math Phys (2009) 289:597–652.

[14] Yamagata K, Fujiwara A, Gill RD. Quantum local asymptotic normality based on a new quantum likelihood ratio, Ann Statist (2013) 41:2197–2217.

[15] Fujiwara A, Yamagata K. Noncommutative Lebesgue decomposition and contiguity with application to quantum local asymptotic normality. Bernoulli (2020) 26:2105–2142.

[16] Fujiwara A, Yamagata K. Efficiency of estimators for locally asymptotically normal quantum statistical models. Ann Statist (to appear). (arXiv:2209.00832)

Appendix

A1 Proof of (3.26)

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

$$\iota_*(\mathcal{T}^{(\mathrm{e})}(X,Y)) = (Y\Omega)(L_X) - (X\Omega)(L_Y), \tag{A.1}$$

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 tha map $\rho \mapsto (Y_\rho\Omega)(L_{X_\rho}) \in \mathcal{L}_{\mathrm{h}}$. In fact, invoking (3.8) and (3.27), we have for any $X, Y, Z \in \mathscr{X}(\mathcal{S})$,

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

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 (A.1). 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 $X_\rho, Y_\rho \in T_\rho(\mathcal{S})$,

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

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 (A.1), we obtain (3.26).

A2 Proof of Lemma 4.2

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

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

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 k

$$\begin{aligned}B^k &:= \int\Big\{\sum_i u_i^k(\hat{\xi}^i - \xi^i(\rho))\Big\}^2\Pi(d\hat{\xi})\\ &= \sum_l p_l\Big\{\sum_i u_i^k(f^i(l, X^l) - \xi^i(\rho))\Big\}^2\\ &= \sum_l p_l(C_l^k + a_l^k)^2, \qquad \text{(A.2)}\end{aligned}$$

where

$$C_l^k := \sum_i u_i^k \frac{w_l^i}{p_l}X^l = \frac{\delta_l^k}{p_k}X^k. \qquad \text{(A.3)}$$

This leads to

$$
\begin{aligned}
(\boldsymbol{u}^k)^T V_\rho(\Pi)\boldsymbol{u}^k &= \mathrm{Tr}(\rho B^k) = \sum_l p_l \mathrm{Tr}(\rho(C_l^k + a_l^k)^2) \\
&= \sum_l p_l \mathrm{Tr}(\rho(C_l^k)^2) + \sum_l p_l (a_l^k)^2, \qquad \text{(A.4)}
\end{aligned}
$$

where we invoked $\mathrm{Tr}(\rho\, C_l^k) = 0$ due to $C_l^k \in \mathrm{span}\{L_{i,\rho}\}_{i=1}^n$. Recalling (4.11) and (A.3), we have

$$
\begin{aligned}
\sum_l p_l \mathrm{Tr}(\rho(C_l^k)^2) &= \frac{1}{p_k} \mathrm{Tr}(\rho(X^k)^2) \\
&= \frac{1}{p_k} \sum_{i,j} u_i^k u_j^k \left\langle L_\rho^i, L_\rho^j \right\rangle_\rho = \frac{1}{p_k} (\boldsymbol{u}^k)^T G_\rho^{-1} \boldsymbol{u}^k, \qquad \text{(A.5)}
\end{aligned}
$$

which, combined with (A.4), yields the desired identity.

A3 Proofs of Propositions 7.9, 7.10 and 7.11

Proof of Prop. 7.9 This proposition is essentially contained in Theorem 7.2 of [3]. Here we give a proof for the reader's convenience.

Given $F \in \mathcal{L}_\mathrm{h}$ 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 (7.18) to the case $\mathcal{M} = \mathcal{S}$, we have $X_\rho = (\mathrm{grad}\, \langle F \rangle)_\rho$, and hence

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

□

Proof of Prop. 7.10 Recalling (7.15) and (7.16), we have

$$
\mathcal{E}(\mathcal{S}) = \{ f \in \mathcal{F}(\mathcal{S}) \,|\, \exists F \in \mathcal{L}_\mathrm{h},\ f = \langle F \rangle \ \text{ and } \ \forall \rho \in \mathcal{S},\ V_\rho(F) = \|(df)_\rho\|_\rho^2 \}. \qquad \text{(A.6)}
$$

Since the condition $V_\rho(F) = \|(df)_\rho\|_\rho^2$ is always satisfied by Prop. 7.9, we have the first equality in (7.10). The second equality follows from Prop. 7.2, and the third follows since under the relation $X \overset{g}{\longleftrightarrow} \omega$ we have

$$
\begin{aligned}
X \text{ is e-parallel } &\Leftrightarrow\ \forall Y, Z \in \mathscr{X}(\mathcal{S}),\ g(\nabla_Y^{(\mathrm{e})} X, Z) = 0 \\
&\Leftrightarrow\ \forall Y, Z \in \mathscr{X}(\mathcal{S}),\ Y g(X, Z) = g(X, \nabla_Y^{(\mathrm{m})} Z) \\
&\Leftrightarrow\ \forall Y, Z \in \mathscr{X}(\mathcal{S}),\ Y \omega(Z) = \omega(\nabla_Y^{(\mathrm{m})} Z) \\
&\Leftrightarrow\ \omega \text{ is m-parallel.} \qquad \text{(A.7)}
\end{aligned}
$$

□

Proof of Prop. 7.11 By Propositions 7.9 and 7.10, the condition imposed on f in (7.25) is equivalent to the existence of $F \in \mathcal{L}_\mathrm{h}$ satisfying (7.15). □

A4 A result on the relationship between autoparallelity and total geodesicness

In Remark 8.6 we noted that condition (ii) of Prop. 8.3 implies the equivalence between autoparallelity and total geodesicness. This is restated in the following proposition.

Proposition A.1. *Suppose that an affine connection ∇ is given on a manifold S whose torsion satisfies*

$$\forall X, Y \in \mathscr{X}(S),\ \mathcal{T}^{(\nabla)}(X,Y) \in \mathrm{span}_{\mathcal{F}(S)}\{X, Y\}. \tag{A.8}$$

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

We present a proof below, which is almost parallel to the proof of Theorem 8.4 in Chap. VII of [4] cited as a result due to E. Cartan.

Proof 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

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

and that $(\xi^1, \ldots, \xi^n) := (\tilde{\xi}^1|_M, \ldots, \tilde{\xi}^n|_M)$ 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 $i, j \in \{1, \ldots, n+r\}$. For arbitrary i, j, it follows from the assumption (A.8) that $\mathcal{T}^{(\nabla)}(\tilde{\partial}_i, \tilde{\partial}_j) = \sum_k (\Gamma_{ij}^k - \Gamma_{ji}^k)\, \tilde{\partial}_k \in \mathrm{span}_{\mathcal{F}(S)}\{\tilde{\partial}_i, \tilde{\partial}_j\}$, which implies that $\Gamma_{ij}^k - \Gamma_{ji}^k = 0$ for any $k \notin \{i, j\}$. Hence we have

$$\forall i, j \in \{1, \ldots, n\},\ \forall k \in \{n+1, \ldots, n+r\},\ \Gamma_{ij}^k = \Gamma_{ji}^k. \tag{A.9}$$

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

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

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

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

and letting $t=0$, we obtain

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

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 (A.9), it follows that

$$\forall i,j \in \{1,\dots,n\},\ \forall k \in \{n+1,\dots,n+r\},\ \Gamma_{ij}^k|_M = 0. \tag{A.10}$$

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

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

which concludes that M is ∇-autoparallel in S. $\square$

Note that the only difference from the proof of [4] is whether (A.9) is derived from $\mathcal{T}^{(\nabla)} = 0$ or from the weaker assumption (A.8).

A5 Proof of Lemma 9.4

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

$$\frac{1}{2}[[A,B],\rho] = -\,2\left\langle A\right\rangle_\rho (\rho \circ B) + 2\left\langle B\right\rangle_\rho (\rho \circ A), \tag{A.11}$$

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

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

Hence, the well-known formula for the vector triple product proves (A.11).

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

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

where the second equality follows from (A.11). Thus we obtain (9.20).